Examen 14 enero

Marta Venegas Pardo

14/1/2022

```
library(TSA)
```

```
## Warning: package 'TSA' was built under R version 4.1.2
```

```
##
## Attaching package: 'TSA'
```

```
## The following objects are masked from 'package:stats':
##
##     acf, arima
```

```
## The following object is masked from 'package:utils':
##
##     tar
```

```
library(tseries)
```

```
## Warning: package 'tseries' was built under R version 4.1.2
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
# install.packages('forecast',dependencies = TRUE)
library(forecast)
```

```
## Warning: package 'forecast' was built under R version 4.1.2
```

```
## Registered S3 methods overwritten by 'forecast':
##   method       from
##   fitted.Arima TSA
##   plot.Arima   TSA
```

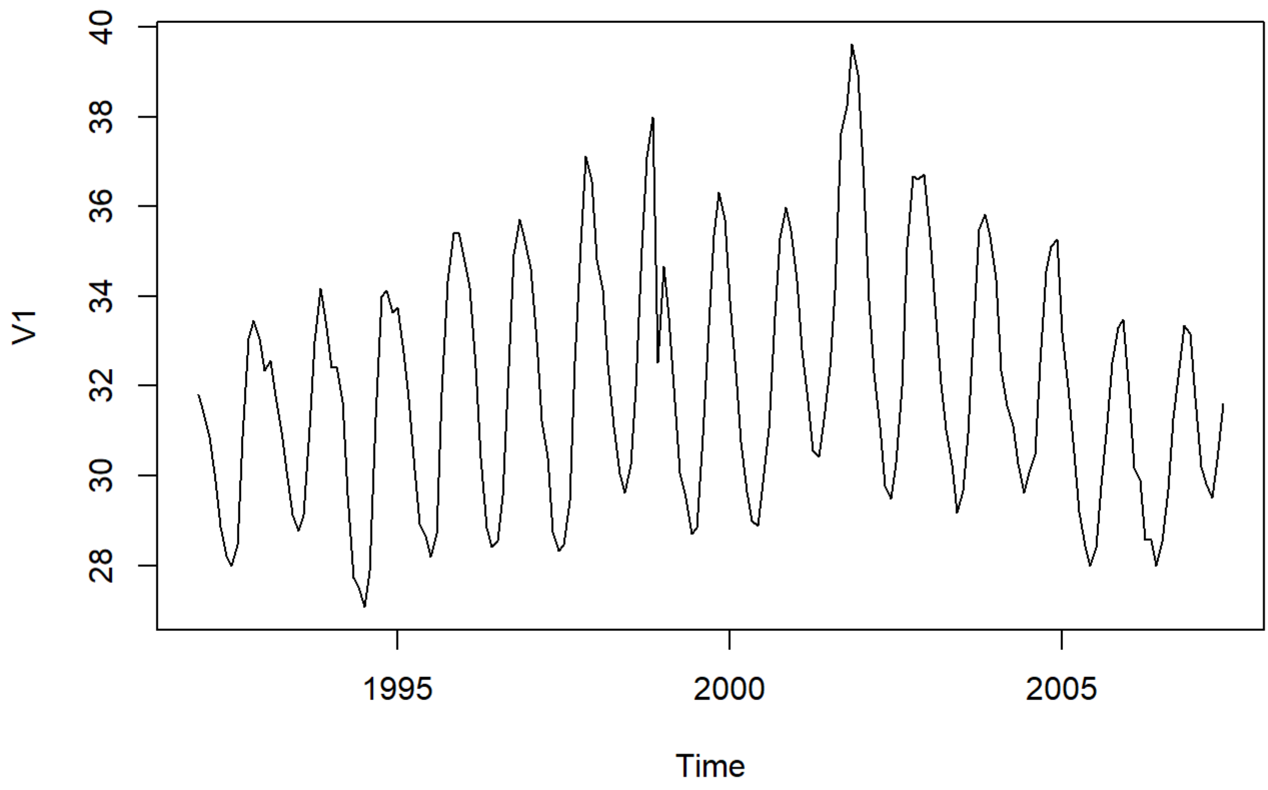
Lectura y representación de los datos

```
datos <-read.table("Luxemburgo.txt")
serie <- ts(datos,start = c(1992,1),frequency = 12)
serie
```

```
##         Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec
## 1992 31.81 31.40 30.86 29.88 28.85 28.22 28.00 28.46 30.87 33.06 33.45 33.05
## 1993 32.35 32.56 31.79 30.97 30.11 29.13 28.76 29.12 31.06 33.00 34.17 33.46
## 1994 32.41 32.43 31.66 29.57 27.73 27.51 27.07 27.87 31.29 34.00 34.12 33.64
## 1995 33.74 32.75 31.71 30.14 28.93 28.66 28.19 28.71 31.85 34.33 35.40 35.42
## 1996 34.78 34.19 32.42 30.39 28.83 28.41 28.54 29.55 32.44 34.94 35.72 35.25
## 1997 34.63 32.99 31.22 30.39 28.75 28.33 28.46 29.45 32.55 35.17 37.13 36.55
## 1998 34.83 34.15 32.48 31.10 30.04 29.62 30.25 31.99 35.06 37.09 38.00 32.51
## 1999 34.66 33.52 31.66 30.07 29.50 28.71 28.83 30.89 32.95 35.35 36.32 35.72
## 2000 33.91 32.30 30.76 29.65 29.00 28.88 29.85 31.04 33.54 35.28 35.99 35.47
## 2001 34.36 32.82 31.68 30.55 30.42 31.25 32.42 34.13 37.63 38.24 39.61 38.94
## 2002 36.92 33.91 32.30 31.09 29.78 29.49 30.28 31.98 35.09 36.67 36.61 36.71
## 2003 35.33 33.80 32.06 31.03 30.24 29.18 29.67 30.93 33.51 35.49 35.82 35.33
## 2004 34.38 32.34 31.56 31.11 30.27 29.62 30.07 30.47 32.52 34.56 35.11 35.27
## 2005 33.19 31.99 30.75 29.19 28.43 28.00 28.40 29.78 31.15 32.53 33.30 33.47
## 2006 31.83 30.18 29.92 28.57 28.56 27.98 28.52 29.69 31.26 32.39 33.35 33.17
## 2007 31.75 30.21 29.81 29.51 30.36 31.60
```

Apartado 1

```
plot(serie)
```



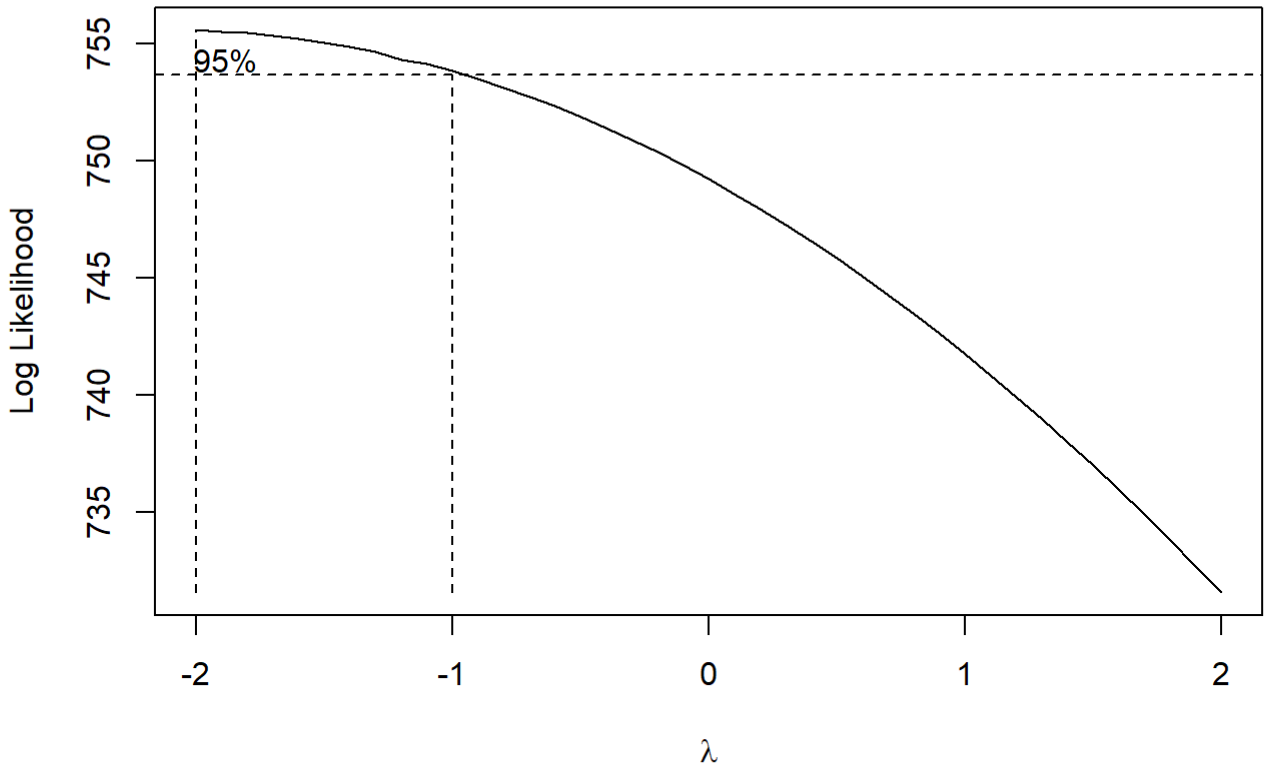

```
#ndiffs(serie)
#nsdiffs(serie)
```

Observando la gráfica podemos ver que las oscilaciones van creciendo cada año, lo que indicaría que la varianza no es constante en el tiempo (no homogeneidad de la varianza).

