



Universidad de Buenos Aires
Facultad de Ciencias Económicas



Universidad de Buenos Aires Facultad de Ciencias Económicas

“Análisis de Series Temporales Univariadas”

Modelos ARIMA (p,d,q)

Trabajo Final Estadística Actuarial

Autor: Sofía Florencia Scaiano

Profesor: Ezequiel Rodrigo Javier Del Rosso

Contenidos

I)	Introducción.....	3
II)	Marco Teórico.....	4
III)	Análisis de Resultados	5
IV)	Conclusiones	8
V)	Bibliografía.....	9

I) Introducción

“El enfoque de los modelos ARIMA se ha popularizado a partir de 1970 con el libro de Box y Jenkins “Time Series Analysis: Forecasting y Control”, de forma que actualmente se considera prácticamente sinónimas las expresiones modelos ARIMA y enfoque Box-Jenkins. La finalidad última de la construcción de los modelos ARIMA es la predicción y, más concretamente, la predicción a corto plazo, habiéndose obtenido resultados satisfactorios en distintas áreas (...) El estudio teórico de los modelos ARIMA debe complementarse con su aplicación práctica a series reales, ya que esta aplicación es la que permite profundizar en el proceso de inferencia con este tipo de modelos. La elaboración de modelos ARIMA es impensable sin la ayuda de un ordenador dada la laboriosidad de los cálculos (...) En la actualidad el panorama ha cambiado sustancialmente con la introducción en el mercado de (...) programas que permiten instrumentar la metodología ARIMA.” (Uriel, 1985)

Se expone a continuación, el análisis de una serie de tiempo de 200 (doscientas) observaciones, mediante la Metodología de Box Jenkins y con el soporte de RStudio para luego exteriorizar sus resultados y conclusiones.

II) Marco Teórico

El análisis por desarrollar en el presente trabajo se sustenta en la teoría de los procesos estocásticos, la estacionariedad y los modelos ARIMA. La divulgación y popularización de estos últimos se debe fundamentalmente a la obra de Box y Jenkins (1976), aunque también fueron estudiados anteriormente por Wold (1938) y Barlett (1946).

Una serie temporal se puede contemplar como una realización de un proceso estocástico, y éste será débilmente estacionario de primer orden si se verifica que, la esperanza matemática permanece constante a lo largo del tiempo, y es de segundo orden si la varianza es finita y permanece también constante a lo largo del tiempo y si la autocovarianza entre dos periodos distintos de tiempo, únicamente se explica por el lapso transcurrido entre esos dos periodos.

Entre los modelos lineales, se encuentran los modelos autorregresivos (AR), los modelos de medias móviles (MA), los modelos mixtos autorregresivos-medias móviles (ARMA). Luego se encuentran los procesos no estacionarios, llamados ARIMA (Modelos integrados).

$$\{ \text{AR (p)} \} \quad y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

$$\{ \text{MA (q)} \} \quad y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

$$\{ \text{ARMA (p,q)} \} \quad y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

En el tratamiento de series temporales, se suele designar a un proceso puramente aleatorio con la denominación de “ruido blanco” (ε_t) y son una pieza clave en la construcción de procesos estocásticos ARIMA. Estas variables cumplen con las siguientes propiedades:

