Series Temporales - Trabajo Final

Esteban Schab - Anabella De Battista

Febrero 2021

Contents

Ejercicio 1 - Modelos ARMA

Se requiere:

a. Identificación del modelo empleando los primeros 140 registros mediante ACF, PACF y AIC.

b. Realizar un pronóstico out-of sample, sobre los últimos 10 datos. Determinar la precisión del pronóstico de los dos mejores modelos calculados en en apartado (a), mediante el error cuadrático medio (mean squared error) y el error absoluto medio (mean absolute error).

Modelo 2

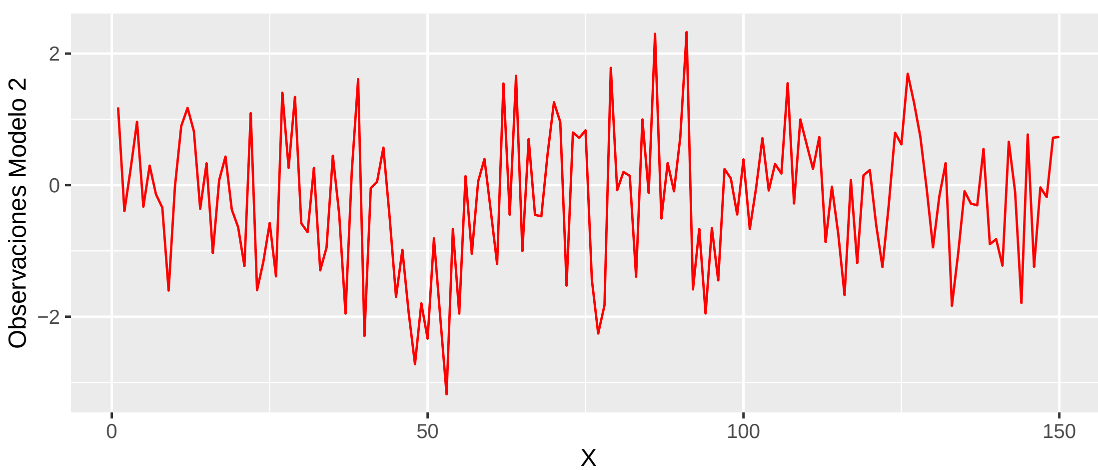
```r
modelo2 <- read.csv("./data/modelo2.csv", header = TRUE, sep = ",", quote = "\"",
                    dec = ".", fill = TRUE)
m2_ts <- ts(modelo2[,2])
```

Estadísticas principales

```r
summary(m2_ts)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -3.17920 -0.94755 -0.09789 -0.21507  0.53178  2.32752
```

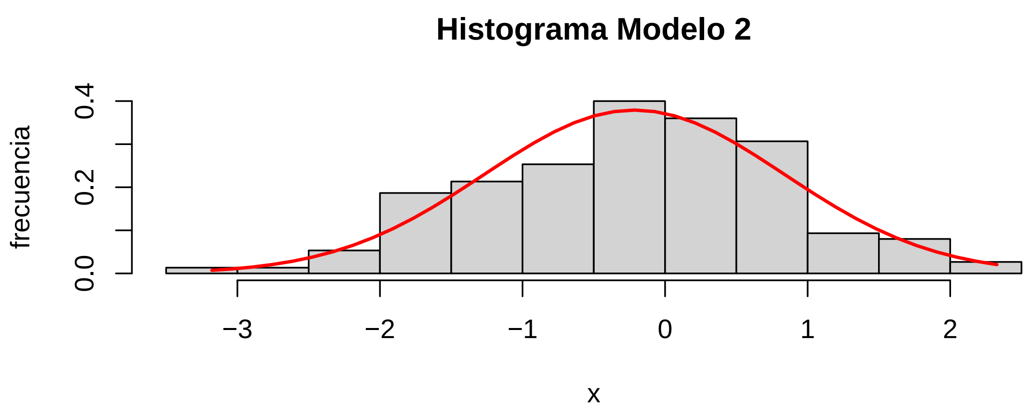
Gráfico de la serie temporal



En base al gráfico la serie aparenta ser constante en media y en varianza, por lo que se podría pensar que es estacionaria.