Sin embargo, no apreciamos tencencia al observar la gráfica, por lo que podemos considerar que los datos son constantes en media.

Vamos a hacer una transformación para conseguir homogeneidad en la varianza de los datos. Buscamos en la familia de transformaciones de Box Cox

```
bc=BoxCox.ar(y=serie, # lambda=seq(-3,3,0.01)
             )
```




```
bc$mle
```

```
## [1] -2
```

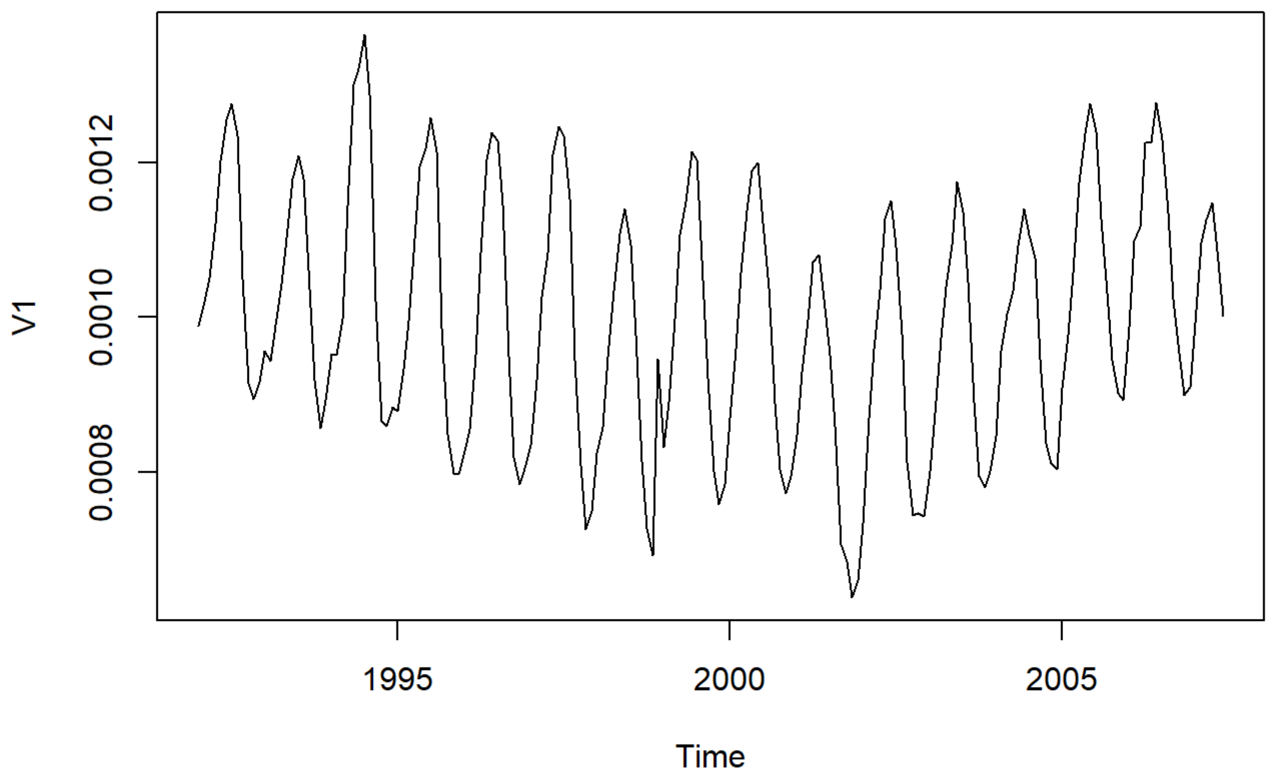
```
bc
```

```
## $lambda
##  [1] -2.0 -1.9 -1.8 -1.7 -1.6 -1.5 -1.4 -1.3 -1.2 -1.1 -1.0 -0.9 -0.8 -0.7 -0.6
## [16] -0.5 -0.4 -0.3 -0.2 -0.1  0.0  0.1  0.2  0.3  0.4  0.5  0.6  0.7  0.8  0.9
## [31]  1.0  1.1  1.2  1.3  1.4  1.5  1.6  1.7  1.8  1.9  2.0
##
## $loglike
##  [1] 755.5663 755.5181 755.4418 755.3376 755.1977 755.0448 754.8557 754.6385
##  [9] 754.3072 754.1190 753.8149 753.4853 753.1229 752.7339 752.3162 751.8694
## [17] 751.3940 750.8900 750.3570 749.7956 749.2053 748.5863 747.9392 747.2635
## [25] 746.5595 745.8272 745.0670 744.2786 743.4625 742.6187 741.7474 740.8487
## [33] 739.9229 738.9701 737.9907 736.9850 735.9523 734.8936 733.8090 732.6988
## [41] 731.5632
##
## $mle
## [1] -2
##
## $ci
## [1] -2 -1
```

Nos sugiere una transformación con lambda = -2.

La transformación es: $\dfrac{x^{\lambda}-1}{\lambda}$

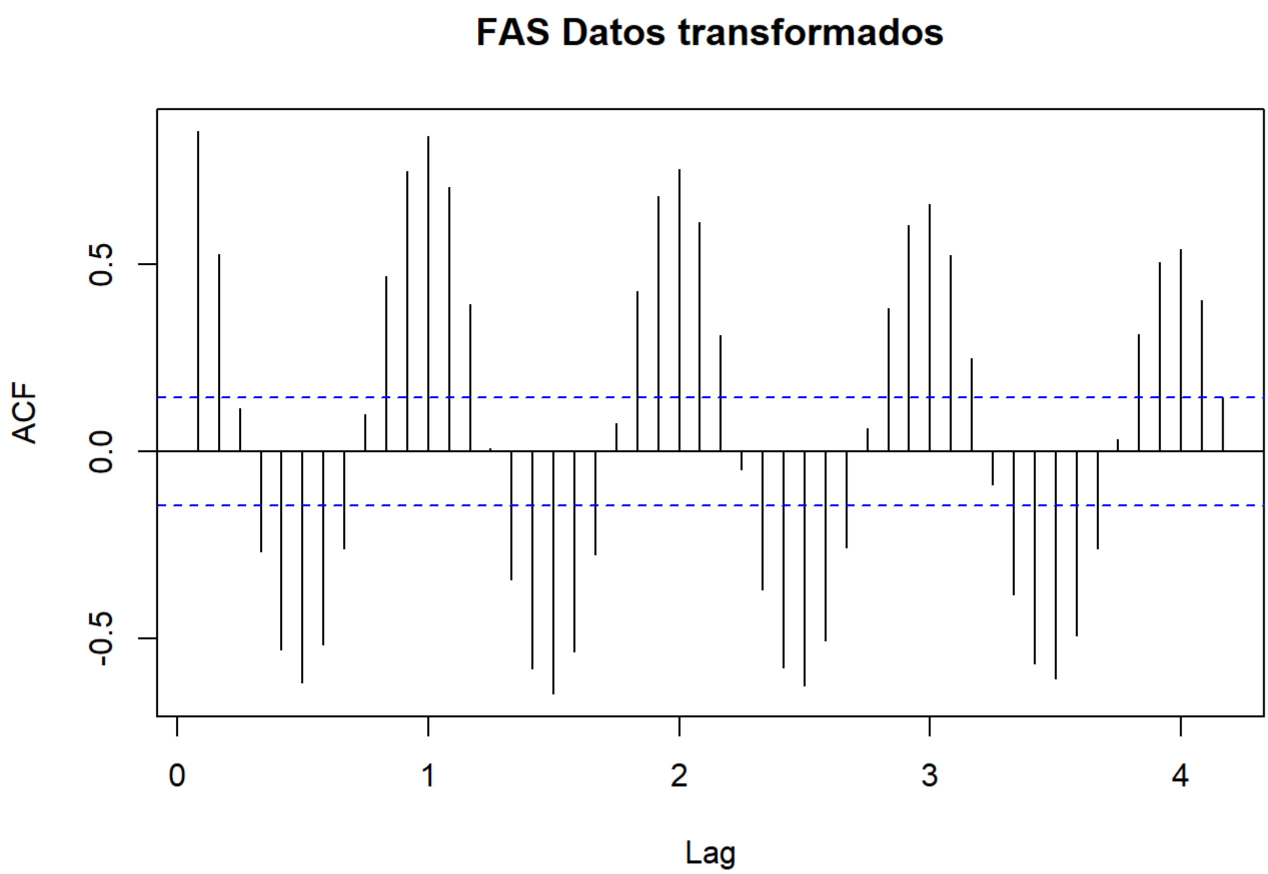
```
lambda <- bc$mle
# serieTransf = (serie^(lambda)-1)/lambda
serieTransf=serie^lambda
plot(serieTransf)
```



Ya el gráfico parece indicar que la varianza se ha estabilizado en el tiempo.

