

Series de Tiempo 2018

Maestría en Estadística Aplicada, UNR
Unidad 6

Luis Damiano

damiano.luis@gmail.com

2018-05-10

- Identificación
 - Transformaciones
 - Selección de orden
 - Desplazamiento
- Ejercicio: Ventas en supermercados
- Ejercicio: Producción de automóviles

Identificación

Típicamente, la identificación de una serie de tiempo incluye los siguientes pasos:¹

1. Análisis exploratorio.

- ¿Qué dice la teoría subyacente a los datos?
- ¿Es estacionario en la media? ¿Tiene tendencia? ¿De qué tipo?
- ¿Es estacionario en la varianza? ¿De qué forma se relaciona la varianza con la media?
- ¿Tiene estacionalidad? ¿Es estacionaria? ¿Es constante a lo largo del tiempo? ¿Es aditiva o multiplicativa?
- ¿Tiene valores atípicos?
- ¿Presenta cambios (quiebres) en los patrones?

2. Identificar las transformaciones necesarias.

- Eliminar tendencia.
- Estabilizar varianza.

3. Seleccionar los órdenes p y q .

4. Identificar la existencia de tendencia determinística (desplazamiento o drift) en series diferenciadas.

¹**Típicamente** es la palabra clave de esta oración. Esta receta es una mera guía indicativa para hacer los primeros análisis. En la práctica, los datos reales desafían todos los protocolos.

1 Análisis exploratorio

Recordar el análisis exploratorio visto en la Unidad 5.

2 Identificación de la raíz unitaria

- Patrón: ACF decae muy lentamente y la PACF se corta abruptamente luego del primer rezago.
- Pruebas de raíz unitarias: Dickey-Fuller, Dickey-Fuller Aumentado, Phillips-Perron.
- Otras reglas prácticas:
 - Si la suma de los parámetros AR es cercana a la unidad, probar de incrementar el orden de la diferenciación d y reducir el orden del componente autorregresivo p .
 - Si la suma de los parámetros MA es cercana a la unidad, probar de reducir el orden de la diferenciación d y reducir el orden del componente de media móvil q .
- ¿Caso muy dudoso? Probar diferenciando.

Identificación de $ARIMA(0, 1, 0)$

```
set.seed(9000)
z <- arima.sim(
  model = list(order = c(0, 1, 0), sd = 1),
  n      = 500
)

library(tseries)
adf.test(z, alternative = "stationary")

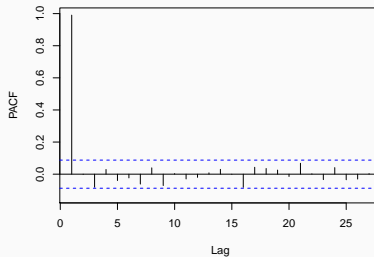
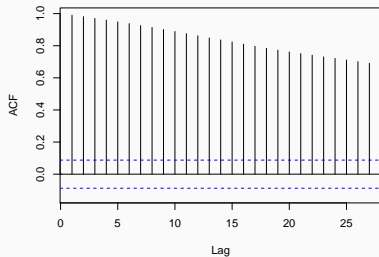
##
## Augmented Dickey-Fuller Test
##
## data: z
## Dickey-Fuller = -1.4611, Lag order = 7, p-value = 0.8064
## alternative hypothesis: stationary

pp.test(z, alternative = "stationary")

##
## Phillips-Perron Unit Root Test
##
## data: z
## Dickey-Fuller Z(alpha) = -5.7519, Truncation lag parameter = 5,
## p-value = 0.7889
## alternative hypothesis: stationary
```

Identificación de $ARIMA(0, 1, 0)$

Simulación de $ARIMA(0, 1, 0)$



Identificación de $ARIMA(2,0,0)$ con raíz (casi) unitaria

```
set.seed(9000)
z <- arima.sim(
  model = list(order = c(2, 0, 0), ar = c(0.7, 0.29), sd = 1),
  n      = 500
)

library(tseries)
adf.test(z, alternative = "stationary")

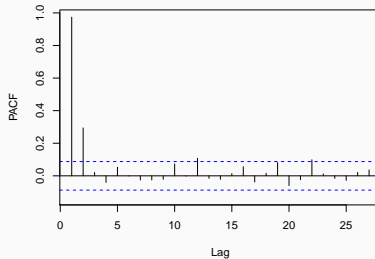
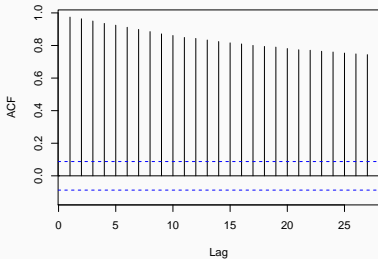
##
## Augmented Dickey-Fuller Test
##
## data: z
## Dickey-Fuller = -3.036, Lag order = 7, p-value = 0.1398
## alternative hypothesis: stationary

pp.test(z, alternative = "stationary")

##
## Phillips-Perron Unit Root Test
##
## data: z
## Dickey-Fuller Z(alpha) = -22.216, Truncation lag parameter = 5,
## p-value = 0.04457
## alternative hypothesis: stationary
```