Histograma

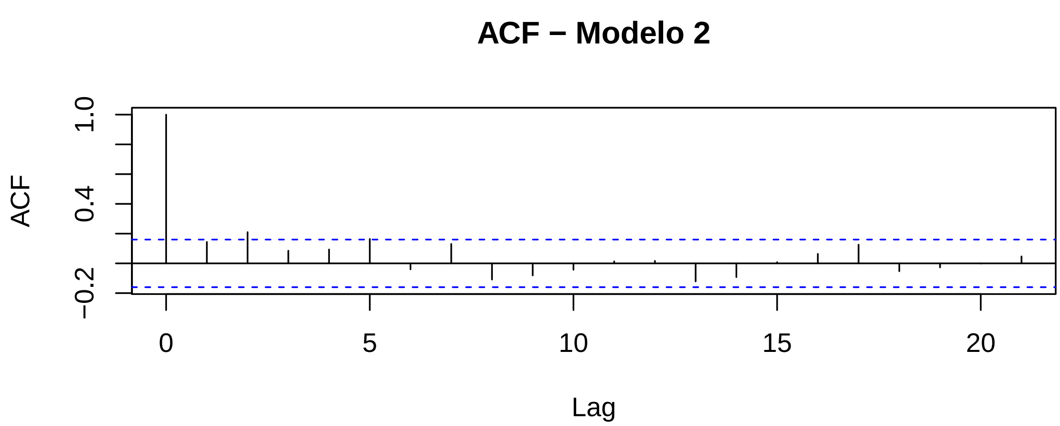
```
hist(m2_ts, prob=TRUE, main = "Histograma Modelo 2", xlab = "x", ylab = "frecuencia")
x <- seq(min(m2_ts), max(m2_ts), length = 40)
f <- dnorm(x, mean = mean(m2_ts), sd = sd(m2_ts))
lines(x, f, col = "red", lwd = 2)
```



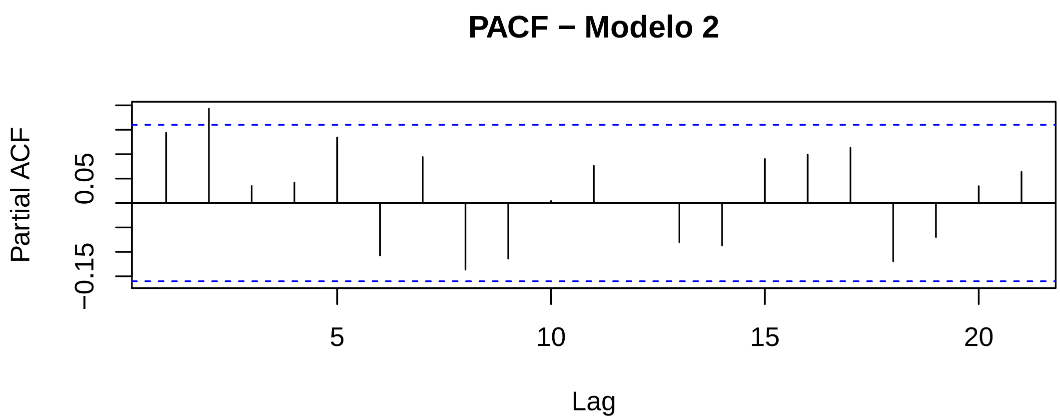
En base al histograma se ve que la distribución de los datos se ajusta a una curva normal.