$$E[\varepsilon_t] = 0$$

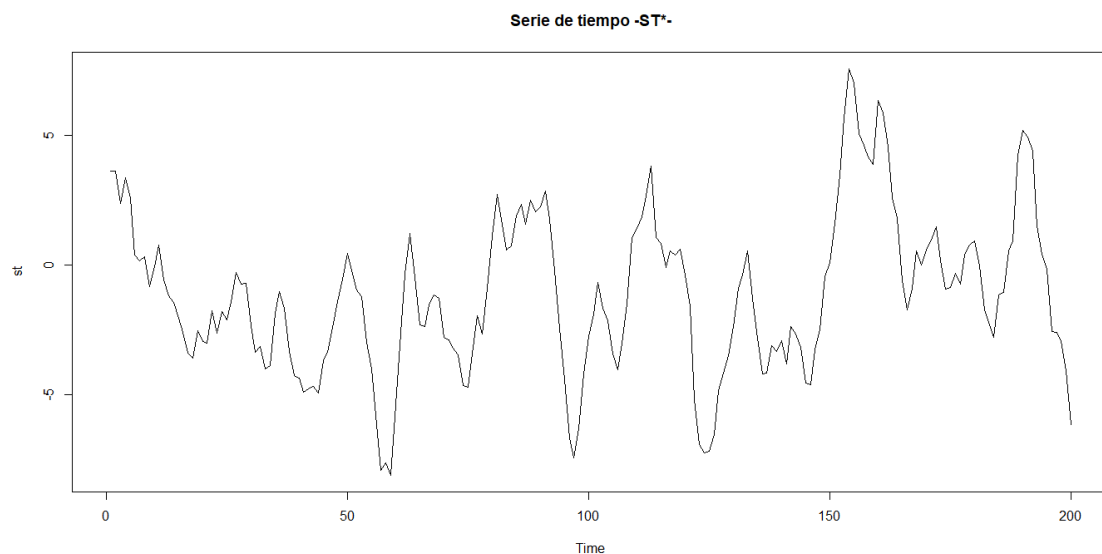
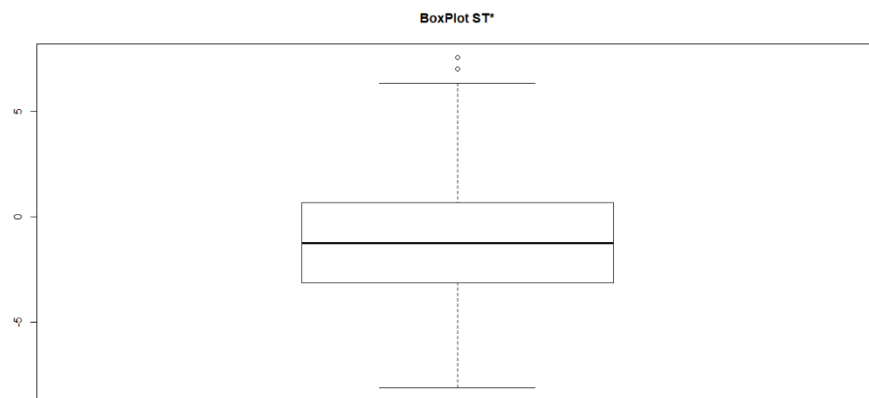
$$E[\varepsilon_t^2] = \sigma^2$$

$$E[\varepsilon_t \varepsilon_{t'}] = 0 \quad \forall t \neq t'$$

III) Análisis de Resultados

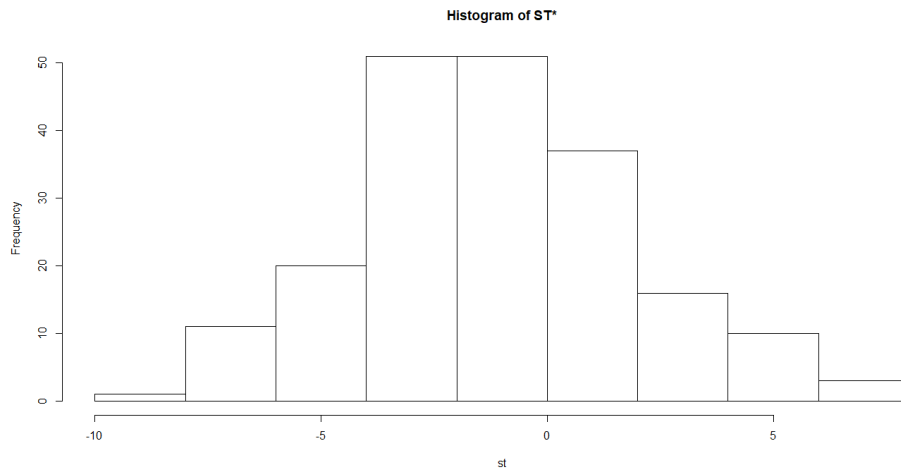
Si se realiza un análisis de estadística descriptiva a la serie de tiempo a estudiar, de ahora en más llámese ST^* , es posible observar en primer lugar, que los valores de la variable, de ahora en más “X”, se encuentran entre -8,0914683 y 7,561628. Luego, la Esperanza Matemática de ST^* , cuyo valor es -1,113829 nos indica que la serie oscila en torno a dicho valor. “El examen visual de la trayectoria de la serie a lo largo del tiempo puede dar una idea de si es o no estacionaria en media. Si existe algún valor en torno al cual la serie va oscilando, pero sin alejarse de forma permanente de dicho valor, entonces se puede considerar que la serie es estacionaria en media” (Uriel, 1985). Lo cual puede llegar a apreciarse en los siguientes gráficos que representan a la serie de tiempo:

