



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	5
1.1 La naturaleza de los datos de serie de tiempo	5
1.1.1 Ejemplo 1: Cambio climático	6
1.1.2 Ejemplo 2: Series temporales financieras	7
1.2 Modelamiento estadístico de las series de tiempo	12
1.2.1 Ruido blanco - White Noise	12
1.2.2 Caminata Aleatoria - Random Walk	13
1.2.3 Promedios móviles - Moving Averages	16
1.2.4 Autorregresiones - Autoregressions	17
1.3 Descomposición de las series de tiempo	19
1.3.1 Descomposición Tendencia - Filtro Hodrick -Prescott	21
1.3.2 Descomposición componente estacional - Transformaciones de Diferencias	21
1.3.3 Descomposición componente estacional - Variables Dummy	22
1.4 Medidas de dependencia	23
1.5 Estacionaridad	25
1.6 Modelos de regresión lineal múltiple en series de tiempo	26
1.6.1 Propiedades de los estimadores de MC	27
1.6.2 Teorema Gauss-Markov	28
1.6.3 Modelos de regresión lineal múltiple en series de tiempo	28
1.6.4 Evaluación Estadística de Regresiones Estimadas	29
1.6.5 Regresiones sin sentido y regresiones espúrias.	29
1.6.6 Ejemplo 3: Regresiones	29
2 Tópico II.- Univariate Time Series Models	34
2.1 Ejemplo de repaso clase anterior-Detrending global temperature	34
2.2 Sobre la descomposición de una serie	40
2.3 Procesos no estacionarios, integrados y el test de raíz unitaria	41
2.3.1 Procesos Integrados	41
2.3.2 Consecuencias de los Procesos Integrados (Raíz Unitaria)	41
2.3.3 Test de Raíz Unitaria	41
2.4 Modelos ARIMA: modelando el corto plazo	44
2.4.1 Sobre el operador de retroceso - backshift operator	44
2.4.2 Ejemplo Proceso Autoregresivo de Orden 1: AR(1)	45
2.4.3 Simulación modelo AR(1)	46
2.4.4 Identificación modelo AR(1)	47
2.4.5 Media Móvil - MA (q)	50

2.4.6	Identificación modelo ARMA	54
2.4.7	Invertibilidad	55
2.4.8	Estacionaridad e Invertibilidad	55
2.4.9	Modelos ARMA (p,q)	56
2.4.10	Modelos ARIMA (p, i, q)	56
2.4.11	Identificación modelo ARMA	56
2.4.12	Modelos SARIMA (p,q)	57
2.4.13	Sobre la Selección de Modelos	58
2.4.14	Ejemplo IPC en Chile	60
2.4.15	Descomposición de la serie	63
2.5	Metodología de Estimación de un Modelo ARIMA	71
2.5.1	Etapas de Estimación de un Modelo ARIMA	71
2.6	Resumen de los pasos de Box-Jenkins	72
2.7	Tarea 2	73
3	Tópico III.- Vector Autoregressive Models	74
3.1	Vector Autoregressive Models	74
3.1.1	Introducción	74
3.1.2	Notación y Conceptos	75
3.1.3	VAR versus Modelos Estructurales	75
3.1.4	La elección de la longitud del rezago óptimo para un VAR	76
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
4	Tópico IV.- Cointegration Analysis	97
4.1	Cointegración	97
4.1.1	Introducción	97
4.1.2	Test de Cointegración de Engle-Granger	97
4.1.3	Modelo de Corrección de Errores	97
4.1.4	Procedimiento de Estimación	104
4.2	Cointegration in a VAR: Vector Error-Correction Models	106
4.2.1	Un modelo VEC de dos variables	106
4.3	Cointegration con Modelos ARDL	108
5	Tópico V.- Non-linear Models: Volatility Forecasting	109
5.1	Modelos ARCH	109
5.2	Modelos GARCH	115

6 Tópico VI - Introduction to Neural Networks	125
6.1 Introducción a las Redes Neuronales	125
6.1.1 Feed forward neural network	125
6.2 Support Vector Machines (SVM)	131
6.2.1 Definiciones del modelo - Funciones de Kernel	131
6.2.2 Definiciones del modelo - Estructura de la SVM	132
6.2.3 Definiciones del modelo - Funciones de perdida	132
6.2.4 Algoritmo de Regresión SVM	132

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 & Jenkins, y sus modelos autorregresivos (ARIMA), se encuentra 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

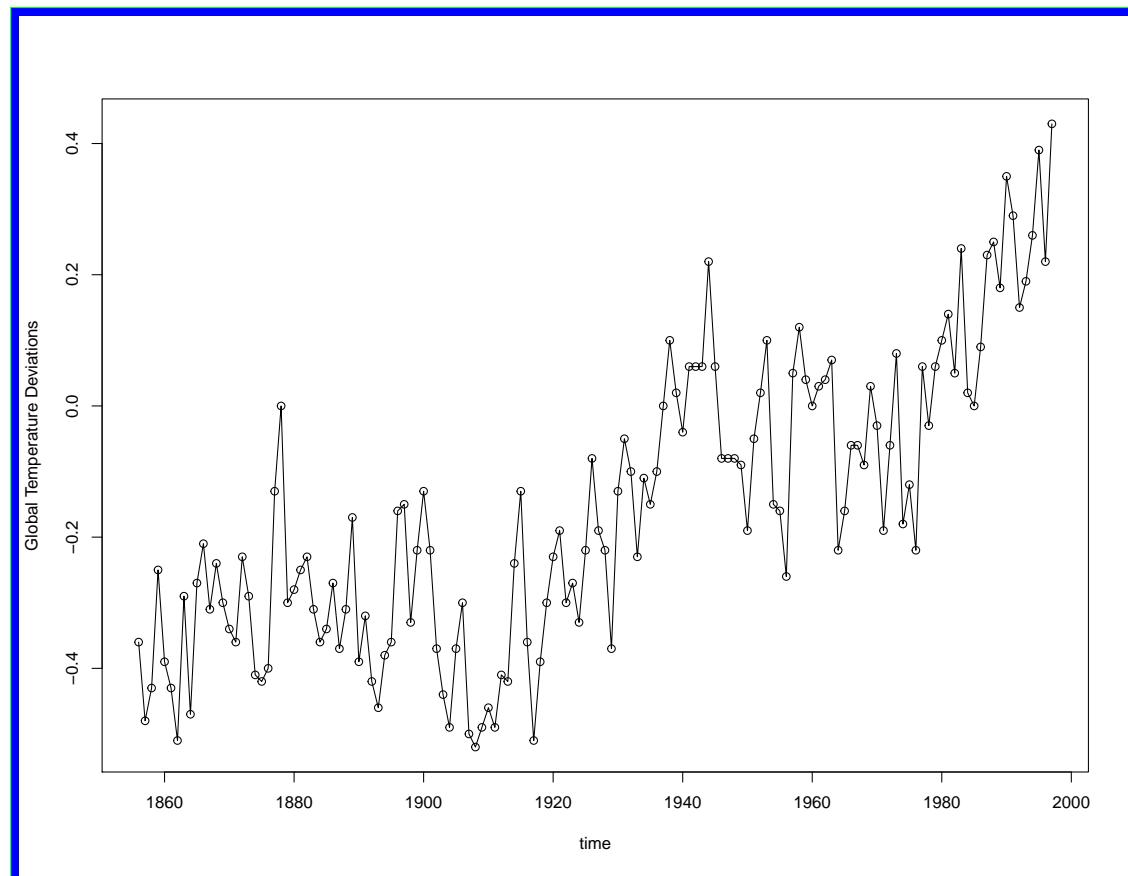


Figure 1.1: Cambio climático: mediciones de temperatura desde 1860

'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

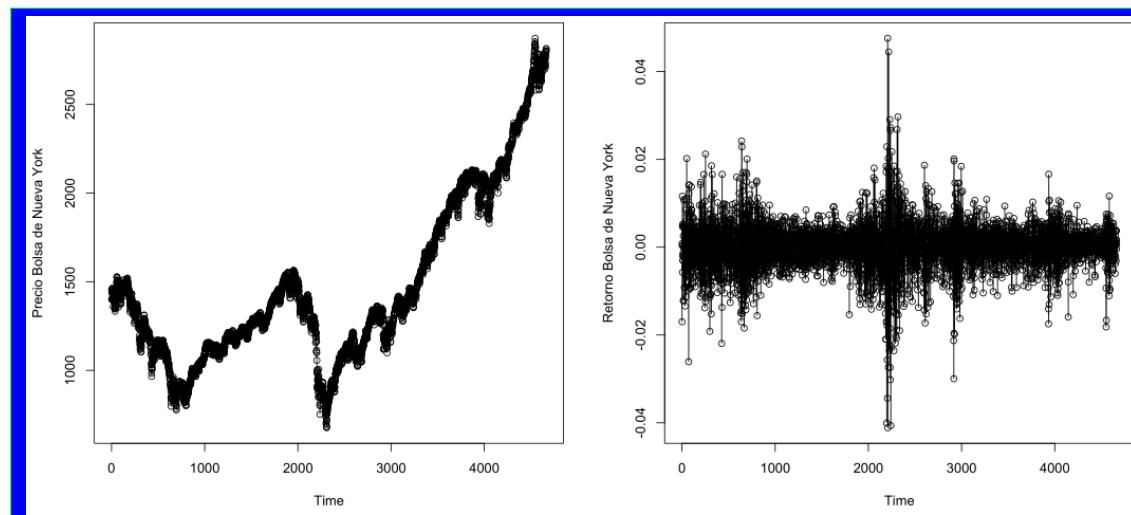


Figure 1.2: Indice Bolsa de Nueva York. Izquierda: precios reales SP500. Derecha: retornos SP500.

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

Ejemplo 2: Histograma de los Precios

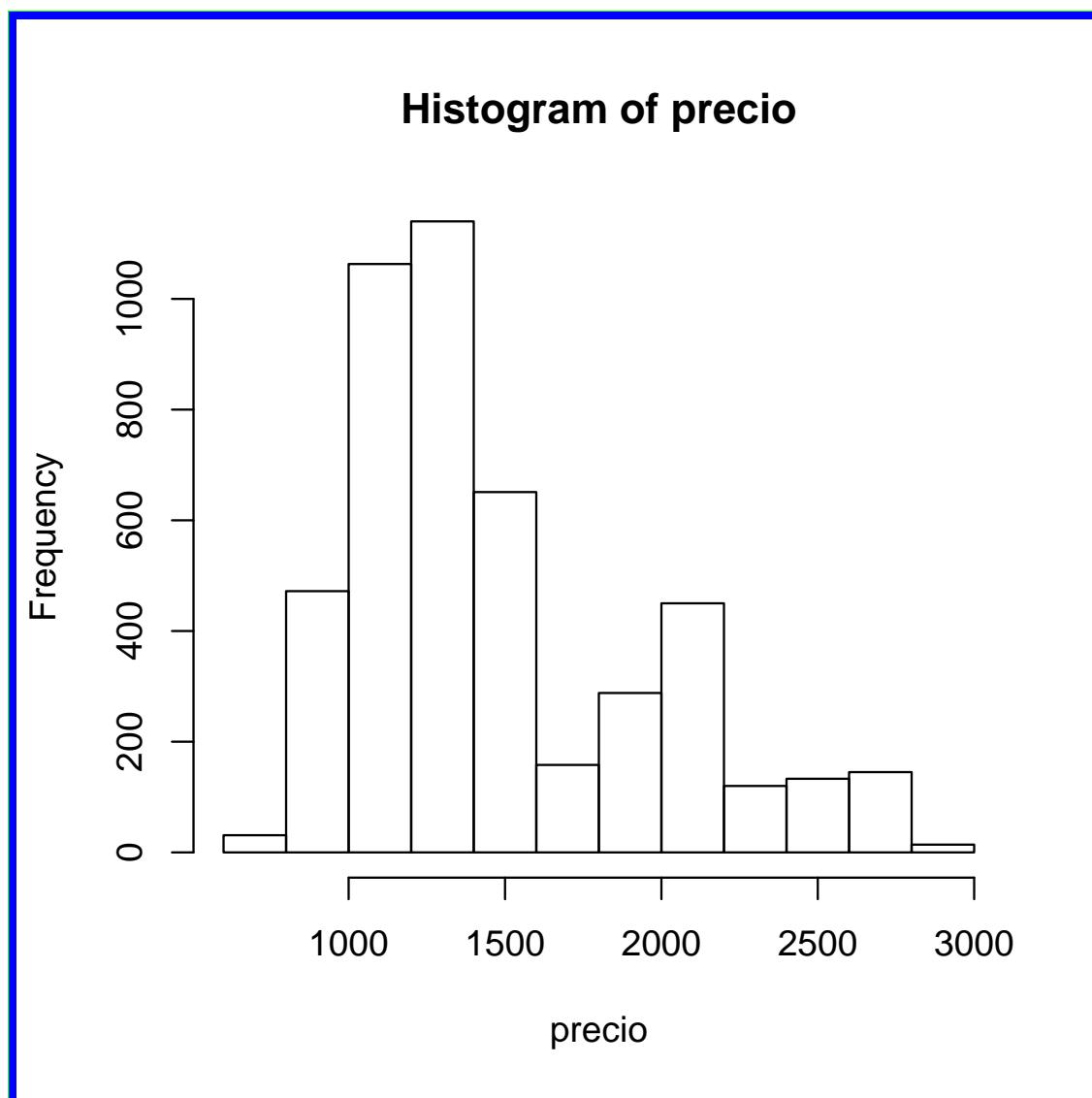


Figure 1.3: Histograma de los precios de la Bolsa de NY

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

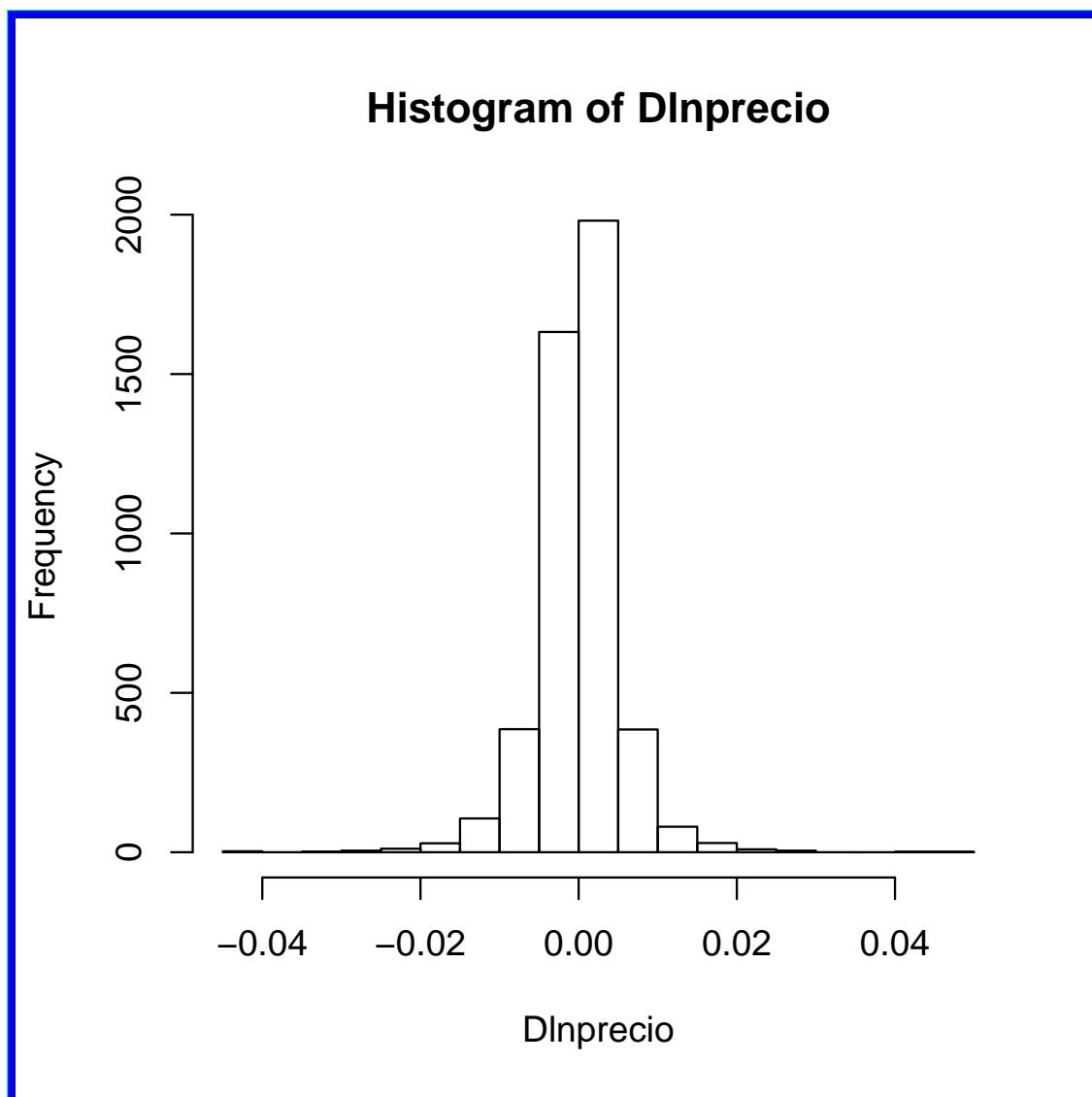


Figure 1.4: Histograma de los retornos de la Bolsa de NY

Ejemplo 2: Histograma de los Precios

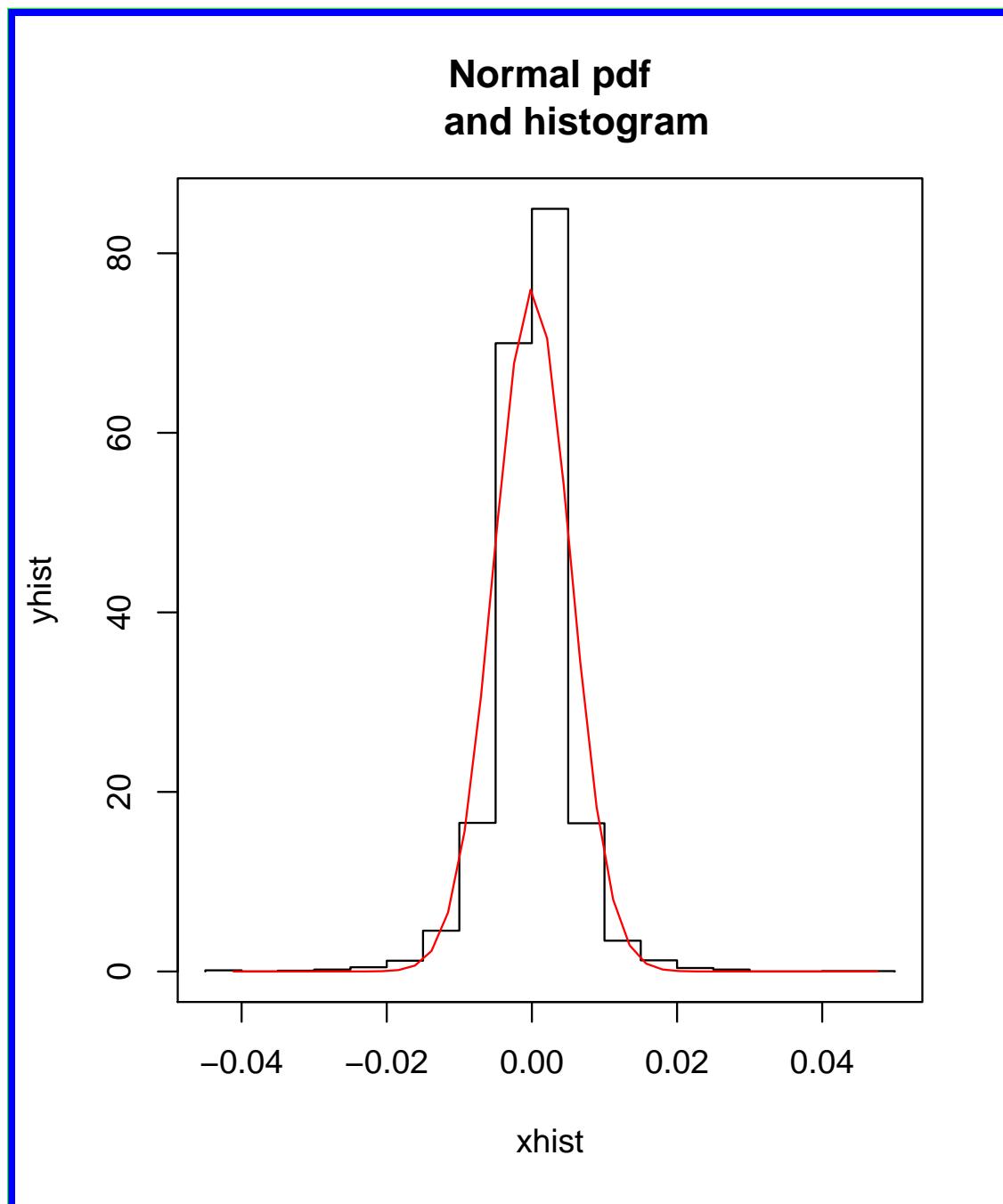


Figure 1.5: Linea negra:Histograma de las diferencias de los log-precios de la Bolsa de NY. Linea roja: ajuste de la distribución normal segúin media y desviación estandar de la muestra. Shapiro-Wilk normality test / data: Dlnprecio / W = 0.90927, p-value < 2.2e-16.

