

Assignment 2

Álvaro Rodríguez y Pablo Sanz

12/1/2021

Presentación de los datos

Los datos a tratar pertenecen al número de desempleados de España a lo largo de los años. Como se puede ver en representación inferior, podemos observar diferentes épocas que ha sufrido el país en su historia más reciente.

Si empezamos por el extremo inferior que data sobre los inicios del año 2000, observamos que el empleo se mantiene más o menos constante en cuanto a tendencia, aunque se aprecia una gran estacionalidad según si estamos en los meses de invierno o verano. Este patrón se repite hasta el año 2008 donde con la gran crisis financiera mundial, el paro subió bruscamente situándose en más de 5.000.000 de personas en el año 2012. Pasado este momento, empieza a decrecer progresivamente observándose siempre aumentos y descensos según los meses de invierno o verano respectivamente hasta finales de 2019.

En este momento, a raíz de la aparición del COVID-19, la serie rompe todos los esquemas, aumentando en muy pocos meses casi 1.000.000 de desempleados debido a la crisis sanitaria y a todas las restricciones impuestas. Unos meses después, estos altos niveles de desempleo empiezan a decrecer a gran ritmo hasta situarse en el momento actual donde el ritmo de bajada ha disminuido.



Predicción entregada

En primer lugar explicaremos el proceso que nos llevó a dar como resultado que el paro en el mes de Novimembre es de $3.240.623$.

Análisis exploratorio

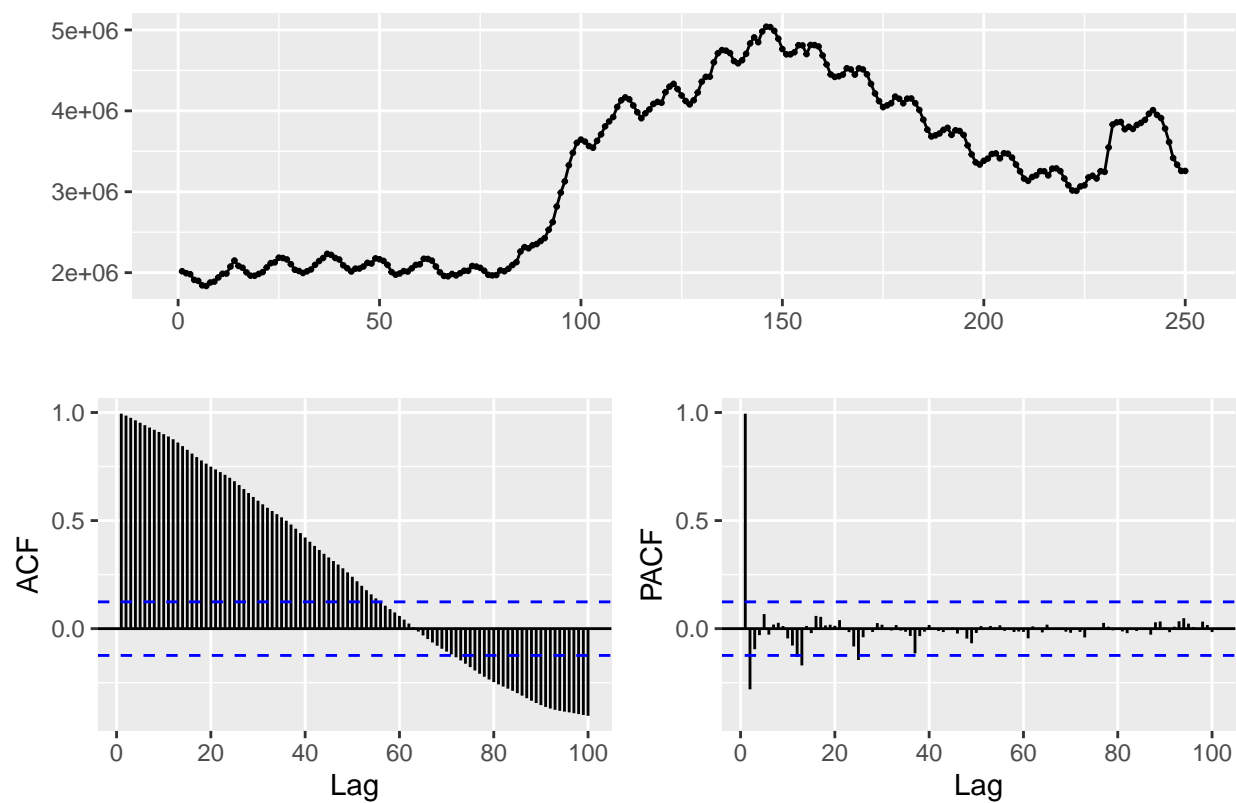
Lo primero que hicimos tras ver los datos del paro fue ver si la serie es estacionaria en varianza. Esto se ve en el siguiente gráfico:

Que refleja si es necesario realizar una transformación Box-Cox. Sin embargo, vemos que la línea azul no es muy creciente y que R-square no es muy alto ($R\text{-square}=0.2419$). Por lo tanto, no es necesario realizar una estabilización de varianza.