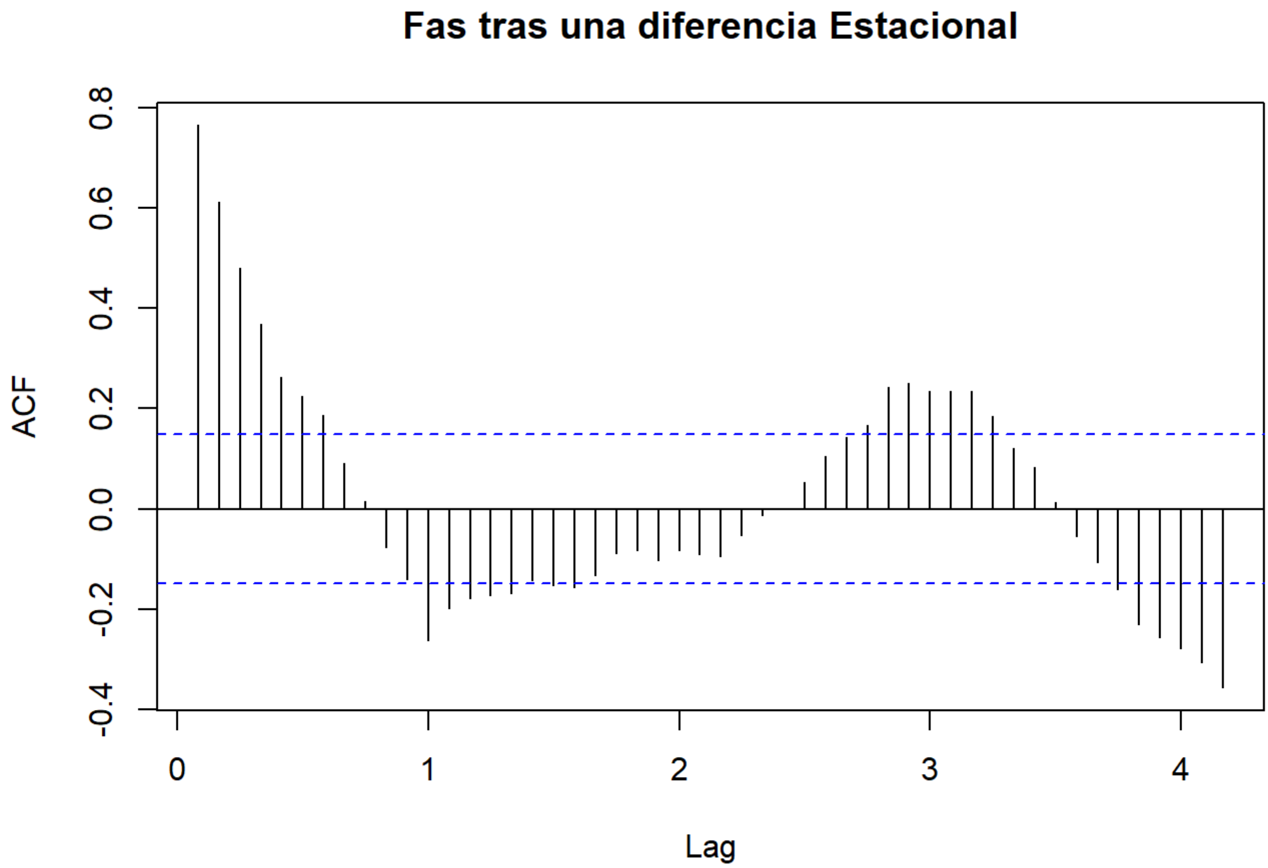
apartado 2

```
acf(serieTransf, main="FAS Datos transformados",lag.max = 50)
```



Observamos que la FAC decrece muy lentamente, se trata de un modelo integrado. Vemos mucha dependencia de la componente estacional. Vamos a hacer una diferencia estacional a los datos.

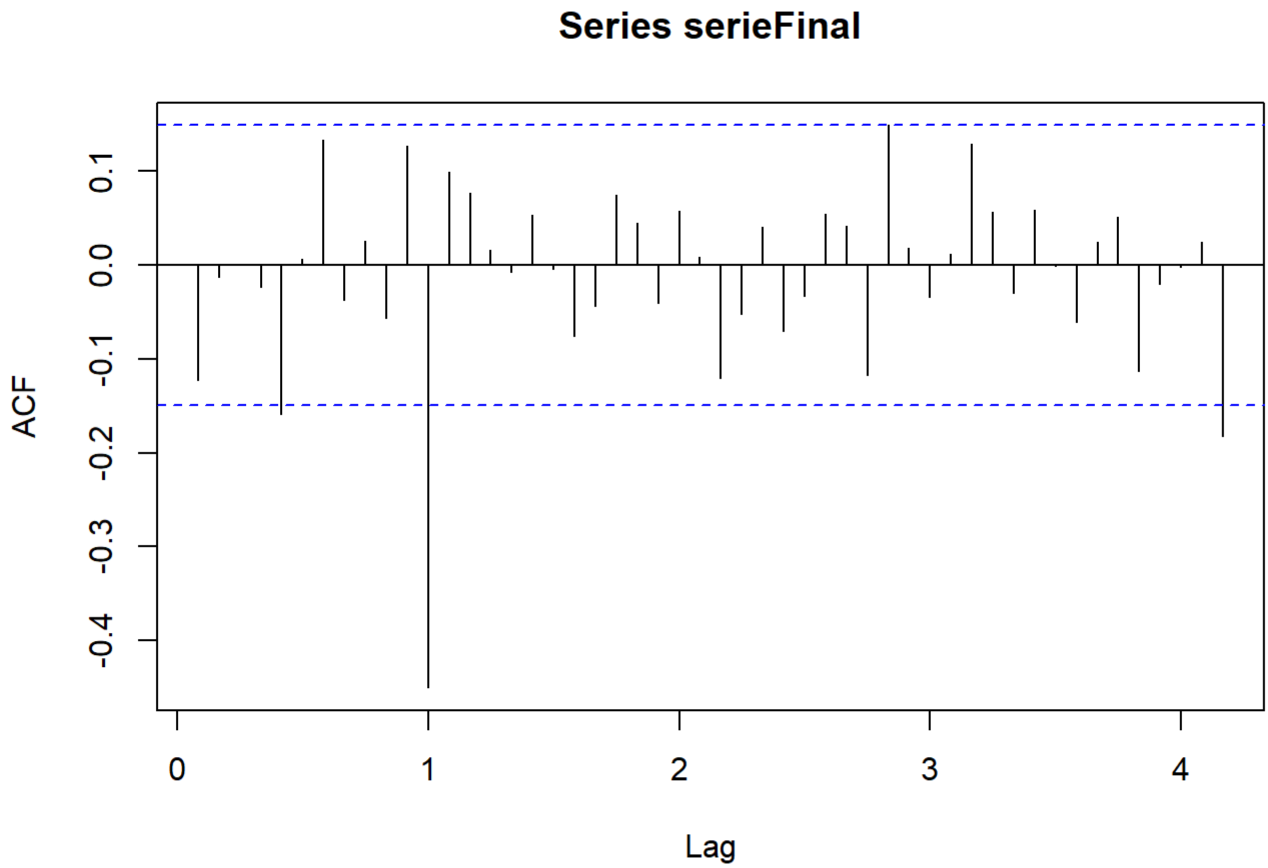
```
serieTransfDiff <- diff(serieTransf,lag=12,diff=1)
acf(serieTransfDiff, main="Fas tras una diferencia Estacional", lag.max = 50)
```



Hacemos una diferencia regular, parece que tambien existe una dependencia de la componente regular