'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")
```

Ruido blanco - White Noise

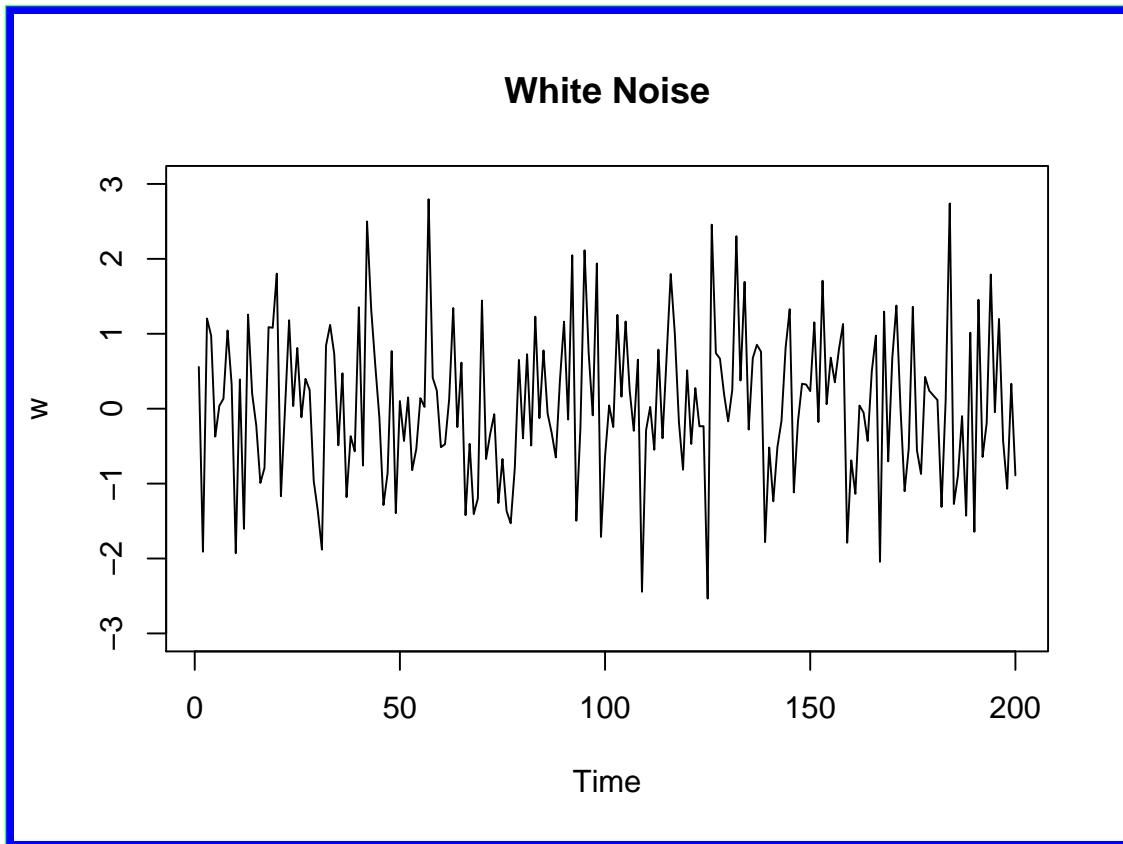


Figure 1.6: Ruido blanco - White Noise

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

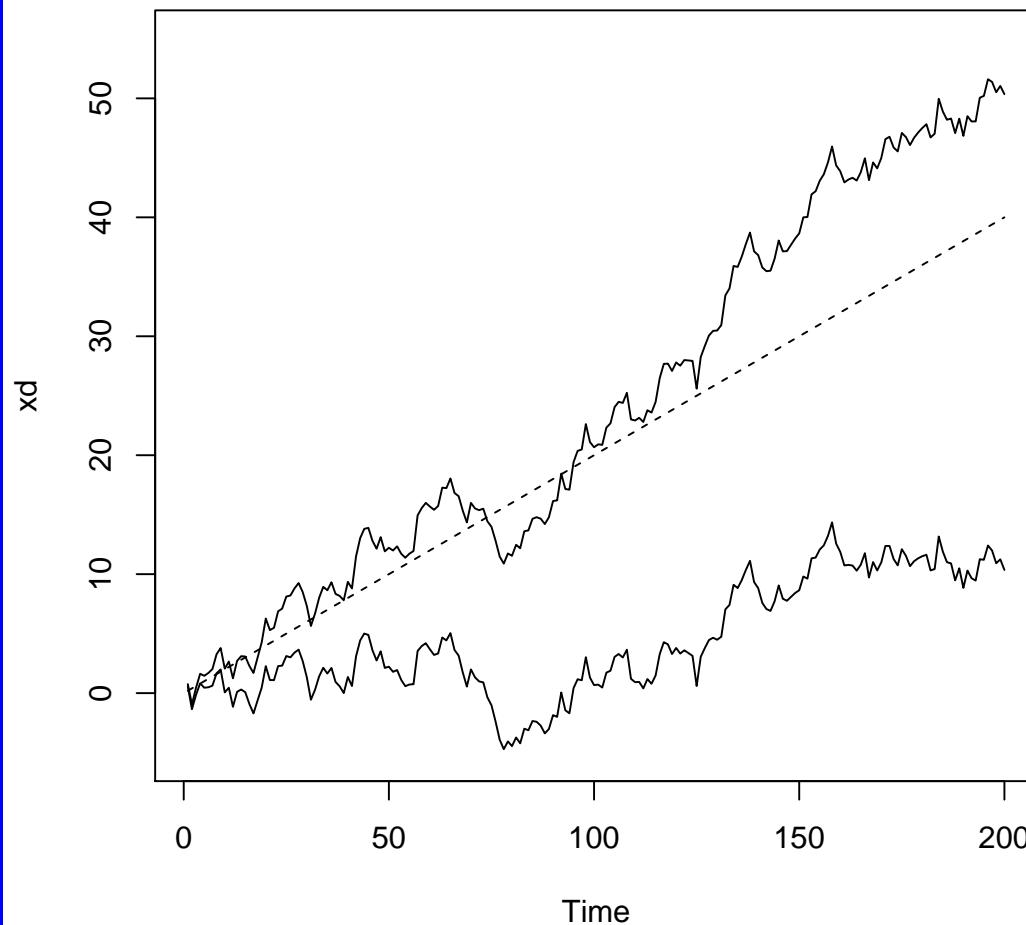


Figure 1.7: Caminata Aleatoria - Random walk, $\sigma_\epsilon = 1$, with drift $\delta = 0.2$ (upper jagged line), without drift, $\delta = 0$ (lower jagged line), and a straight line with slope 0.2 (dashed line).

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")
```

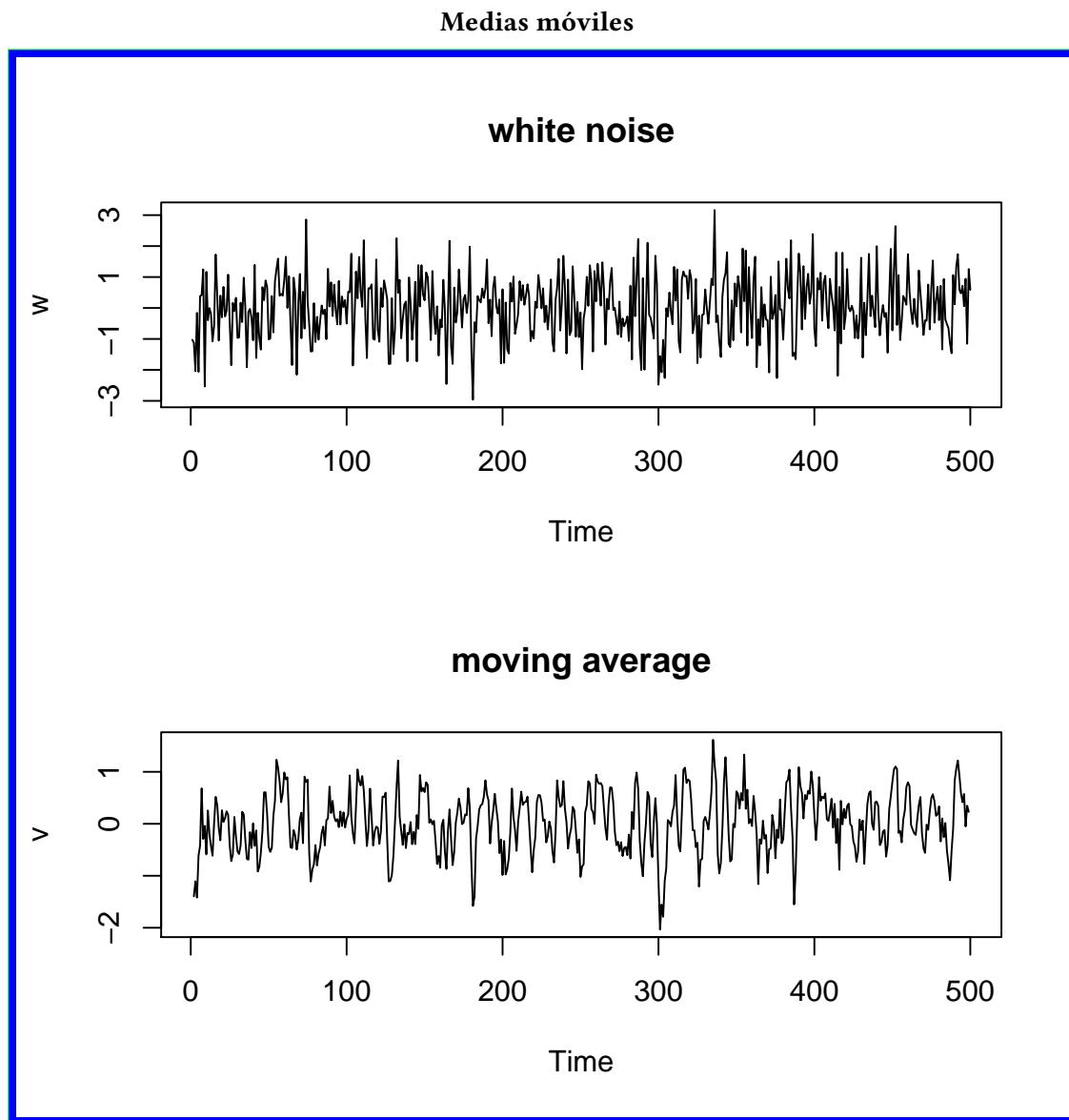


Figure 1.8: Promedios Móviles - Moving Averages

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

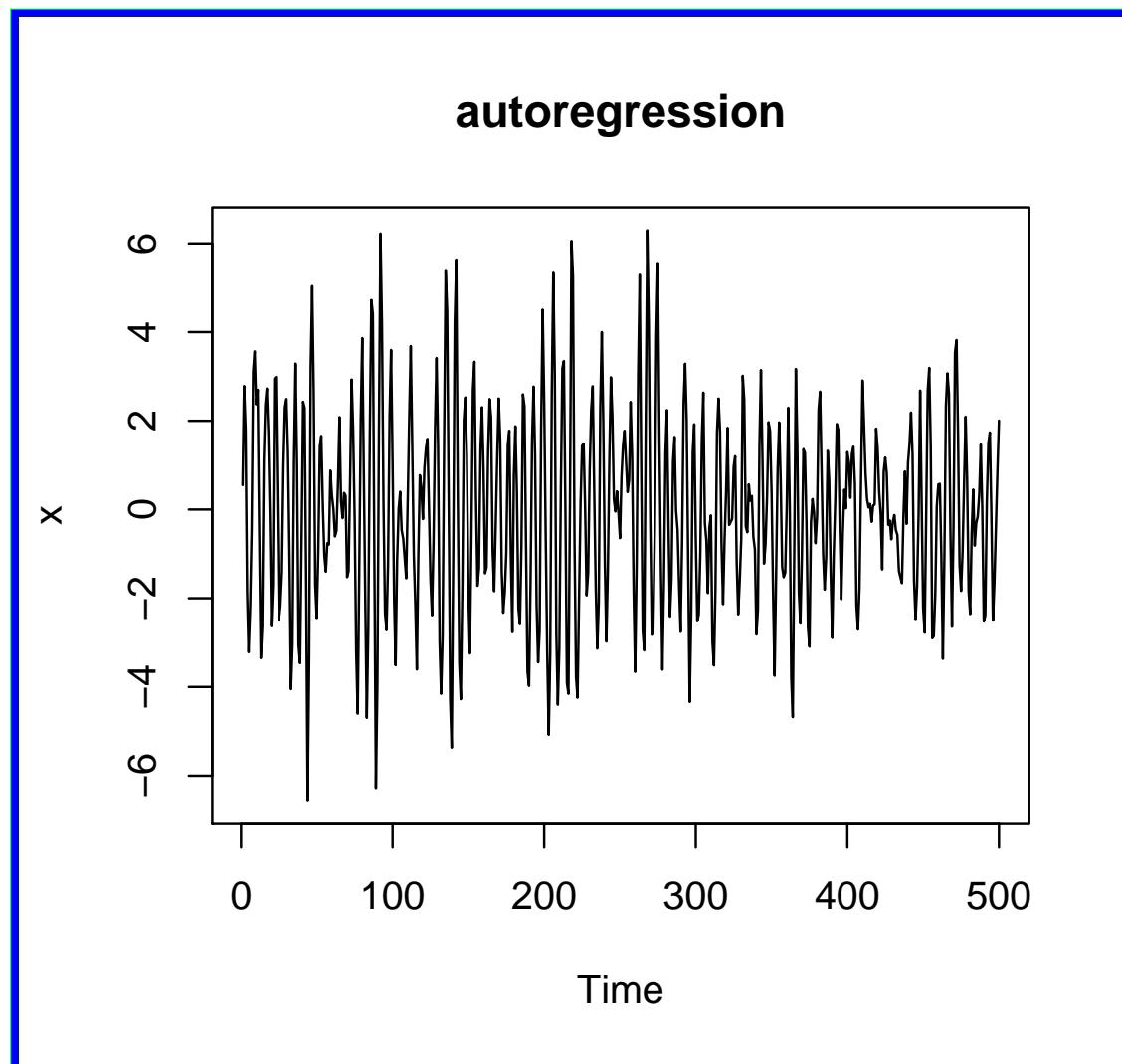


Figure 1.9: Autorregresiones - Autoregressions

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

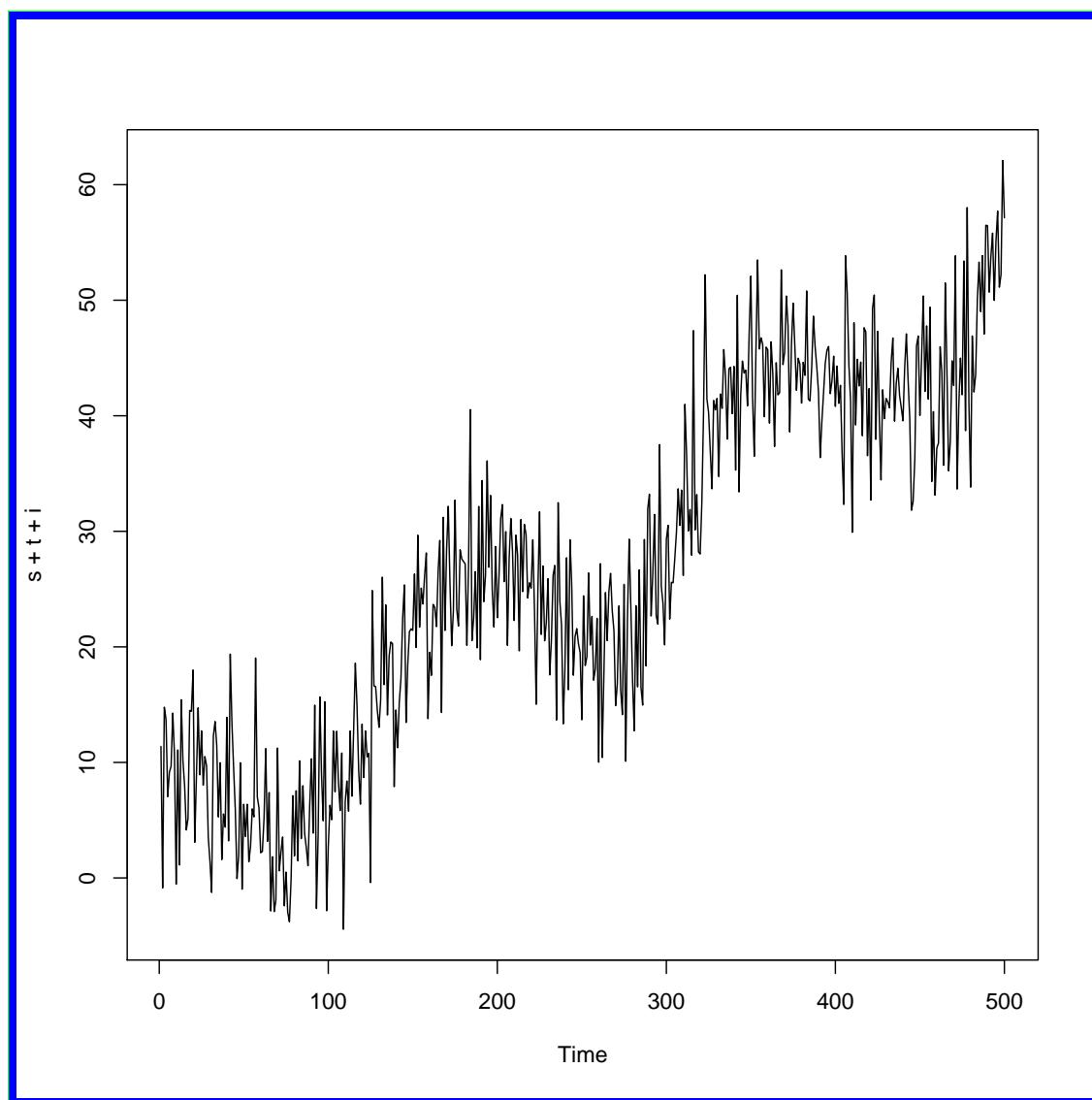


Figure 1.10: Serie de tiempo con sus componentes T, S e I.

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

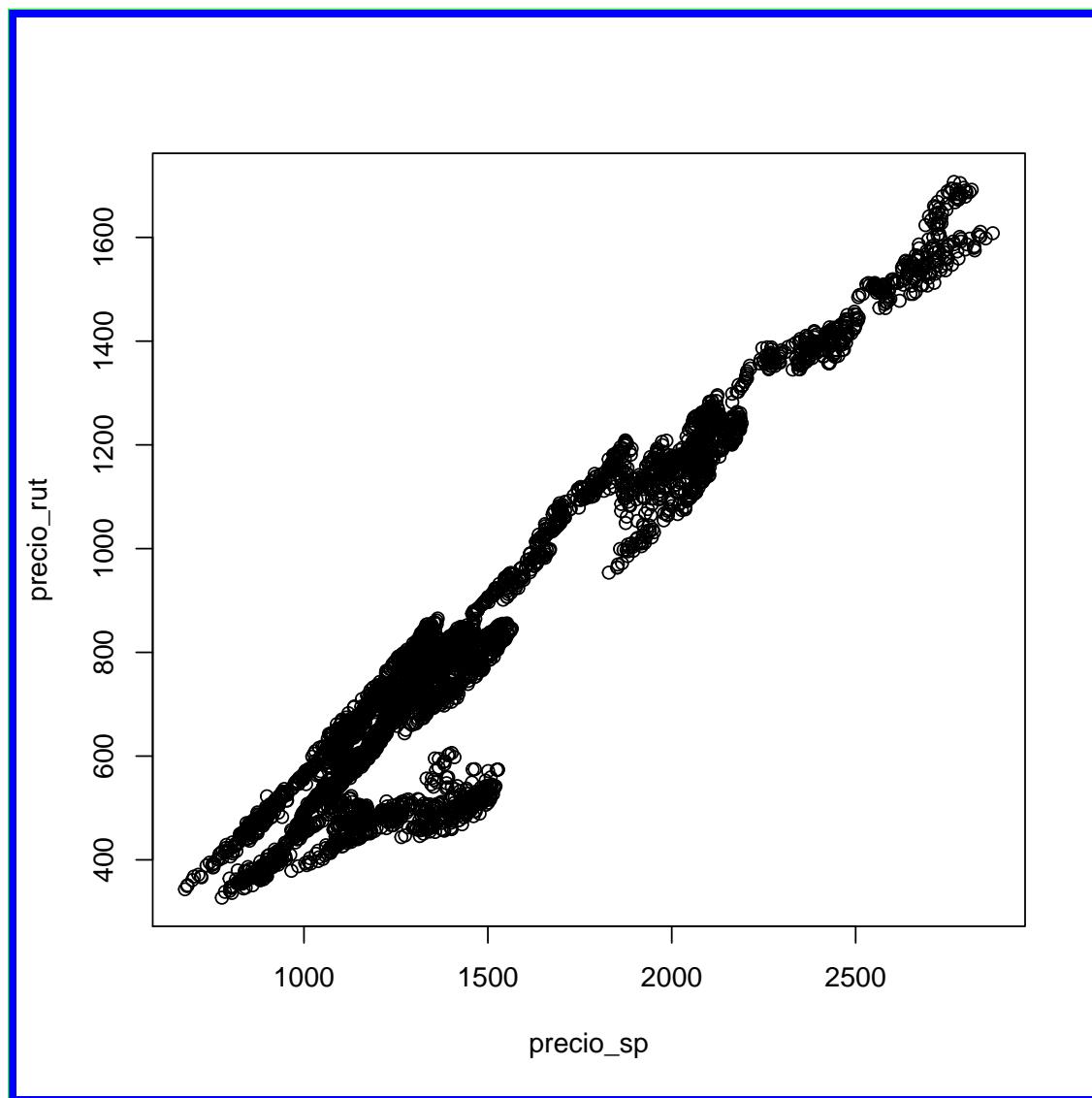


Figure 1.11: Regresión de Russell 2000 en función del S&P500

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

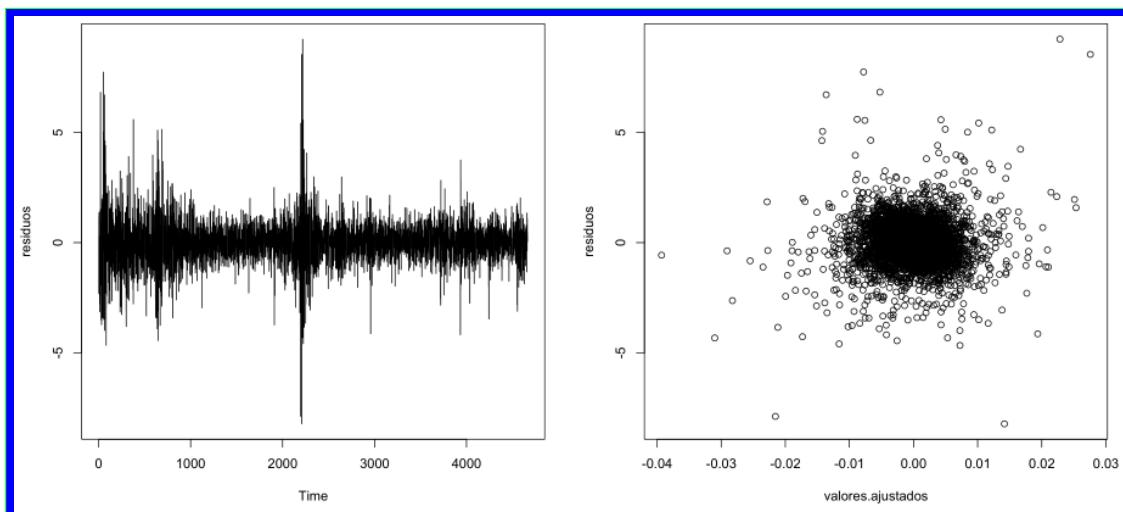


Figure 1.12: Izquierda: Derecha

Normalidad de los residuos

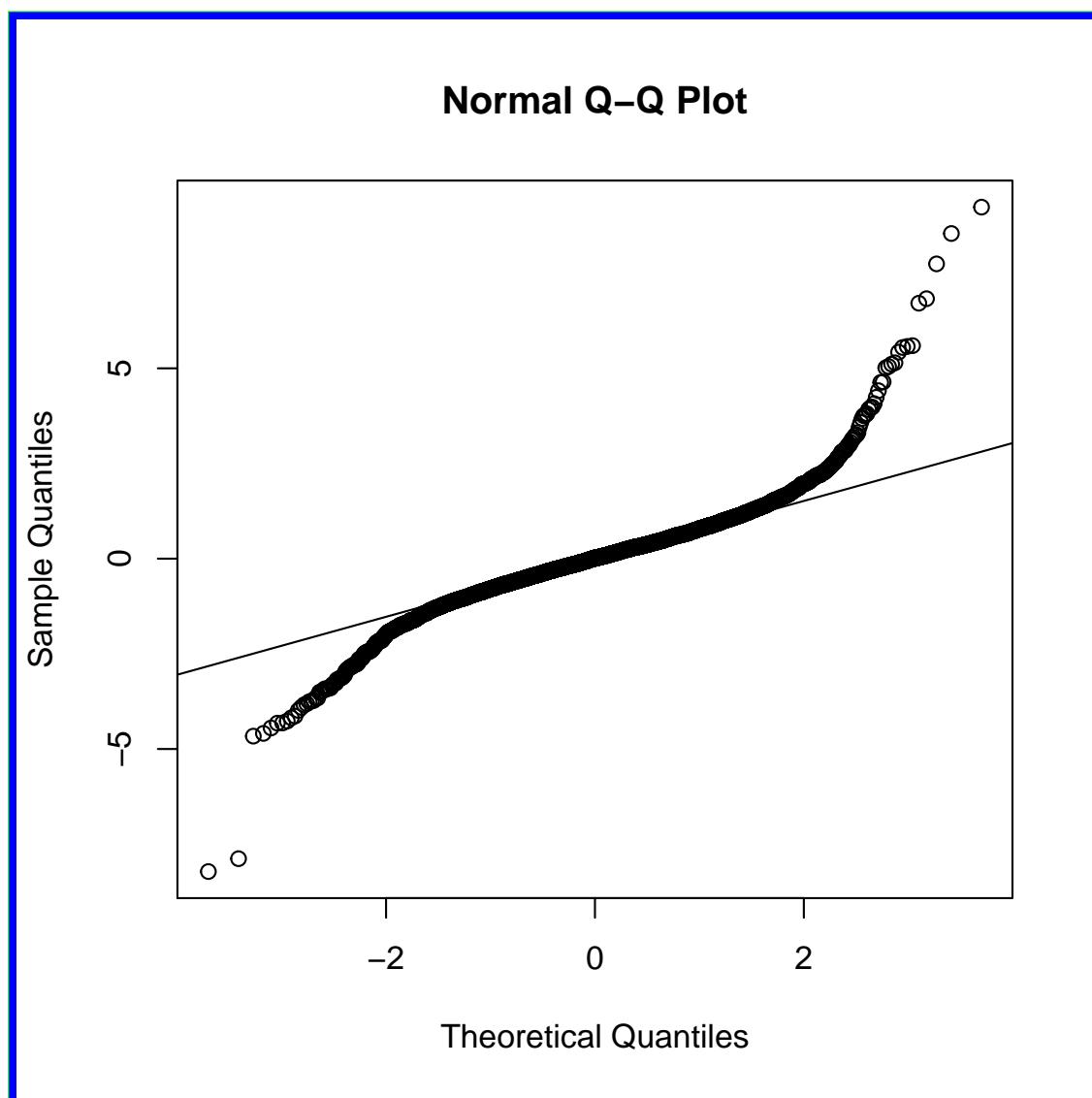


Figure 1.13: Grafico QQ-Plot de los residuos. La linea punteada indica el comportamiento de una distribución normal, mientras los puntos alejados de ella en los extremos dan cuenta de como se alejan los residuos del comportamiento normal.