Seasonal ARIMA

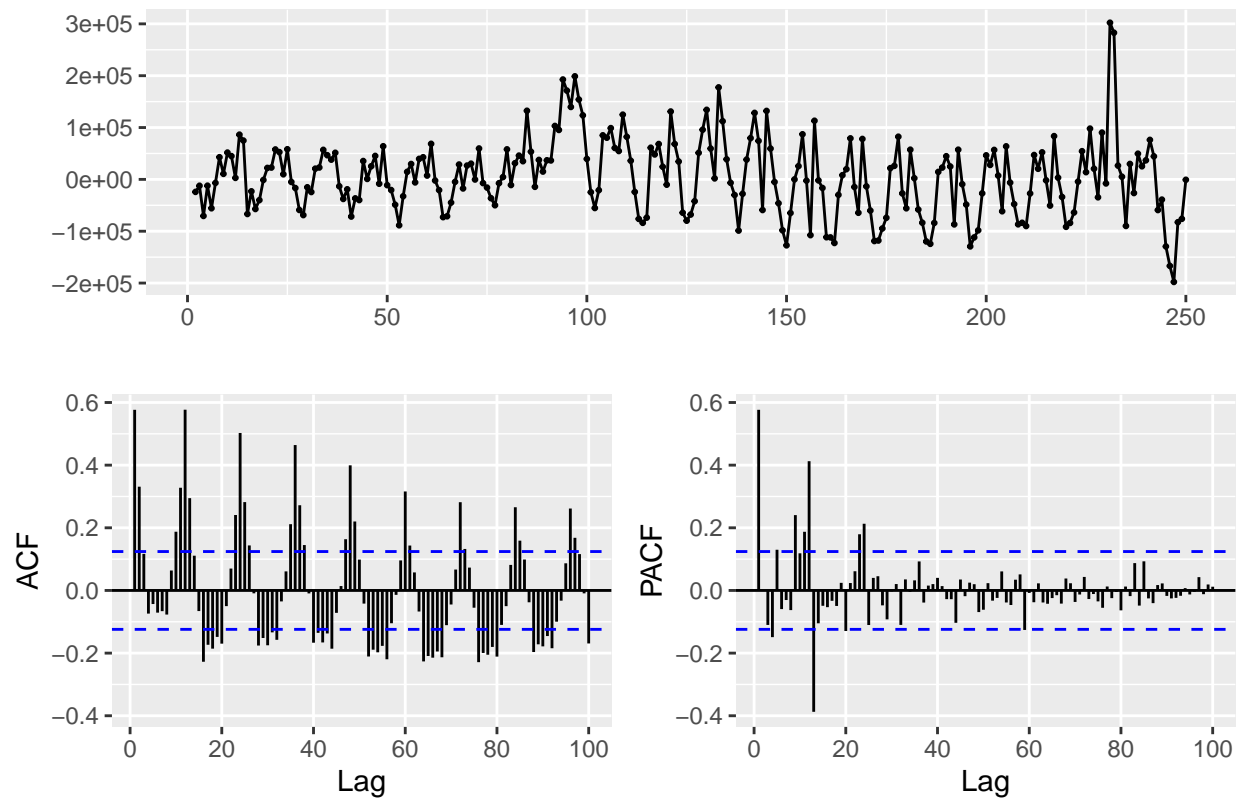
En primer lugar representamos la serie temporal junto con su ACF y PACF para inspeccionar la serie regular y estacional. Vemos que la serie necesita una diferenciación en la parte regular ya que el ACF dismunye progresivamente por el tiempo.

```
ggtsdisplay(y, lag.max = 100)
```