Identificación de $ARIMA(2, 0, 0)$ con raíz (casi) unitaria

Simulación de $ARIMA(2, 0, 0)$ con $\phi_1 + \phi_2 = 1$



3 Seleccionar los órdenes p y q

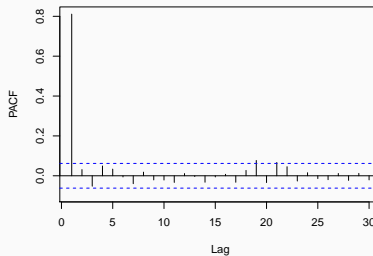
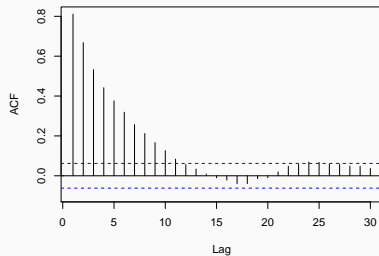
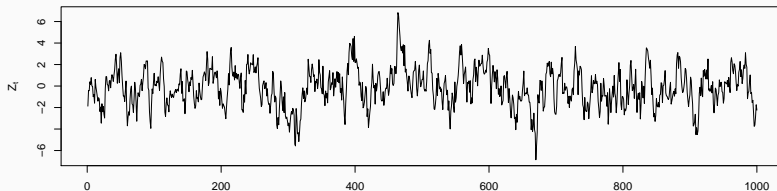
Selección de los órdenes p y q

Proceso	ACF	PACF
$AR(p)$	Decae exponencialmente (raíz real) Sinusoidal (raíz compleja)	Se corta en el rezago p
$MA(q)$	Se corta en el rezago q	Decae exponencialmente (raíz real) Sinusoidal (raíz compleja)
$ARMA(p, q)$	Decae luego de $q - p$	Decae luego de $p - q$

En ciertas oportunidades, los términos AR y MA se cancelan. Cuando se identifica un modelo con ambos componentes, probar de disminuir ambos órdenes en una unidad. Por ejemplo, partiendo de $ARIMA(1, 1, 2)$, probar $ARIMA(0, 1, 1)$.

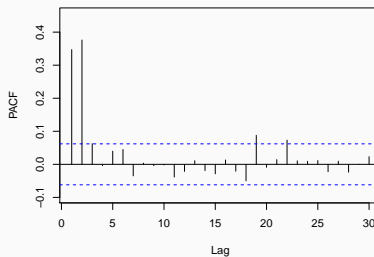
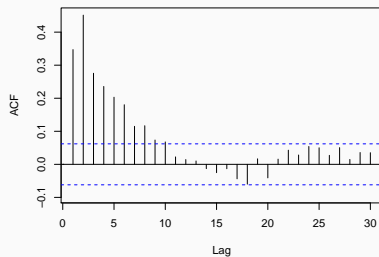
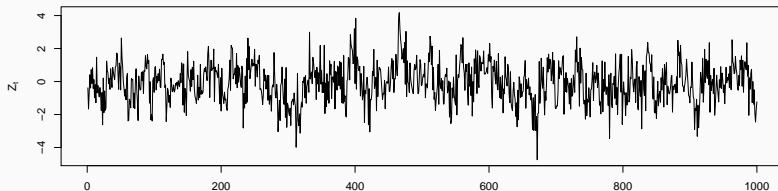
Identificación de $ARIMA(1,0,0)$

Simulación de $ARIMA(1, 0, 0)$ con $\phi_1 = 0.8$



Identificación de $ARIMA(3, 0, 0)$

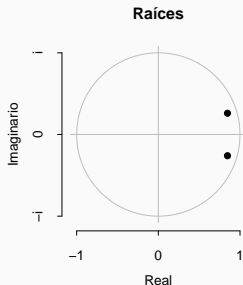
Simulación de $ARIMA(3, 0, 0)$ con $\phi_1 = 0.2$ $\phi_2 = 0.3$ $\phi_3 = 0.1$