Gráficas ACF y PACF

```
acf(m2_ts, main = "ACF - Modelo 2")
```



```r
pacf(m2_ts, main = "PACF - Modelo 2")
```



De acuerdo al gráfico de autocorrelación se puede apreciar que la serie es estacionaria ya que la función decae a medida que aumentan los rezagos en el tiempo. No existe componente estacional.

Si se analizan las gráficas ACF y PACF se deduce que es posible plantear varios modelos tentativos para el análisis, realizando combinaciones de p y q entre 0 y 2.

Previamente se descarta que exista componente estacional en la serie de datos.

```r
fit <- tbats(m2_ts)
s <- !is.null(fit$seasonal)
s
```

```
## [1] FALSE
```

El hecho de que **s=FALSE** indica que no hay estacionalidad en la serie de tiempo modelo2.

Verificación de que la serie es estacionaria

Aplicamos el test de estacionariedad ADF (Dickey-Fuller).

Planteamiento de Hipótesis Significancia $\alpha = 0.05$ H0: La serie es no estacionaria: tiene raíz unitaria H1: La serie es estacionaria: no tiene raíz unitaria

```r
library(tseries)
adf<-adf.test(m2_ts)
adf$p.value
```

```
## [1] 0.01
```

Se aplica el test KPSS para verificar si la serie es estacionaria.

```r
kpss<-kpss.test(m2_ts)
kpss$p.value
```

```
## [1] 0.1
```

Como el p-valor del Test de Dickey-Fuller da menor que α y en el caso del test KPSS da mayor que α, concluimos que la serie *es estacionaria.*

Configuración de parámetros para procesos ARMA, variando p y q.

```r
p.maximo <- 2  # orden maximo p
q.maximo <- 2  # orden maximo q
cantidad.modelos <- (p.maximo+1)*(q.maximo+1)

#combinaciones de p y q
orden.modelos <- permutations(p.maximo+1,2,0:p.maximo,repeats.allowed = TRUE)
resumen<- matrix(0,cantidad.modelos,6)  #matriz resumen (loglik, sigma2 y aic)
colnames(resumen) <- c("Loglik","Sigma2","AIC","AIC diff",
                       "Likelihood_model","Akaike weights")
 #nombre de las filas
nombres <- matrix(0,1,cantidad.modelos)
for (modelo.numero in 1:cantidad.modelos){
  nombres[1,modelo.numero] <- paste("p=",as.character(orden.modelos[modelo.numero,1]),
                                    "q=",as.character(orden.modelos[modelo.numero,2]))
}
rownames(resumen) <- nombres
```

Prueba de los modelos con los primeros 140 valores de la serie.

```r
fit_m2_ts <- window(m2_ts, start=1, end=140)

residuos.modelos <- matrix(0,length(fit_m2_ts),cantidad.modelos)
colnames(residuos.modelos)<- nombres
celda <- 0
for(p in 0:p.maximo){
  for(q in 0:q.maximo){
    celda= celda+1
    AR2 <- arima(fit_m2_ts, order = c(p,0,q))
    residuos.modelos[,celda] <- AR2$residuals
    resumen[celda,1] <- AR2$loglik
    resumen[celda,2] <- AR2$sigma2
    resumen[celda,3] <- AR2$aic
  }}
```