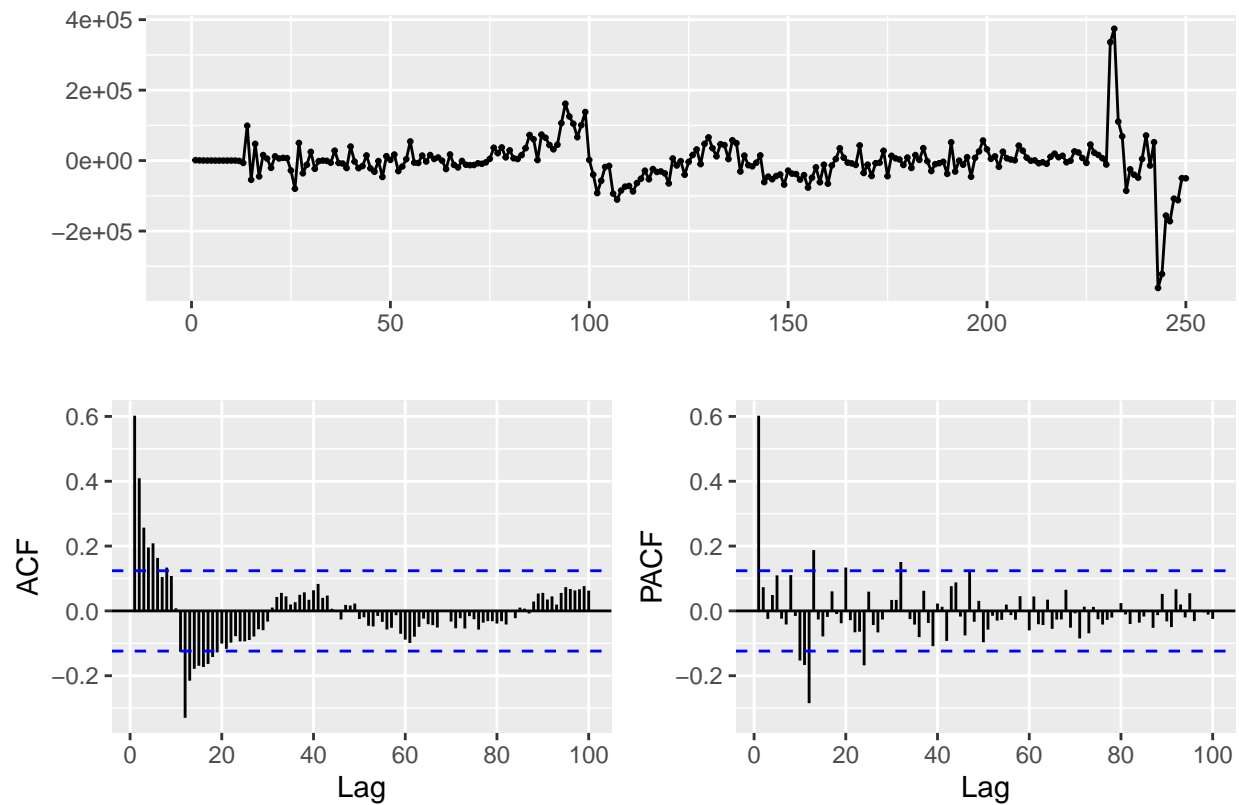
Tras haber diferenciado en la parte regular, volvemos a mostrar el ACF y el PACF. Vemos que cada 12 puntos aparece un pico y por tanto deducimos que tenemos que diferenciar en la parte estacional también.

```
Bz <- diff(y,differences = 1)
ggtsdisplay(Bz,lag.max = 100)
```



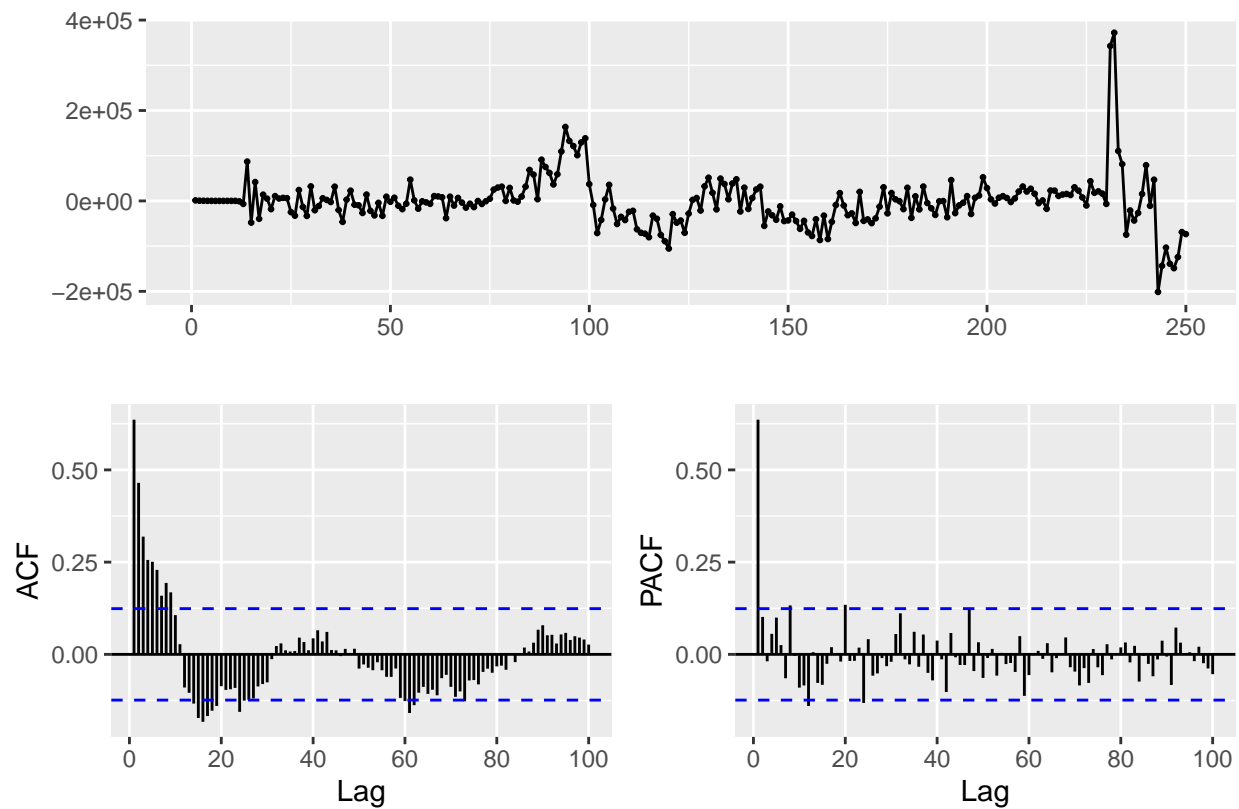
Tras haber diferenciado en la parte regular y estacional vemos los siguientes ACF y PACF.

```
arima.fit <- Arima(y,
  order=c(0,1,0),
  seasonal = list(order=c(0,1,0),period=12),
  include.constant = FALSE)
ggtsdisplay(residuals(arima.fit),lag.max = 100)
```



