Introducción a la Inteligencia Predictiva en R con aplicaciones:

Series Temporales.

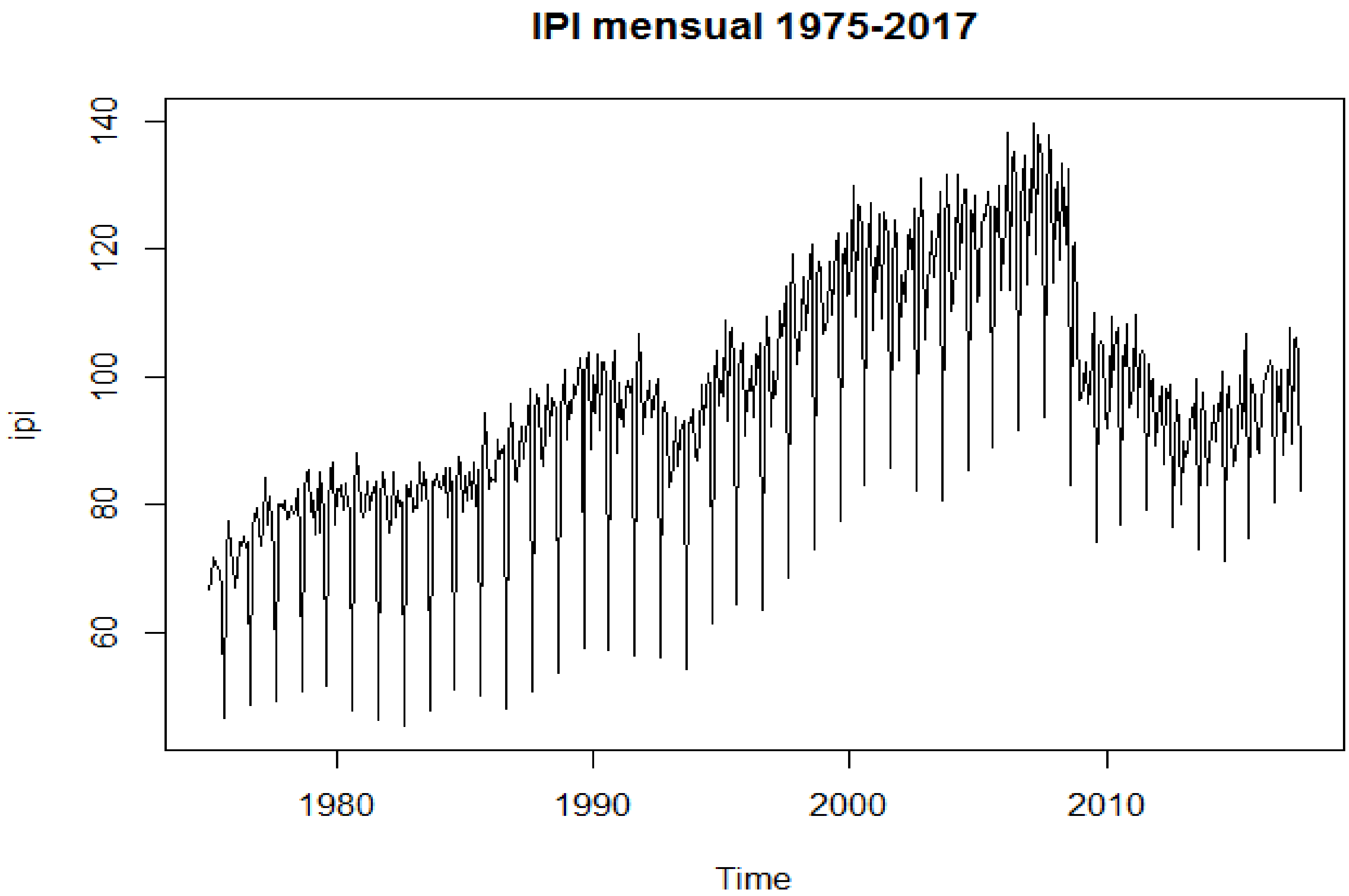
Diego J. Pedregal
Universidad de Castilla-La Mancha
Diego.Pedregal@uclm.es