Identificación de $ARIMA(2,0,0)$ con raíces complejas

El siguiente es un proceso $AR(2)$:

$$Z_t - 1.7Z_{t-1} + 0.8Z_{t-2} = a_t, \quad t = 0, \pm 1, \dots, \quad a_t \sim \mathcal{N}(0, 1).$$

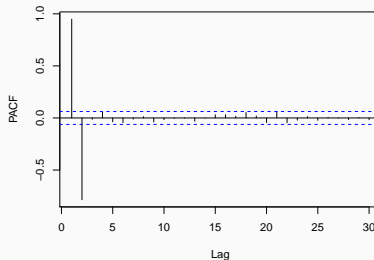
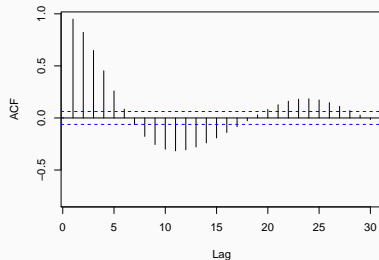
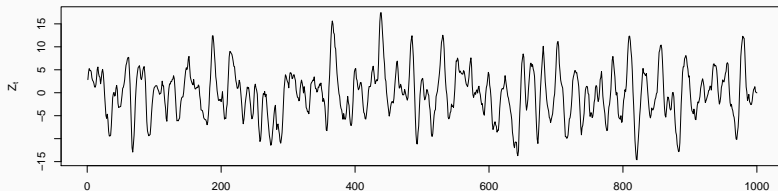


Ejercicio en clases

Calcular analíticamente las raíces del polinomio característico

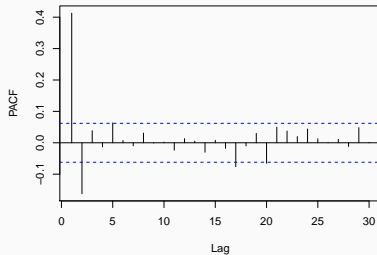
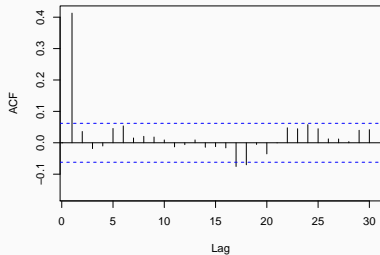
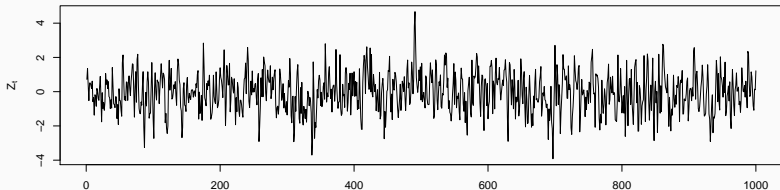
Identificación de $ARIMA(2,0,0)$ con raíces complejas

Simulación de $ARIMA(2,0,0)$ con $\phi_1 = 1.7$ $\phi_2 = -0.8$



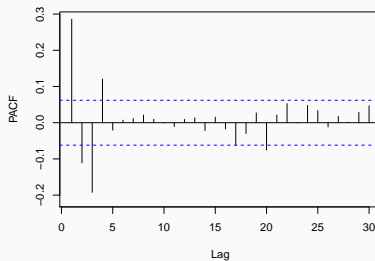
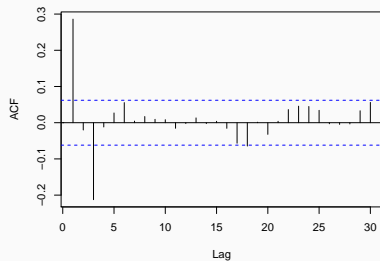
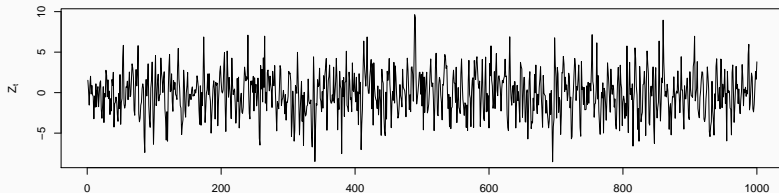
Identificación de $ARIMA(0,0,1)$