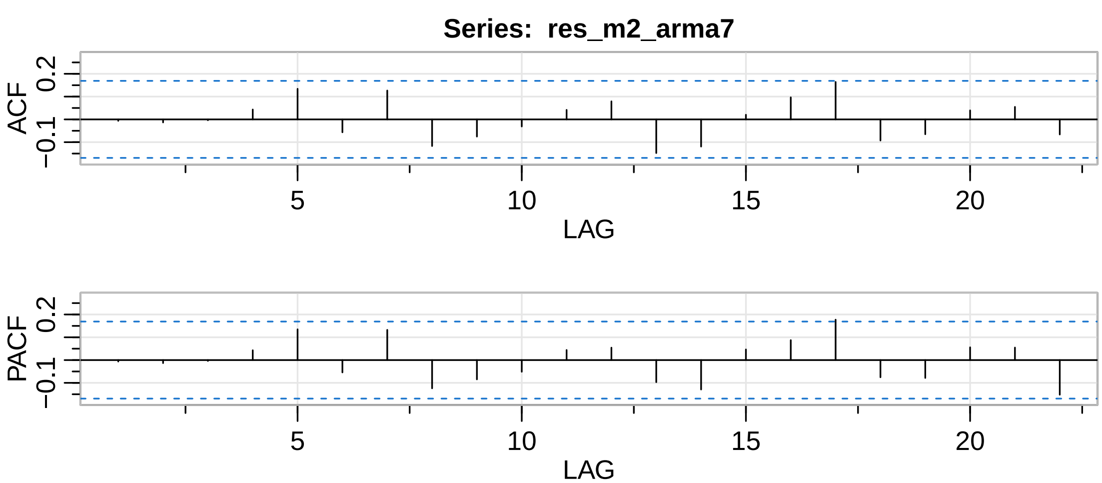
```r
resumen[,4]<-resumen[,3]-min(resumen[,3])
resumen[,5]<-exp(-0.5*resumen[,4])
resumen[,6]<-resumen[,5]/sum(resumen[,5])

# resumen de resultados
round(resumen,digits=4)
```

```
##              Loglik Sigma2      AIC AIC diff Likelihood_model Akaike weights
## p= 0 q= 0 -206.6337 1.1208 417.2675   5.2932           0.0709         0.0187
## p= 0 q= 1 -205.2406 1.0986 416.4813   4.5071           0.1050         0.0277
## p= 0 q= 2 -202.5139 1.0561 413.0277   1.0535           0.5905         0.1560
## p= 1 q= 0 -204.6809 1.0897 415.3618   3.3876           0.1838         0.0486
## p= 1 q= 1 -202.2367 1.0518 412.4735   0.4993           0.7791         0.2058
## p= 1 q= 2 -201.8525 1.0459 413.7051   1.7309           0.4209         0.1112
## p= 2 q= 0 -201.9871 1.0480 411.9742   0.0000           1.0000         0.2642
## p= 2 q= 1 -201.7812 1.0448 413.5625   1.5882           0.4520         0.1194
## p= 2 q= 2 -201.6831 1.0434 415.3663   3.3921           0.1834         0.0484
```

En base a los resultados se aprecia que para el dasaset *modelo2*, los dos mejores modelos son: arma7: ARMA(2,0) cuyo valor para *AIC diff* es = 0.0000 y arma5: ARMA(1,1) cuyo valor para *AIC diff* es = 0.4993.

Para validar el modelo se grafica el correlograma de los residuos para comprobar que son ruido blanco.

```r
# Calculo ACF y PACF para el mejor modelo ARMA para el dataset modelo2
res_m2_arma7 <- residuos.modelos[,7]
acf2(res_m2_arma7)
```




```
##       [,1]  [,2] [,3] [,4] [,5]  [,6] [,7]  [,8]  [,9] [,10] [,11] [,12] [,13]
## ACF  -0.01 -0.01    0 0.04 0.13 -0.06 0.13 -0.12 -0.08 -0.03  0.04  0.08 -0.15
## PACF -0.01 -0.01    0 0.04 0.13 -0.05 0.13 -0.12 -0.08 -0.05  0.04  0.05 -0.10
```

```
##      [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22]
## ACF  -0.12  0.02  0.10  0.16 -0.09 -0.07  0.04  0.05 -0.07
## PACF -0.13  0.05  0.09  0.18 -0.08 -0.08  0.06  0.05 -0.15
```

En los correlogramas se aprecia que no hay ningún rezago significativo por lo que se puede decir que los residuos son ruido blanco.

Prueba con la función auto.arima

La función auto.arima permite determinar el mejor modelo ARIMA para un conjunto de datos. Al ejecutarla en este caso coinicide con el modelo identificado previamente.