Quantiles	
0%	-8.0914683
25%	-3.1324228
50%	-1.2366516
75%	0.6371156
100%	7.561628

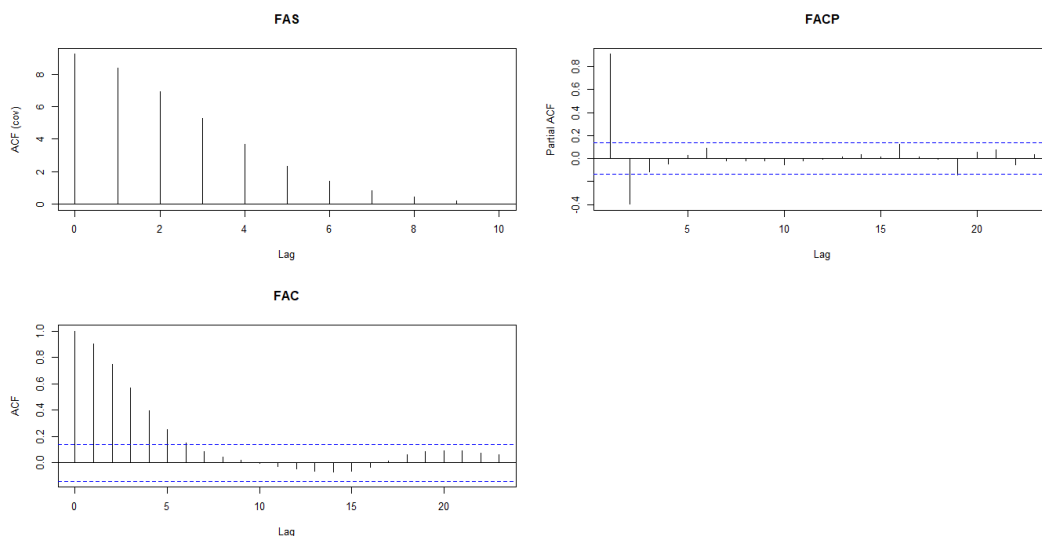


Luego, por el lado de las medidas de dispersión, se observa que la varianza de ST^* es 10,452793. Por lo tanto, el desvío estándar o desvío típico de la serie es 3,233077945. Por

último, las medidas de forma, Kurtosis y Asimetría. La primera arroja un valor de 3,036235, lo cual indica que tiene una distribución leptocúrtica, siendo que las variables se agrupan mayoritariamente alrededor del cero. Por el otro lado, el Coeficiente de Asimetría (0,2738153) nos indica que la distribución es Simétrica “Un criterio generalmente aceptado indica que si el valor numérico del Coeficiente de Asimetría se encuentra entre -0,5 y 0,5 a la distribución de frecuencias se la considera Simétrica” (Capriglioni, 2003). En el siguiente Histograma se puede observar la distribución de las variables de la serie:



Luego de realizar el análisis descriptivo de la serie, se procede a la fase de identificación y validación del modelo óptimo para la serie ST*. El objetivo perseguido al elaborar un modelo ARIMA es encontrar un modelo que sea lo mas adecuado posible para representar el comportamiento de la serie estudiada. Así un modelo ideal sería el que cumpliera los siguientes requisitos: 1) Los residuos del modelo estimado se aproximan al comportamiento de un ruido blanco. 2) El modelo estimado es estacionario e invertible. 3) Los coeficientes son estadísticamente significativos, y están poco correlacionados entre sí.