Simulación de $ARIMA(0, 0, 1)$ con $\theta_1 = 0.5$



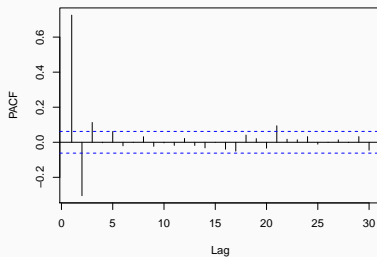
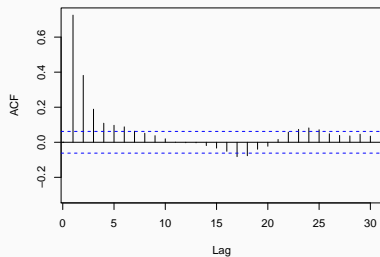
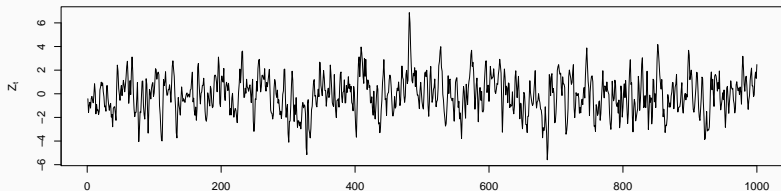
Identificación de $ARIMA(0,0,3)$

Simulación de $ARIMA(0,0,3)$ con $\theta_1 = 1.5$ $\theta_2 = 1.4$ $\theta_3 = -1.2$



Identificación de $ARIMA(1,0,1)$

Simulación de $ARIMA(1, 0, 1)$ con $\phi_1 = 0.5$ $\theta_2 = 0.5$



4 Identificación de tendencia determinística

¿Qué representa el desplazamiento?³

Supongamos un $ARMA(p, q)$ diferenciado d veces (donde Z'_t representa la d -ésima diferencia)

$$Z'_t = c + \phi_1 Z'_{t-1} + \cdots + \phi_p Z'_{t-p} + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q} + a_t, \quad t = 0, \pm 1, \dots$$

La constante c se llama desplazamiento o drift, y tiene un efecto muy importante en los pronósticos de largo plazo².

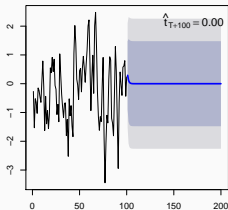
c	d	Pronóstico de largo plazo
0	0	Cero
0	1	Constante distinta de cero
0	2	Línea recta
$\neq 0$	0	Promedio muestral
$\neq 0$	1	Línea recta
$\neq 0$	2	Tendencia cuadrática

²Los pronósticos se desarrollan formalmente en la Unidad 9.

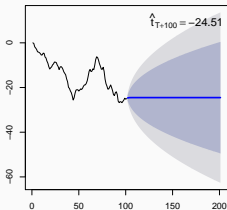
³Hyndman and Athanasopoulos (2018), sec. 8.5. [Ver online](#).

Ejemplos simulados

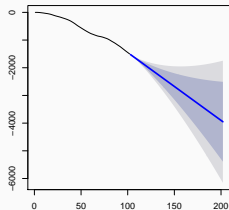
ARIMA(1, 0, 1) con $c = 0$



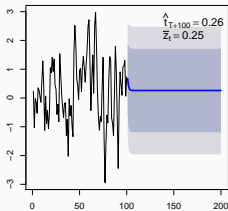
ARIMA(1, 1, 1) con $c = 0$



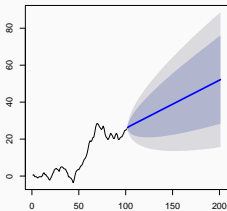
ARIMA(1, 2, 1) con $c = 0$



ARIMA(1, 0, 1) con $c = 0.5$



ARIMA(1, 1, 1) con $c = 0.5$



Ejercicio: Ventas en supermercados

Ejercicio en clases

Identificar la serie de tiempo de ejemplo.

Este ejercicio invita a discutir temas como significancia estadística, parsimonia, y otras cuestiones subjetivas que hacen al análisis de series de tiempo. No esperen una respuesta clara y contundente cuando trabajen con datos reales. Asimismo, no consideren que el modelo elegido por `auto.arima` es la respuesta definitiva al proceso de identificación. El Anexo incluye algunos gráficos útiles.