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)
```

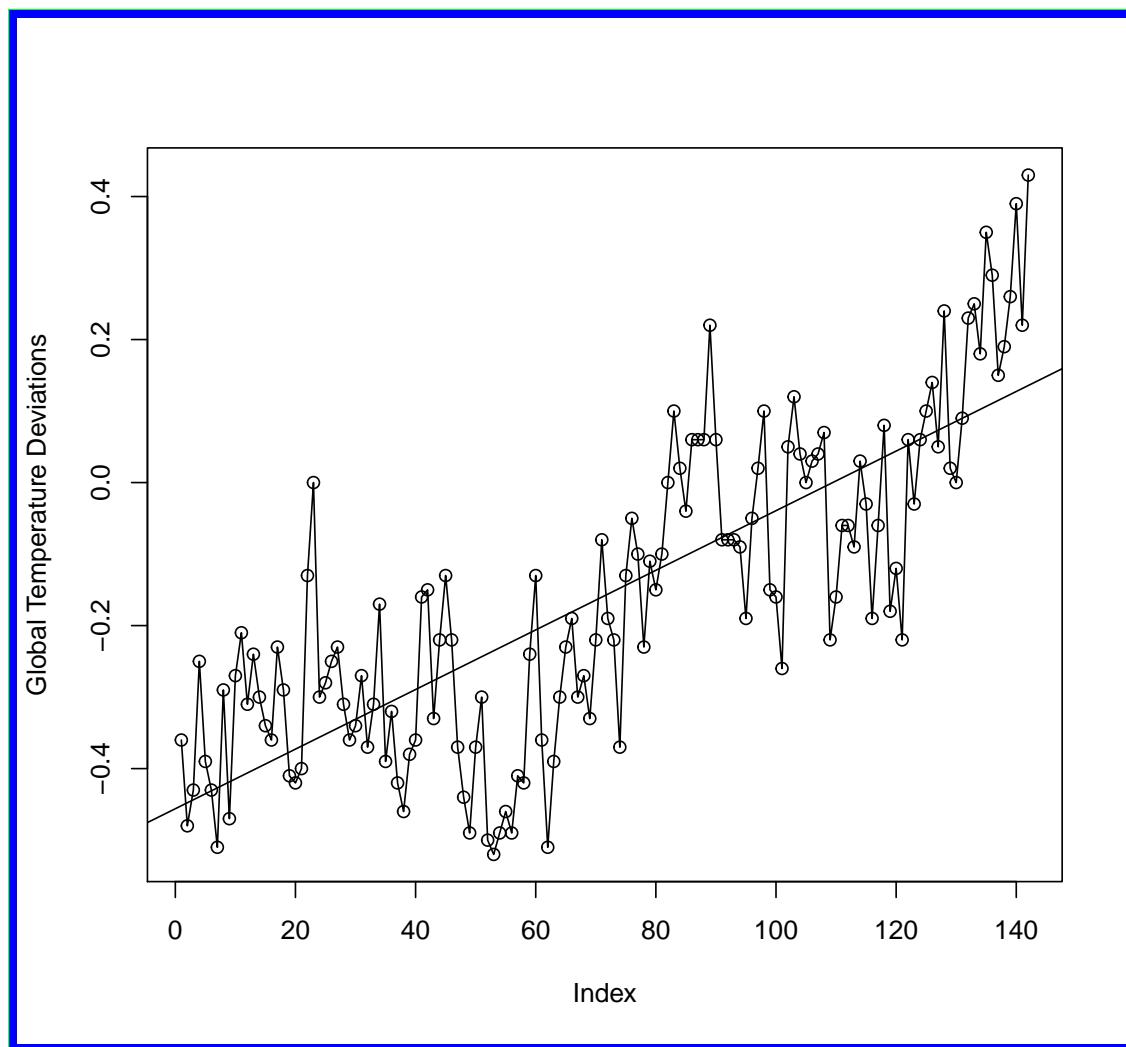
Ejemplo 1: Detrending global temperature

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

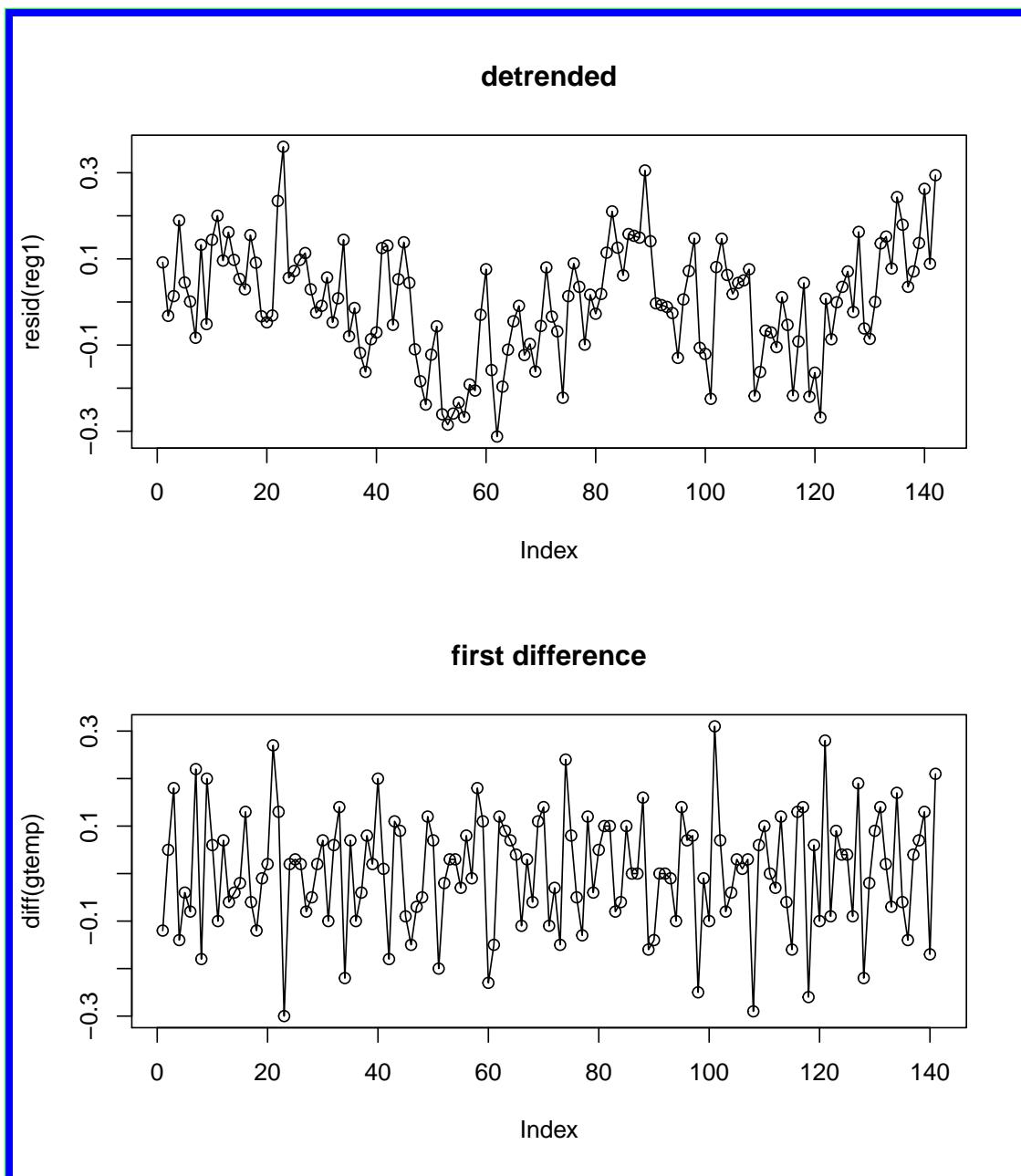


Figure 2.3: Arriba: residuos de la regresión. Abajo: primeras diferencias de la serie original

'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")
```

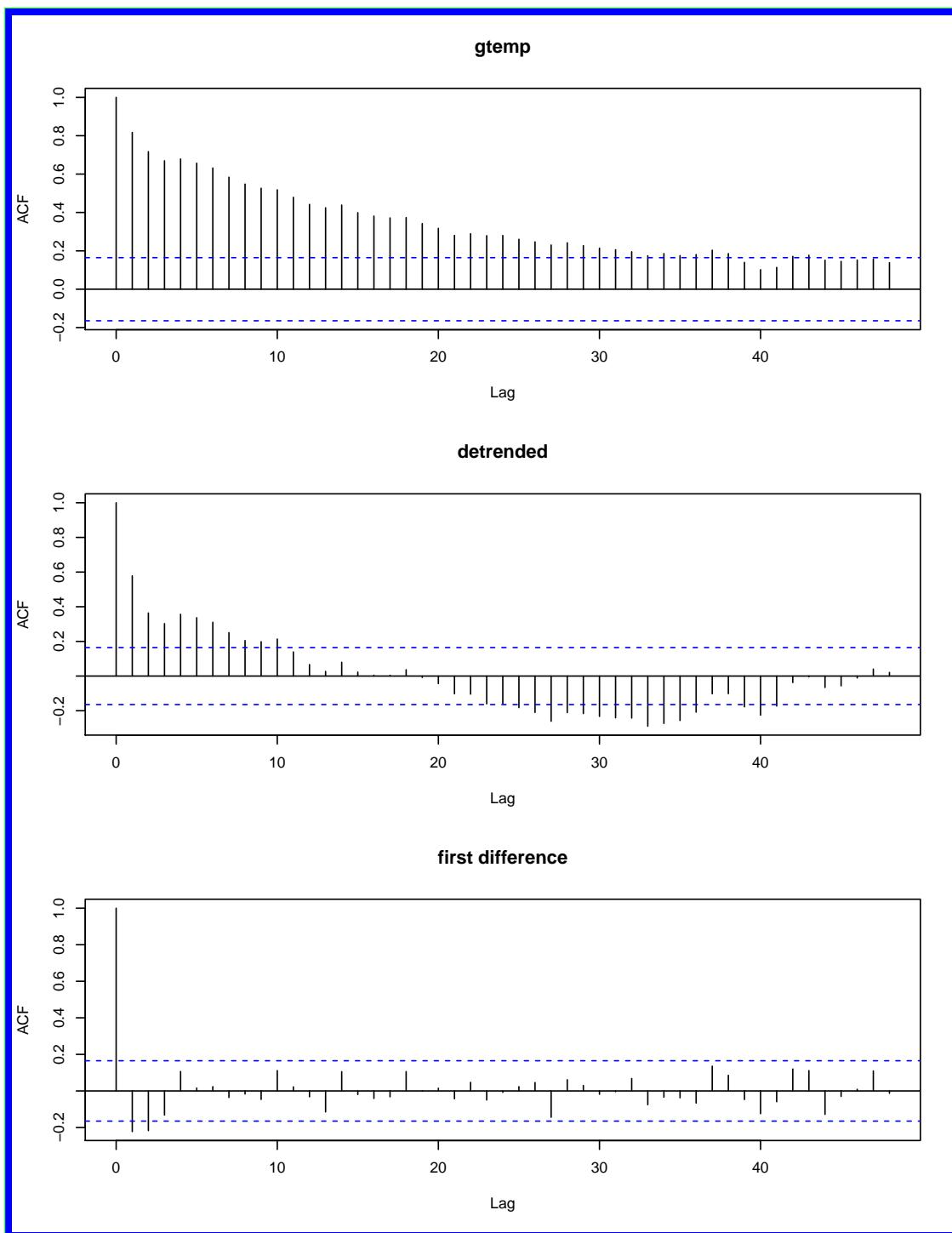
Ejemplo 1: Autocorrelograma

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-Füller 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)

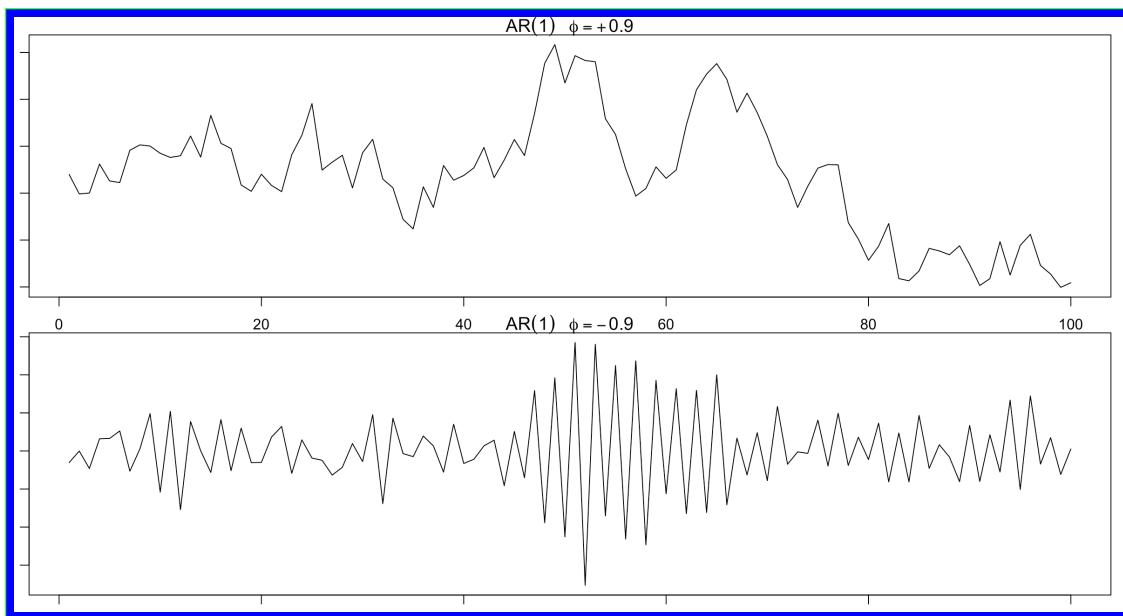


Figure 2.6: Simulaciones procesos AR(1): Arriba: $\phi = +0.9$. Abajo: $\phi = -0.9$

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)

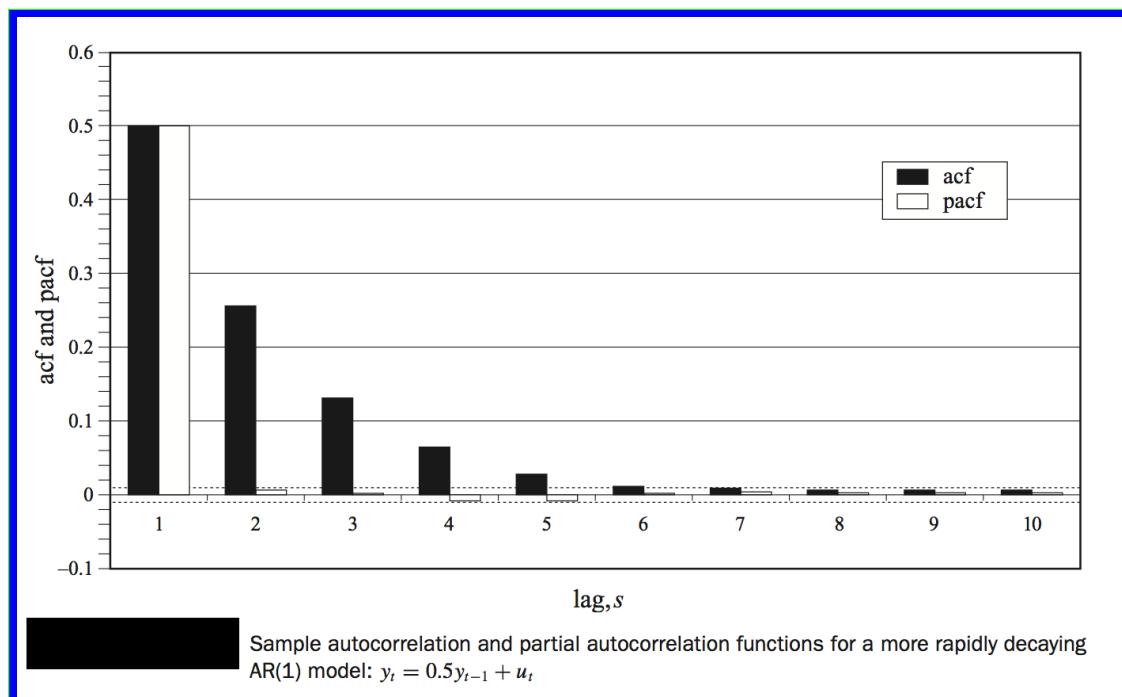


Figure 2.7: Autocorrelogramas para autocorrelación(acf) y autocorrelación parcial(pacf)

Proceso Autoregresivo: AR(1)

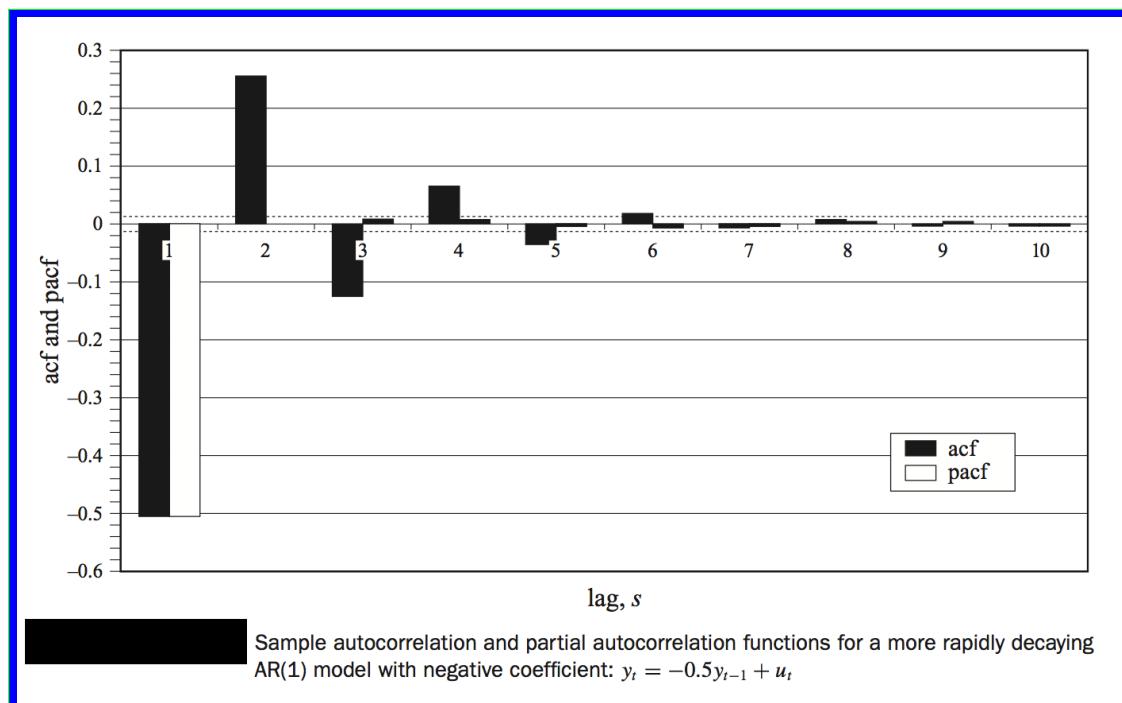


Figure 2.8: Autocorrelogramas para autocorrelación(acf) y autocorrelación parcial(pacf)

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)

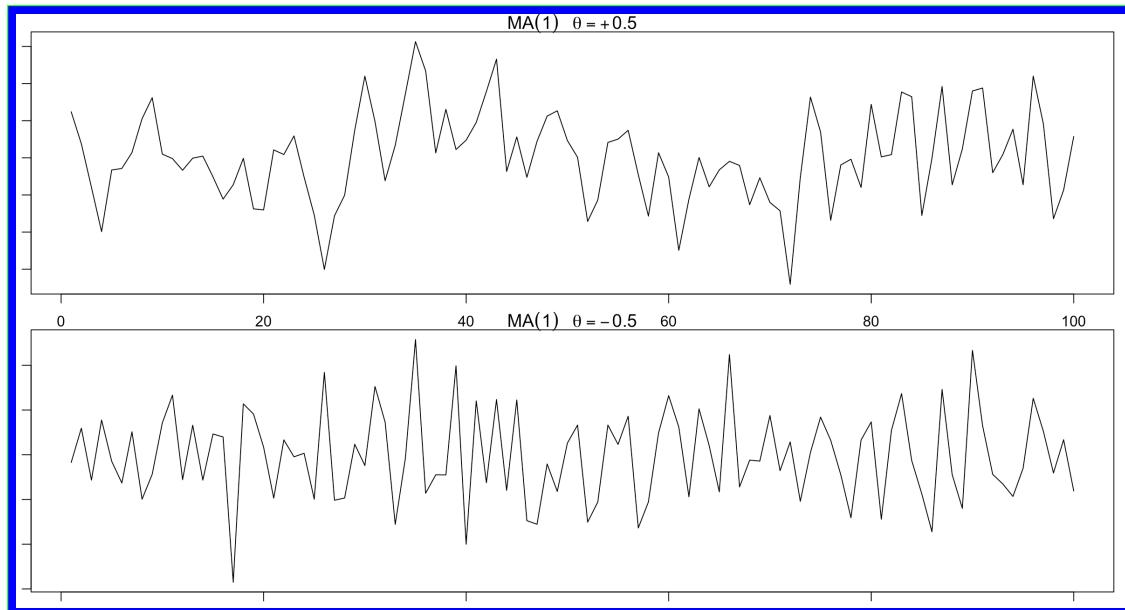


Figure 2.9: Procesos de medias móviles. Arriba: parámetro $\theta = +0.5$. Abajo: parámetro $\theta = -0.5$

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), $\theta = +0.5$

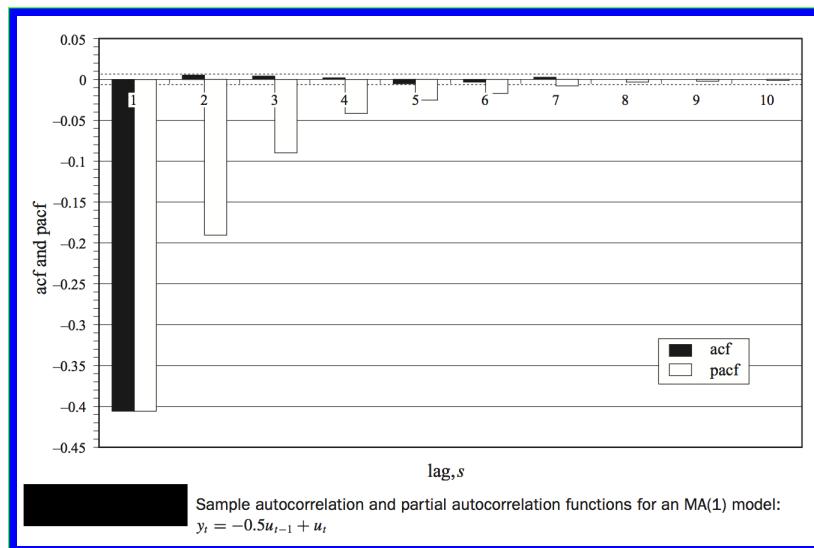


Figure 2.10: Autocorrelogramas para diferentes retardos de proceso de media móvil con $\theta = +0.5$ (lags). ACF(barras negras), PACF(barras blancas).