Del correlograma de la Función de Autocovarianzas (FAS) y de la Función de Autocorrelación (FAC) podemos observar como decrecen exponencialmente, lo cual exterioriza la estacionariedad de la serie y además es correcto señalar que las raíces del modelo son complejas, dando lugar a oscilaciones sinusoidales. Inmediatamente por la Función de Autocorrelación Parcial (FACP) es posible advertir que presenta ésta, valores significativos hasta el orden 2. *“En un modelo $AR(p)$, los coeficientes de autocorrelación presentan un decrecimiento rápido que puede ser geométrico puro, geométrico con alternancia de signos, sinusoidal o una mezcla de los anteriores. El tipo de decrecimiento concreto dependerá del signo y de las características de las raíces de la ecuación en diferencias del proceso generador. Por otra parte, los coeficientes de autocorrelación parcial son igual a cero, para retardos superiores p (...) En los modelos MA el comportamiento es el inverso del presentado en los modelos AR. Así en un modelo $MA(q)$ la FACT se hace cero para retardos superiores a q , por el contrario, la FACPT decrece de forma rápida, pero sin llegar a anularse nunca”.* (Uriel, 1985)

Si bien, no se aprecia una tendencia estocástica en el gráfico de la serie, (acumulación de shocks aleatorios). Se realiza de todas maneras el Test de Dicky-Fuller (test de raíces unitarias), de cuyos resultados se decide Rechazar la Hipótesis Nula, y, por lo tanto, bajo la evidencia empírica se afirma que la serie no posee tendencia estocástica.

Se analizan, luego, por los Criterios de Información, AIC y BIC, los siguientes modelos estimados:

	AR (2)	AR (1)	ARMA (1,1)
AIC	600.68833	648.61485	614.95944
BIC	610.58328	655.21149	624.85440

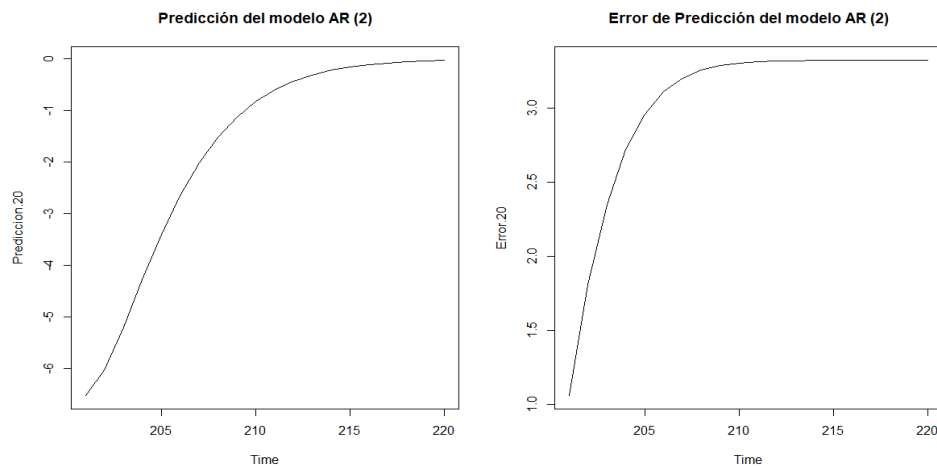
Los modelos seleccionados se estimaron con intercepto igual a cero, y todos sus coeficientes son significativos, testados por el Test de Significatividad Individual. Debido al análisis de las funciones de autocovarianzas, autocorrelaciones y autocorrelaciones parciales, es posible inferir que el modelo adecuado es un AR (2) por lo enunciado anteriormente, de la misma manera, por los Criterios de Información, tanto AIC como BIC, se infiere que el modelo adecuado es el AR (2), ya que el orden adecuado es el que minimice el AIC (Akaike). Mismo algoritmo para el criterio BIC (Schwarz). Con los métodos de identificación mencionados, podemos inferir que el modelo adecuado para la serie es un AR (2) estacionario tal que: $y_t = 1,3742y_{t-1} - 0,4752y_{t-2}$.

Luego, se testea la incorrelación y la normalidad de los residuos mediante el Box Test y el Jarque Bera Test. Bajo la evidencia empírica se puede afirmar que los errores se encuentran incorrelacionados y son normales, por lo tanto, consideramos a los residuos ruidos blancos normales, con media cero y varianza constante.