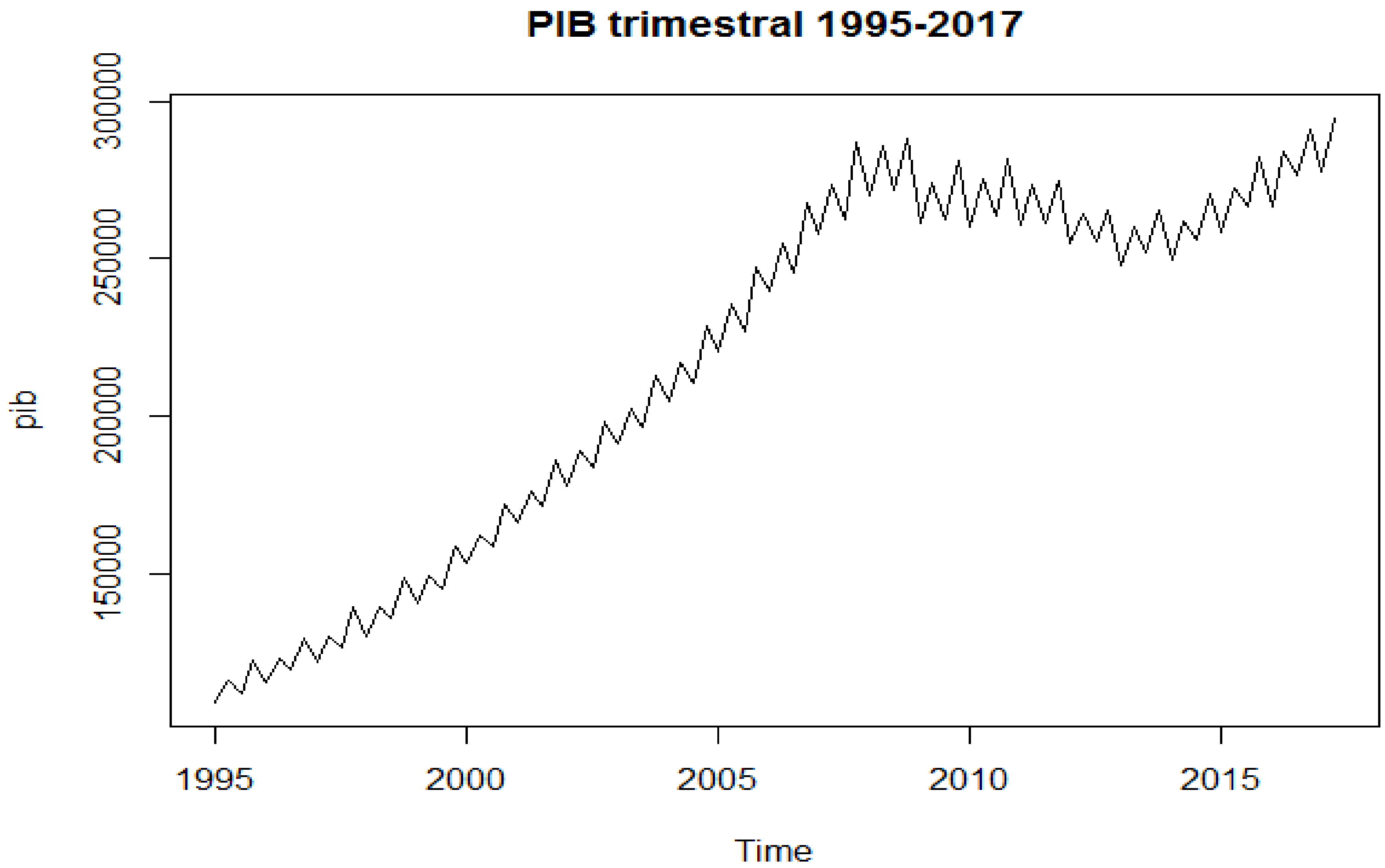
Universidad de Sevilla
15-16 de Noviembre de 2018