2.4.6 Identificación modelo ARMA

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

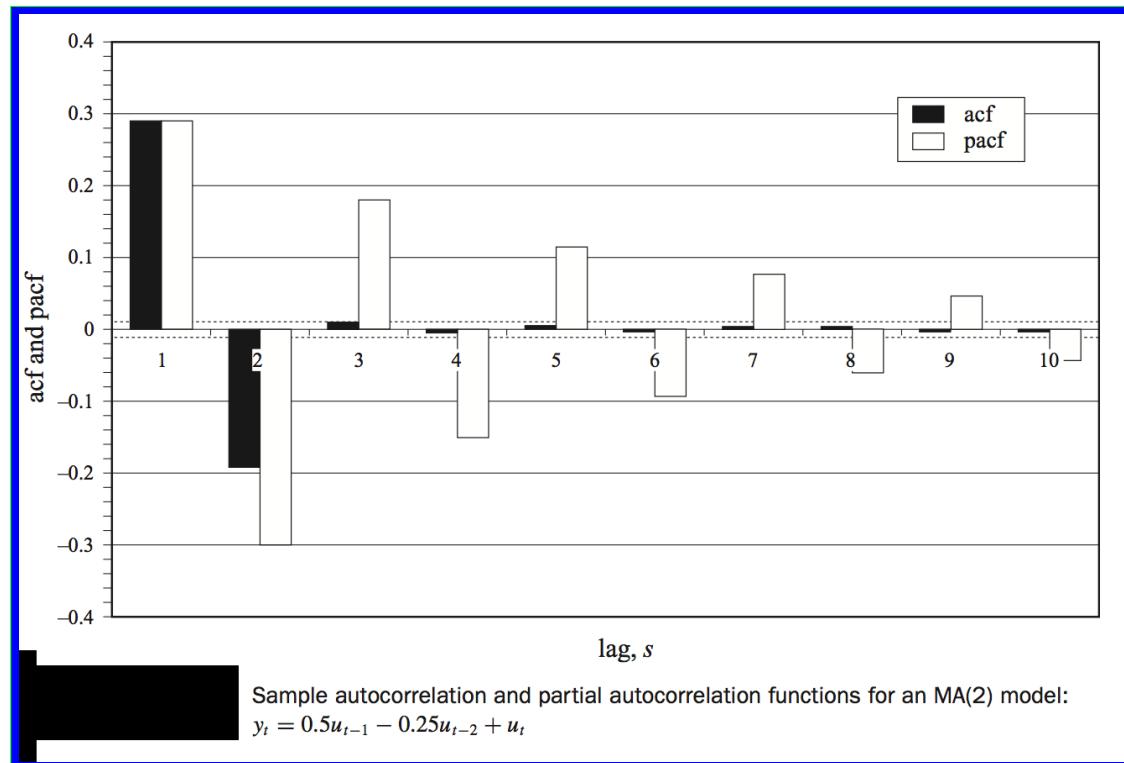


Figure 2.11: Autocorrelogramas para diferentes retardos de proceso de media móvil con $\theta = -0.5$ (lags). ACF(barras negras), PACF(barras blancas).

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)

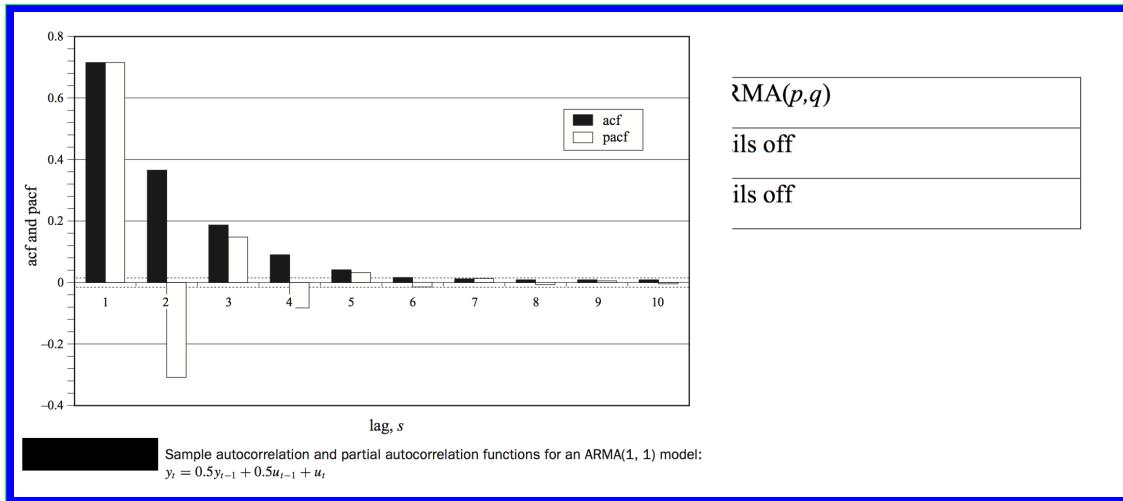


Figure 2.12: Autocorrelogramas para diferentes retardos(lags) para un proceso ARMA(1,1)

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.

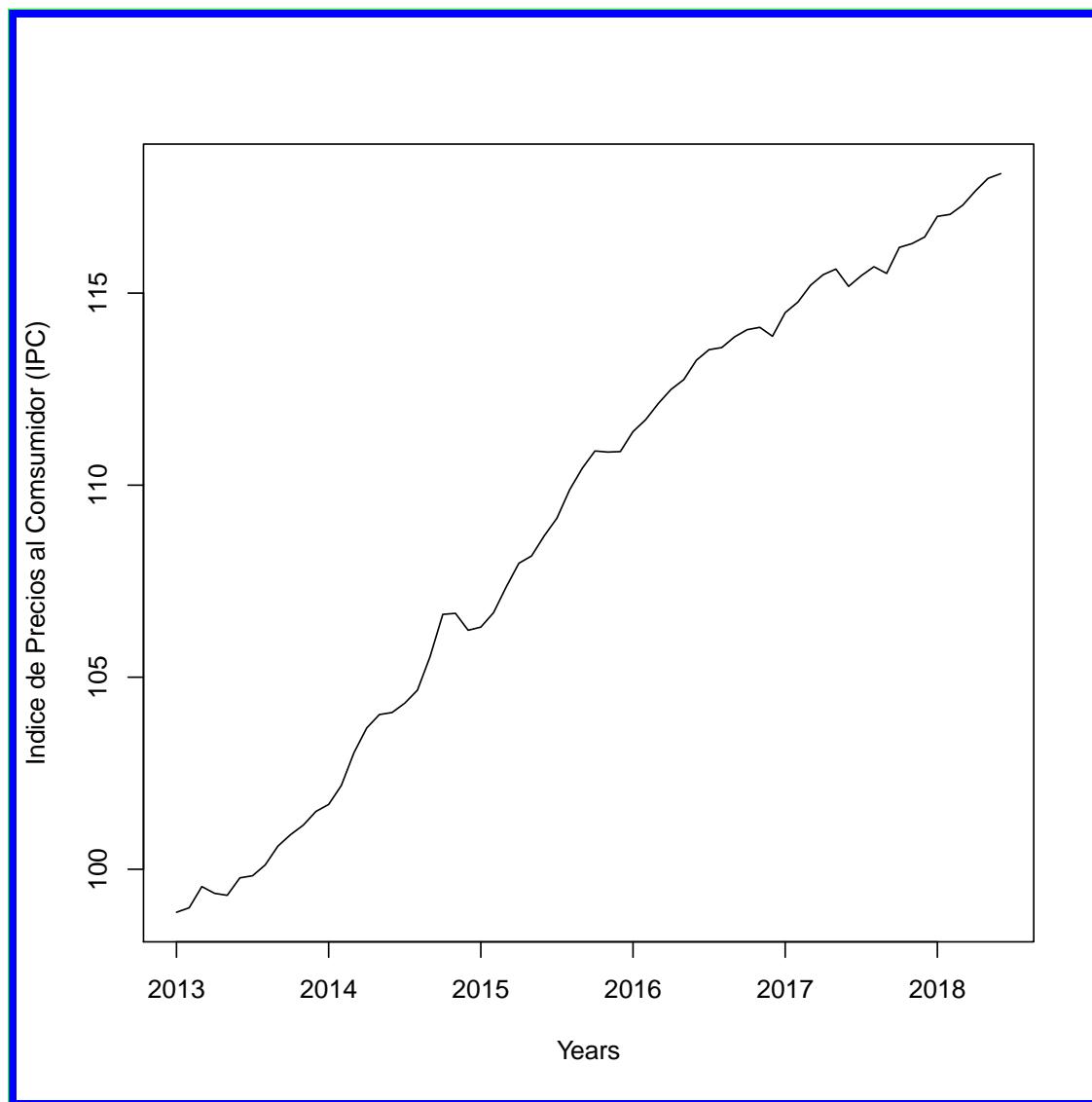


Figure 2.14: Serie de tiempo del índice de precios al consumidor (IPC) en Chile.

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

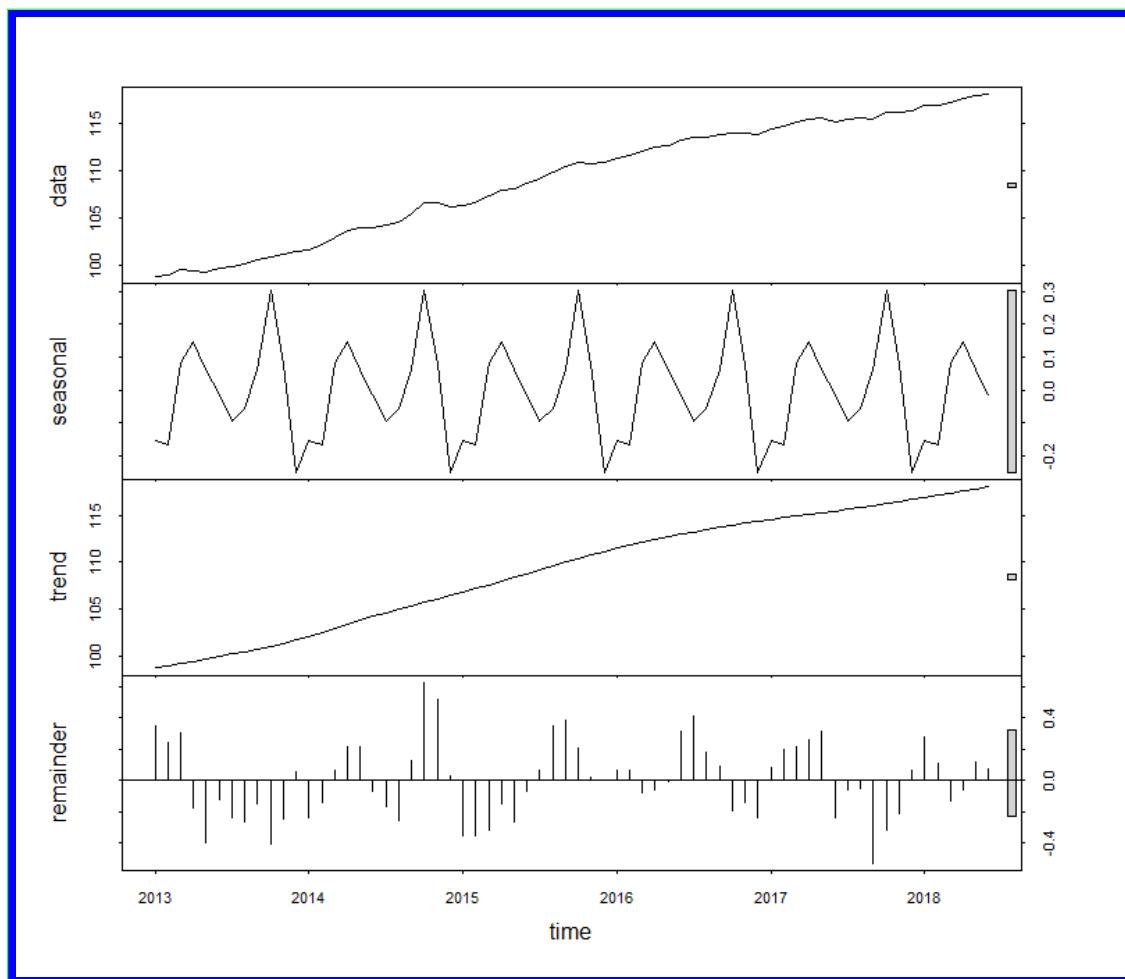


Figure 2.15: Diferentes componentes de la serie: data: serie original de IPC; seasonal: componente estacional; trend: tendencia de la serie; remainder: componentes irregulares.

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)

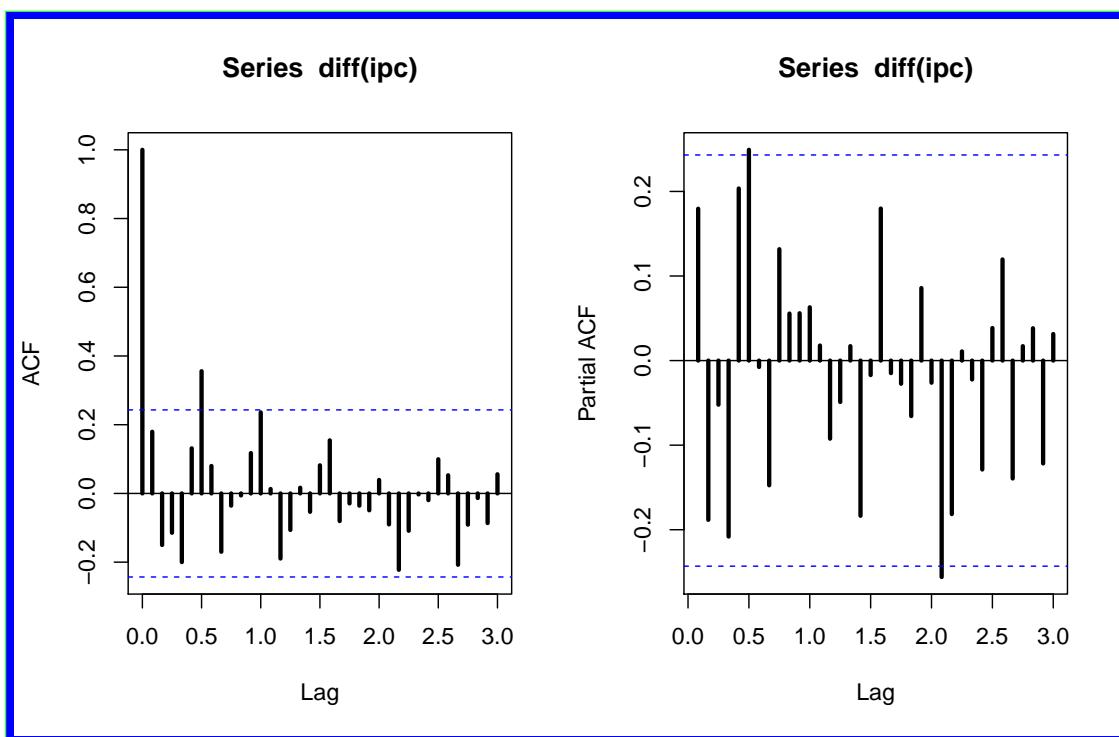


Figure 2.17: Izquierda: autocorrelacion de serie de ipc diferenciada. Derecha: autocorrelación parcial de serie de IPC diferenciada.

'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

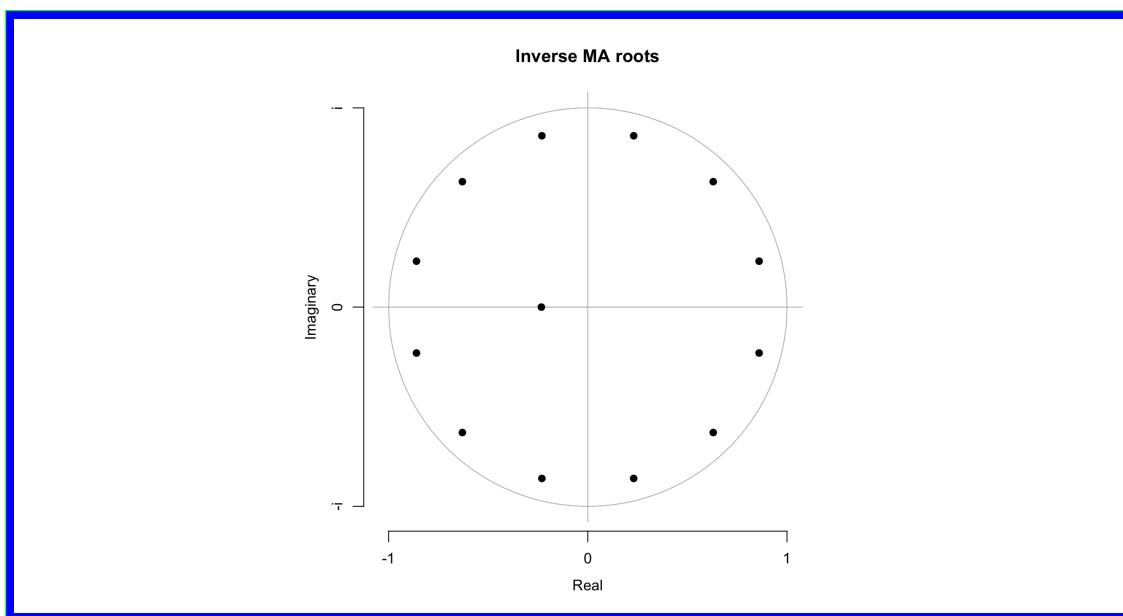


Figure 2.21: Soluciones de ecuación característica - Representación de circulo unitario