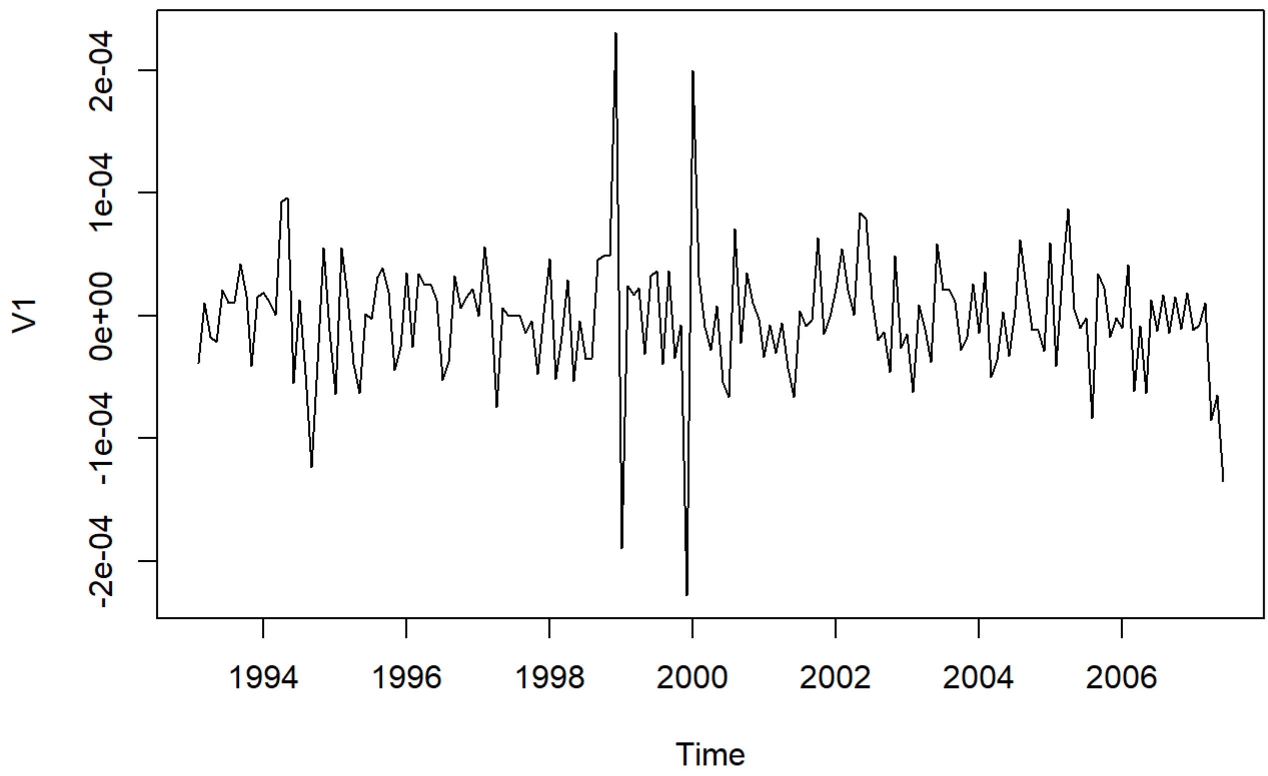
```
serieFinal <- diff(serieTransfDiff,lag=1,diff=1)

acf(serieFinal,lag.max = 50)
```



Podemos apreciar que esta FAC corresponde a un modelo estacionario y no se ven dependencias de la componente regular ni estacional.

```
plot(serieFinal)
```



Observamos un muelle, lo que indicaría estacionariedad en los datos.

Apartado 3

Vamos a hacer el test de Dikey-Fuller para comprobar la estacionariedad de los datos. Test de raiz unitaria que contrasta las hipótesis siguientes:

$$H_0 : \text{El polinomio autoregresivo tiene una raiz unitaria}$$

$$H_1 : \text{Todas las raices del polinomio autoregresivo son estacionarias (en módulo mayores que 1)}$$

```
library(tseries)
adf.test(serieFinal)
```

```
## Warning in adf.test(serieFinal): p-value smaller than printed p-value
```

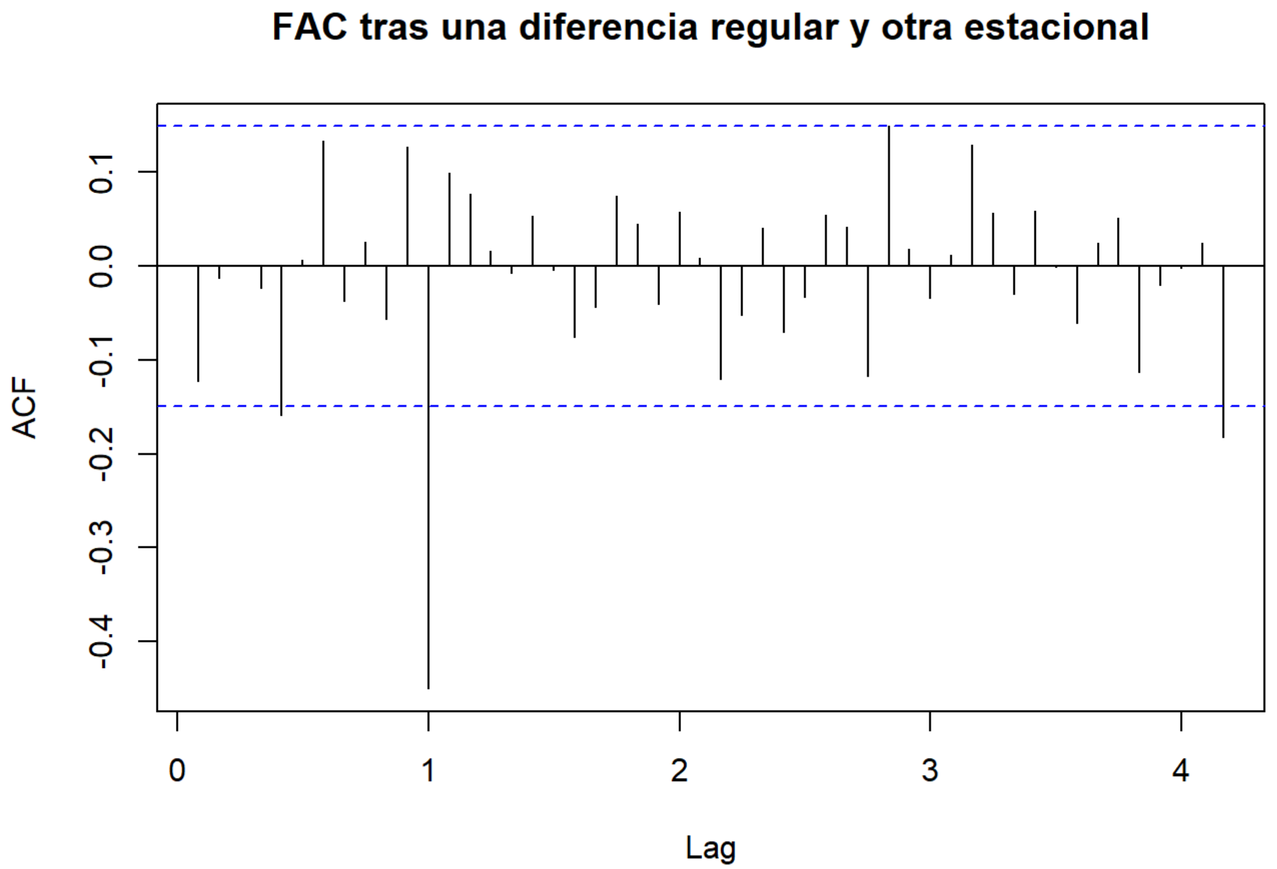
```
##
##  Augmented Dickey-Fuller Test
##
## data:  serieFinal
## Dickey-Fuller = -6.2341, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```

- p-valor del contraste: $\text{p-valor} < 0.01 < \alpha = 0.05$
- Conclusión: rechazo la hipótesis nula a favor de la alternativa, es decir, no existen evidencias diferencias significaticas para aceptar que el polinomio autoregresivo tiene alguna raiz unitaria.

apartado 4

Vamos a identificar la estructura ARIMA a través de la FAC y la FAP