Series temporales

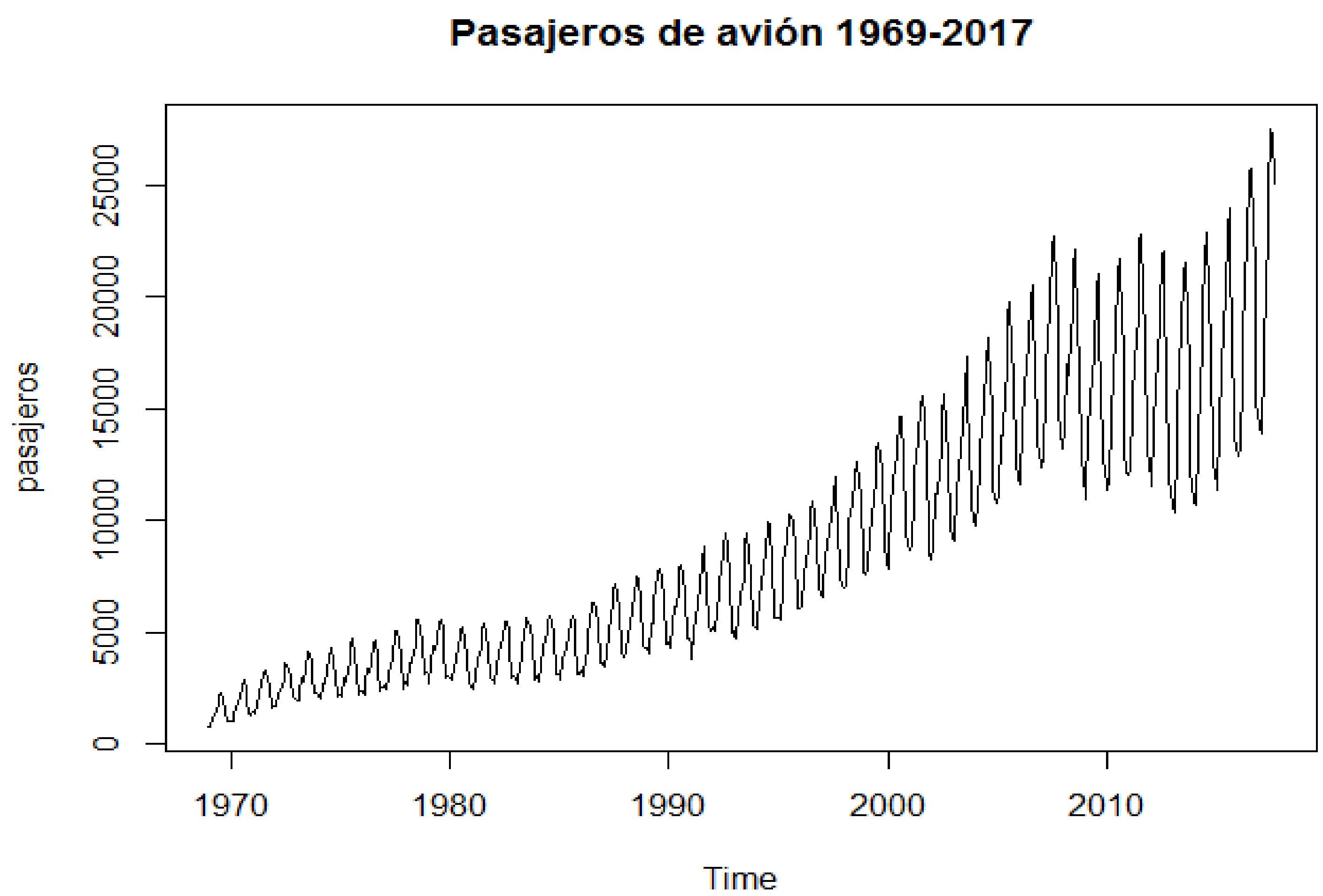
- Una **serie temporal** es una secuencia de observaciones ordenadas en el tiempo
- Asumimos que las observaciones son equidistantes en el tiempo, aunque no tiene por qué ser así
- Ejemplos:
  - Demanda o precios horarios de electricidad en el mercado español
  - Beneficios anuales del BSCH
  - PIB español
  - Inflación
- **Predecir** una serie temporal implica calcular cómo continúa la secuencia de datos en el futuro.