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

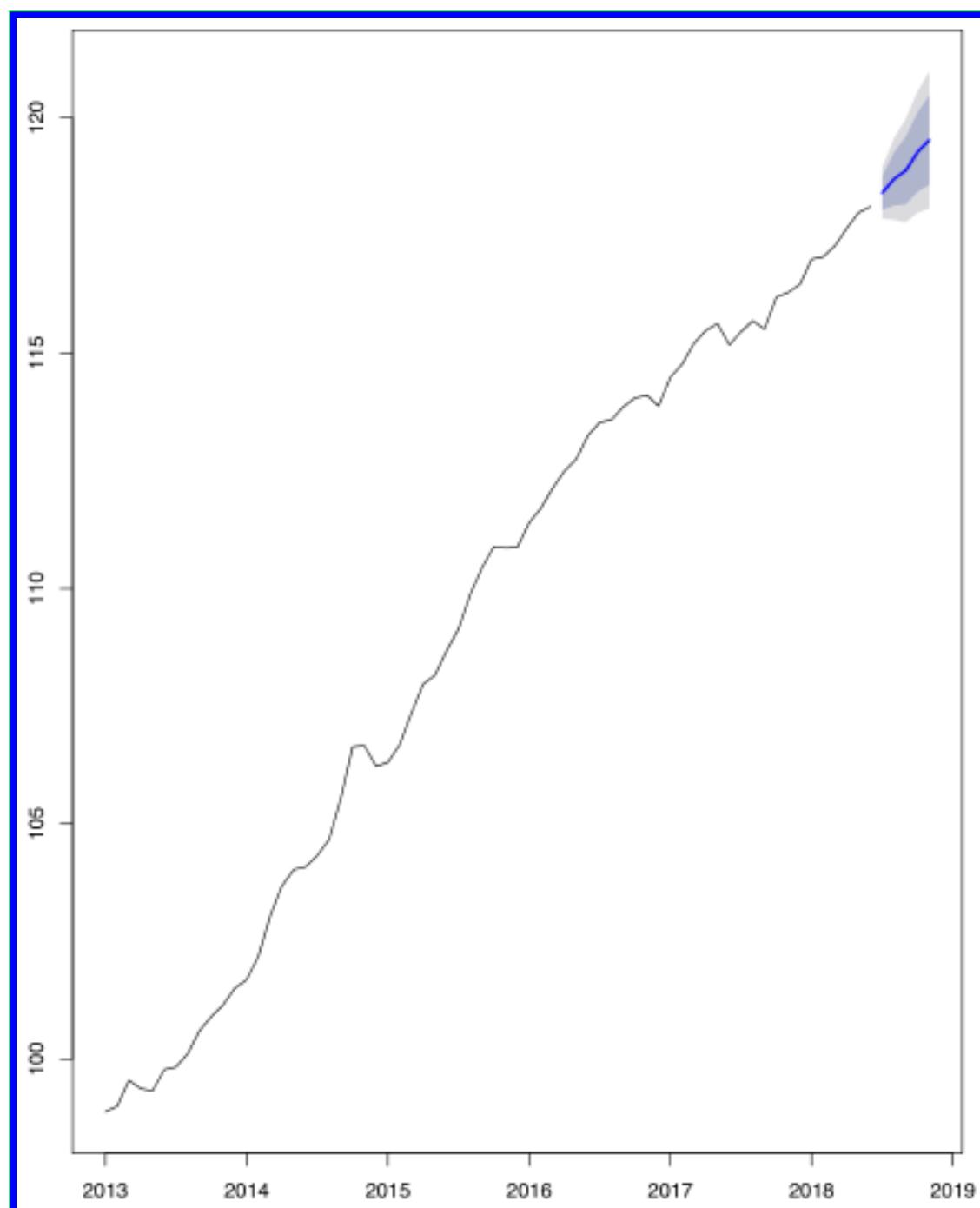


Figure 2.22: Ejemplo IPC en Chile: Forecast de modelo auto.arima

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

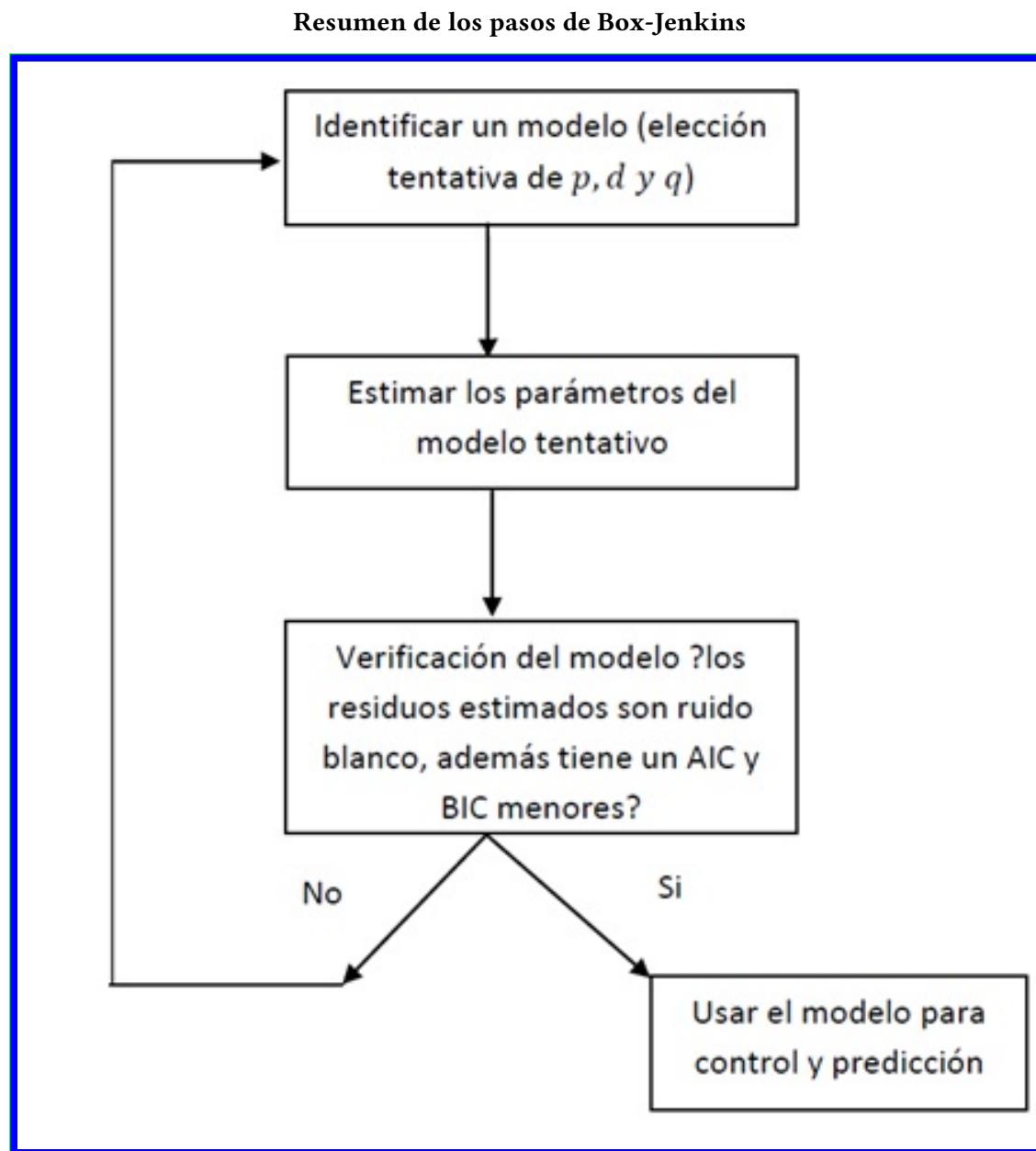


Figure 2.23: Metodología de Box-Jenkins

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

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

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

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

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))
```

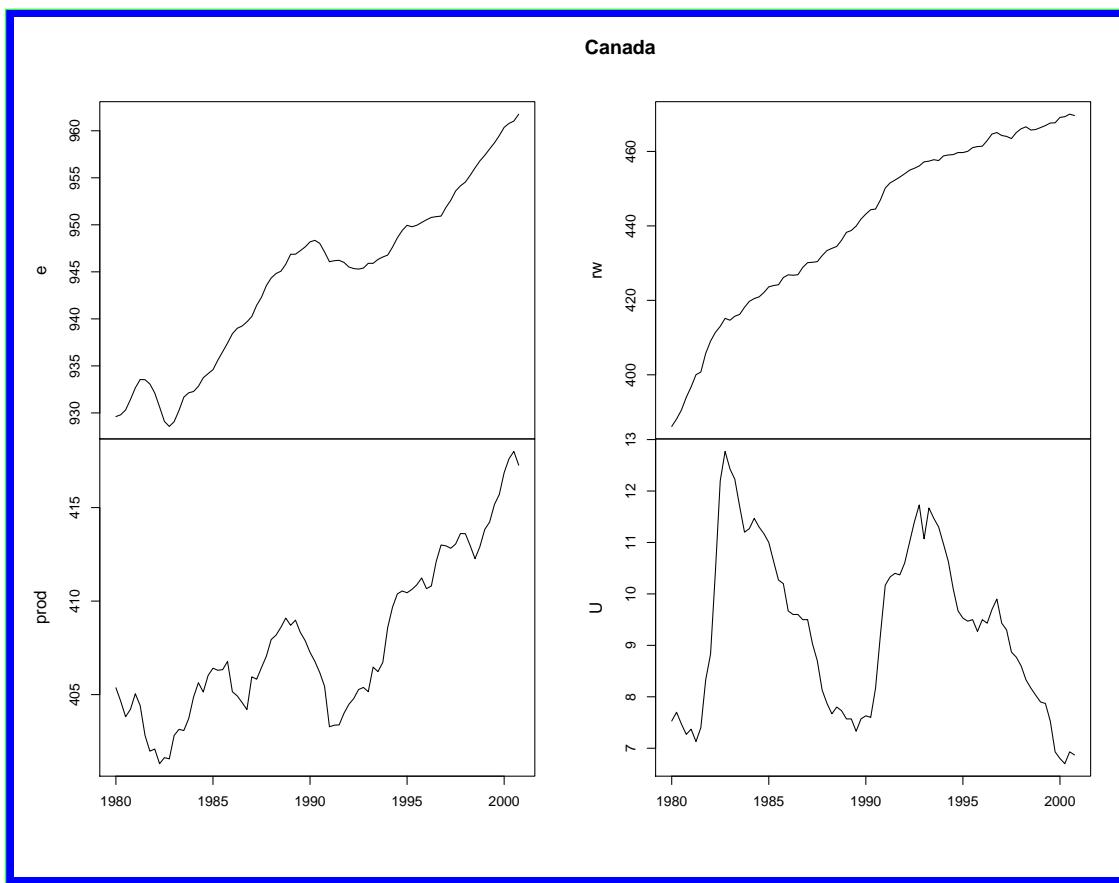
Ejemplo Modelo VAR: Gráficos de las series de tiempo de cada variable

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

	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”.

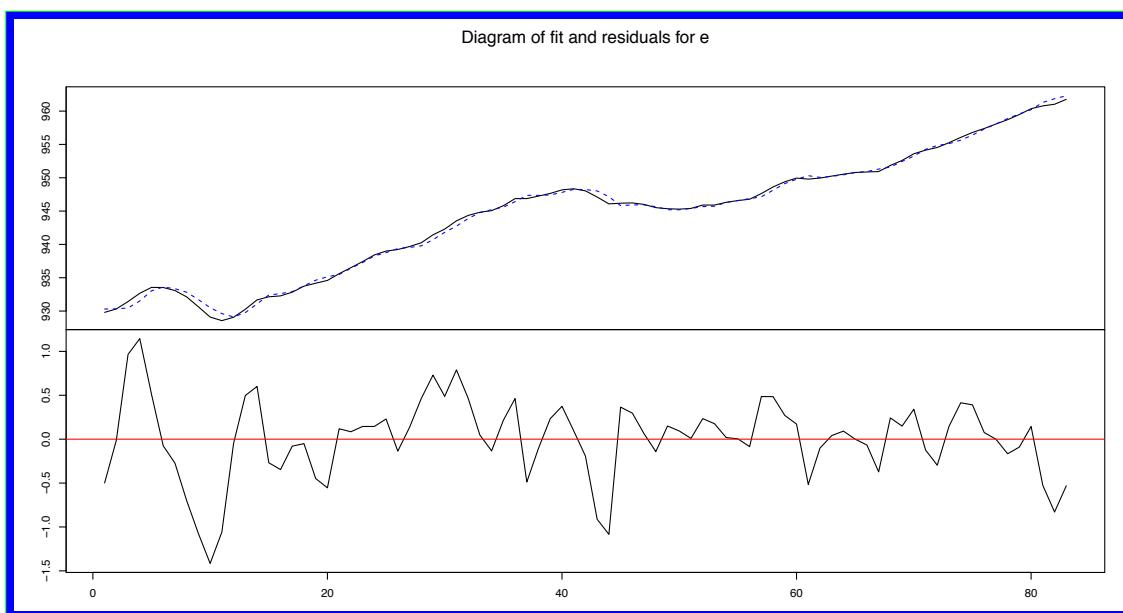
Ejemplo Modelo VAR - instrucción VAR para e

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”

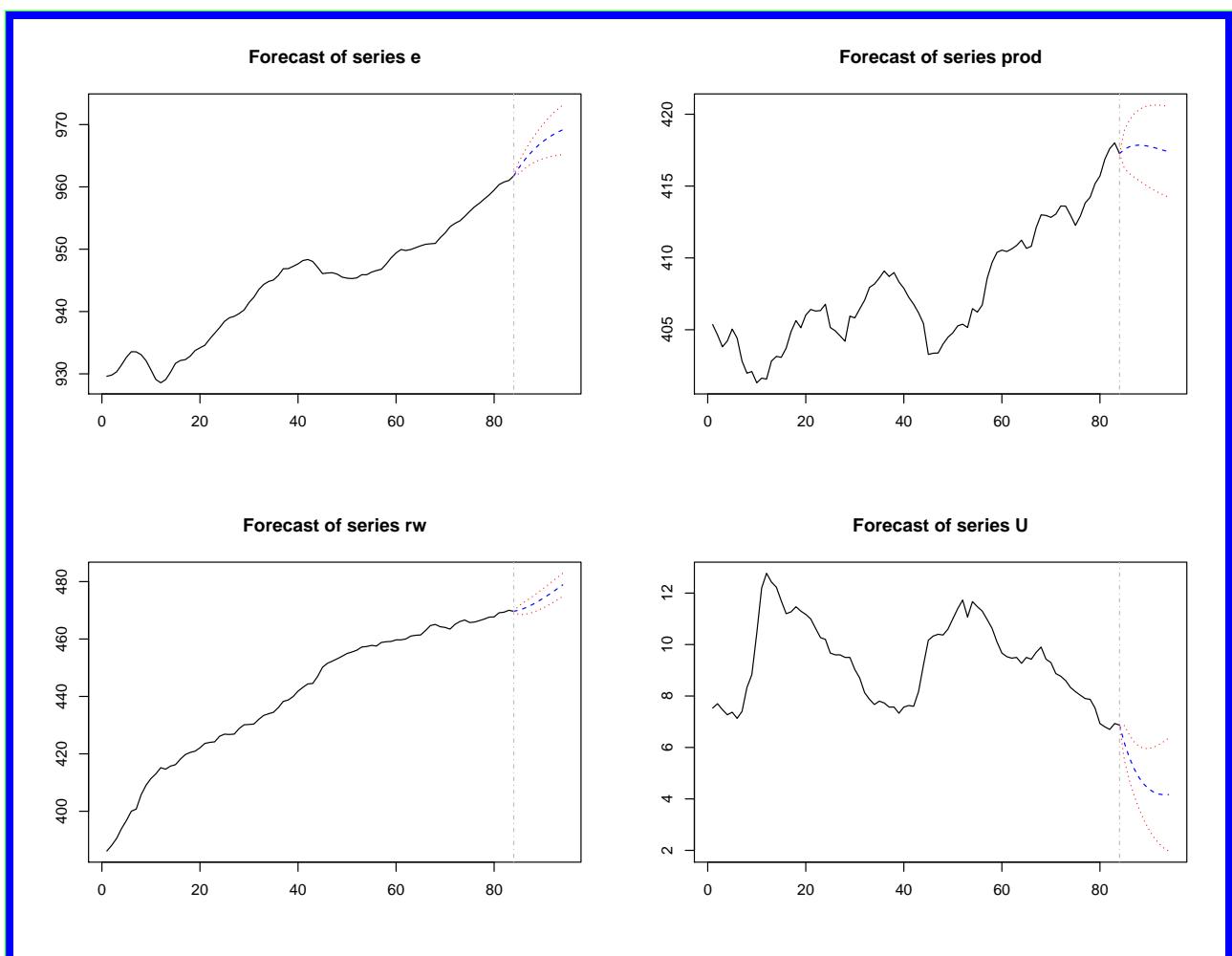


Figure 3.12: Pronosticos de modelo VAR.

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

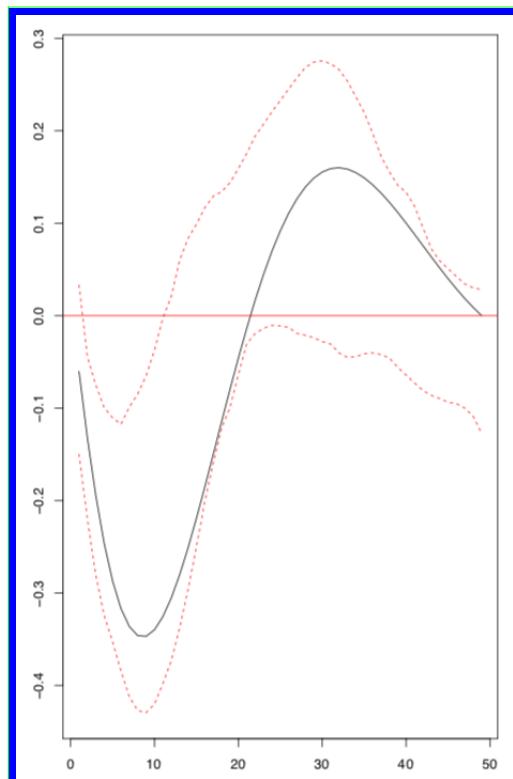


Figure 3.13

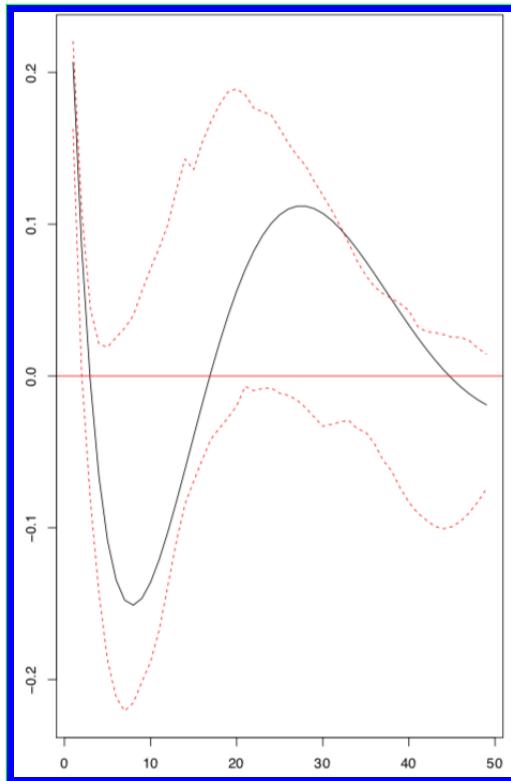
Ejemplo Modelo VAR - fevd

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 párametros.
- 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.

4 Tópico IV.- Cointegration Analysis

4.1 Cointegración

4.1.1 Introducción

Si dos o más series temporales no estacionarias siguen un camino común (o de equilibrio) a largo plazo, podemos hablar de cointegración. La prueba clásica de cointegración se reduce a determinar si una combinación lineal de la serie es estacionaria o no. Si, por ejemplo, dos series temporales están cointegradas por un factor común (vector de cointegración), no es posible utilizar un enfoque VAR estándar. Tenemos que dar cuenta de esta relación y utilizar un modelo de corrección de errores para obtener resultados correctos.

Suponga que $Y_t = I(1)$ y $X_t = I(1)$. Entonces Y_t y X_t están cointegradas, $CI(1, 1)$, si existe un β , tal que $Y_t - \beta X_t = \epsilon_t = I(0)$. Esto implica que hay una relación de largo plazo entre Y_t y X_t , es decir ellas no se “separan” a través del tiempo. De aquí la relación, $Y_t = \beta X_t + \epsilon_t$, tiene sentido.

Si Y_t y X_t no están cointegradas, es decir, ϵ_t también es $I(1)$, entonces, Y_t y X_t se separarán cada vez más a través del tiempo, y por lo tanto no existirá una relación de largo plazo entre dichas variables. Cualquier regresión de Y_t sobre X_t es “Espuria”.

En general, si Y_t y X_t son ambas $I(d)$, entonces Y_t y X_t son $CI(d, b)$ si $Y_t - \beta X_t = \epsilon_t = I(d - b)$, $b > 0$.

Si si Y_t y X_t están cointegradas, esto significa que es posible modelar la relación de largo plazo entre Y_t y X_t . Esta es una estrategia de modelamiento alternativa a eliminar la tendencia a través de diferenciación. El procedimiento de eliminar la tendencia en general, pierde información. Por lo tanto, es posible aplicar un procedimiento de “Cointegración de temporada” más que tratar de eliminar el efecto de temporada diferenciando.

4.1.2 Test de Cointegración de Engle-Granger

Estimar la regresión cointegrada, $Y_t = \beta X_t + \epsilon_t = I(0)$ usando MCO para obtener los residuos ϵ_t . Aplicar el test de Dickey-Fuller (DF) y/o el test de Dickey-Fuller Aumentado (ADF) para examinar si los residuos tienen raíces unitarias, ver [Engle and Granger \(1987\)](#).

Si la hipótesis de raíces unitarias no es rechazada, esto implica que ϵ_t es $I(1)$, lo cual implica que Y_t y X_t NO están cointegradas. Es importante notar que los valores críticos para los test DF y ADF no son validos para ser usados para Cointegración. Engle y Granger han calculado los valores críticos apropiados.

4.1.3 Modelo de Corrección de Errores

Si Y_t y X_t están cointegradas, entonces hay una relación de largo plazo entre estas series y la dinámica de corto plazo de esta relación puede ser descrita por el modelo de corrección de errores (ECM).

Relación de largo plazo:

$$Y_t = \beta X_t + \epsilon_t \quad (4.1)$$

El modelo de corrección de errores, es decir la dinámica de corto plazo viene dada por:

$$\Delta Y_t = \alpha \Delta X_t + \phi[Y_{t-1} - \beta X_{t-1}] + \epsilon_t \quad (4.2)$$

donde ϵ_t = white noise, es decir $I(0)$.

Interpretación: el cambio actual en Y_t consiste de dos componentes:

- (i) $\alpha \Delta X_t$: la respuesta de corto plazo a los cambios actuales en X_t , y
- (ii) $\phi[Y_{t-1} - \beta X_{t-1}]$: la corrección parcial de la desviación previa de Y_t de su nivel deseado de largo plazo.

Procedimiento de Estimación

El procedimiento de dos pasos de Engle-Granger

- (i) Estimar la regresión de Cointegración para obtener una estimación del parámetro de largo plazo, y entonces,
- (ii) Usar los residuos para estimar el modelo de corrección de errores.

Cointegration in practice Ejemplo Modelo IPSA - CU - S&P

Analizaremos la relación de largo plazo entre la bolsa chilena, el IPSA, la bolsa americana, el S&P 500, y el precio del cobre, cu. En particular probaremos la hipótesis de cointegración. Primero presentamos el test de estacionaridad para las tres variables.

‘Código R: Ejemplo Modelo IPSA - CU - S&P

```
library(tseries)# df, adf
library(dynlm)
mydata<-read.csv("cointegration.csv")
ipsa <- ts(mydata$IPSA,frequency=12, start = c(2010,3))
cu <- ts(mydata$CU,frequency=12, start = c(2010,3))
sp <- ts(mydata$s.p,frequency=12, start = c(2010,3))
summary(mydata)
adf.test(diff(ipsa));
adf.test(diff(sp));
adf.test(diff(cu))
```

Ejemplo Modelo IPSA - CU - S&P

```
> adf.test(diff(ipsa))

Augmented Dickey-Fuller Test

data: diff(ipsa)
Dickey-Fuller = -3.1289, Lag order = 3, p-value = 0.1304
alternative hypothesis: stationary

> adf.test(diff(sp))

Augmented Dickey-Fuller Test

data: diff(sp)
Dickey-Fuller = -3.1821, Lag order = 3, p-value = 0.1095
alternative hypothesis: stationary

> adf.test(diff(cu))

Augmented Dickey-Fuller Test

data: diff(cu)
Dickey-Fuller = -2.8209, Lag order = 3, p-value = 0.2513
alternative hypothesis: stationary
```

Ahora ejecutamos el modelo de cointegración de dos etapas Engle-Granger. Primero buscamos una relación de largo plazo de las variables, y encontramos que la relación IPSA y CU es la más fuerte.

‘Código R:

```
ipsa.reg1 <- dynlm(ipsa ~ cu + sp)
summary(ipsa.reg1)
ipsa.reg2 <- dynlm(ipsa ~ L(cu, 1:3))
summary(ipsa.reg2)
ipsa.reg3 <- dynlm(ipsa ~ cu)
summary(ipsa.reg3)
residuos <- ipsa.reg3[["residuals"]]
plot(residuos);
```

Ejemplo Modelo IPSA - CU - SP

```
Time series regression with "ts" data:  
Start = 2010(1), End = 2013(5)  
  
Call:  
dynlm(formula = ipsa ~ cu + sp)  
  
Residuals:  
    Min      1Q  Median      3Q     Max  
-8.2716 -3.4712  0.0346  2.6565  7.7781  
  
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) 29.811617   7.095477   4.201 0.000155 ***  
cu          0.922684   0.108395   8.512 2.46e-10 ***  
sp          0.003932   0.004235   0.928 0.359016  
---  
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Residual standard error: 4.199 on 38 degrees of freedom  
Multiple R-squared:  0.6581,    Adjusted R-squared:  0.6401  
F-statistic: 36.58 on 2 and 38 DF,  p-value: 1.391e-09
```

Ejemplo Modelo IPSA - CU

```
Time series regression with "ts" data:  
Start = 2010(1), End = 2013(5)  
  
Call:  
dynlm(formula = ipsa ~ cu)  
  
Residuals:  
    Min      1Q  Median      3Q     Max  
-8.9276 -3.2706  0.1624  2.8326  7.0641  
  
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) 35.6604    3.2600 10.939 1.90e-13 ***  
cu          0.9021    0.1059  8.518 1.96e-10 ***  
---  
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Residual standard error: 4.192 on 39 degrees of freedom  
Multiple R-squared:  0.6504,   Adjusted R-squared:  0.6414  
F-statistic: 72.55 on 1 and 39 DF,  p-value: 1.957e-10
```

```
Augmented Dickey-Fuller Test
```

```
data: residuos
Dickey-Fuller = -3.0126, Lag order = 3, p-value = 0.1753
alternative hypothesis: stationary
```

(a)

```
Time series regression with "ts" data:
Start = 2010(1), End = 2013(5)

Call:
dynlm(formula = ipsa ~ cu)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.9276 -3.2706  0.1624  2.8326  7.0641 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 35.6604    3.2600 10.939 1.90e-13 ***
cu          0.9021    0.1059  8.518 1.96e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.192 on 39 degrees of freedom
Multiple R-squared:  0.6504,    Adjusted R-squared:  0.6414 
F-statistic: 72.55 on 1 and 39 DF,  p-value: 1.957e-10
```

(b)

Figure 4.1: (a) Test de Dickey-Füller aumentado. (b) Grafica Q-Q sobre los residuos.

El ECM usando una regresión dinámica

'Código R: Modelo de corrección de errores(ECM) con regresión dinámica

```
ipsa_reg4 <- dynlm(diff(ipsa) ~ diff(cu) + lag(residuos))
summary(ipsa_reg4)
```

Modelo ECM para el IPSA, en función de CU y de los residuos

```
Time series regression with "ts" data:
Start = 2010(2), End = 2013(4)

Call:
dynlm(formula = diff(ipsa) ~ diff(cu) + lag(residuos))

Residuals:
    Min      1Q  Median      3Q     Max 
-5.4594 -1.4556 -0.2161  1.6634  4.5257 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.1216    0.4236   0.287   0.7757    
diff(cu)     0.9680    0.1288   7.514 7.03e-09 ***
lag(residuos) 0.2994    0.1131   2.648   0.0119 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.621 on 36 degrees of freedom
Multiple R-squared:  0.6679,    Adjusted R-squared:  0.6495 
F-statistic: 36.21 on 2 and 36 DF,  p-value: 2.408e-09
```

4.1.4 Procedimiento de Estimación

Testing for and Estimating Cointegrating Systems Using the Johansen Technique Based on VARs

El test de Johansen ver [Johansen \(1988\)](#), es una prueba de cointegración que permite más de una relación de cointegración, a diferencia del método de Engle-Granger.

Hay dos tipos de prueba de Johansen, ya sea con traza (trace) o con valor propio (eigenvalue), y las inferencias que pueden ser un poco diferente.

Esta prueba se basa en la estimación de máxima verosimilitud y dos estadísticos: valores propios máximos y una estadística de seguimiento. Esto está relacionado con el rango de la matriz. Si el rango es cero, no hay una relación de cointegración. Si el rango es uno, hay uno, si son dos hay dos y así sucesivamente.

En R este procedimiento está automatizado como veremos... El ECM usando una regresión dinámica

‘Código R:Ejemplo Modelo IPSA - CU - S&P: procedimiento de Johansen

```
library(urca)
a<-data.frame(ipsa,cu)
cointegration <- ca.jo(a, type="trace", ecdet="trend", spec="transitory")
summary(cointegration)
```

Cointegration in practice-Procedimiento de Johansen

```
# Johansen-Procedure #
#####
Test type: trace statistic , with linear trend in cointegration
Eigenvalues (lambda):
[1] 2.620162e-01 1.516308e-01 5.551115e-17
Values of teststatistic and critical values of test:
      test 10pct 5pct 1pct
r <= 1 | 6.41 10.49 12.25 16.26
r = 0 | 18.26 22.76 25.32 30.45
Eigenvectors, normalised to first column:
(These are the cointegration relations)

      ipsa.l1     cu.l1   trend.l1
ipsa.l1  1.000000000  1.000000  1.000000
cu.l1    -0.865537128 -4.617804  1.348603
trend.l1  0.006460006 -2.079750 -1.048632