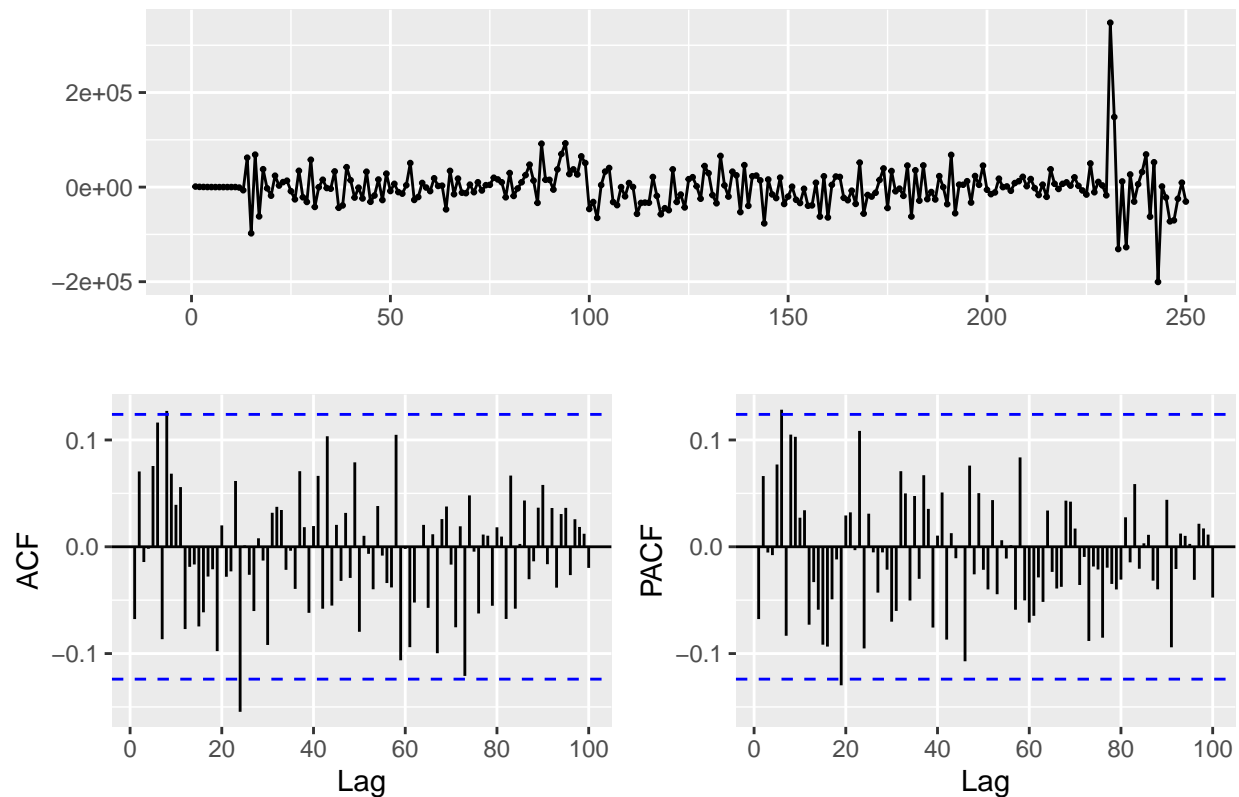
Podemos ver ahora que la parte estacional ha mejorado porque ya no se observan picos en los puntos 12, 24, 36, ... Sin embargo, observamos cómo el ACF va decreciendo progresivamente y en PACF hay un grande en el punto 12, por lo que aplicamos un AR(1)

```
arima.fit <- Arima(y,
  order=c(0,1,0),
  seasonal = list(order=c(1,1,0),period=12),
  include.constant = FALSE)
ggtsdisplay(residuals(arima.fit),lag.max = 100)
```