Algunos pasos:

- Descargar los datos desde <https://bit.ly/2GXzXoa>.
- De la Sección A 1.11, leer los datos mensuales para la columna *Ventas totales*.⁴
- Graficar y describir la serie original. ¿Es estacionaria en la media y en la varianza? ¿Observan tendencia y estacionalidad? ¿De qué tipo?
- Graficar y describir las ACF y PACF muestrales. ¿Observan algunos de los patrones estudiados?

⁴ Hay una copia local en `data/INDECSuper.txt` en caso de que el sitio esté fuera de línea.

Ejercicio: Producción de automóviles

Ejercicio en clases

Identificar la serie de tiempo de ejemplo.

El Anexo incluye algunos gráficos útiles.

Algunos pasos:

- Descargar los datos desde <https://bit.ly/2GXzXoa>.
- De la Sección A 1.22, leer los datos mensuales para la columna *Automóviles*.⁵
- Graficar y describir la serie original. ¿Es estacionaria en la media y en la varianza? ¿Observan tendencia y estacionalidad? ¿De qué tipo?
- Graficar y describir las ACF y PACF muestrales. ¿Observan algunos de los patrones estudiados?

⁵ Hay una copia local en `data/haciendasAutos.txt` en caso de que el sitio esté fuera de línea.

Anexo: Ventas en supermercados

Lectura & procesamiento

```
# https://bit.ly/2GXzXoa
df <- read.table(
  file = "data//INDECSuper.txt",
  header = TRUE,
  sep = "\t"
)

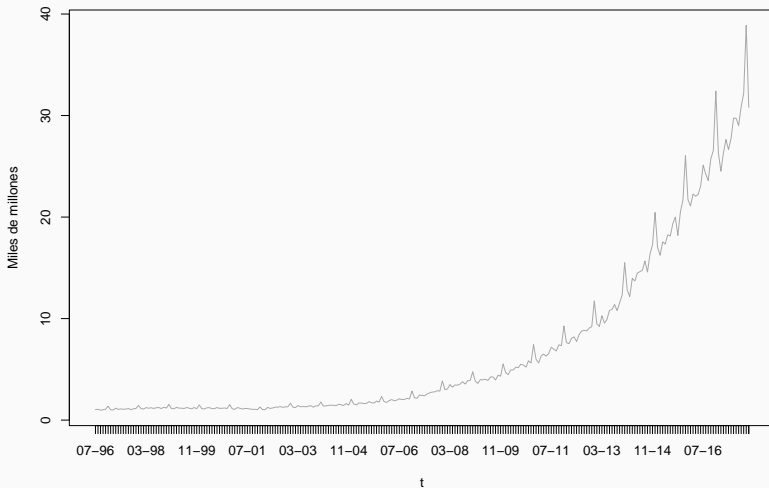
df[, 1] <- as.POSIXct(df[, 1], format = "%Y-%m-%d")

z <- xts(x = df[, 2] / 1000, order.by = df[, 1])
z_ts <- ts(z, frequency = 12) # stl requiere un objeto del tipo ts

t(head(z, 9))

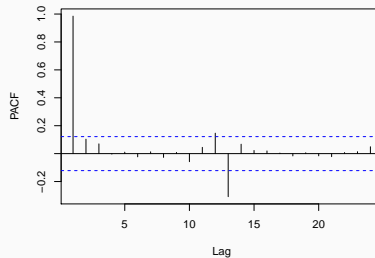
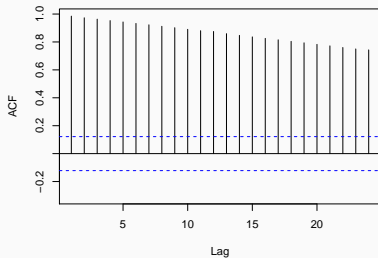
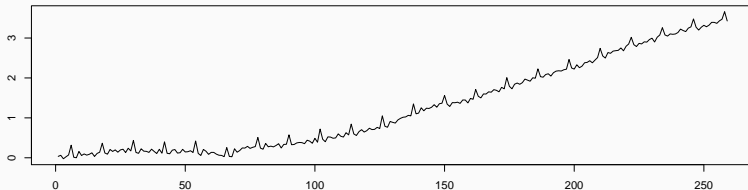
##      1996-07-01 1996-08-01 1996-09-01 1996-10-01 1996-11-01 1996-12-01
## x           1.036         1.064         0.975         1.025         1.073         1.371
##      1997-01-01 1997-02-01 1997-03-01
## x           1.013           1         1.172
```