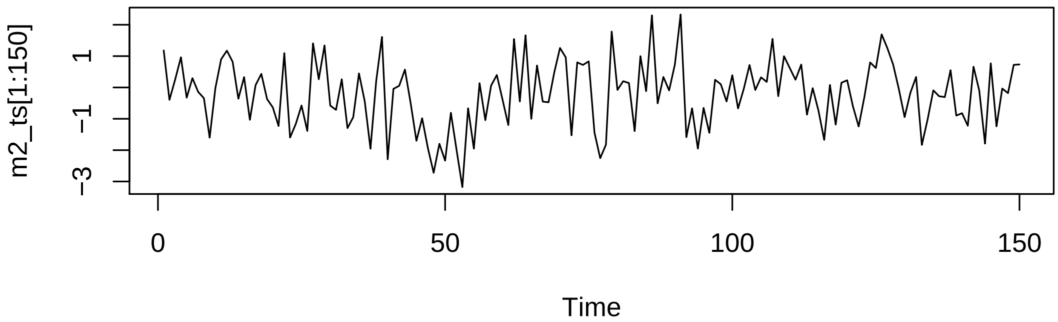
```r
auto.arima(fit_m2_ts, stepwise = FALSE, approximation = FALSE)
```

```
## Series: fit_m2_ts
## ARIMA(2,0,0) with non-zero mean
##
## Coefficients:
##          ar1     ar2     mean
##       0.1338  0.1946  -0.2174
## s.e.  0.0830  0.0830   0.1281
##
## sigma^2 estimated as 1.071:  log likelihood=-201.99
## AIC=411.97   AICc=412.27   BIC=423.74
```

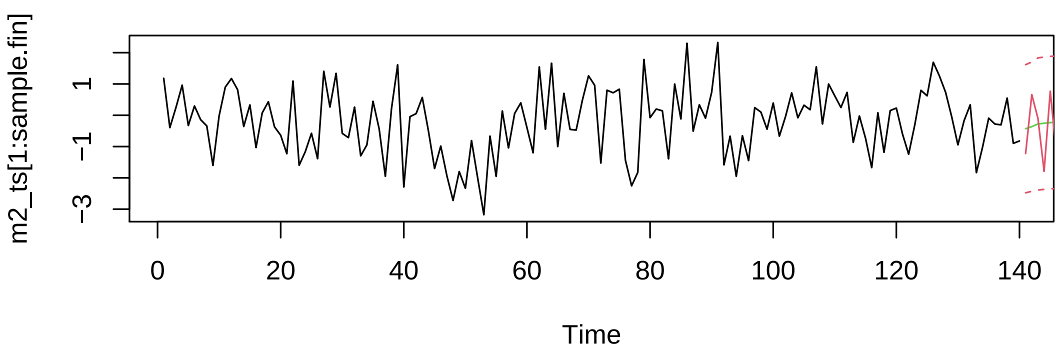
Por lo tanto estamos en condiciones de realizar un pronóstico para las siguientes 10 observaciones de la serie.

Pronóstico para Modelo 2 con ARMA(2,0)

```r
#parametros p y q del mejor modelo
p <- 2
q <- 0
#fiteo el modelo con 140 datos (para despues hacer el out-of-sample forecast)
sample.fin <- 140
AR2_20 <- arima(m2_ts[1:sample.fin], order = c(p,0,q))
ts.plot(m2_ts[1:150])
```