Ahora podemos ver que la parte estacional parece ajustada más o menos, sin embargo; la parte regular quiere una corrección. Como va decreciendo poco a poco en ACF y vemos un gran valor en el PACF, aplicamos un AR(1)

```
arima.fit <- Arima(y,
                    order=c(1,1,0),
                    seasonal = list(order=c(1,1,0),period=12),
                    include.constant = FALSE)
ggtsdisplay(residuals(arima.fit),lag.max = 100)
```



```
summary(arima.fit) # summary of training errors and estimated coefficients
```

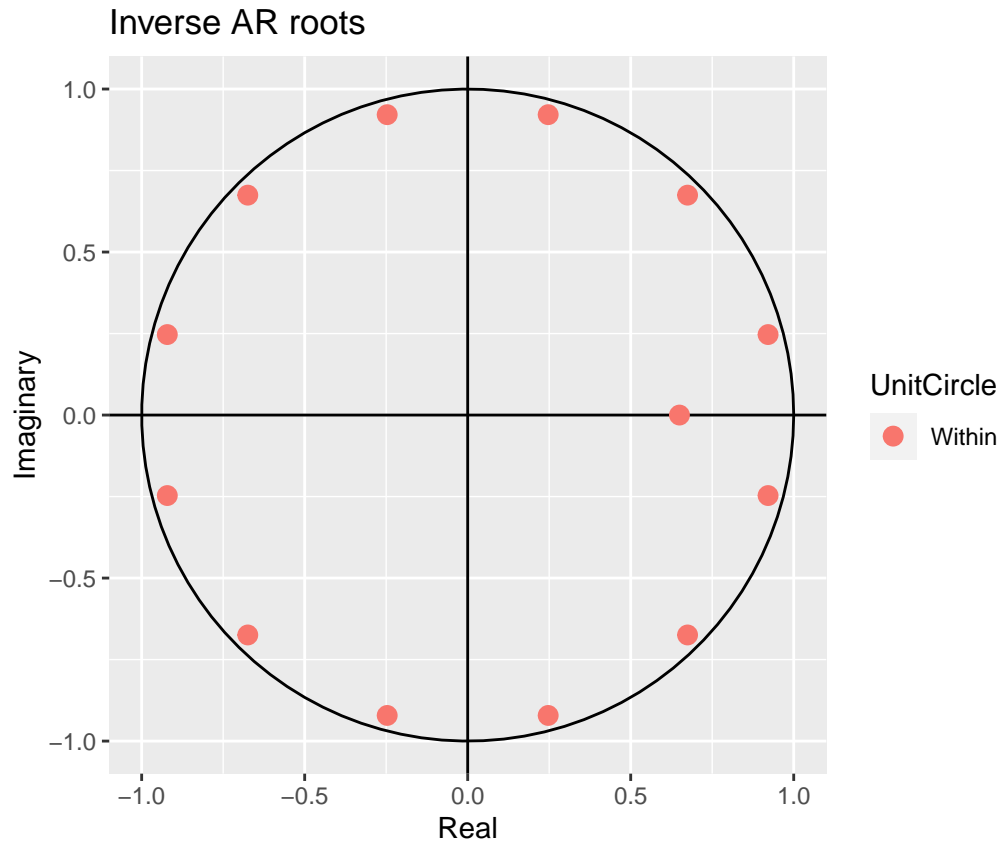
```
## Series: y
## ARIMA(1,1,0)(1,1,0)[12]
##
## Coefficients:
##      ar1      sar1
##      0.6496 -0.5673
## s.e.  0.0497  0.0663
##
## sigma^2 estimated as 1.962e+09:  log likelihood=-2873.46
## AIC=5752.93  AICc=5753.03  BIC=5763.33
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -585.4926 42948.38 27682.11 0.02010853 0.8720577 0.4961341
##              ACF1
## Training set -0.06769942
```

```
coeftest(arima.fit) # statistical significance of estimated coefficients
```

```
##
## z test of coefficients:
##
```