Weights W:
(This is the loading matrix)

      ipsa.l1     cu.l1   trend.l1
ipsa.d -0.2750092  0.06383922 -2.407584e-17
cu.d    0.1063474  0.05211698 -2.296213e-17
```

4.2 Cointegration in a VAR: Vector Error-Correction Models

En nuestro análisis VAR, hemos supuesto que las variables del modelo son estacionarias y ergódicas. Por otro lado, recién vimos que las variables que son individualmente no estacionarias pueden estar cointegradas. Para el caso simple de dos variables y una relación de cointegración, vimos que un modelo de corrección de errores es la especificación econométrica apropiada.

En este modelo, la ecuación se diferencia y se incluye un término de corrección de errores, que mide la desviación del período anterior del equilibrio a largo plazo.

Ahora consideramos cómo las variables cointegradas se pueden utilizar en un VAR utilizando un modelo vectorial de corrección de errores (**Vector error correction modelo - VEC model**).

4.2.1 Un modelo VEC de dos variables

Si dos series I(1), digamos Y_t y X_t , están cointegradas, entonces existe unos únicos α_0 y α_1 tal que $v_t = y_t - \alpha_0 - \alpha_1 x_t$ es $I(0)$. En el modelo de cointegración de una sola ecuación, vimos que el modelo de corrección de errores, tenía la siguiente forma

$$\begin{aligned} \Delta y_t &= \beta_0 + \beta_1 \Delta x_t + \lambda v_{t-1} + \epsilon_t = \\ &\beta_0 + \beta_1 \Delta x_t + \lambda(y_{t-1} - \alpha_0 - \alpha_1 x_{t-1}) + \epsilon_t \end{aligned} \quad (4.3)$$

Todos los términos en la ecuación anterior son $I(0)$, siempre que los coeficientes α (el "vector de cointegración") sean conocidos o al menos consistentemente estimados. El término v_{t-1} es la magnitud por la cual y estaba por encima o por debajo de su valor de equilibrio a largo plazo en el período anterior. El coeficiente λ (que esperamos que sea negativo) representa la cantidad de "corrección" de este período, ($t - 1$), desequilibrio que ocurre en el período t .

Por ejemplo, si λ es -0.25, entonces un cuarto de la brecha entre y_{t-1} y su valor de equilibrio tenderían (todo lo demás igual) a invertirse (porque el signo es negativo) en el período t .

El modelo VEC amplía este modelo de corrección de errores de una sola ecuación para permitir que y y x evolucionen conjuntamente a lo largo del tiempo como en un sistema VAR. En el caso de dos variables, solo puede haber una relación de cointegración y la ecuación y del sistema VEC es similar a la vista anteriormente, excepto que reflejamos la especificación VAR al poner diferencias desfasadas de y y x en el lado derecho.

Con solo una diferencia de retraso (puede haber más), se puede escribir un modelo VEC bivariado como:

$$\Delta y_t = \beta_{y0} + \beta_{yy1} \Delta y_{t-1} + \beta_{yx1} \Delta x_{t-1} + \lambda_y (y_{t-1} - \alpha_0 - \alpha_1 x_{t-1}) + v_t^y \quad (4.4)$$

$$\Delta x_t = \beta_{x0} + \beta_{xy1} \Delta y_{t-1} + \beta_{xx1} \Delta x_{t-1} + \lambda_x (y_{t-1} - \alpha_0 - \alpha_1 x_{t-1}) + v_t^x \quad (4.5)$$

Nuevamente, los términos en ambas ecuaciones son $I(0)$ si las variables son cointegradas con un vector de cointegración $(1, -\alpha_0, -\alpha_1)$, en otras palabras, si $y_t - \alpha_0 - \alpha_1 x_t$ es estacionario.

Los coeficientes λ son nuevamente los coeficientes de corrección de errores, midiendo la respuesta de cada variable

al grado de desviación del equilibrio a largo plazo en el período anterior. Nosotros esperamos que $\lambda_y < 0$ por la misma razón que antes: si y_{t-1} está por encima de su valor de largo plazo en relación con x_{t-1} entonces el término de corrección de errores entre paréntesis es positivo y esto debería producir una baja en y en el período t .

El signo esperado de λ_x depende del signo de α_1 . Esperamos que $\delta \Delta x_t / \delta x_{t-1} = -\lambda_x \alpha_1 < 0$ por la misma razón que esperamos que $\delta \Delta y_t / \delta y_{t-1} = \lambda_y < 0$: si x_{t-1} está por encima de su relación de largo plazo con y , entonces esperamos que Δx_t sea negativo, otras cosas constantes.

Un ejemplo simple y concreto puede ayudar a aclarar el papel de los términos de corrección de errores en un modelo VEC. Supongamos que la relación de cointegración a largo plazo sea $y_t = x_t$, de modo que $\alpha_0 = 0$ y $\alpha_1 = 1$. El

término de corrección de error entre paréntesis en cada ecuación del sistema VAR es ahora $y_{t-1} - x_{t-1}$, la diferencia entre y y x en el período anterior.

Supongamos que debido a shocks previos, $y_{t-1} = x_{t-1} + 1$ o que y está por encima de su relación de equilibrio a largo plazo con x en una unidad (o de manera equivalente, x está por debajo de su relación de equilibrio de largo plazo con y en una unidad).

Para avanzar hacia el equilibrio de largo plazo en el período t , esperamos (si no hay otros cambios) $\Delta y_t < 0$ y $\Delta x_t > 0$. Δy_t cambia en respuesta a este equilibrio mediante $\lambda_y(y_{t-1} - x_{t-1}) = \lambda_y$, para que ocurra un ajuste estable, $\lambda_y < 0$; y es demasiada alta, por lo que debe disminuir en respuesta al desequilibrio. El cambio correspondiente en $\lambda_x(y_{t-1} - x_{t-1}) = \lambda_x$. Como x es "demasiado bajo", el ajuste estable requiere que la respuesta en x sea positiva, por lo que necesitamos $\lambda_x > 0$. Tenga en cuenta que si la relación a largo plazo entre y y x fuera inversa ($\alpha_1 < 0$), entonces x necesitaría disminuir para avanzar hacia el equilibrio y necesitaríamos $\lambda_x < 0$. El signo esperado en λ_x depende del signo de α_1 .

El ECM usando una regresión dinámica

'Código R: Cointegration in practice, Ejemplo Modelo IPSA - CU - S&P

```
a<-data.frame(ipsa,cu)
library(tsDyn)
#Fit a VECM with Engle-Granger 2OLS estimator:
vecm.eg<-VECM(a, lag=2)
#Fit a VECM with Johansen MLE estimator:
vecm.jo<-VECM(a, lag=2, estim="ML")
```

Un modelo VEC de dos variables

```
> summary(vecm.eg)
#####
##Model VECM
#####
Full sample size: 41   End sample size: 38
Number of variables: 2 Number of estimated slope parameters 12
AIC 189.8528    BIC 211.1414    SSR 1118.398
Cointegrating vector (estimated by 2OLS):
  ipsa      cu
r1  1 -2.036934

          ECT      Intercept     ipsa -1      cu -1      ipsa -2      cu -2
Equation ipsa 0.0046(0.1046) -0.2102(0.7902) 0.1548(0.2937) -0.2737(0.3914) 0.1995(0.2989) -0.0884(0.4012)
Equation cu   0.0956(0.0763) -0.2381(0.5762) -0.0847(0.2141)  0.0532(0.2854)  0.0170(0.2180)  0.0882(0.2926)
```

```
> summary(vecm.jo)
#####
##Model VECM
#####
Full sample size: 41   End sample size: 38
Number of variables: 2 Number of estimated slope parameters 12
AIC 176.5604    BIC 197.849    SSR 1026.44
Cointegrating vector (estimated by ML):
  ipsa      cu
r1  1 -0.7961372

          ECT      Intercept     ipsa -1      cu -1      ipsa -2      cu -2
Equation ipsa -0.5588(0.2333)* 21.8032(9.2154)* 0.3078(0.2732) -0.2420(0.3478) 0.4583(0.2837) -0.2088(0.3525)
Equation cu   -0.0017(0.1892) -0.0627(7.4738) -0.0353(0.2216) -0.0418(0.2820)  0.0985(0.2301) -0.0319(0.2859)
```

4.3 Cointegration con Modelos ARDL

La técnica de cointegración ARDL, ver [Pesaran et al. \(2001\)](#), no requiere pruebas previas para raíces unitarias a diferencia de otros métodos. En consecuencia, la técnica de cointegración ARDL es preferible cuando lidiamos con variables que están integradas con diferente orden, $I(0)$, $I(1)$ o una combinación de las dos.

La relación a largo plazo de las variables subyacentes se detecta a través del estadístico F (prueba de Wald). En este enfoque, la relación a largo plazo de la serie se establece cuando el Fstatistic excede la banda de valor crítico. La gran ventaja de este enfoque radica en su identificación de los vectores de cointegración donde hay múltiples vectores cointegrantes.

‘Código R: Ejemplo Modelo IPSA - CU - S&P

```
# ARDL

ipsa_reg4 <- auto.ardl(ipsa ~ cu)
ipsa_reg4 <- ardl(ipsa ~ cu)
```

Instructor: Marcelo Villena Chamorro PhD.

5 Tópico V.- Non-linear Models: Volatility Forecasting

5.1 Modelos ARCH

Una excursión al mundo no-lineal Motivación: los modelos lineales (y de series de tiempo) estructurales, no pueden explicar una serie de características importantes comunes a muchos datos financieros

- (i) Leptokurtosis
- (ii) Volatility clustering or volatility pooling
- (iii) Leverage effects

Ejemplo de una serie financiera: Retornos diarios S&P 500 - Ene2000-Dic2010

‘Código R: Ejemplo 1: S&P500

```
library(tseries) # df, adf
library(dynlm)   #time series regression
library(quantmod) #getSymbols

#Descargar data de precios ajustados
spy <- getSymbols("SPY", from="2000-01-01", to="2010-12-01",
                  auto.assign = FALSE)$SPY.Adjusted
#Grafica de precios
plot(spy)
#Calculo de retorno
Rspy<-log(spy)-lag(log(spy))
#Grafica de retornos
plot(Rspy)
```

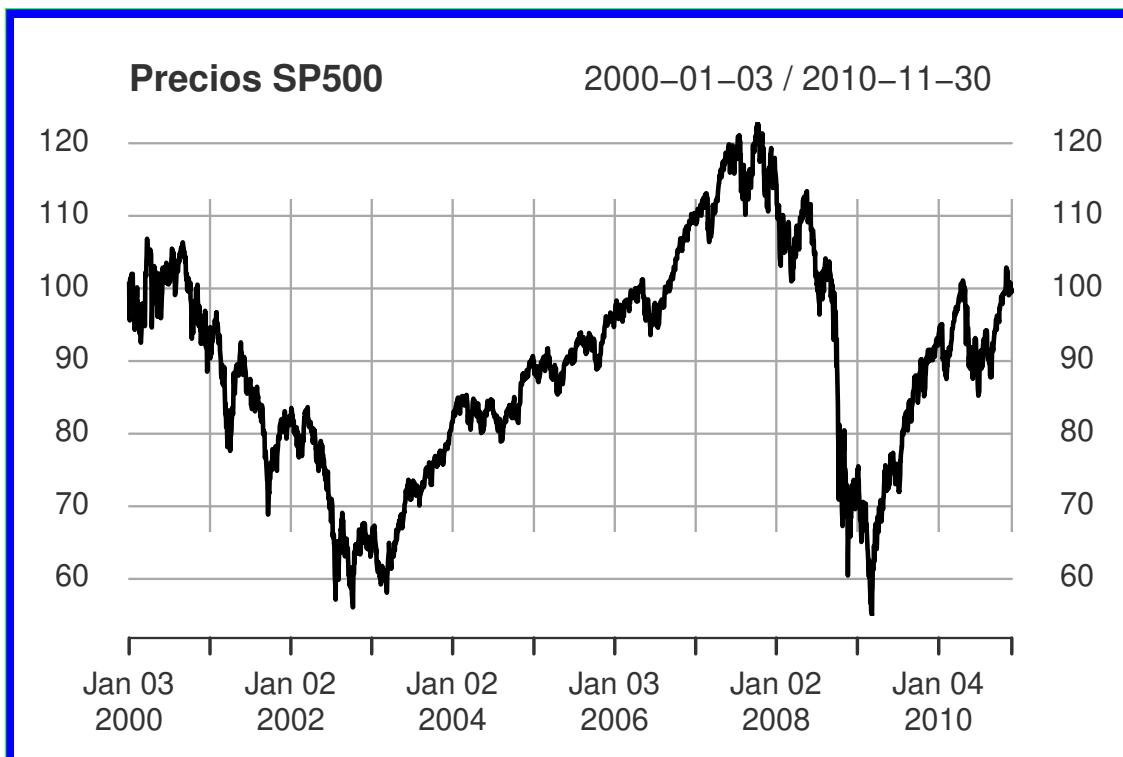


Figure 5.1: Precios diarios S&P 500

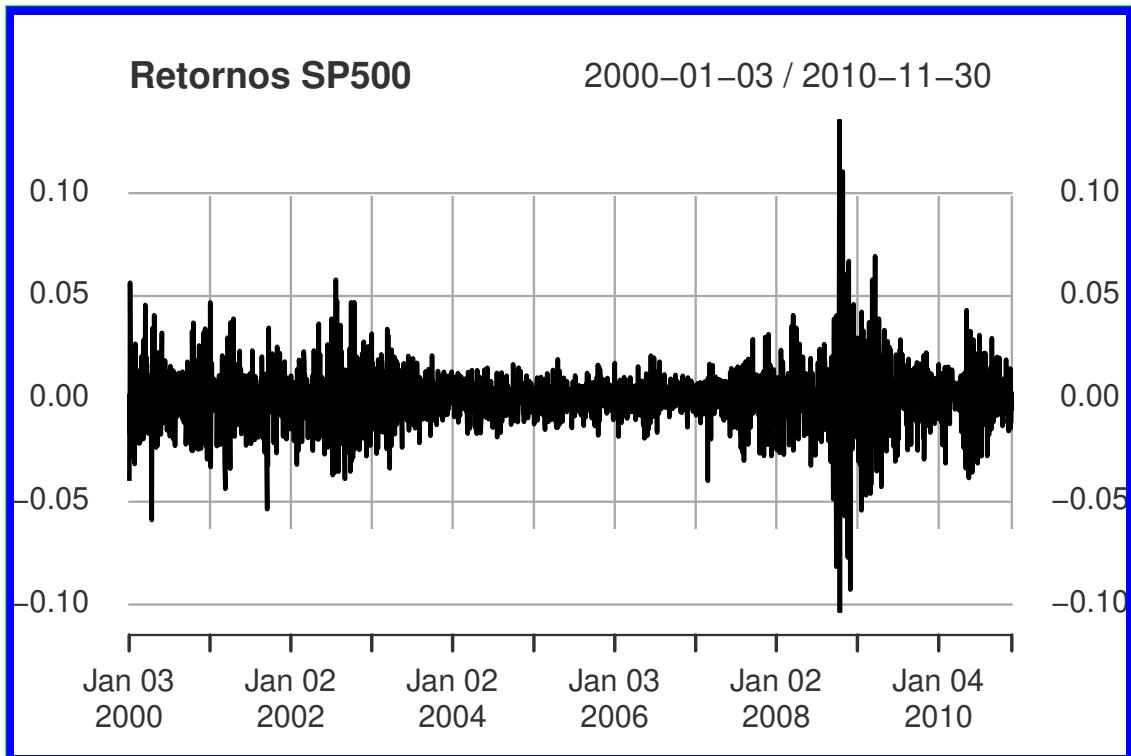


Figure 5.2: Retornos diarios S&P 500

Una excursión al mundo no-lineal

Nuestro modelo estructural “tradicional” podría ser algo como:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + \mu_t \quad (5.1)$$

donde μ_t = white noise, es decir $\mu_t \sim N(0, \sigma^2)$.

Campbell et al. (1997) definen un proceso de generación de datos no-lineal como uno que se puede escribir como:

$$Y_t = f(\mu_t, \mu_{t-1}, \mu_{t-2}, \dots) \quad (5.2)$$

donde μ_t es un término de error aleatorio (iid) y f es una función no lineal.

También dan una definición un poco más específica:

$$Y_t = g(\mu_{t-1}, \mu_{t-2}, \dots) + \mu_t \sigma^2(\mu_{t-1}, \mu_{t-2}, \dots) \quad (5.3)$$

donde g es una función de términos de error pasados solamente y σ^2 es un término de varianza.

Modelos con $g(\cdot)$ no lineales, son "no-lineales en la media", mientras que los no lineales en $\sigma^2(\cdot)$ son "no-lineales en varianza".

Tipos de modelos no lineales

El paradigma lineal es muy útil. Muchas relaciones aparentemente no lineales se pueden linealizar, a través de una transformación adecuada. Por otra parte, es probable que muchas relaciones en finanzas sean intrínsecamente no lineales.

Hay muchos tipos de modelos no lineales, e.g.

- ARCH / GARCH
- switching models
- bilinear models
- Redes neuronales

Pruebas de no linealidad

Las herramientas "tradicionales" de análisis de series temporales (ACF, análisis espectral, etc.) pueden no encontrar evidencia de que podamos utilizar un modelo lineal, pero los datos aún pueden no ser independientes.

Se han desarrollado Pruebas de Portmanteau para la dependencia no-lineal. La más simple es el RESET de Ramsey, que toma la forma:

$$\hat{\mu}_t = \beta_0 + \beta_1 \hat{y}^2 + \beta_2 \hat{y}^3 + \dots + \beta_{p-1} \hat{y}^p + v_t \quad (5.4)$$

Se dispone de muchas otras pruebas de no linealidad, por ejemplo, el "test BDS" y la prueba de biespectro. Un modelo no-lineal en particular que ha demostrado ser muy útil en finanzas es el modelo ARCH, desarrollado por Engle (1982).

Heteroscedasticidad Revisada

Como vimos anteriormente, un ejemplo de un modelo estructural es:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + \mu_t \quad (5.5)$$

donde μ_t = white noise, es decir $\mu_t \sim N(0, \sigma^2)$.

La suposición de que la varianza de los errores es constante se conoce como homocedasticidad, i.e. $Var(\mu_t) = \sigma^2$

¿Qué pasa si la varianza de los errores no es constante?: **heterocedasticidad**

Lo que implican que las estimaciones de error estándar podrían estar equivocadas.

En la práctica la varianza de los errores NO es constante en el tiempo, por ejemplo para los datos financieros.

Autoregressive Conditionally Heteroscedastic Models - Modelos ARCH

Utilicemos un modelo que no asume que la varianza es constante.

Recuerde la definición de la varianza de μ_t :

$$\sigma_t^2 = \text{Var}(\mu_t | \mu_{t-1}, \mu_{t-2}, \dots) = E((\mu_t - E(\mu_t))^2 | \mu_{t-1}, \mu_{t-2}, \dots) \quad (5.6)$$

Normalmente suponemos que $E(\mu_t) = 0$, de aquí:

$$\sigma_t^2 = \text{Var}(\mu_t | \mu_{t-1}, \mu_{t-2}, \dots) = E(\mu_t^2 | \mu_{t-1}, \mu_{t-2}, \dots) \quad (5.7)$$

¿De qué dependerá el valor actual de la varianza de los errores? R: **Del cuadrado de los términos de error anteriores.**

Esto nos lleva al modelo conocido como ARCH, “autoregressive conditionally heteroscedastic model”(modelo condicionalmente heterocedástico autorregresivo):

$$\sigma_t^2 = \alpha_0 + \alpha_1 \mu_{t-1}^2 \quad (5.8)$$

En particular el modelo anterior representa un ARCH(1) model.

El modelo completo es:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + \mu_t, \mu_t \sim N(0, \sigma^2) \quad (5.9)$$

donde $\sigma_t^2 = \alpha_0 + \alpha_1 \mu_{t-1}^2$

Podemos extender fácilmente esto para el caso general en que la varianza del error depende de q rezagos de errores al cuadrado:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \mu_{t-1}^2 + \alpha_2 \mu_{t-2}^2 + \dots + \alpha_q \mu_{t-q}^2 \quad (5.10)$$

Este es un modelo ARCH(q).

En lugar de llamar a la varianza, σ_t^2 en la literatura se le suele llamar h_t , por lo que el modelo es en definitiva:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + \mu_t, \mu_t \sim N(0, h) \quad (5.11)$$

donde

$$h_t = \alpha_0 + \alpha_1 \mu_{t-1}^2 + \alpha_2 \mu_{t-2}^2 + \dots + \alpha_q \mu_{t-q}^2 \quad (5.12)$$

Otra forma de representar a los ARCH Models Por ejemplo, considere un ARCH (1). En lugar de la representación anterior, podemos escribir

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + \mu_t, \mu_t = v_t \sigma_t \quad (5.13)$$

donde

$$\sigma_t = \sqrt{\alpha_0 + \alpha_1 \mu_{t-1}^2} \quad (5.14)$$

Las dos formas representan diferentes maneras de expresar exactamente el mismo modelo. La primera forma es más fácil de entender, mientras que la segunda representa mejor la simulación de un modelo ARCH.

La prueba del “efecto ARCH” 1.- En primer lugar, se corre una regresión lineal, por ejemplo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + \mu_t \quad (5.15)$$

y se guardan los residuos,

2. A continuación, los residuos se elevan al cuadrado, y se corre una regresión sobre los q rezagos propios para la prueba de ARCH de orden q , es decir, ejecutar la regresión

$$\hat{\mu}_t^2 = \gamma_0 + \gamma_1 \hat{\mu}_{t-1}^2 + \gamma_2 \hat{\mu}_{t-2}^2 + \dots + \gamma_q \hat{\mu}_{t-q}^2 + v_t \quad (5.16)$$

3. El test estadístico se define como TR2 (el número de observaciones multiplicado por el coeficiente de correlación múltiple) a partir de la última regresión, y se distribuye como una $\chi^2(q)$.

4. Las hipótesis nula y alternativa son:

$$H_0 : \gamma_1 = \gamma_2 = \gamma_3 = \dots = \gamma_q = 0.$$

$$H_1 : \gamma_1 \neq 0 \text{ o } \gamma_2 \neq 0 \text{ o } \gamma_3 \neq 0 \dots \gamma_q \neq 0.$$

Si el valor de la prueba estadística es mayor que el valor crítico de la distribución $\chi^2(q)$, se rechaza la hipótesis nula.

Tenga en cuenta que la prueba ARCH también se aplica directamente a la rentabilidad, en lugar de a los residuos de la etapa 1 anterior.

Principales problemas de los modelos ARCH ¿Cómo decidimos el mejor q ?

El valor requerido de q podría ser muy grande

Las restricciones de no negatividad pueden ser violadas.

Cuando se estima un modelo ARCH, requerimos $\alpha_i > 0 \forall i = 1, 2, \dots, q$ (ya que la varianza no puede ser negativa)

Una extensión natural de un modelo ARCH (q), que evita algunos de estos problemas es el modelo GARCH que veremos a continuación.

5.2 Modelos GARCH

Generalised ARCH - GARCH Models

Introducido por [Bollerslev \(1986\)](#) deja que la varianza condicional sea dependiente de sus propios rezagos. De esta forma, la ecuación de la varianza es ahora:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \mu_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (5.17)$$

Se trata de un GARCH (1,1), que es equivalente a un ARMA (1,1) de la ecuación de la varianza.

También podríamos escribir

$$\sigma_{t-1}^2 = \alpha_0 + \alpha_1 \mu_{t-2}^2 + \beta \sigma_{t-2}^2 \quad (5.18)$$

$$\sigma_{t-2}^2 = \alpha_0 + \alpha_1 \mu_{t-3}^2 + \beta \sigma_{t-3}^2 \quad (5.19)$$

Sustituyendo (5.18) en (5.17)

$$\sigma_t^2 = \alpha_0 + \alpha_1 \mu_{t-1}^2 + \beta(\alpha_0 + \alpha_1 \mu_{t-2}^2 + \beta \sigma_{t-2}^2) \quad (5.20)$$

$$\sigma_t^2 = \alpha_0(1 + \beta) + \alpha_1 \mu_{t-1}^2(1 + \beta L) + \beta \sigma_{t-1}^2 \quad (5.21)$$

Si seguimos reemplazando términos, el modelo GARCH(1,1) puede escribirse como un modelo ARCH de orden infinito.

Así, podemos extender el GARCH (1,1) a un GARCH (p, q):

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \mu_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \quad (5.22)$$

En general, un modelo GARCH(1,1) es suficiente para capturar la volatilidad clusterizada de los datos.

$$\sigma_t^2 = \alpha_0 + \alpha_1 \mu_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (5.23)$$

¿Por qué es mejor un modelo GARCH que un modelo ARCH?

- Es más parsimonioso - evita el sobreajuste
- Menos probabilidades de violar restricciones de no negatividad

La varianza incondicional bajo la especificación GARCH Calculando la varianza incondicional, podemos estimar la desviación estándar que estamos buscando. De esta forma, derivaremos la varianza incondicional a partir de los modelos ARCH y GARCH. Además, se presenta una nota sobre la variación diaria de escala.

ARCH Unconditional Variance

Suponemos un proceso que podría representarse mediante un modelo econométrico, por ejemplo:

$$y_t = \mu + \epsilon_t \quad (5.24)$$

with $\epsilon_t \sim (0, \sigma_t^2)$

Asumimos que la varianza condicional sigue un modelo del tipo ARCH (1), es decir:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 \quad (5.25)$$

Usando el operador de expectativa incondicional, tenemos:

$$\mathbb{E}(\sigma_t^2) = \sigma_t^2$$

$$\mathbb{E}(\alpha_0) = \alpha_0$$

$$\mathbb{E}(\epsilon_{t-1}^2) = \sigma_t^2$$

Tenemos:

$$\sigma_t^2 (1 - \alpha_1) = \alpha_0 \quad (5.26)$$

$$\Rightarrow \sigma_t^2 = \frac{\alpha_0}{1 - \alpha_1} \quad (5.27)$$

Si generalizamos a un modelo ARCH (q), obtenemos:

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 \\ &= \alpha_0 + \sum_{k=1}^q \alpha_k \epsilon_{t-k}^2 \end{aligned} \quad (5.28)$$

donde:

$$\mathbb{E}(\epsilon_{t-1}^2) = \mathbb{E}(\epsilon_{t-2}^2) = \cdots = \mathbb{E}(\epsilon_{t-q}^2) = \sigma_t^2$$

entonces:

$$\begin{aligned} \sigma_t^2 &= \frac{\alpha_0}{1 - \alpha_1 - \alpha_2 - \cdots - \alpha_q} \\ &= \frac{\alpha_0}{1 - \sum_{k=1}^q \alpha_k} \end{aligned} \quad (5.29)$$

GARCH Unconditional Variance

Supongamos el mismo proceso dado previamente, pero esta vez la varianza también depende de sus propios p lags:

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 + \cdots + \beta_p \sigma_{t-p}^2 \\ &= \alpha_0 + \sum_{k=1}^q \alpha_k \epsilon_{t-k}^2 + \sum_{l=1}^p \alpha_l \sigma_{t-l}^2\end{aligned}\quad (5.30)$$

La ecuación anterior da un modelo GARCH (p, q). Utilizando

$$\mathbb{E}(\sigma_{t-1}^2) = \mathbb{E}(\sigma_{t-2}^2) = \cdots = \mathbb{E}(\sigma_{t-p}^2) = \sigma_t^2$$

De esta forma tenemos:

$$\begin{aligned}\sigma_t^2 &= \frac{\alpha_0}{1 - \alpha_1 - \alpha_2 - \cdots - \alpha_q - \beta_1 - \beta_2 - \cdots - \beta_p} \\ &= \frac{\alpha_0}{1 - \sum_{k=1}^q \alpha_k - \sum_{l=1}^p \beta_l}\end{aligned}\quad (5.31)$$

Scaling Volatility

El cálculo de la volatilidad y la escala en diferentes horizontes de tiempo es posible solo en los casos en que los cambios en el registro del precio del activo v_t se distribuyen de forma independiente e idéntica (iid).

$$v_t = v_{t-1} + \varepsilon_t \quad \varepsilon_t \sim (0, \sigma^2) \quad (5.32)$$

Entonces 1-day return es:

$$v_t - v_{t-1} = \varepsilon_t$$

con desviación estándar σ .

Similarmente, el h-day return es:

$$v_t - v_{t-h} = \sum_{i=0}^{h-1} \varepsilon_{t-i} \quad (5.33)$$

con varianza $h\sigma^2$ y desviación estándar $\sqrt{h\sigma^2}$.

Sin embargo, los rendimientos de los activos financieros de alta frecuencia claramente no son iid ... pero sigue siendo una buena aproximación.

Estimación de modelos ARCH / GARCH

Dado que el modelo ya no es de la forma lineal que acostumbramos, no podemos usar MCO.

Utilizamos otra técnica conocida como de máxima verosimilitud.

El método funciona mediante la búsqueda de los valores más probables de los parámetros, dados los datos reales.

Más específicamente, construimos una función de verosimilitud y la maximizamos.

Los pasos a seguir en la estimación de un modelo ARCH o GARCH son los siguientes:

1. Especificar las ecuaciones apropiadas para la media y la varianza - por ejemplo, un AR (1) - GARCH (1,1):

$$y_t = \alpha + \phi y_{t-1} + \mu_t, \mu_t \sim N(0, \sigma^2) \quad (5.34)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \mu_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (5.35)$$

2. Especifique la función de verosimilitud para maximizar:

$$L = -(T/2)\log(2\pi) - (1/2) \sum_{t=1}^T \log(\sigma_t^2) - (1/2) \sum_{t=1}^T (y_t - \alpha - \phi y_{t-1})/\sigma_t^2 \quad (5.36)$$

3. El computador maximizar la función, y calcula los parámetros y sus errores estándares...

Extensiones al modelo GARCH básico

Los principales problemas de los modelos GARCH (p, q) son:

- Restricciones de no negatividad pueden ser violadas
- Los Modelos GARCH no pueden dar cuenta de los efectos de apalancamiento

En este contexto, desde que se desarrolló el modelo GARCH, se han propuesto un gran número de extensiones y variantes. Tres de los ejemplos más importantes son los modelos GARCH-M, EGARCH, y GJR.

De hecho, posibles soluciones a los dos problemas planteados anteriormente, pueden ser abordados por el modelo GARCH exponencial (EGARCH) o el modelo GJR, que plantean modelos GARCH asimétricos.

GARCH - in Mean

Basados en el problema clásico de cobertura de riesgo, si esperamos que un riesgo se compense con una mayor rentabilidad, ¿por qué no dejar que el retorno de un valor determinado sea parcialmente determinado por su riesgo?

[Engle et al. \(1987\)](#) sugirieron la especificación ARCH-M:

$$Y_t = \mu + \delta \sigma_{t-1} + \mu_t, \mu_t \sim N(0, \sigma_t^2) \quad (5.37)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \mu_{t-1}^2 + \beta \mu_{t-1}^2 \quad (5.38)$$

δ puede interpretarse como una especie de prima por riesgo.

Es posible combinar todos o algunos de estos modelos en conjunto para obtener modelos más complejos, híbridos - por ejemplo, un modelo del tipo ARMA-EGARCH (1,1)-M.

Modelo EGARCH

Sugerido por [Nelson \(1991\)](#) La ecuación de varianza está dada por:

$$\log(\sigma_t^2) = \omega + \beta \log(\sigma_{t-1}^2) + \gamma \frac{\mu_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \alpha \left[\frac{|\mu_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right] \quad (5.39)$$

Ventajas del modelo

Dado que modelamos $\log(\sigma_t^2)$, incluso si los parámetros son negativos, σ_t^2 será positivo. Podemos tomar cuenta el efecto de apalancamiento: si la relación entre volatilidad y rentabilidad es negativa, γ , será negativo.

Modelo GJR

Debido a [Glosten et al. \(1993\)](#):

$$\sigma_t^2 = \alpha_0 + \alpha_1 \mu_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma \mu_{t-1}^2 I_{t-1} \quad (5.40)$$

Donde:

$$I_{t-1} = 1 \text{ si } \mu_{t-1} < 0$$

$$I_{t-1} = 0 \text{ si } \mu_{t-1} \geq 0$$

Para un efecto de apalancamiento veríamos $\gamma > 0$.

Requerimos $\alpha_1 + \gamma \geq 0$ y $\alpha_1 \geq 0$ para no-negatividad. **Modelos GARCH Multivariados**

Los modelos GARCH multivariados se utilizan para estimar y pronosticar covarianzas y correlaciones. La formulación básica es similar a la del modelo GARCH, pero cuando a las varianzas, así como también a las covarianzas, se les permite variar en el tiempo.

Hay 3 clases principales de formulación GARCH multivariante, que son ampliamente utilizadas: VECH, diagonal VECH y BEKK.

Los modelos GARCH multivariantes (MGARCH) generalizan los modelos GARCH univariantes y permiten incorporar relaciones entre los procesos de volatilidad de varias series. Queremos saber por ejemplo, cómo los cambios en la volatilidad de una acción afectan a la volatilidad de otra acción. Estas relaciones se pueden parametrizar de varias maneras.