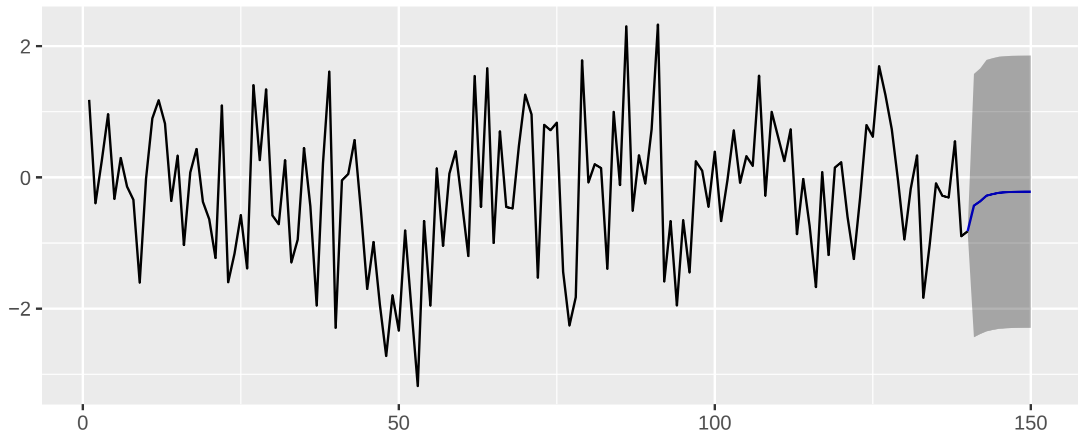

```
#plotting the m2_ts series plus the forecast and 95% prediction intervals
ts.plot(m2_ts[1:sample.fin], xlim = c(1, sample.fin))
AR2_20_forecast <- predict(AR2_20, n.ahead = length(m2_ts)-sample.fin)$pred
AR2_20_forecast_se <- predict(AR2_20, n.ahead = length(m2_ts)-sample.fin)$se
points(AR2_20_forecast, type = "l", col = 3)
points(AR2_20_forecast - 2*AR2_20_forecast_se, type = "l", col = 2, lty = 2)
points(AR2_20_forecast + 2*AR2_20_forecast_se, type = "l", col = 2, lty = 2)
points(sample.fin+1:150,m2_ts[sample.fin+1:150], type = "l", col = 2)
```



Otro método para pronóstico con ARMA(2,0)

```
library(forecast)
arima7<- arima(m2_ts[1:sample.fin], order = c(2,0,0))
```

```r
forecast7<-forecast(arima7, level = c(95), h = 10)
autoplot(forecast7)
```



Comprobar los errores de los pronósticos con ARMA(2,0)

```r
error(forecast=AR2_20_forecast, true=m2_ts[141:150], method=c("mse"))
```

```
## [1] 0.793164
```

```r
error(forecast=AR2_20_forecast, true=m2_ts[141:150], method=c("mae"))
```

```
## [1] 0.766441
```

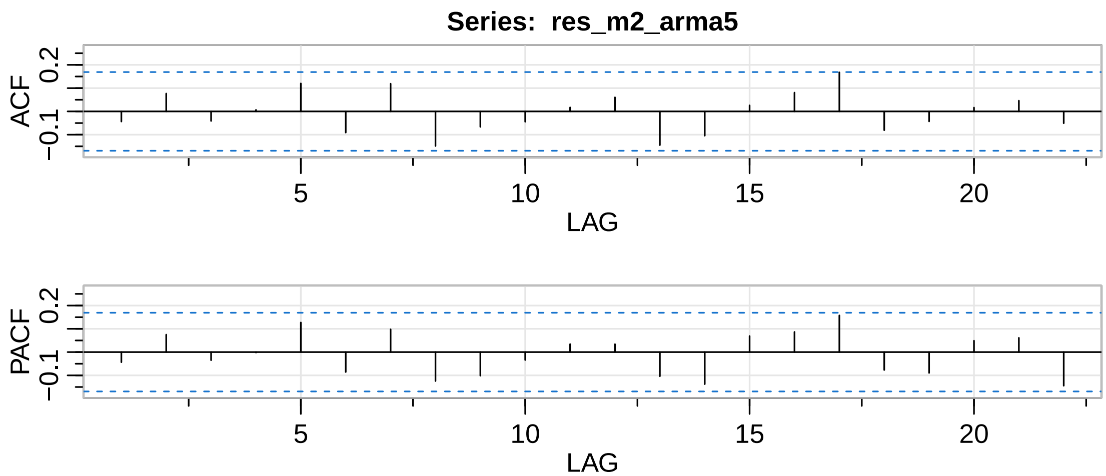
```r
error(forecast=AR2_20_forecast, true=m2_ts[141:150], method=c("mape"))
```

```
## [1] 151.0832
```