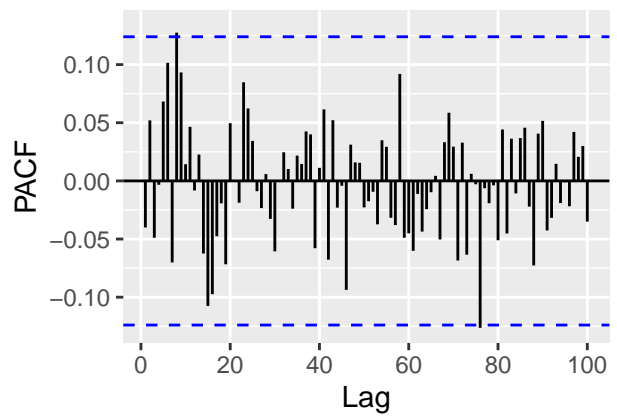
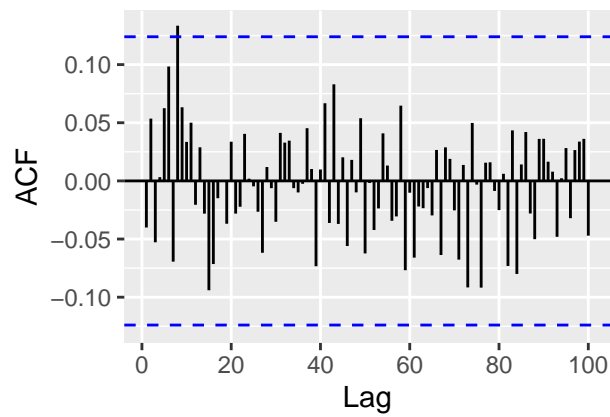
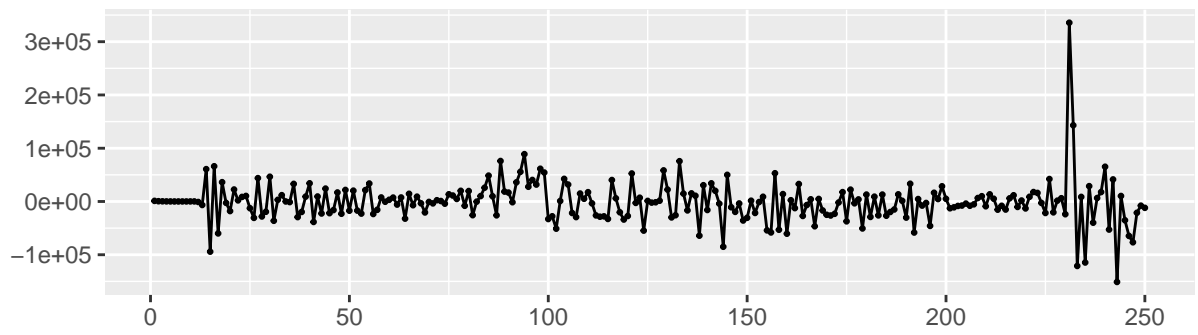
```
##      Estimate Std. Error z value Pr(>|z|)
## ar1   0.649599   0.049729 13.0629 < 2.2e-16 ***
## sar1 -0.567347   0.066343 -8.5517 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
autoplot(arima.fit) # root plot
```



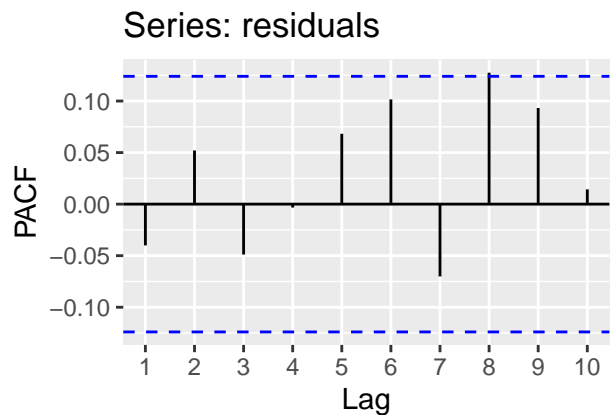
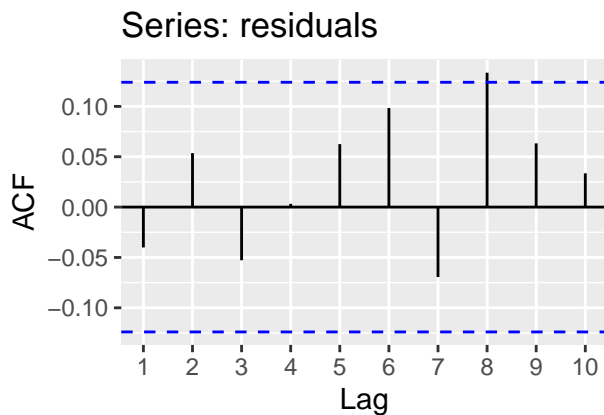
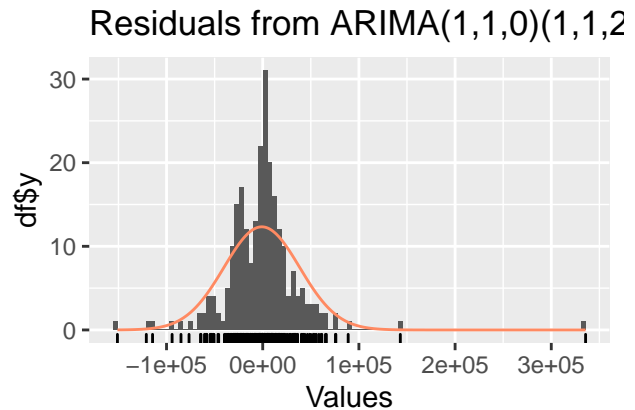
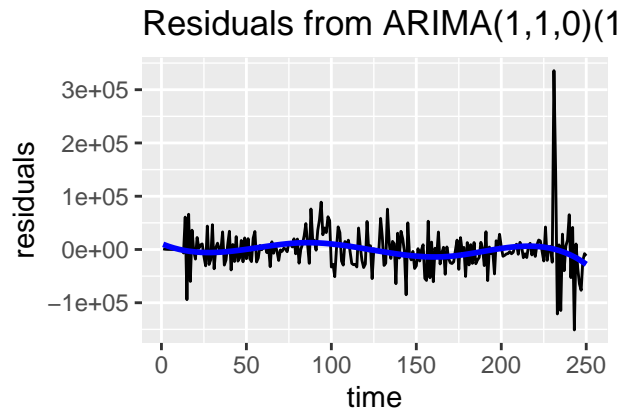
No obstante, vemos que los residuos no son perfectamente ruido blanco en la parte estacional, por este motivo intentamos ajustarlo con un MA(2).

```
arima.fit <- Arima(y,
  order=c(1,1,0),
  seasonal = list(order=c(1,1,2),period=12),
  include.constant = FALSE)
ggtsdisplay(residuals(arima.fit),lag.max = 100)
```