```
acf(serieFinal,lag.max = 50, main ="FAC tras una diferencia regular y otra estacional")
```

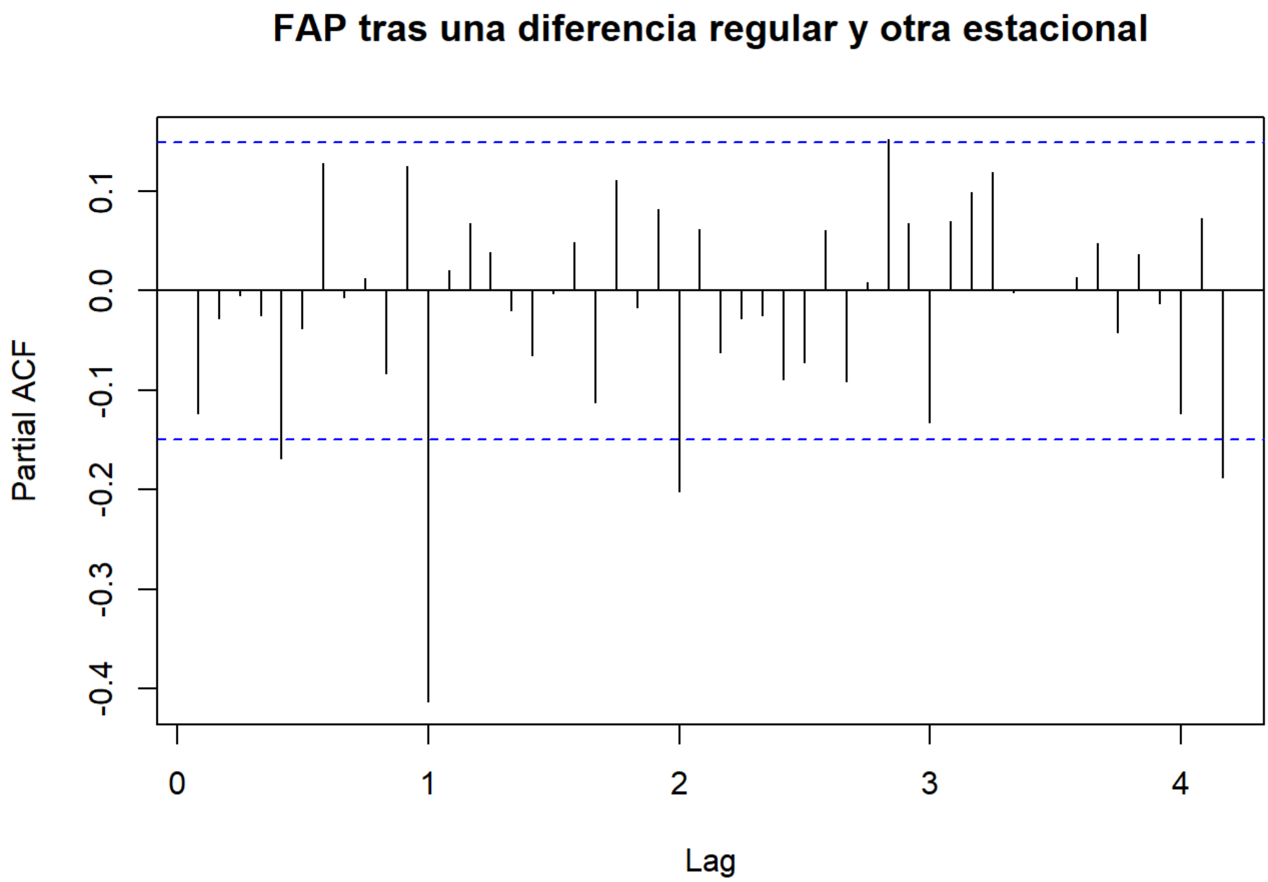


En los retrardos estacionales observamos una autocorrelación en el primer retardo, podríamos pensar que la parte estacional tiene una estructura MA(1)

En los primeros retardos observamos una autocorrelación, lo que podría indicar que en la parte regular el modelo tenga estructura MA(1), pero debemos comprobarlo con la FAP. Tambien puede ocurrir que sea un modelo solo con parte estacional.

Alrededor del retardo 4 vemos una autocorrelación que se sale de las bandas, pero no nos preocupa ya que estas bandas son un IC al 95%, y por tanto, cabe esperar que algunas se salgan.

```
pacf(serieFinal,lag.max = 50, main ="FAP tras una diferencia regular y otra estacional")
```



en los retardos estacionales podemos ver dos autocorrelaciones que decrecen rápidamente, lo que avalaría aún más la hipótesis de un MA(1) en la parte estacional.

En la parte regular, encontramos unaúnica autocorrelación fuera de la banda, que nos lleva a pensar que podría tratarse de un modelo AR(1), o un modelo sin componente regular.

Vuelve a haber una autocorrelación fuera de las bandas debido a la aleatoriedad, se trata de un IC

Modelos candidatos:

- ARIMA(1,0)X(0,1)12
- ARIMA(0,1)x(0,1)12
- ARIMA(1,1)x(0,1)12
- ARIMA(0,0)X(0,1)12

Estos modelos son:

- SARIMA(1,1,0)X(0,1,1)12
- SARIMA(0,1,1)x(0,1,1)12
- SARIMA(1,1,1)x(0,1,1)12
- SARIMA(0,1,0)X(0,1,1)12

Ajuste de los modelos

SARIMA(1,1,0)X(0,1,1)12

```
ajuste1<-arima(serieFinal,order=c(1,1,0), # p. regular
      seasonal = list(order=c(0,1,1), period=12) # p. estacional
      )
ajuste1
```

```
##
## Call:
## arima(x = serieFinal, order = c(1, 1, 0), seasonal = list(order = c(0, 1, 1),
##     period = 12))
##
## Coefficients:
##           ar1     sma1
##       -0.5520  -1.0000
## s.e.   0.0659   0.0568
##
## sigma^2 estimated as 4.167e-09:  log likelihood = 1300.51,  aic = -2597.02
```

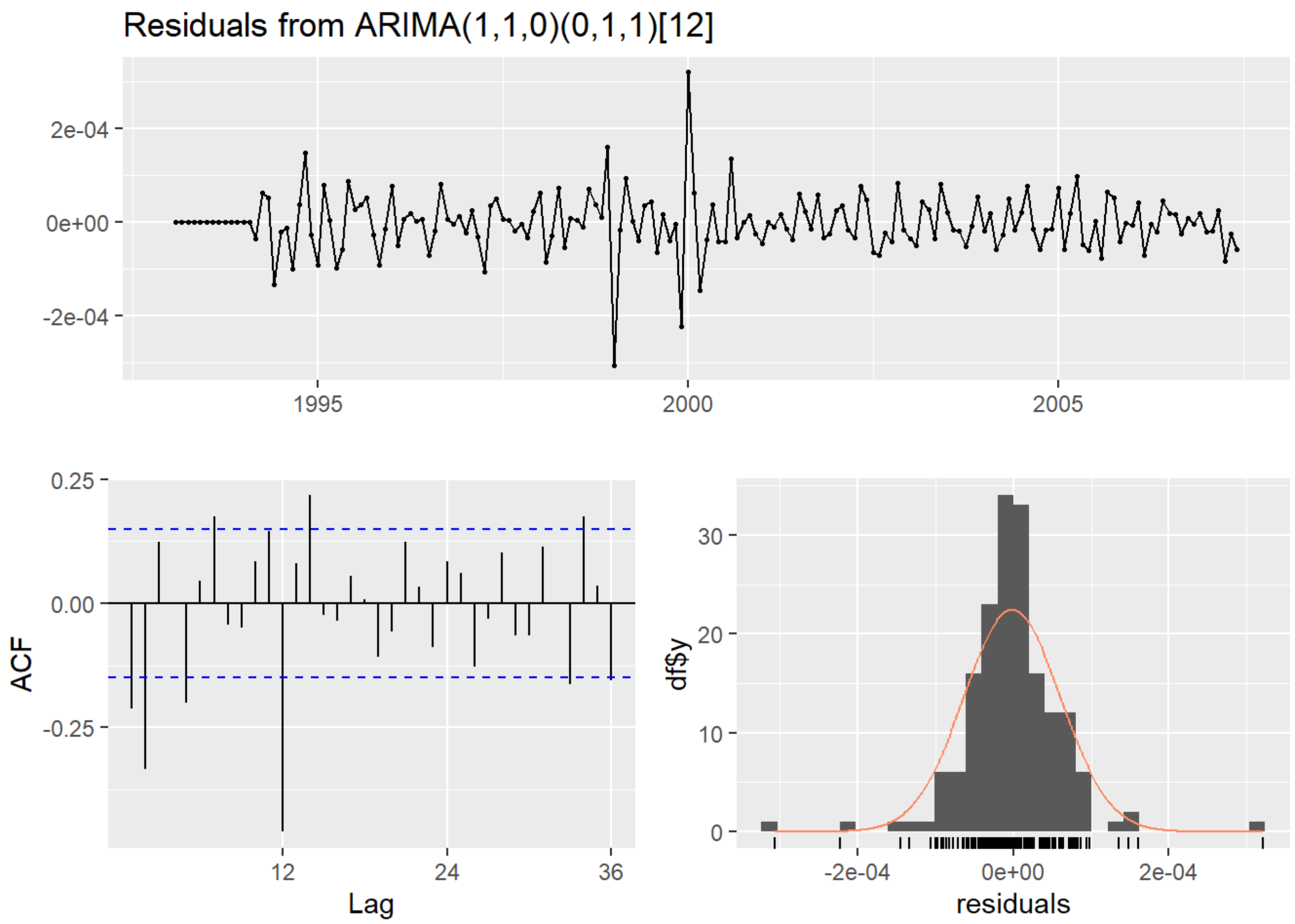
```
confint(ajuste1)
```

```
##           2.5 %     97.5 %
## ar1  -0.6810936 -0.4228787
## sma1 -1.1113563 -0.8886402
```

No hay coeficientes significativamente nulos, sn todos no nulos. Valor del AIC = -2597.02

Vamos a ver si se cumple que los resíduos siguen un proceso ruido blanco.