Prueba del segundo mejor modelo ARMA(1,1)

Para validar el modelo se grafica el correlograma de los residuos para comprobar que son ruido blanco.

```
# Calculo ACF y PACF para el mejor modelo ARMA para el dataset modelo2
res_m2_arma5 <- residuos.modelos[,5]
acf2(res_m2_arma5)
```



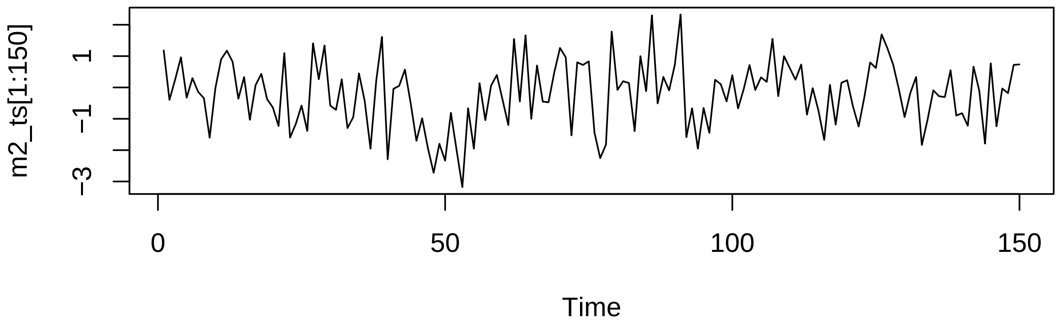

```
##       [,1] [,2]  [,3] [,4] [,5]  [,6] [,7]  [,8]  [,9] [,10] [,11] [,12] [,13]
## ACF  -0.04 0.08 -0.04 0.01 0.12 -0.09 0.12 -0.15 -0.07 -0.04  0.02  0.06 -0.15
## PACF -0.04 0.08 -0.04 0.00 0.13 -0.09 0.10 -0.12 -0.10 -0.03  0.03  0.03 -0.10
##      [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22]
## ACF  -0.10  0.03  0.08  0.17 -0.08 -0.04  0.02  0.05 -0.05
## PACF -0.14  0.07  0.09  0.16 -0.08 -0.09  0.05  0.06 -0.14
```

En los correlogramas se aprecia que no hay ningún rezago significativo por lo que se puede decir que los residuos son ruido blanco.

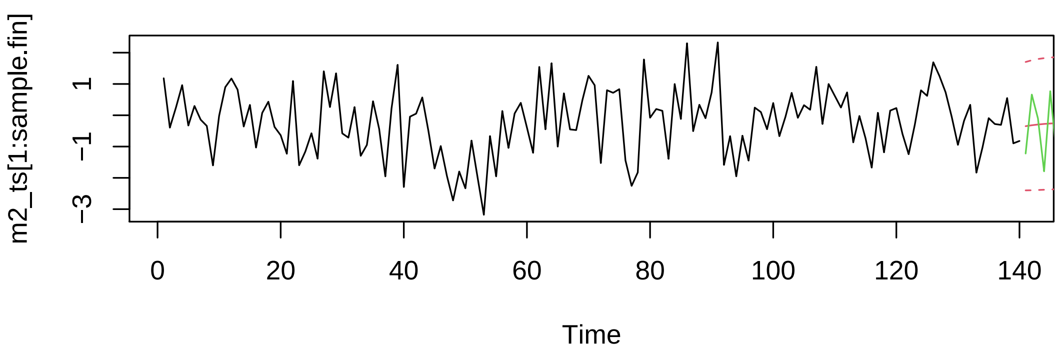
Por lo tanto estamos en condiciones de realizar un pronóstico para las siguientes 10 observaciones de la serie.