Comprobamos definitivamente los residuos

```
CheckResiduals.ICAI(arima.fit, bins = 100)
```



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,0)(1,1,2)[12]
## Q* = 12.58, df = 6, p-value = 0.05021
##
## Model df: 4.   Total lags used: 10
```

```
summary(arima.fit) # summary of training errors and estimated coefficients
```

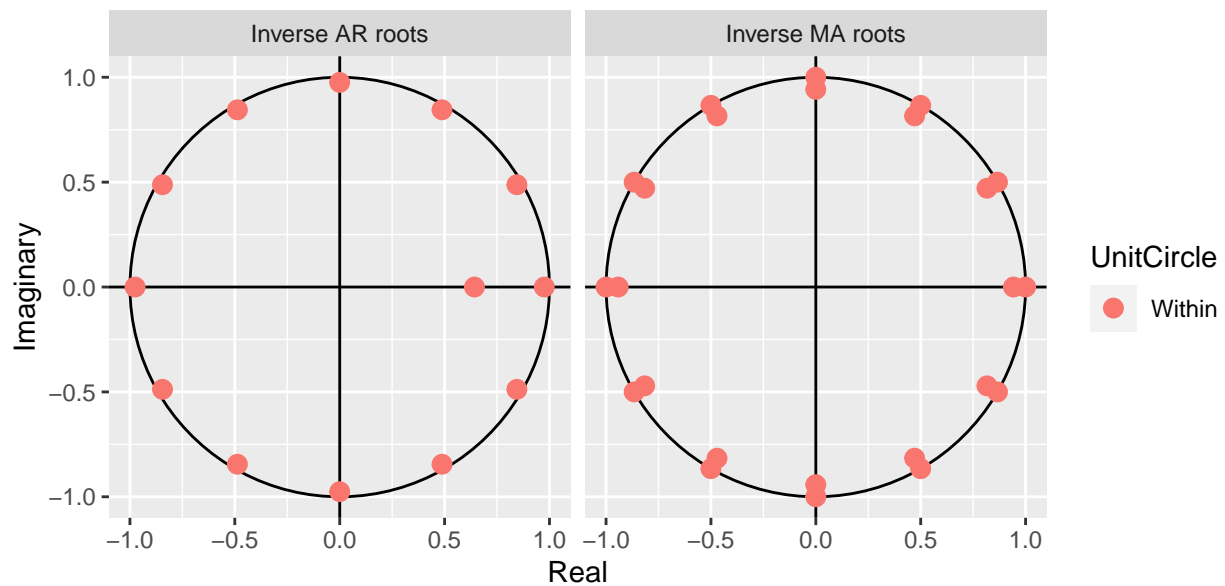
```
## Series: y
## ARIMA(1,1,0)(1,1,2)[12]
##
## Coefficients:
##      ar1      sar1      sma1      sma2
##      0.6424  0.7419  -1.4893  0.4893
## s.e.  0.0502  0.1498   0.1930  0.1664
##
## sigma^2 estimated as 1.657e+09:  log likelihood=-2862.14
## AIC=5734.29  AICc=5734.55  BIC=5751.63
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -945.3542 39302.49 24359.89 0.004428724 0.7701857 0.4365914
##
##              ACF1
```