Ventas en supermercados

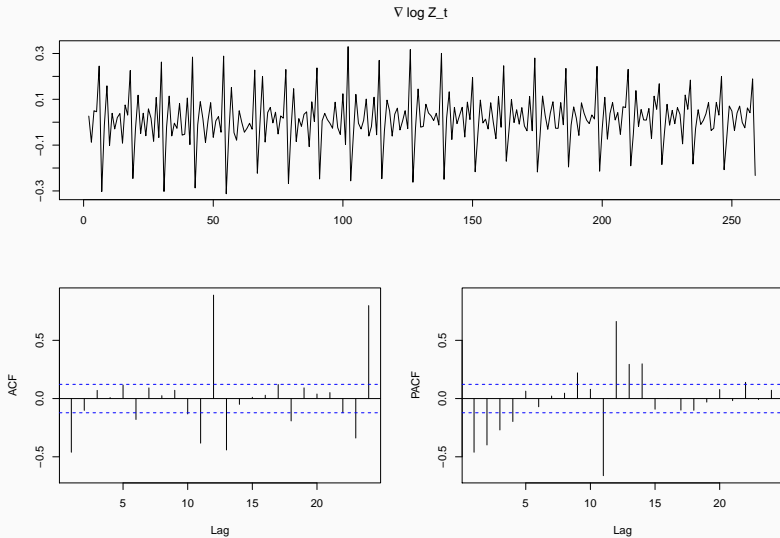


Serie transformada (ln)

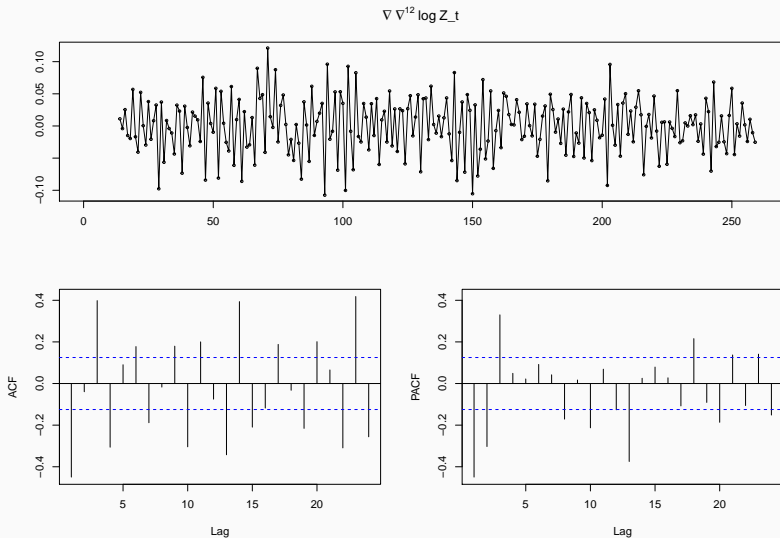
log Z_t



Primera diferencia de la serie transformada (ln)



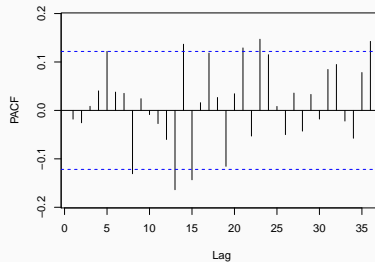
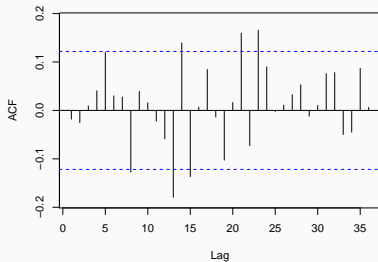
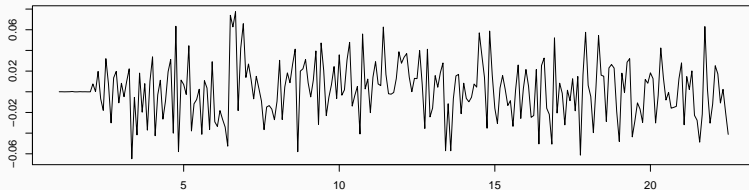
Primera diferencia principal y estacional de la serie transformada (ln)




```
## Series: log(z_ts)
## ARIMA(3,1,0)(2,1,2)[12]
##
## Coefficients:
##          ar1      ar2      ar3    sar1      sar2      sma1      sma2
##      -0.3977  -0.0677  0.3190  0.813   -0.5900  -1.3668  0.6808
## s.e.   0.0697   0.0662  0.0618  0.100   0.0872   0.1265  0.0892
##
## sigma^2 estimated as 0.0007995:  log likelihood=522.97
## AIC=-1029.93   AICc=-1029.32   BIC=-1001.89
```