```
getSymbols("ECH", from="2008-01-01")
Returns = diff(log(Ad(ECH)))
adf.test(Ad(ECH))
adf.test(Returns)
Returns[as.character(head(index(Ad(ECH)),1))] = 0
fit.garch <- garch(Returns, trace=FALSE)
print(fit.garch)
coeftest(fit.garch)
sigmaGarch<- fit.garch[["coef"]][["a0"]]/(1-fit.garch[["coef"]][["a1"]]-fit.garch[["coef"]][["b1"]]) sigmaAvg<- var(Returns)
```

'Código R:GARCH vs volatilidad promedio retorno

```

library(rugarch)
library(lmtest)
getSymbols("ECH", from="2008-01-01")
Returns = diff(log(Ad(ECH)))
adf.test(Ad(ECH))
adf.test(Returns[-1])
Returns[as.character(head(index(Ad(ECH)),1))] = 0
fit.garch <- garch(Returns, trace=FALSE)
print(fit.garch)
coeftest(fit.garch)
sigmaGarch<-fit.garch[["coef"]][["a0"]]/(1-fit.garch[["coef"]][["a1"]]-fit.garch[["coef"]][["b0"]])
sigmaAvg<-var(Returns)
sigmaGarch
sigmaAvg

```

> adf.test(Ad(ECH))

Augmented Dickey-Fuller Test

```

data: Ad(ECH)
Dickey-Fuller = -2.0624, Lag order = 14, p-value = 0.5519
alternative hypothesis: stationary

```

> adf.test(Returns[-1])

Augmented Dickey-Fuller Test

```

data: Returns[-1]
Dickey-Fuller = -13.642, Lag order = 14, p-value = 0.01
alternative hypothesis: stationary

```

Figure 5.3: GARCH vs volatilidad promedio retorno, Dickey-Füller Aumentado

```
> fit.garch <- garch(Returns, trace=FALSE)
> print(fit.garch)

Call:
garch(x = Returns, trace = FALSE)

Coefficient(s):
      a0          a1          b1
4.464e-06 1.045e-01 8.768e-01

> coeftest(fit.garch)

z test of coefficients:

   Estimate Std. Error z value Pr(>|z|)
a0 4.4644e-06 7.5166e-07 5.9394 2.861e-09 ***
a1 1.0446e-01 7.3513e-03 14.2096 < 2.2e-16 ***
b1 8.7676e-01 8.9253e-03 98.2323 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> sigmas
      sigmaAVG    sigmaGARCH
ECH.Adjusted 0.0002426729 0.0002376655
```

‘Código R: GARCH vs volatilidad promedio retorno

```
windowLength = 40
foreLength = length(Returns) - windowLength
sigmaAvgV = vector(mode='character', length=foreLength)
sigmaGarchV <- vector(mode='character', length=foreLength)
```

‘Código R: GARCH Ventanas móviles

```
for (d in 1:foreLength){
  ReturnsOffset = Returns[(d):(windowLength+d)]
  fit.garch <- garch(ReturnsOffset, trace=FALSE)
  sigmaGarch <- fit.garch[['coef']][['a0']] / (1 - fit.garch[['coef']][['a1']] -
    fit.garch[['coef']][['b1']])
  sigmaAvg <- var(ReturnsOffset)
  print(sigmaGarch); print(sigmaAvg)
  sigmaGarchV[d] <- sigmaGarch
  sigmaAvgV[d] <- sigmaAvg
}

plot.ts(sigmaAvgV, col = 'red', main='Volatility_Average_vs_GARCH(1,1)')
lines.default(sigmaGarchV, col = 'black')
```

Retornos diarios ECH - 1Ene2008-today

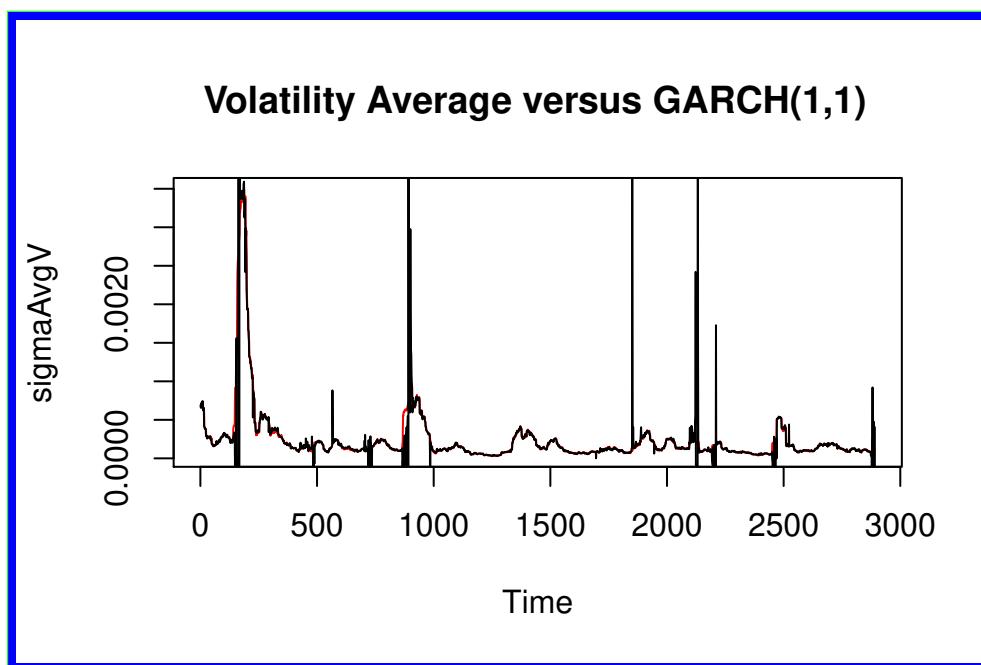


Figure 5.4: Volatilidad promedio vs volatilidad calculada por GARCH

Volatilidad calculada en la web V-Lab

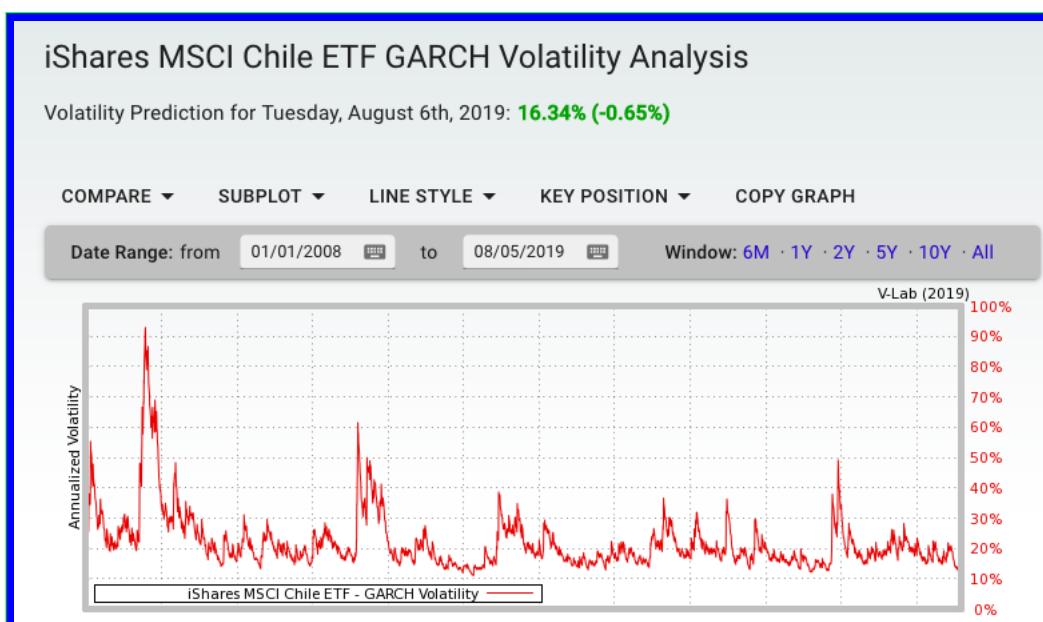


Figure 5.5: <https://vlab.stern.nyu.edu/analysis/VOL.ECH:US-R.GARCH>

5 Tópico V.- Non-linear Models: Volatility Forecasting

Instructor: Marcelo Villena Chamorro PhD.

6 Tópico VI - Introduction to Neural Networks

6.1 Introducción a las Redes Neuronales

Un tema popular en el análisis de datos moderno es la red neuronal, el cual se puede clasificar como un método semiparamétrico. La literatura sobre redes neuronales es enorme, y su aplicación se extiende a muchas áreas científicas con diversos grados de éxito. [Cheng and Titterington \(1994\)](#) proporcionan información sobre redes neuronales desde un punto de vista estadístico.

Primeramente, nos enfocaremos en las redes neuronales feed-forward (Feed forward neural network) en las cuales las entradas están conectadas a una o más neuronas, o nodos, en la capa de entrada, y estos nodos están conectados a otras capas hasta que alcanzan la capa de salida.

Las redes neuronales son un marco de aprendizaje automático (machine learning framework), que intenta imitar el patrón de aprendizaje de las redes neuronales biológicas naturales. Las redes neuronales biológicas tienen neuronas interconectadas con dendritas que reciben entradas, y luego, basadas en estas entradas, producen una señal de salida a través de un axón a otra neurona. Tratamos de imitar este proceso mediante el uso de Redes neuronales artificiales (Artificial Neuronal Networks: ANN), que a partir de ahora llamaremos redes neuronales. El proceso de crear una red neuronal comienza con la forma más básica, un sólo perceptrón.

Posteriormente, introduciremos las Support Vector Machine (SVM). En el aprendizaje automático (machine learning), los Support Vector Machine (SVM) son modelos de aprendizaje supervisados mediante algoritmos que permiten analizar los datos utilizados para la clasificación y el análisis de regresión.

En particular, dado un conjunto de ejemplos de entrenamiento, cada uno marcado como perteneciente a una u otra de dos categorías, un algoritmo de entrenamiento SVM construye un modelo que asigna nuevos datos a una de las dos categorías, convirtiéndolo en un clasificador lineal binario no probabilístico.

Un modelo SVM es una representación de los datos como puntos en el espacio, mapeados de manera que los ejemplos de las categorías, están divididas por un espacio libre lo más ancho posible. A continuación, se mapean nuevos datos en ese mismo espacio y se predice que pertenecen a una categoría según el lado del espacio en el que caen.

6.1.1 Feed forward neural network

Desarrollaremos un ejemplo de una red de feed-forward simple para análisis de series temporales univariadas con una capa oculta. La capa de entrada tiene dos nodos, y la capa oculta tiene tres. Los nodos de entrada se conectan hacia adelante a todos y cada uno de los nodos de la capa oculta, y estos nodos ocultos se conectan al único nodo en la capa de salida. Llamamos a la red una red de feed-forward.

Las redes neuronales más complicadas, incluidas las que tienen conexiones de retroalimentación, han proliferado, pero las redes de feed-forward son las más relevantes para nuestro estudio.

Una red neuronal feed-forward con una capa oculta para el análisis de series temporales univariadas.

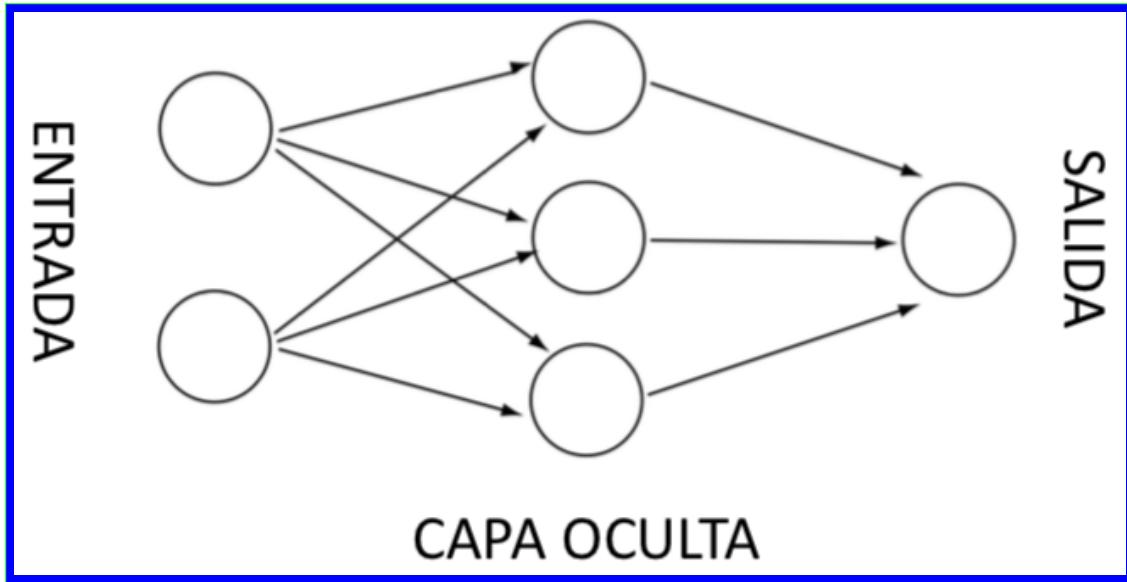


Figure 6.1: Esquema de red neuronal.

El perceptrón recibe entradas, las multiplica por un poco de peso y luego las pasa a una función de activación para producir una salida. Es así como, una red neuronal procesa información de una capa a la siguiente mediante una "función de activación". Consideré una red de feed-forward con una capa oculta. El j -ésimo nodo en la capa oculta se define como

$$h_j = f_j \left(\alpha_{0j} + \sum_{i \rightarrow j} w_{ij} x_i \right) \quad (6.1)$$

donde x_i es el valor del i -ésimo nodo de entrada, $f_j(\cdot)$ es una función de activación que generalmente se toma como la función logística:

$$f_j(z) = \frac{\exp(z)}{1 + \exp(z)}$$

α_{0j} se llama sesgo, la suma $i \rightarrow j$ significa sumando todos los nodos de entrada que alimentan a j , y w_{ij} son los pesos.

A modo ilustrativo, el j -ésimo nodo de la capa oculta de la red de avance 2-3-1 en la figura es:

$$h_j = \frac{\exp(\alpha_{0j} + w_{1j}x_1 + w_{2j}x_2)}{1 + \exp(\alpha_{0j} + w_{1j}x_1 + w_{2j}x_2)}, j = 1, 2, 3. \quad (6.2)$$

Para la capa de salida, el nodo se define como:

$$o = f_o \left(\alpha_{0o} + \sum_{j \rightarrow o} w_{jo} h_j \right) \quad (6.3)$$

donde la función de activación $f_o(\cdot)$ es lineal o una función de escalón unitario (Heaviside function). Si $f_o(\cdot)$ es lineal, entonces:

$$o = \alpha_{0o} + \sum_{j \rightarrow o} w_{jo} h_j$$

donde k es la cantidad de nodos en la capa oculta. Por una función de escalón unitario, queremos decir $f_o(z) = 1$ si $z > 0$ y $f_o(z) = 0$ en caso contrario. Una neurona con una función de escalón unitario se llama neurona umbral, con "1" que indica que la neurona envía su mensaje. Por ejemplo, la salida de la red 2-3-1 en la figura es:

$$o = \alpha_{0o} + w_{1o} h_1 + w_{2o} h_2 + w_{3o} h_3$$

Si la función de activación es lineal, tenemos:

$$o = \begin{cases} 1, & \text{if } \alpha_{0o} + w_{1o} h_1 + w_{2o} h_2 + w_{3o} h_3 > 0. \\ 0, & \text{if } \alpha_{0o} + w_{1o} h_1 + w_{2o} h_2 + w_{3o} h_3 \leq 0. \end{cases}$$

Si $f_o(\cdot)$ es una función de escalón unitario.

Al combinar las capas, la salida de una red neuronal feed-forward se puede escribir como:

$$o = f_o \left[\alpha_{0o} + \sum_{j \rightarrow o} w_{jo} f_j \left(\alpha_{0j} + \sum_{i \rightarrow j} w_{ij} x_i \right) \right] \quad (6.4)$$

Si uno también permite conexiones directas desde la capa de entrada a la capa de salida, entonces la red se convierte en:

$$o = f_o \left[\alpha_{0o} + \sum_{i \rightarrow o} \alpha_{io} x_i + \sum_{j \rightarrow o} w_{jo} f_j \left(\alpha_{0j} + \sum_{i \rightarrow j} w_{ij} x_i \right) \right] \quad (6.5)$$

donde la primera suma se suma a los nodos de entrada. Cuando la función de activación de la capa de salida es lineal, las conexiones directas desde los nodos de entrada al nodo de salida representan una función lineal entre las entradas y la salida. En consecuencia, en este caso particular, el modelo presentado en la ecuación 6.5, es una generalización de modelos lineales. Para la red 2-3-1 de nuestra figura, si la función de activación de salida es lineal, entonces la ecuación 6.4 se convierte en:

$$o = \alpha_{0o} + \sum_{j=1}^3 w_{jo} h_j$$

donde h_j se obtiene de la ecuación 6.2.

La red tiene 13 parámetros. Si utilizamos la ecuación 6.5, la red se convierte en:

$$o = \alpha_{0o} + \sum_{i=1}^2 \alpha_{io} x_i + \sum_{j=1}^3 w_{jo} h_j$$

donde h_j se obtiene de la ecuación 6.2. El número de parámetros de la red aumenta a 15.

Nos referimos a la función en la ecuación 6.4 o 6.5 como una función semiparamétrica porque se conoce su forma funcional, pero se desconoce el número de nodos y sus desviaciones y pesos.

Las conexiones directas desde la capa de entrada a la capa de salida en la ecuación 6.5, significa que la red puede omitir la capa oculta. Nos referimos a dicha red como una red de feed-forward de capa de salto.

Las redes feed-forward se conocen como perceptrones multicapa en la literatura de redes neuronales. Pueden

aproximar cualquier función continua uniformemente en conjuntos compactos aumentando la cantidad de nodos en la capa oculta, ver [Hornik et al. \(1989\)](#), [Chen and Chen \(1995\)](#). Esta propiedad de las redes neuronales es la propiedad de aproximación universal de los perceptrones multicapa.

En resumen, las redes neuronales feed-forward con una capa oculta se pueden ver como una forma de parametrizar una función no lineal continua general.

Training and Forecasting

La aplicación de redes neuronales implica dos pasos.

El primer paso es entrenar (training) a la red (es decir, para construir una red, incluida la determinación del número de nodos y la estimación de sus sesgos y ponderaciones).

El segundo paso es la inferencia, especialmente el pronóstico extra-muestral (forecasting).

Al comparar el resultado comparado de cada pronóstico, se selecciona la red que supera a las demás, definiéndose como la mejor red para realizar inferencias. En una aplicación de serie de tiempo, supongamos que $\{(r_t, x_t) | t = 1, \dots, T\}$ representan los datos disponibles para el entrenamiento de la red, donde x_t denota el vector de entradas, y r_t es la serie de interés (por ejemplo, retornos-log de un activo).

Para una red determinada, supongamos que o_t sea la salida de la red, con una entrada de x_t ; ver el modelo presentado en la ecuación [6.5](#).

El entrenamiento de una red neuronal equivale a elegir sus sesgos y pesos de forma de minimizar algunos criterios de ajuste, por ejemplo, el mínimo cuadrado de su error

$$S^2 = \sum_{t=1}^T (r_t - o_t)^2$$

Este es un problema de estimación no lineal que puede ser resuelto por varios métodos iterativos. Para garantizar la continuidad (smoothness) de la función ajustada, se pueden agregar algunas restricciones adicionales al problema de minimización anterior.

En la literatura de redes neuronales, el algoritmo de aprendizaje Back Propagation (BP) es el método más popular para el entrenamiento de una red. El método BP, introducido por ?, funciona hacia atrás comenzando con la capa de salida, y usa una regla de gradiente (gradient rule) para modificar los sesgos y los pesos iterativamente.

Una vez que se construye una red neuronal feed-forward, se puede usar para calcular pronósticos extramuestrales.

En resumen, una vez que tenemos la salida, podemos compararla con una serie conocida y ajustar los pesos de la mejor forma posible (los pesos usualmente comienzan con valores de inicialización aleatorios). Seguimos repitiendo este proceso hasta que hayamos alcanzado un número máximo de iteraciones permitidas, o una tasa de error aceptable.

Para crear una red neuronal, simplemente comenzamos a agregar capas de perceptrones, creando un modelo de perceptrón multicapa de una red neuronal. Tendremos una capa de entrada que tomará directamente las entradas de funciones y una capa de salida que creará las salidas resultantes. Las capas intermedias se conocen como capas ocultas porque no ven directamente las entradas o salidas de características

Pre-procesamiento de datos

Es importante normalizar los datos antes de entrenar una red neuronal. La red neuronal puede tener dificultades para converger antes de la cantidad máxima de iteraciones permitidas si los datos no están normalizados. Hay muchos métodos diferentes para la normalización de datos. Por lo general, es mejor escalar los datos de 0 a 1, o de -1 a 1.

Ejemplo 1

Para ilustrar las potenciales aplicaciones de redes neuronales en finanzas, modelaremos los retornos mensuales, incluyendo dividendos, de la firma IBM de enero de 1926 a diciembre de 1999.

Dividimos los datos en dos sub-muestras. La primera sub-muestra, que consiste en retornos de enero de 1926 a

diciembre de 1997 para 864 observaciones, se usa para modelar.

Usando el modelo presentado en la ecuación 6.5, con tres entradas y dos nodos en la capa oculta, obtenemos una red 3-2-1 para la serie. Las tres entradas son r_{t-1} , r_{t-2} y r_{t-3} , y los sesgos y ponderaciones se presentan a continuación:

$$r_t = 3.22 - 1.81f_1(r_{t-1}) - 2.28f_2(r_{t-1}) - 0.09r_{t-1} - 0.05r_{t-2} - 0.12r_{t-3}$$

donde, $r_{t-1} = (r_{t-1}, r_{t-2}, r_{t-3})$

y las dos funciones logísticas son:

$$f_1(r_{t-1}) = \frac{\exp(-8.34 - 18.97r_{t-1} + 2.17r_{t-2} - 19.17r_{t-3})}{1 + \exp(-8.34 - 18.97r_{t-1} + 2.17r_{t-2} - 19.17r_{t-3})}$$

$$f_2(r_{t-1}) = \frac{\exp(39.25 - 22.17r_{t-1} - 17.34r_{t-2} - 5.98r_{t-3})}{1 + \exp(39.25 - 22.17r_{t-1} - 17.34r_{t-2} - 5.98r_{t-3})}$$

El error estándar de los residuos para el modelo anterior es 6.56. Para comparación, también construyó un modelo AR para los datos, obteniéndose el siguiente modelo:

$$r_t = 1.101 + 0.077r_{t-1} + a_t$$

con $\sigma_a = 6.61$

El error estándar residual es ligeramente mayor que el del modelo feed-forward.

Comparación de Pronóstico

Las rentabilidades mensuales de las acciones de IBM en 1998 y 1999 forman la segunda submuestra y se utilizan para evaluar el rendimiento de las predicciones fuera de muestra de las redes neuronales.

Como punto de referencia para la comparación, utilizamos la media muestral de r_t en la primera submuestra como el pronóstico de 1 paso para todas las rentabilidades mensuales en la segunda submuestra. Esto equivale a suponer que el precio mensual del precio de las acciones de IBM sigue una caminata aleatoria con drift.

El error de pronóstico cuadrático medio (MSE) del modelo de referencia es 91.85. Para el modelo AR (1), el MSE de los pronósticos de 1 paso adelante es 91.70. Por lo tanto, el modelo AR (1) supera ligeramente al benchmark. Para la red, el MSE es 91.74.

Ejemplo 2 - Pronosticando el S&P 500

En el presente ejemplo pronosticaremos el S&P 500 usando modelos univariados. En particular, usaremos un modelo lineal, como es el ARIMA, y uno no-lineal, una feed-forward neural network.

'Código R: Pronóstico de SP500