IV) Conclusiones

“Las tres primeras fases de la elaboración de un modelo ARIMA constituyen un proceso iterativo cuyo resultado final es la obtención de un modelo estimado que sea compatible con la estructura de los datos. Una vez que se ha conseguido este resultado, la fase siguiente consiste en utilizar este modelo estimado en la predicción de valores futuros de la variable objeto de estudio”. (Uriel, 1985)

Una vez validado el modelo elegido, se procede a la predicción en 1, 2, 3 y 20 períodos (t) y se obtienen los gráficos del predictor y del error correspondiente:



Del gráfico del predictor, es posible observar que a medida que crece la cantidad de períodos estimados, el valor tiende a ser nulo, ya que el modelo pierde capacidad predictiva, por tal razón, es que el error de predicción aumenta a medida que crece la cantidad de periodos estimados, como puede observarse en el gráfico.

A continuación, se expone el Intervalo de Confianza para el modelo seleccionado con una predicción de veinte períodos:

Intervalo de Confianza del modelo para la predicción de 20 períodos							
Periodo	Lo.94	Lo.95	Lo.99	Point.Forecast	Hi.94	Hi.95	Hi.99
1	-8.52350466	-8.60769225	-9.26258648	-6.523520514	-4.52353637	-4.43934877	-3.78445455
2	-9.43889354	-9.58197241	-10.6949811	-6.03987185	-2.64085016	-2.49777129	-1.38476263
3	-9.62037651	-9.80645629	-11.2539684	-5.199813327	-0.77925014	-0.59317037	0.85434176
4	-9.39252977	-9.60793698	-11.2835868	-4.275255795	0.84201818	1.05742539	2.73307526
5	-8.9738114	-9.20826976	-11.0321183	-3.403952891	2.16590561	2.40036398	4.22421252
6	-8.49861385	-8.74497511	-10.6614161	-2.645986993	3.20663987	3.45300112	5.36944208
7	-8.04196132	-8.29551541	-10.2679093	-2.018459566	4.00504219	4.25859627	6.23099017
8	-7.64007243	-7.89784652	-9.90306778	-1.516319043	4.60743434	4.86520843	6.8704297
9	-7.30552383	-7.5657088	-9.58968432	-1.124496739	5.05653035	5.31671533	7.34069085
10	-7.03765497	-7.29918451	-9.33361936	-0.824686088	5.38828279	5.64981233	7.68424718
11	-6.82928871	-7.09155188	-9.13169367	-0.598891382	5.63150594	5.89376912	7.93391091
12	-6.67080171	-6.93345727	-8.97665141	-0.43108278	5.80863615	6.07129171	8.11448585
13	-6.55239839	-6.81526001	-8.86005711	-0.30778418	5.93683003	6.19969165	8.24448875
14	-6.46523649	-6.72820452	-8.77382932	-0.218094579	6.02904734	6.29201536	8.33764016
15	-6.40186478	-6.66488688	-8.71093239	-0.153438056	6.09498867	6.35801077	8.40405628
16	-6.35628021	-6.6193294	-8.6655856	-0.107210045	6.14186012	6.40490931	8.45116551
17	-6.32379796	-6.58686052	-8.63322076	-0.074410079	6.1749778	6.43804037	8.48440061
18	-6.30084783	-6.56391691	-8.61032782	-0.051305191	6.19823745	6.46130652	8.50771744
19	-6.28475892	-6.54783113	-8.5942664	-0.035141897	6.21447512	6.47754733	8.5239826
20	-6.27356268	-6.53663637	-8.5830832	-0.023910374	6.22574193	6.48881562	8.53526245

V) Bibliografía

Capriglioni, C. (2003). Estadística Tomo 1.

Uriel, E. (1985). Análisis de Series Temporales modelos ARIMA. Valencia, España.