```
checkresiduals(ajuste1)
```



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,0)(0,1,1)[12]
## Q* = 109.62, df = 22, p-value = 1.291e-13
##
## Model df: 2.   Total lags used: 24
```

- p-valor del constaste: p-value = 1.291e-13. no puedo aceptar la incorrelación. Los resíduos nosiguien un ruido blanco. Tenemos que desechar este modelo. ajuste 2

SARIMA(0,1,1)x(0,1,1)12

```
ajuste2<-arima(serieFinal,order=c(0,1,1), # p. regular
      seasonal = list(order=c(0,1,1), period=12) # p. estacional
      )
ajuste2
```

```
##
## Call:
## arima(x = serieFinal, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1),
##     period = 12))
##
## Coefficients:
##           ma1     sma1
##       -1.0000  -0.9999
## s.e.   0.0349   0.0618
##
## sigma^2 estimated as 2.699e-09:  log likelihood = 1331.39,  aic = -2658.78
```

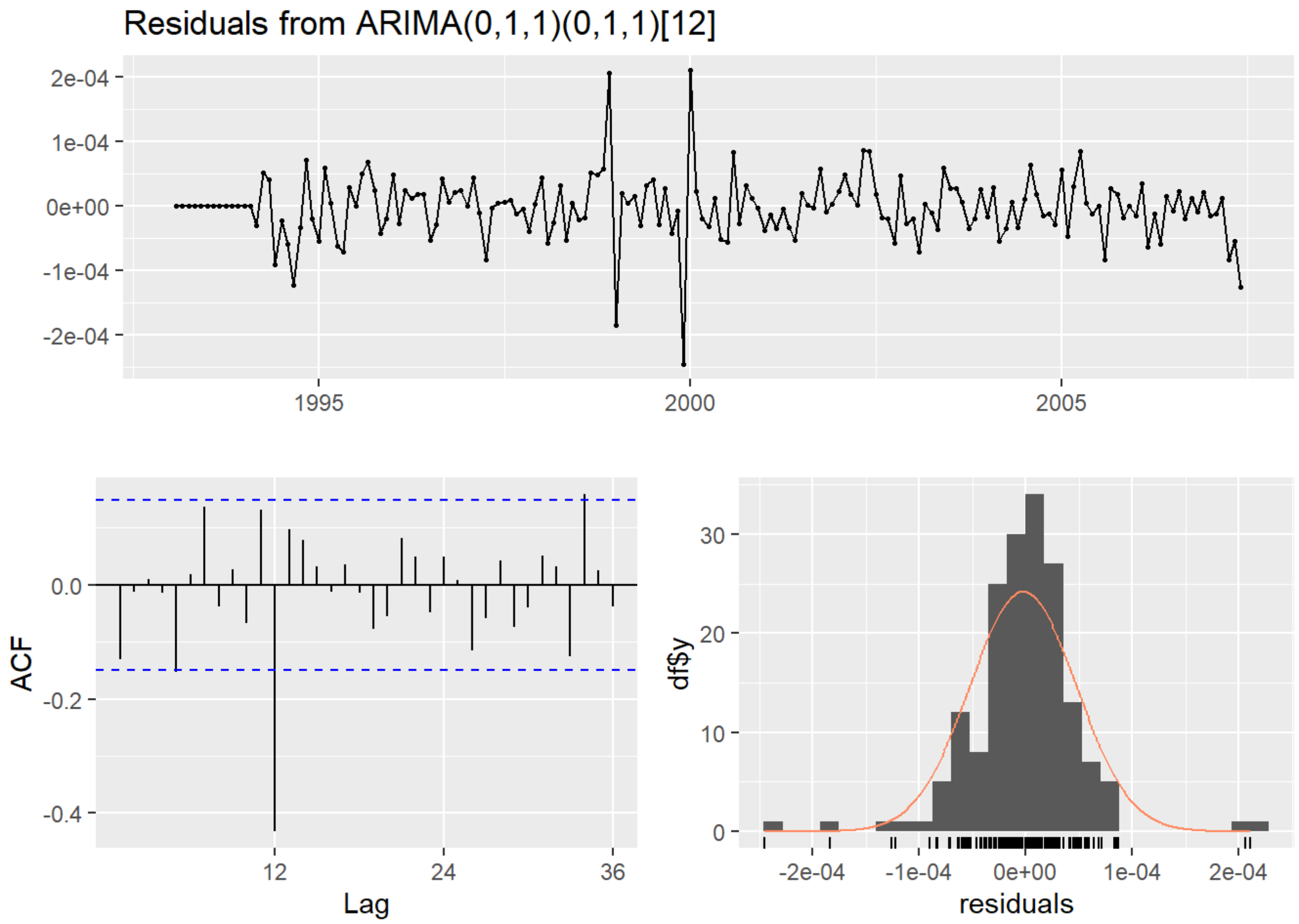
```
confint(ajuste2)
```

```
##          2.5 %     97.5 %
## ma1  -1.068374 -0.9316141
## sma1 -1.121013 -0.8787217
```

No hay coeficientes significativamente nulos, sn todos no nulos. Valor del AIC = -2658.78

Vamos a ver si se cumple que los resíduos siguen un proceso ruido blanco.

```
checkresiduals(ajuste2)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,1)(0,1,1)[12]
## Q* = 58.405, df = 22, p-value = 3.836e-05
##
## Model df: 2.   Total lags used: 24
```

- p-valor del constaste: p-value = 3.836e-05. no puedo aceptar la incorrelación. Los resíduos nosiguien un ruido blanco. Tenemos que desechar este modelo.

```
Box.test(serieFinal,lag=12,type = c("Box-Pierce"),fitdf =2 )
```

```
##
##  Box-Pierce test
##
## data:  serieFinal
## X-squared = 49.018, df = 10, p-value = 4.042e-07
```

Misma conclusión.

ajuste 3: SARIMA(1,1,1)x(0,1,1)12

```
ajuste3<-arima(serieFinal,order=c(1,1,1), # p. regular
      seasonal = list(order=c(0,1,1), period=12) # p. estacional
      )
ajuste3
```

```
##
## Call:
## arima(x = serieFinal, order = c(1, 1, 1), seasonal = list(order = c(0, 1, 1),
##     period = 12))
##
## Coefficients:
##          ar1      ma1     sma1
##       -0.132  -1.0000  -1.0000
## s.e.   0.080   0.0412   0.0632
##
## sigma^2 estimated as 2.65e-09:  log likelihood = 1332.74,  aic = -2659.48
```

```
confint(ajuste3)
```

```
##           2.5 %      97.5 %
## ar1  -0.2887844  0.02469459
## ma1  -1.0808225 -0.91917516
## sma1 -1.1238300 -0.87616978
```

El coeficiente AR(1) es significativamente nulo, por lo que lo tengo que quitar y me queda el modelo del ajuste 1 MA(1)xMA(1)

Ajuste 4: SARIMA(0,1,0)X(0,1,1)12

```
ajuste4<-arima(serieFinal,order=c(0,1,0), # p. regular
      seasonal = list(order=c(0,1,1), period=12) # p. estacional
      )
ajuste4
```