```
## Training set -0.0400685
```

```
coeftest(arima.fit) # statistical significance of estimated coefficients
```

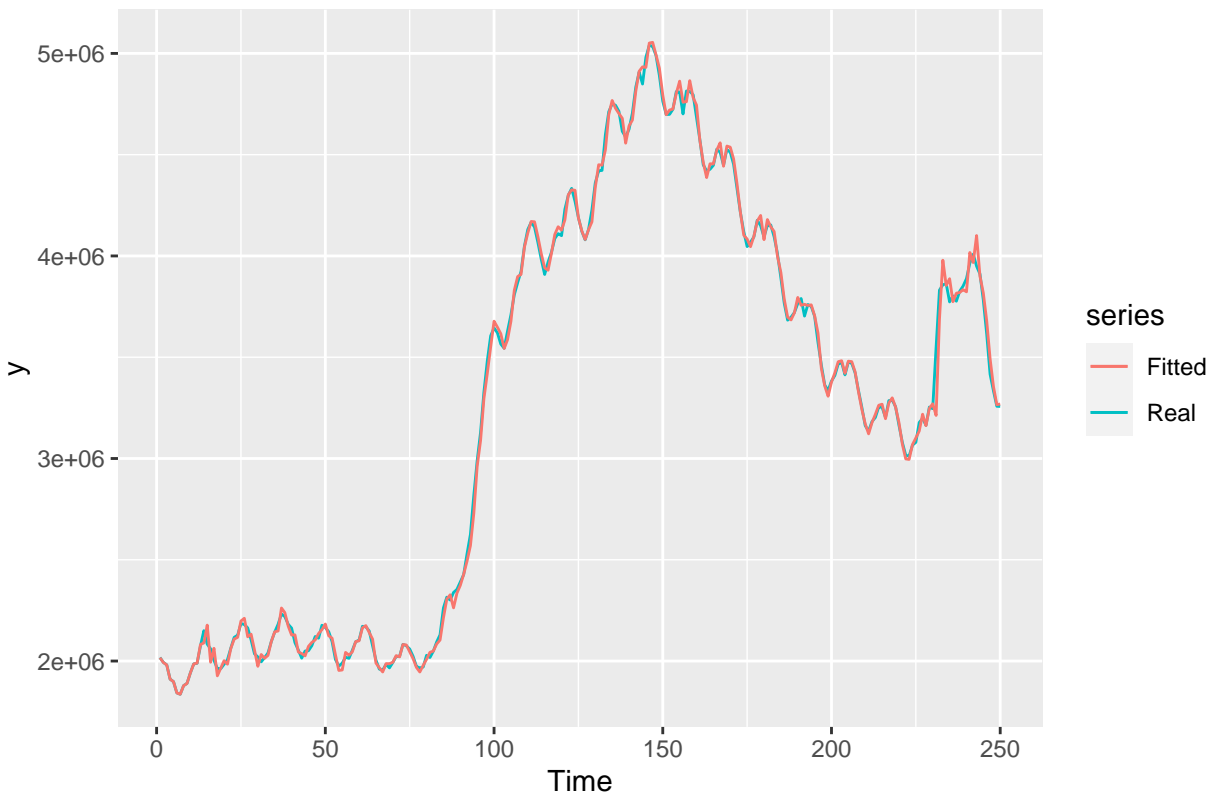
```
##  
## z test of coefficients:  
##  
##      Estimate Std. Error z value Pr(>|z|)  
## ar1    0.642436   0.050165 12.8064 < 2.2e-16 ***  
## sar1    0.741882   0.149813  4.9521 7.343e-07 ***  
## sma1   -1.489325   0.192980 -7.7175 1.186e-14 ***  
## sma2    0.489340   0.166420  2.9404 0.003278 **  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
autoplot(arima.fit) # root plot
```



Vemos que los el ruido generado es ruido blanco. Tras ello comprobamos que el modelo genere datos predichos que se adapten a los datos reales. Eso lo podemos ver en el siguiente gráfico:

```
autoplot(y, series = "Real")+  
  forecast::autolayer(arima.fit$fitted, series = "Fitted")
```



Como podemos observar, la curva azul y la roja son muy parecidas y se adapta bien a todos los cambios. Lo que es muy satisfactorio. Por ello predecimos obteniendo el resultado enviado para el assignment

```
y_est <- forecast(arima.fit, h=1)
y_est[("mean")]
```

```
## $mean
## Time Series:
## Start = 251
## End = 251
## Frequency = 1
## [1] 3240623
```

Sin embargo, observamos cómo el ACF va decreciendo progresivamente en forma senoidal y en PACF hay dos picos grande en el punto 12 y 13, por lo que aplicamos un AR(2)