VI) Anexos

```
library(urca)
```

```
library(moments)
```

```
library(forecast)
```

```
library(tseries)
```

```
serie <- read.csv("serie20.csv", header=T)
```

```
ts(serie)
```

```
x<-serie$x
```

```
st <- ts(x)
```

```
par(mfrow=c(1,1))
```

```
plot(st,main=("Serie de tiempo -ST*-"))
```

```
#Análisis Descriptivo
```

```
Quantiles <- quantile(st,na.rm=T)
```

```
Promedio <- sapply(st,mean,na.rm=T)
```

```
Momentos <- all.moments(st)
```

```
Kurtosis <- kurtosis(st)
```

```
Asimetria <- skewness(st)
```

```
summary(st)
```

```
boxplot(st,main="BoxPlot ST*")
```

```
#Grafico FAS FAC FACP
```

```
par(mfrow=c(1,3))
```

```
layout(matrix(1:4,2,2))
```

```
acf(st,lag.max=10,type="covariance",plot=T,main="FAS")
```

```
acf(st,main="FAC")
```

```
pacf(st,main="FACP")
```

```
#Comparacion de modelos
```

```
none.df<-ur.df(x,type="none",lags=5,selectlags=c("AIC"))
```

```
drift.df<-ur.df(x,type="drift",lags=5,selectlags=c("AIC"))
```

```
trend.df<-ur.df(x,type="trend",lags=5,selectlags=c("AIC"))
```

```
summary(none.df) #Rechazo Ho
```

```
summary(drift.df) #Rechazo Ho
```

```
summary(trend.df) #Rechazo Ho
```

```
ar2 <- arima(x,order=c(2,0,0),include.mean=FALSE)
```

```
ar1 <- arima(x,order=c(1,0,0),include.mean=FALSE)
```

```
arma11 <- arima(x,order=c(1,0,1),include.mean=FALSE)
```

```
ar2
```

```
ar1
```

```
arma11
```

```

AICar2 <- AIC(ar2)
AICar1 <- AIC(ar1)
AICarma11 <- AIC(arma11)
BICar2 <- BIC(ar2)
BICar1 <- BIC(ar1)
BICarma11 <- BIC(arma11)

CI <- data.frame(AICar2,AICar1,AICarma11,BICar2,BICar1,BICarma11)
CIm <- matrix(CI,nrow=2,ncol=3,byrow=T)
m <- c("AR(2)","AR(1)","ARMA(1,1)")
ci <- c("AIC","BIC")
rownames(CIm) <- ci
colnames(CIm) <- m
CIm #Elijo menor AIC y menor BIC -> AR(2)

#Prediccion
Prediccion.1 <- predict(ar2,n.ahead=1)$pred
Error.1 <- predict(ar2,n.ahead=1)$se
Prediccion.2 <- predict(ar2,n.ahead=2)$pred
Error.2 <- predict(ar2,n.ahead=2)$se
Prediccion.3 <- predict(ar2,n.ahead=3)$pred
Error.3 <- predict(ar2,n.ahead=3)$se
Prediccion.20 <- predict(ar2,n.ahead=20)$pred
Error.20 <- predict(ar2,n.ahead=20)$se

par(mfrow=c(2,2))
layout(matrix(1:2,1,2))
plot(Prediccion.20,main="Predicción del modelo AR (2)")
plot(Error.20, main="Error de Predicción del modelo AR (2)")

#Intervalos de Confianza
IC.P1 <- forecast(ar2,h=1,level= c(0.94,0.95,0.99))

```

```
IC.P2 <- forecast(ar2,h=2,level= c(0.94,0.95,0.99))
```

```
IC.P3 <- forecast(ar2,h=3,level= c(0.94,0.95,0.99))
```

```
IC.P20 <- forecast(ar2,h=20,level= c(0.94,0.95,0.99))
```

```
#DataFrame
```

```
IC.DF <- data.frame(IC.P20)
```

```
IC.DF <- as.data.frame (IC.DF[,c(2,4,6,1,3,5,7)])
```

```
View(IC.DF)
```

```
write.csv(IC.DF,"C:\\Users\\soofi\\Documents\\tp final EA\\ic.csv")
```

```
#Análisis de la serie
```

```
Box.test(residuals(ar2)) #No Rechazo Ho -> Incorrelacion/Independencia
```

```
jarque.bera.test(residuals(ar2)) #No Rechazo Ho -> Residuos normales
```