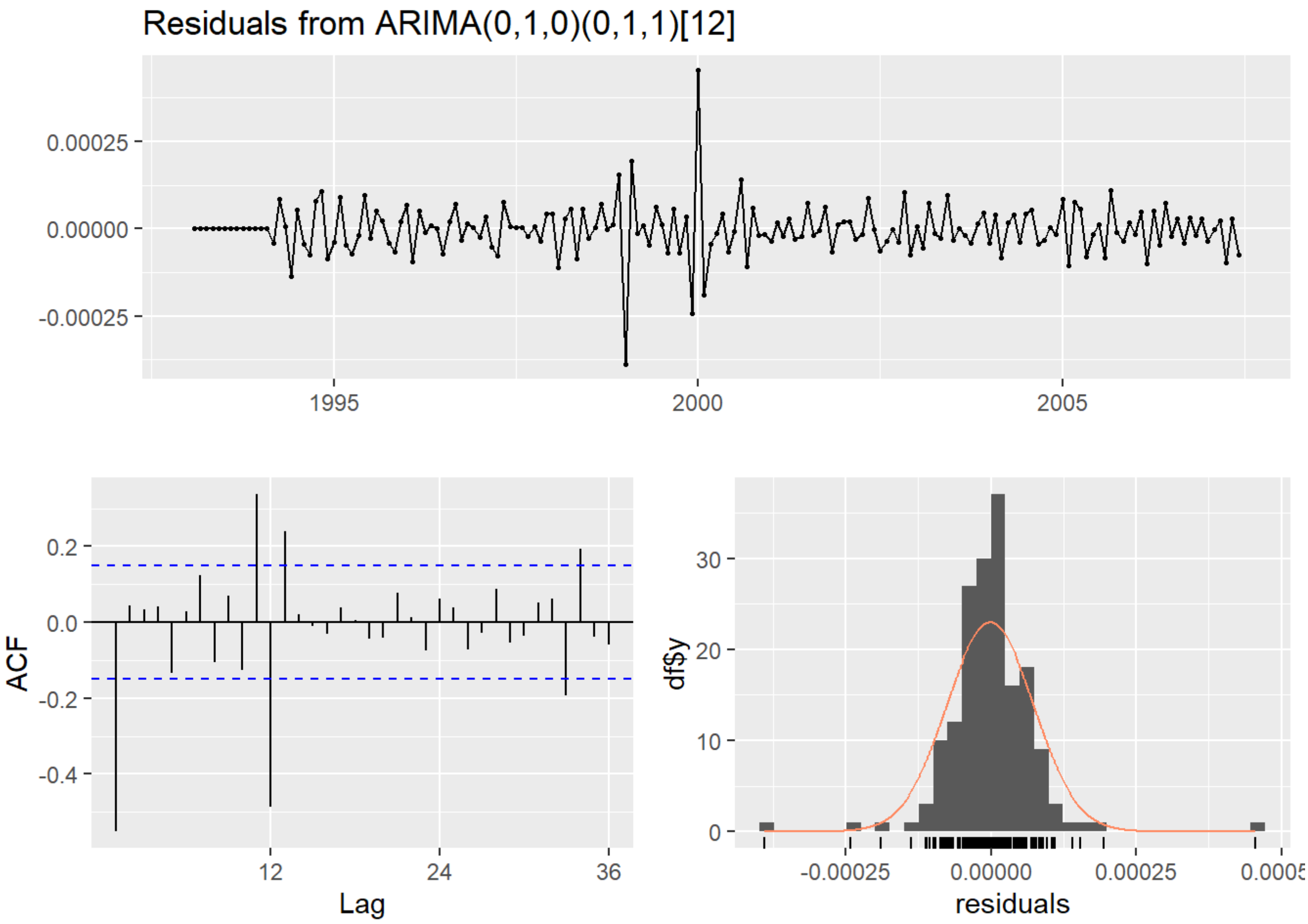
```
##
## Call:
## arima(x = serieFinal, order = c(0, 1, 0), seasonal = list(order = c(0, 1, 1),
##     period = 12))
##
## Coefficients:
##          sma1
##       -1.0000
## s.e.   0.0561
##
## sigma^2 estimated as 6.001e-09:  log likelihood = 1271.5,  aic = -2541.01
```

```
confint(ajuste4)
```

```
##          2.5 %     97.5 %
## sma1 -1.110006 -0.8899784
```

AIC = aic = -2541.01 El coeficiente estimado es significativamente no nulo. Vamos a ver si los resíduos siguen un proceso ruido blanco.

```
checkresiduals(ajuste4)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,0)(0,1,1)[12]
## Q* = 147.3, df = 23, p-value < 2.2e-16
##
## Model df: 1.   Total lags used: 24
```

El p-valor es muy pequeño, p-valor < 2.2e-16. Por tanto, no acepto la incorrelación de los resíduos.

Ajuste de autoarima

```
ajuste5 <- auto.arima(serieFinal,stepwise = FALSE)
confint(ajuste5)
```

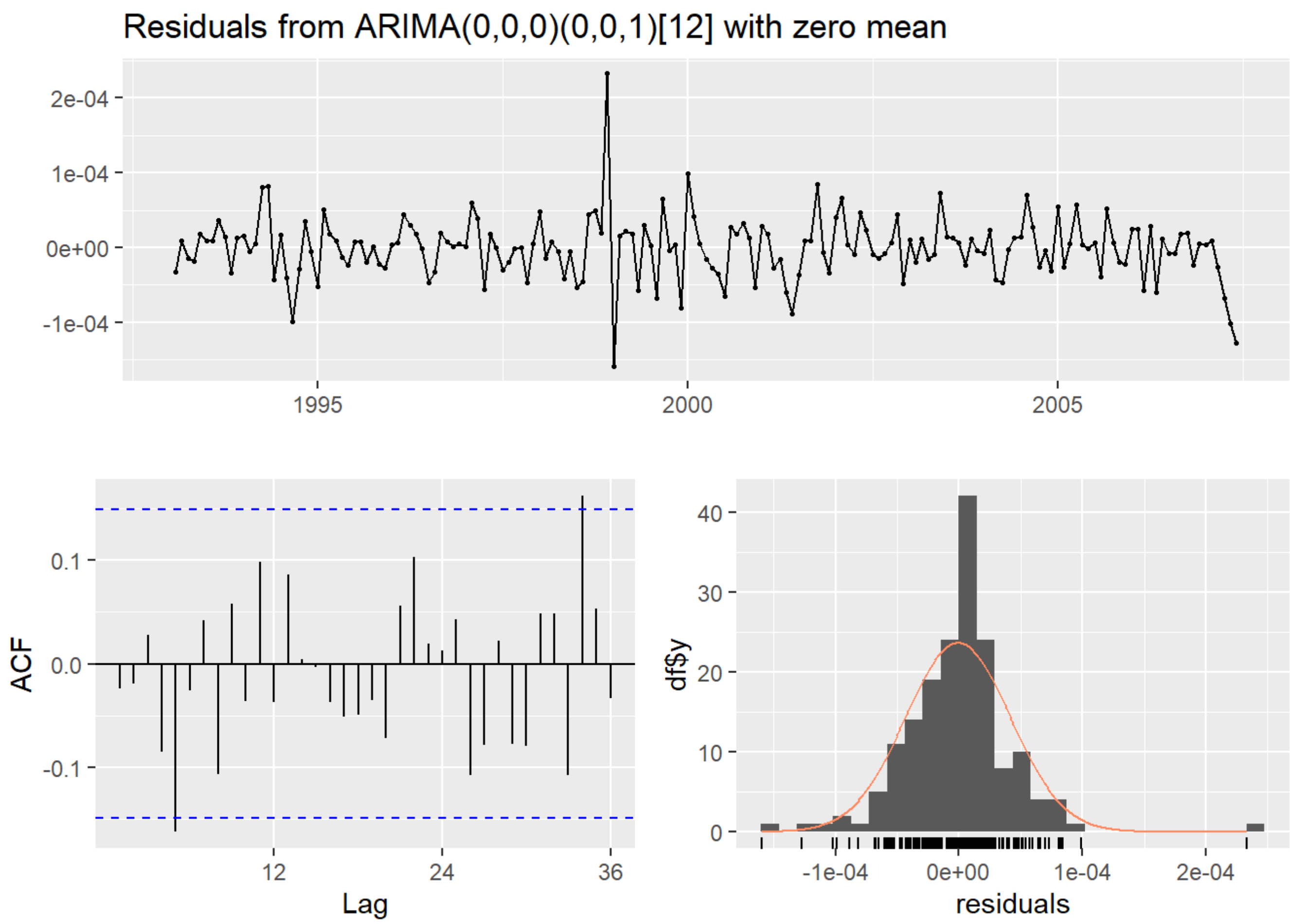
```
##           2.5 %     97.5 %
## sma1 -0.7625081 -0.4957249
```

```
ajuste5
```

```
## Series: serieFinal
## ARIMA(0,0,0)(0,0,1)[12] with zero mean
##
## Coefficients:
##          sma1
##       -0.6291
## s.e.   0.0681
##
## sigma^2 = 1.807e-09:  log likelihood = 1493.4
## AIC=-2982.81   AICc=-2982.74   BIC=-2976.5
```

No hay coeficientes significativamente nulos. Se trata de un modelo MA(1)12=ARIMA(0,1,0)x(0,1,1) Valor del AIC=-2982.81

```
checkresiduals(ajuste5)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,0,0)(0,0,1)[12] with zero mean
## Q* = 18.511, df = 23, p-value = 0.7292
##
## Model df: 1.   Total lags used: 24
```

- p-valor= 0.7292. Acepto la incorrelación de los resíduos.