Residuos de un modelo ajustado

Residuos de SARIMA(3, 1, 0)(2, 1, 2)₁₂ sobre $\log Z_t$



Anexo: Producción de automóviles

Lectura & procesamiento

```
# https://bit.ly/2GXzXoa
df <- read.table(
  file = "data/haciendaAutos.txt",
  header = TRUE,
  sep = "\t"
)

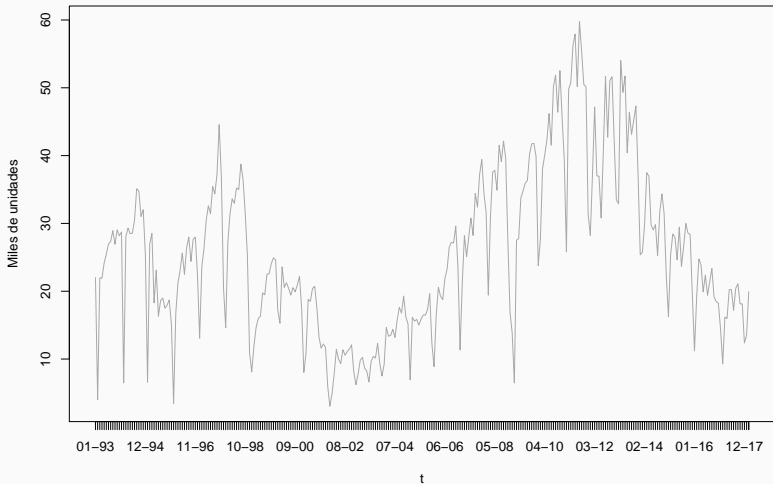
df[, 1] <- as.POSIXct(df[, 1], format = "%Y-%m-%d")

z <- xts(x = df[, 2] / 1000, order.by = df[, 1])
z_ts <- ts(z, frequency = 12) # stl requiere un objeto del tipo ts

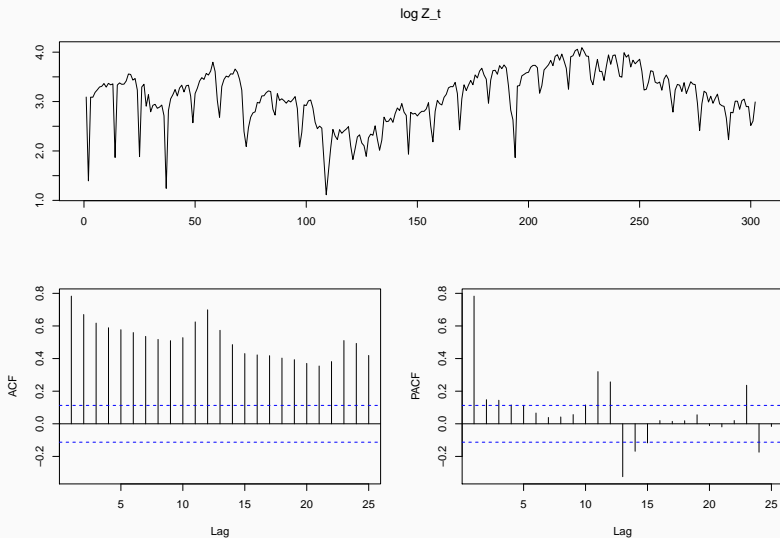
t(head(z, 9))

##      1993-01-01 1993-02-01 1993-03-01 1993-04-01 1993-05-01 1993-06-01
## x          22.01      4.033      21.971      21.919      24.172      25.468
##      1993-07-01 1993-08-01 1993-09-01
## x          26.967      27.342      28.936
```