```
library(neuralnet)
library(nnet)
library(forecast)
getSymbols("SPY", from = "2000-01-01",
          to = "2017-12-01", src = "yahoo", adjust = TRUE, periodicity = "monthly")
Returns = diff(log(Ad(SPY)))
Returns[as.character(head(index(Ad(SPY)), 1))] = 0
```

nnetar (**Neural Network Time Series Forecast**) es una red neuronal del tipo feed-forward que considera valores rezagados de y como entradas, y una única capa oculta. Las entradas son retardos de 1 a p. Se ajustan diversas

redes, cada una con pesos iniciales aleatorios. Luego los resultados se promedian al calcular los pronósticos. La red está diseñada para la predicción en un solo paso. Los pronósticos de pasos múltiples se calculan recursivamente. Para datos no estacionales, el modelo ajustado se denota como un modelo NNAR (p, k), donde k es el número de nodos ocultos. Esto es análogo a un modelo AR (p) pero con funciones no lineales. Para datos estacionales, el modelo ajustado se llama modelo NNAR (p, P, k) [m], que es análogo a un modelo ARIMA ($p, 0,0$) ($P, 0,0$) [m] pero con funciones no lineales.

'Código R:

```
train>Returns[1:209]
test>Returns[210:215]
nn <- nnetar(train)
fcast <- forecast(nn, h=length(test))
autoplot(fcast)
plot(fcast)
fcast
test
accuracy(fcast)

arimaModel <- auto.arima(train)
accuracy(arimaModel)
```

Forecasts from NNAR(3,2)

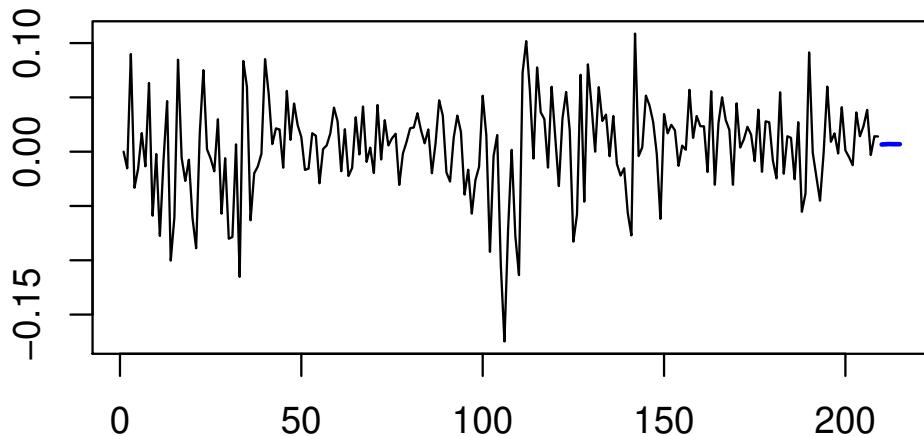


Figure 6.2

Claramente para el período 2000-2017 gana la rede neuronal autoregresiva, considerando las no-linealidades presentes.

```
> accuracy(fcast)
      ME      RMSE      MAE  MPE MAPE      MASE      ACF1
Training set -1.014823e-05 0.03718881 0.02862862 -Inf  Inf 0.6467296 0.05227047
```

Figure 6.3

```
> accuracy(arimaModel)
      ME      RMSE      MAE  MPE MAPE      MASE      ACF1
Training set 0.003710983 0.04289159 0.03243669 -Inf  Inf 0.7327552 -0.01633621
```

Figure 6.4

6.2 Support Vector Machines (SVM)

Fueron creadas por Boser, Guyon y Vapnik en 1992 [Boser et al. \(1992\)](#). La formulación original esta motivada por la resolución de problemas de clasificación, donde la idea básica consiste en **mapear los datos desde el espacio original a un espacio de mayor dimensión a través de una transformación no lineal escogida a priori, para luego construir el hiperplano de separación óptimo en el nuevo espacio**. De esta manera, mediante la resolución de un problema lineal en el nuevo espacio, se tiene un modelo no lineal en el espacio original. En base a la misma filosofía, el método se extendió luego a problemas de regresión y de clustering. Desde su creación, SVM ha acaparado gran atención teórica, siendo aplicado con gran éxito a problemas prácticos de predicción de series de tiempo de distinta naturaleza.

Dentro de las principales características de SVM se cuentan: i) la posibilidad de resolver un problema convexo, sin entrampamientos en óptimos locales, ii) la representación de la solución en base a una fracción del total de puntos disponibles (estos puntos son los llamados Support Vectors), iii) la capacidad de generalización a nuevos datos, debido a que el algoritmo SVM se basa en el principio de minimización del riesgo estructural propuesto en la Teoría de Aprendizaje Estadístico de Vapnik, y iv) la capacidad de modelar fenómenos no lineales mediante la ya mencionada transformación de los datos desde el espacio original a un espacio de mayor dimensión, espacio en el cual se obtiene un modelo lineal que equivale a un modelo lineal en el espacio original. En base a la misma filosofía, el método se extendió luego a problemas de regresión y de clustering. Desde su creación, SVM ha acaparado gran atención teórica, siendo aplicado con gran éxito a problemas prácticos de predicción de series de tiempo de distinta naturaleza.

Dentro de las principales características de SVM se cuentan: i) la posibilidad de resolver un problema convexo, sin entrampamientos en óptimos locales, ii) la representación de la solución en base a una fracción del total de puntos disponibles (estos puntos son los llamados Support Vectors), iii) la capacidad de generalización a nuevos datos, debido a que el algoritmo SVM se basa en el principio de minimización del riesgo estructural propuesto en la Teoría de Aprendizaje Estadístico de Vapnik, y iv) la capacidad de modelar fenómenos no lineales mediante la ya mencionada transformación de los datos desde el espacio original a un espacio de mayor dimensión, espacio en el cual se obtiene un modelo lineal que equivale a un modelo lineal en el espacio original.

6.2.1 Definiciones del modelo - Funciones de Kernel

Un kernel se define como una función K , tal que $\forall x, y \in K$

$$K(x, z) = \langle \Phi(x)\Phi(z) \rangle$$

donde X es el espacio de los datos de entrada (finito, generalmente R^n); y Φ es una función de mapeo de los datos de entrada desde X a un espacio F de mayor dimensión, donde $\langle \bullet, \bullet \rangle$ es el producto interno de F .

Se puede probar que $K(x, z)$ es una función de kernel si y sólo si la matriz $M = (K(x_i, x_j))_{i,j=1}^n$ es semidefinida positiva.

Algunos de los kernels más comunes son:

- Lineal: $K(x, x') = \langle x, x' \rangle$
- Polinomial: $K(x, x') = (\langle x, x' \rangle + 1)^d$
- RBF: $K(x, x') = \exp(-\|x - x'\|^2/\sigma^2)$

6.2.2 Definiciones del modelo - Estructura de la SVM

El modelo de SVM se puede ver como capas de nodos, en donde:

- La primera capa consta de n nodos, que corresponden al vector de entrada
- La segunda capa consta de N nodos, que es la transformación no lineal a base de support vectors

- La tercera capa contiene 1 solo nodo, que es la predicción
- Cada capa se conecta de forma completa con la siguiente
- Los nodos que llegan al nodo de output se ponderan por constantes, que son a determinar por el modelo, y luego se suman

Durante el proceso de aprendizaje, la primera capa selecciona las bases $K(x_i; X)$, $i = 1, \dots, N$; dentro del conjunto de bases posibles, en tanto que la segunda capa construye una función lineal en el nuevo espacio, lo que es equivalente a encontrar un modelo no lineal en el espacio de entrada. Las N bases seleccionadas son aquellas inducidas por los puntos denominados Support Vectors.

6.2.3 Definiciones del modelo - Funciones de perdida

El modelo que se busca es de la forma $y = f(x) + e$, donde $f(x)$ es una función no lineal y e el error. Luego, uno desea minimizar el valor de $y_i - f(x_i) = e$; para cada i , y para esto se usa una función de pérdida.

Las más comunes son:

- Cuadrática:

$$L(f(x), y) = (f(x) - y)^2$$

- ϵ - sensible:

$$L(f(x), y, \epsilon) = \begin{cases} 0 & \text{si } |f(x) - y| < \epsilon \\ |f(x) - y| & \text{si no} \end{cases}$$

- Huber:

$$L(f(x), y, \epsilon) = \begin{cases} \frac{1}{2}(f(x) - y)^2 & \text{si } |f(x) - y| < \mu \\ \mu|f(x) - y| - \frac{\mu^2}{2} & \text{si no} \end{cases}$$

6.2.4 Algoritmo de Regresión SVM

El problema de optimización que encuentra los pesos del modelo, usando función de pérdida ϵ sensible es:

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 \\ \text{s.a.} \quad & y_i - \langle w, \Phi(x_i) \rangle - b \leq \epsilon \\ & -y_i + \langle w, \Phi(x_i) \rangle + b \geq \epsilon \end{aligned} \tag{6.6}$$

Dado que puede que no exista solución para el problema anterior, se suele reformular como:

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\zeta_i + \zeta_i^*) \\ \text{s.a.} \quad & y_i - \langle w, \Phi(x_i) \rangle - b \leq \epsilon + \zeta_i \\ & -y_i + \langle w, \Phi(x_i) \rangle + b \geq \epsilon + \zeta_i \\ & \zeta_i, \zeta_i^* \geq 0, i = 1, 2, \dots, l \end{aligned} \tag{6.7}$$

El problema es que puede ser que no exista solución, por lo que se reformula como: donde C es un parámetro a fijar, que representa el trade-off entre la complejidad y la exactitud del modelo, y el parámetro ϵ representa el rango de tolerancia a los errores en el modelo. Este problema tiene solución, y además es convexo, por lo que los métodos de optimización convergen bien a la solución, y el planteamiento del dual es bastante más sencilla que el problema primal.

Una vez encontrados los pesos w , entonces nuestro modelos es:

$$y = \sum_{i=1}^N w_i K(X_i, x)$$

Ejemplo 3 - Calculando el beta de Mastercard

'Código R: Packages requeridos

```
# load required library
rm(list=ls())
library(quantmod)
library(e1071)
library(dynlm)
library(forecast)
library(TTR)
library(tseries)
library(lmtest)
```

'Código R: Descarga de precios de índices S&P500(SPY) y Mastercard(MA).

```
getSymbols("SPY", from = "2006-05-01", to = "2017-12-01", src = "yahoo", adjust =
TRUE, periodicity = "monthly")
Returns = diff(log(Ad(SPY)))
Returns[as.character(head(index(Ad(SPY)), 1))] = 0

getSymbols("MA", from = "2006-05-01", to = "2017-12-01", src = "yahoo", adjust =
TRUE, periodicity = "monthly")
Returns = diff(log(Ad(MA)))
Returns[as.character(head(index(Ad(MA)), 1))] = 0
```

'Código R: Función RMSE y modelos de regresión lineal, estándar y SVM

```
Y <- Returns$MA.Adjusted
X <- Returns$SPY.Adjusted

rmse <- function(error){
  sqrt(mean(error^2))
}

# Create a linear regression model
model1 <- lm(Y~X)
print(model1)
coefTest(model1)
error1 <- model1$residuals
predictionRMSE.OLS <- rmse(error1)

# Create a SVM regression model
model2 <- svm(Y ~ X)
print(model2)
error2 <- model2$residuals
predictionRMSE.SVM <- rmse(error2)
```

Para el timespam 2007-2012 tenemos:
predictionRMSE.OLS

```
[1] 0.08724145  
predictionRMSE.SVM  
[2] 0.08654265
```

Sin embargo, para el timespam 2012-2017 tenemos:

```
predictionRMSE.OLS  
[1] 0.03777724  
predictionRMSE.SVM  
[2] 0.03829508
```

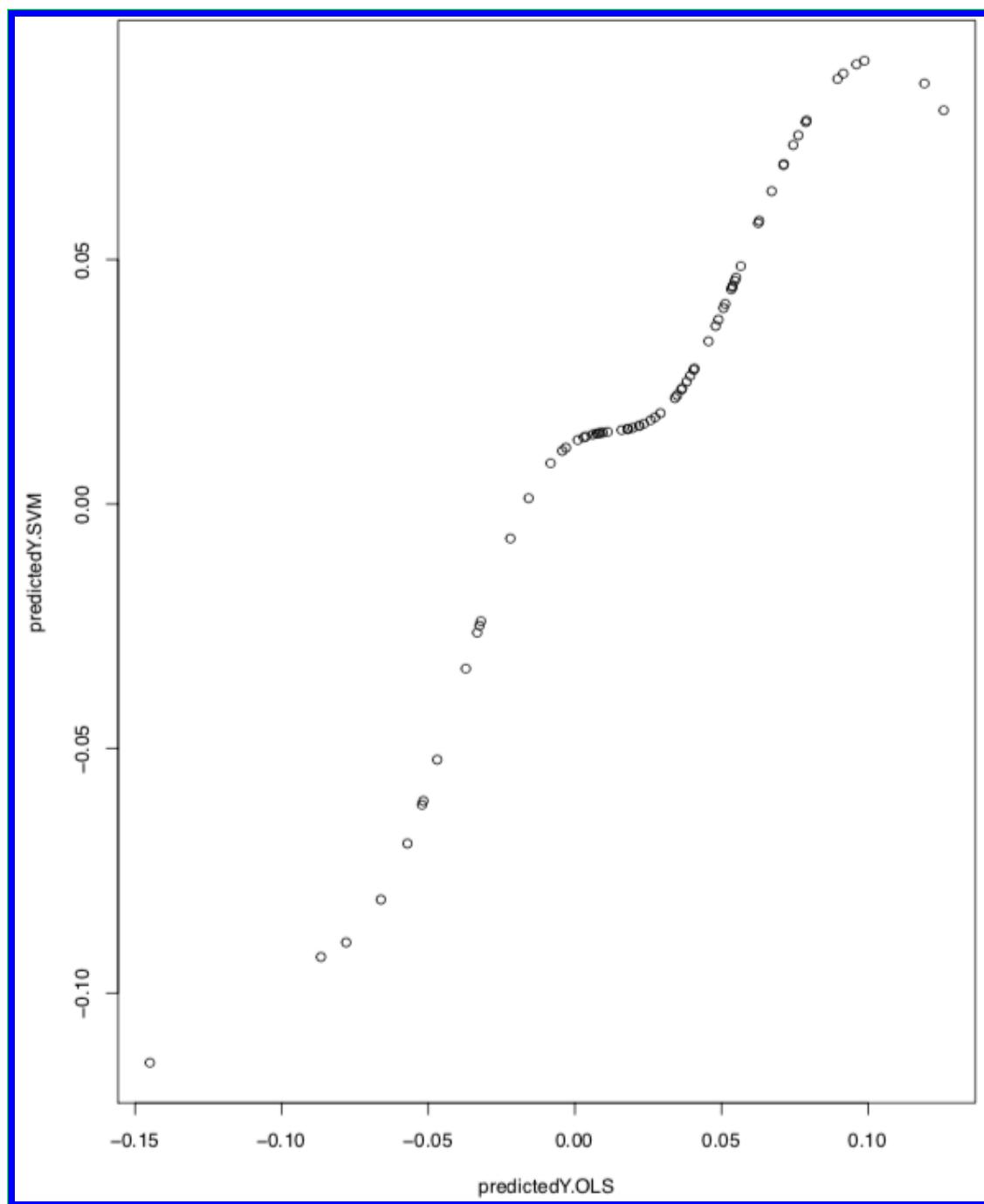


Figure 6.5

Instructor: Marcelo Villena Chamorro PhD.

Bibliography

- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics* 31(3), 307–327.
- Boser, B. E., I. M. Guyon, and V. N. Vapnik (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pp. 144–152. ACM.
- Box, G. E. P. and G. Jenkins (1970). *Time Series Analysis, Forecasting, and Control*. Francisco Holden-Day.
- Campbell, J. Y., A. W. Lo, A. C. MacKinlay, et al. (1997). *The econometrics of financial markets*, Volume 2. princeton University press Princeton, NJ.
- Chen, T. and H. Chen (1995). Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems. *IEEE Transactions on Neural Networks* 6(4), 911–917.
- Cheng, B. and D. M. Titterington (1994). Neural networks: A review from a statistical perspective. *Statistical science*, 2–30.
- 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.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987–1007.
- Engle, R. F. and C. W. Granger (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, 251–276.
- Engle, R. F., D. M. Lilien, and R. P. Robins (1987). Estimating time varying risk premia in the term structure: The arch-m model. *Econometrica: journal of the Econometric Society*, 391–407.
- Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance* 48(5), 1779–1801.
- Granger, C. W. and P. Newbold (1974). Spurious regressions in econometrics. *Journal of econometrics* 2(2), 111–120.
- Hornik, K., M. Stinchcombe, and H. White (1989). Multilayer feedforward networks are universal approximators. *Neural networks* 2(5), 359–366.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of economic dynamics and control* 12(2-3), 231–254.
- Lütkepohl, H. and M. Krätsig (2004). *Applied time series econometrics*. Cambridge university press.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347–370.
- Pesaran, M. H., Y. Shin, and R. J. Smith (2001). Bounds testing approaches to the analysis of level relationships. *Journal of applied econometrics* 16(3), 289–326.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, 1–48.