Pronóstico para Modelo 2 con ARMA(1,1)

```
#parametros p y q del mejor modelo
p <- 1
q <- 1
#fiteo el modelo con 140 datos (para despues hacer el out-of-sample forecast)
sample.fin <- 140
AR2_11 <- arima(m2_ts[1:sample.fin], order = c(p,0,q))
ts.plot(m2_ts[1:150])
```



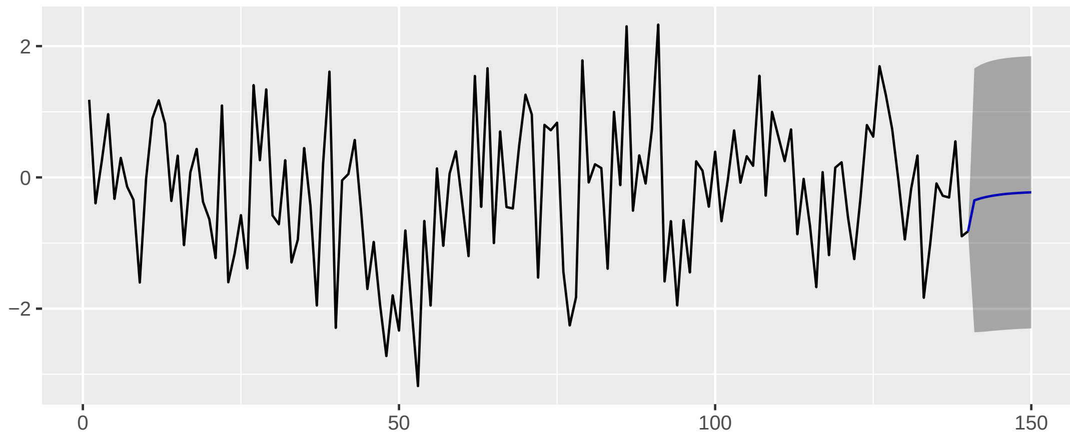

```
#plotting the m2_ts series plus the forecast and 95% prediction intervals
ts.plot(m2_ts[1:sample.fin], xlim = c(1, sample.fin))
AR2_11_forecast <- predict(AR2_11, n.ahead = length(m2_ts)-sample.fin)$pred
AR2_11_forecast_se <- predict(AR2_11, n.ahead = length(m2_ts)-sample.fin)$se
points(AR2_11_forecast, type = "l", col = 2)
points(AR2_11_forecast - 2*AR2_11_forecast_se, type = "l", col = 2, lty = 2)
points(AR2_11_forecast + 2*AR2_11_forecast_se, type = "l", col = 2, lty = 2)
points(sample.fin+1:150,m2_ts[sample.fin+1:150], type = "l", col = 3)
```



Otro método para pronóstico con ARMA(1,1)

```
library(forecast)
arima5<- arima(m2_ts[1:sample.fin], order = c(1,0,1))
```

```r
forecast5<-forecast(arima5, level = c(95), h = 10)
autoplot(forecast5)
```



Comprobar los errores de los pronósticos con ARMA(1,1)

```r
error(forecast=AR2_11_forecast, true=m2_ts[141:150], method=c("mse"))
```

```
## [1] 0.797546
```

```r
error(forecast=AR2_11_forecast, true=m2_ts[141:150], method=c("mae"))
```

```
## [1] 0.776066
```

```r
error(forecast=AR2_11_forecast, true=m2_ts[141:150], method=c("mape"))
```

```
## [1] 160.3067
```