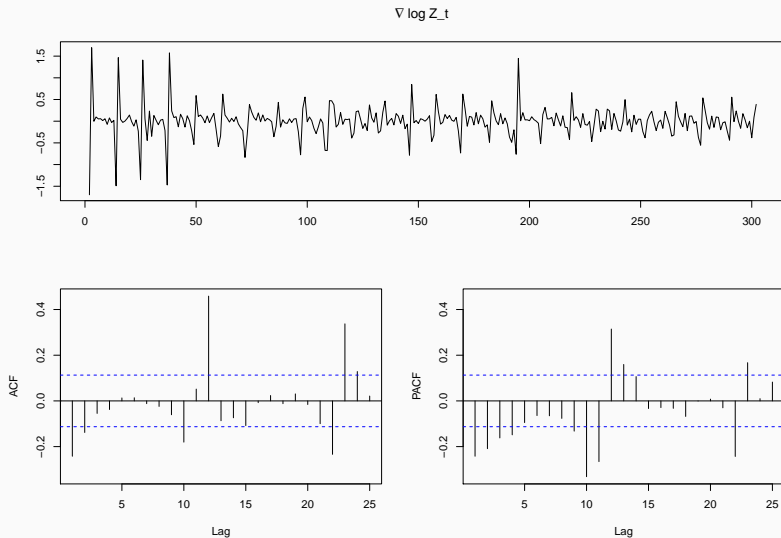
Producción de automóviles



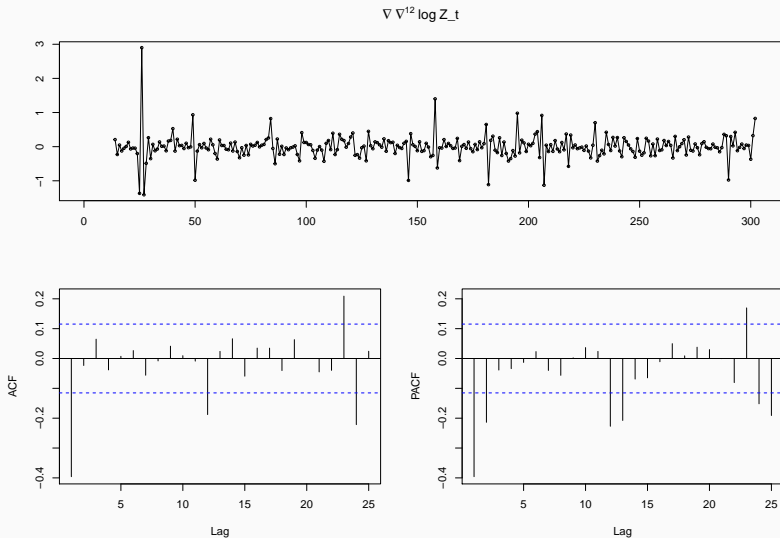
Serie transformada (ln)



Primera diferencia de la serie transformada (ln)



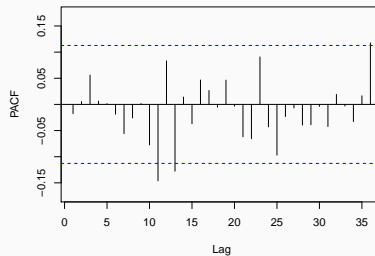
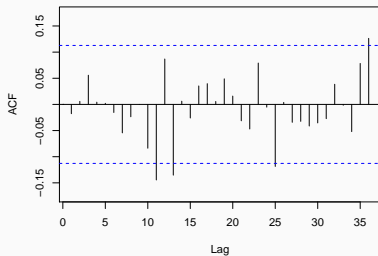
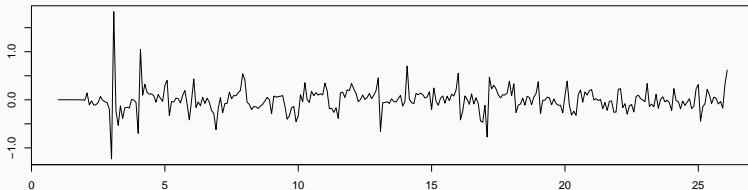
Primera dif. principal y estacional de la serie transformada (ln)




```
## Series: z_ts
## ARIMA(0,1,1)(1,1,1)[12]
## Box Cox transformation: lambda= 0
##
## Coefficients:
##          ma1      sar1      sma1
##      -0.5766  0.2350  -1.0000
## s.e.   0.0486  0.0666  0.0604
##
## sigma^2 estimated as 0.06379:  log likelihood=-27.68
## AIC=63.36   AICc=63.5   BIC=78.02
```

Residuos de un modelo ajustado

Residuos de SARIMA(2, 1, 0)(1, 0, 0)₁₂ sobre $\log Z_t$



Hyndman, Rob J, and George Athanasopoulos. 2018. *Forecasting: Principles and Practice*. <https://otexts.org/fpp2/>.