Modelo final

Este es mi modelo final MA(1)12

```
ajusteFinal<-ajuste5
ajusteFinal
```

```
## Series: serieFinal
## ARIMA(0,0,0)(0,0,1)[12] with zero mean
##
## Coefficients:
##          sma1
##       -0.6291
## s.e.   0.0681
##
## sigma^2 = 1.807e-09:  log likelihood = 1493.4
## AIC=-2982.81   AICc=-2982.74   BIC=-2976.5
```

$Y_t = y_{t-1} + \alpha_{t-1} + \alpha_t + 0.6291\alpha_{t-12}$ o lo que es lo mismo
$Y_t(1-L) = (1-L^{12})(1+0.6291\Theta^{12})\alpha_t$

Apartado 5

Predicción para 12 meses siguientes e IC para el precio de la leche

```
pred= predict(ajusteFinal,n.ahead = 12)
pred
```

```
## $pred
##                Jan           Feb           Mar           Apr           May
## 2007
## 2008 -2.616978e-06 -5.413757e-06  1.627452e-05  4.234052e-05  6.428053e-05
##                Jun           Jul           Aug           Sep           Oct
## 2007                4.973175e-06  4.691409e-06 -1.149151e-05 -1.204203e-05
## 2008  8.006345e-05
##                Nov           Dec
## 2007  1.475274e-05 -3.263789e-06
## 2008
##
## $se
##               Jan          Feb          Mar          Apr          May
## 2007
## 2008 4.250447e-05 4.250446e-05 4.250446e-05 4.250446e-05 4.250446e-05
##               Jun          Jul          Aug          Sep          Oct
## 2007              4.250447e-05 4.250447e-05 4.250447e-05 4.250447e-05
## 2008 4.250446e-05
##               Nov          Dec
## 2007 4.250447e-05 4.250447e-05
## 2008
```

```
inf<-pred$pred-qnorm(0.95)*pred$se
sup<-pred$pred+qnorm(0.95)*pred$se

cbind.data.frame(pred,inf,sup,colnames=c("Predicción","Inf","Sup"))
```

```
##             pred           se           inf          sup   colnames
## 1   4.973175e-06 4.250447e-05 -6.494045e-05 7.488680e-05 Predicción
## 2   4.691409e-06 4.250447e-05 -6.522222e-05 7.460504e-05        Inf
## 3  -1.149151e-05 4.250447e-05 -8.140514e-05 5.842212e-05        Sup
## 4  -1.204203e-05 4.250447e-05 -8.195566e-05 5.787160e-05 Predicción
## 5   1.475274e-05 4.250447e-05 -5.516089e-05 8.466637e-05        Inf
## 6  -3.263789e-06 4.250447e-05 -7.317742e-05 6.664984e-05        Sup
## 7  -2.616978e-06 4.250447e-05 -7.253061e-05 6.729665e-05 Predicción
## 8  -5.413757e-06 4.250446e-05 -7.532737e-05 6.449986e-05        Inf
## 9   1.627452e-05 4.250446e-05 -5.363910e-05 8.618813e-05        Sup
## 10  4.234052e-05 4.250446e-05 -2.757310e-05 1.122541e-04 Predicción
## 11  6.428053e-05 4.250446e-05 -5.633088e-06 1.341941e-04        Inf
## 12  8.006345e-05 4.250446e-05  1.014983e-05 1.499771e-04        Sup
```

Apartado 6

```
dec=decompose(serieTransf,type="additive")
serie_dec= serie-dec$seasonal # serie desestacionalizada
```

Ahora ajustamos a una recta la serie desestracionalizada

```
time = time(serie_dec)
tendencia = lm(serie_dec~time)
summary(tendencia)
```

```
##
## Call:
## lm(formula = serie_dec ~ time)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.6889 -2.2986 -0.2316  1.9519  7.5499
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -50.25893   87.78613  -0.573    0.568
## time          0.04112    0.04390   0.937    0.350
##
## Residual standard error: 2.679 on 184 degrees of freedom
## Multiple R-squared:  0.004746,   Adjusted R-squared:  -0.0006628
## F-statistic: 0.8775 on 1 and 184 DF,  p-value: 0.3501
```

- a = -50.25893
- b= 0.04112

```
cc=1:12/12
cc=time[186]+cc
cc=data.frame(time=cc)
cc # M móviles centradas
```

```
##        time
## 1  2007.500
## 2  2007.583
## 3  2007.667
## 4  2007.750
## 5  2007.833
## 6  2007.917
## 7  2008.000
## 8  2008.083
## 9  2008.167
## 10 2008.250
## 11 2008.333
## 12 2008.417
```

```
# prediccion de la tendencia
pred_tendencia = predict(tendencia,cc)
dec$figure # IVE
```

```
##  [1] -1.272146e-04 -6.137250e-05  1.447858e-05  1.029164e-04  1.747403e-04
##  [6]  2.091159e-04  1.864864e-04  1.032607e-04 -6.233325e-05 -1.664404e-04
## [11] -2.027266e-04 -1.709108e-04
```

```
season = c(dec$figure[7:12],dec$figure[1:6])
p2 = ts(pred_tendencia+season, freq=12,start(2007,7))
p2
```

```
##        Jan      Feb      Mar      Apr      May      Jun      Jul      Aug
## 7                                              32.29350 32.29684 32.30010
## 8 32.31724 32.32074 32.32425 32.32775 32.33121
##        Sep      Oct      Nov      Dec
## 7 32.30342 32.30681 32.31027 32.31374
## 8
```

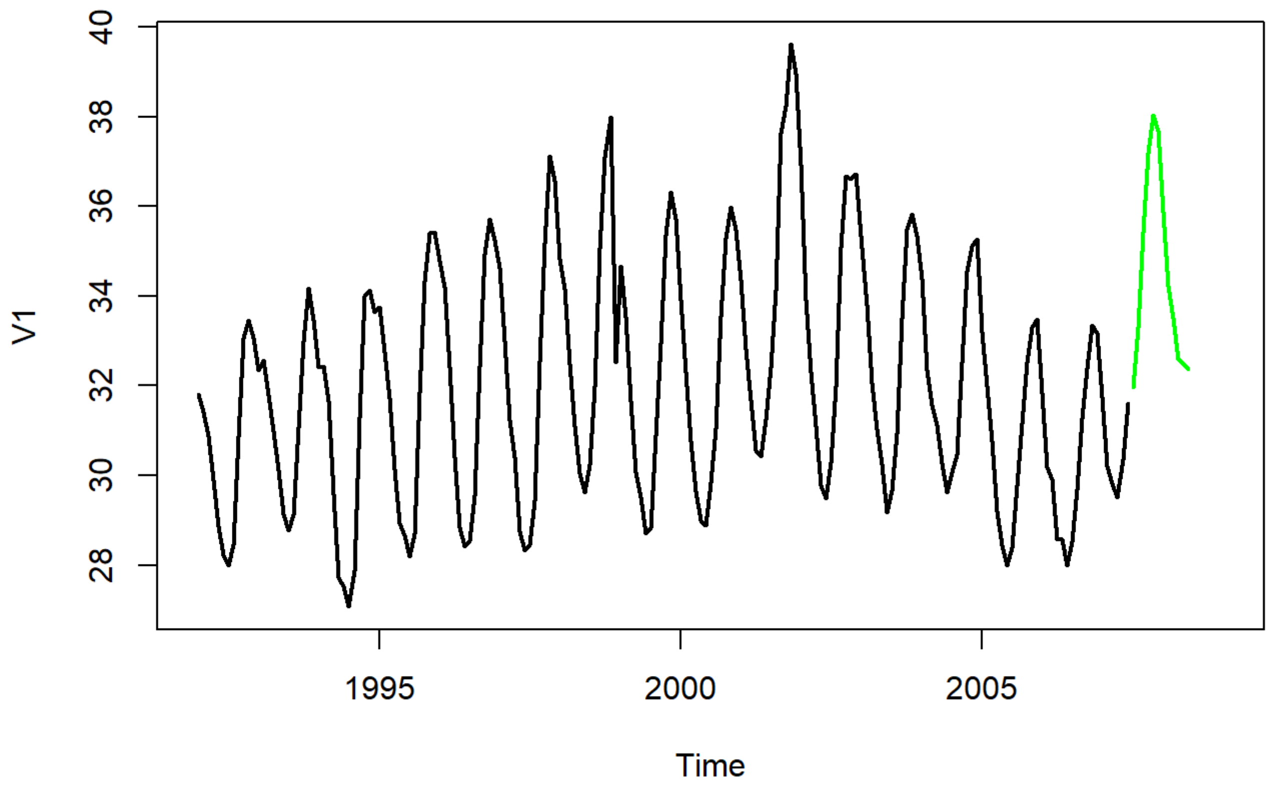
Apartado 7

```
serie_suavizado <- HoltWinters(serie)
p3 <- predict(serie_suavizado,n.ahead=12)
p3
```

```
##           Jan      Feb      Mar      Apr      May      Jun      Jul      Aug
## 2007                                                        31.95330 33.31204
## 2008 35.87067 34.20815 33.50482 32.59302 32.49789 32.35586
##           Sep      Oct      Nov      Dec
## 2007 35.42010 37.19227 38.03618 37.67472
## 2008
```

Apartado 8

```
# y = x^lambda -> datos = y^(1/lambda)
p1<-pred$pred
plot(serie,xlim=c(1992,2009),lwd=2)
lines(p1,col="red",lwd=2)
lines(p2,col="blue",lwd=2)
lines(p3,col="green",lwd=2)
```



Holt wilters predice una bajada del predio medio mensual de la leche cruda en los próximos meses y el según el ajuste al modelo se prevee una subida.