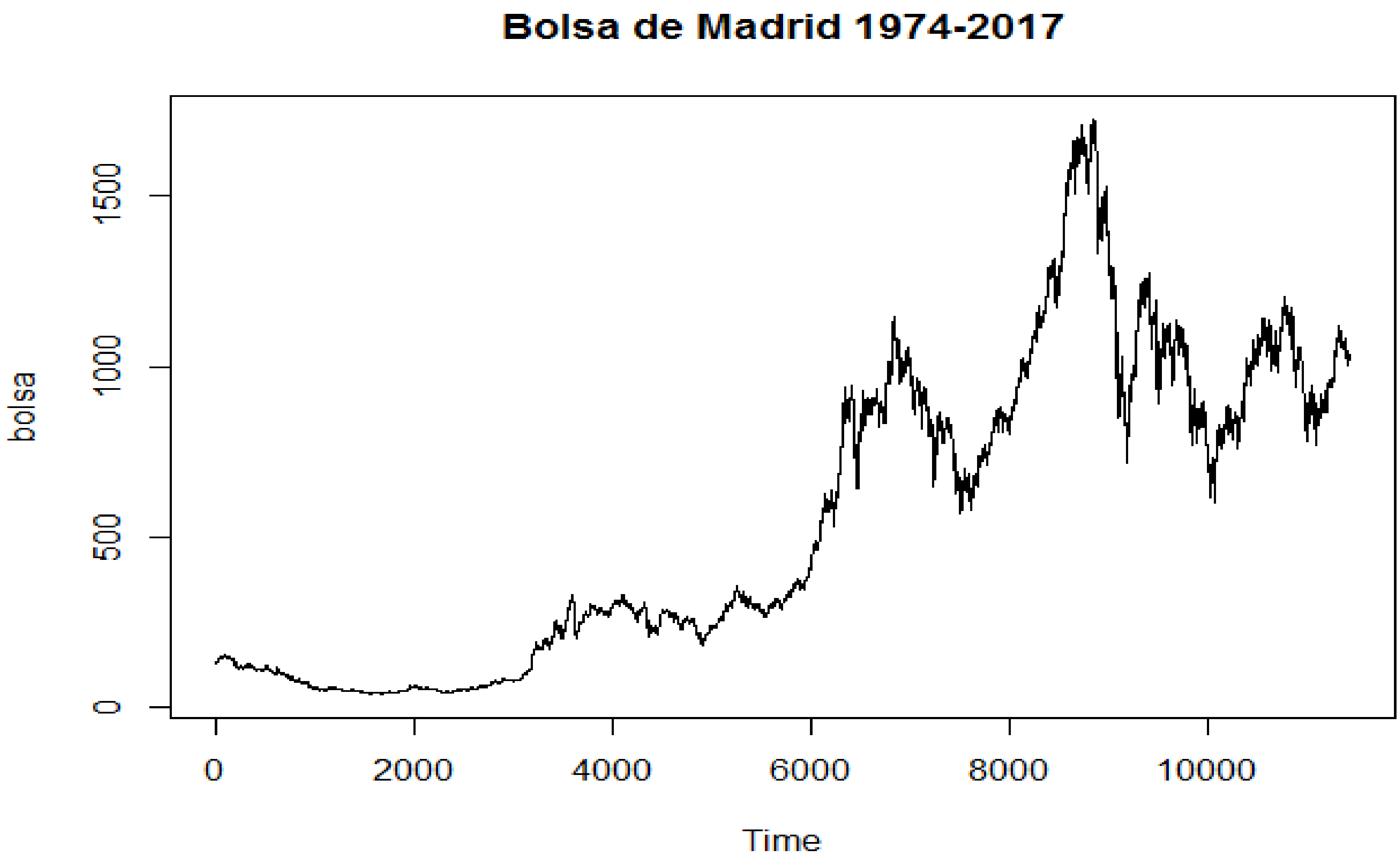
Series temporales




PIB trimestral 1995-2017
pib
150000
200000
250000
300000
1995
2000
2005
2010
2015
Time

Series temporales




Bolsa de Madrid 1974-2017
bolsa
0
500
1000
1500
0
2000
4000
6000
8000
10000
Time

Métodos simples de predicción

- Método de la **media**: la predicción es la media de los datos, $\hat{y}_{T+h|T} = \bar{y} = (y_1 + y_2 + \cdots + y_T)/T$
- Método **Naïve** (Paseo Aleatorio): la predicción es el último valor observado, $\hat{y}_{T+h|T} = y_T$
- **Naïve estacionalidad**: la predicción es el valor observado en la misma posición estacional, $\hat{y}_{T+h|T} = y_{T+h-s}$

PRÁCTICA #2.1

Medidas de precisión

- **Errores de predicción**: es la diferencia entre el valor observado y su predicción basada en todas las observaciones: e_t =*predicción – valores reales*
- Medidas con unidades:
  - ME= $\frac{1}{T}\sum_{t=1}^{T} e_t$
  - MAE= $\frac{1}{T}\sum_{t=1}^{T}|e_t|$
  - RMSE= $\sqrt{\frac{1}{T}\sum_{t=1}^{T} e_t^2}$
- Medidas sin unidades:
  - MAPE= $100\frac{1}{T}\sum_{t=1}^{T}|e_t|/|y_t|$
  - PRMSE= $100\frac{1}{T}\sum_{t=1}^{T}(e_t)^2/|y_t|$

PRÁCTICA #2.2

Autocorrelación

- Covarianza y correlación miden el grado de relación lineal entre dos variables
- Autocovarianzas y autocorrelaciones miden la dependencia lineal de una variable con sus retardos
- Autocovarianzas:

$$c_k = \frac{1}{T} \sum_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y})$$

- Autocorrelación:

$$r_k = \frac{c_k}{c_0}$$

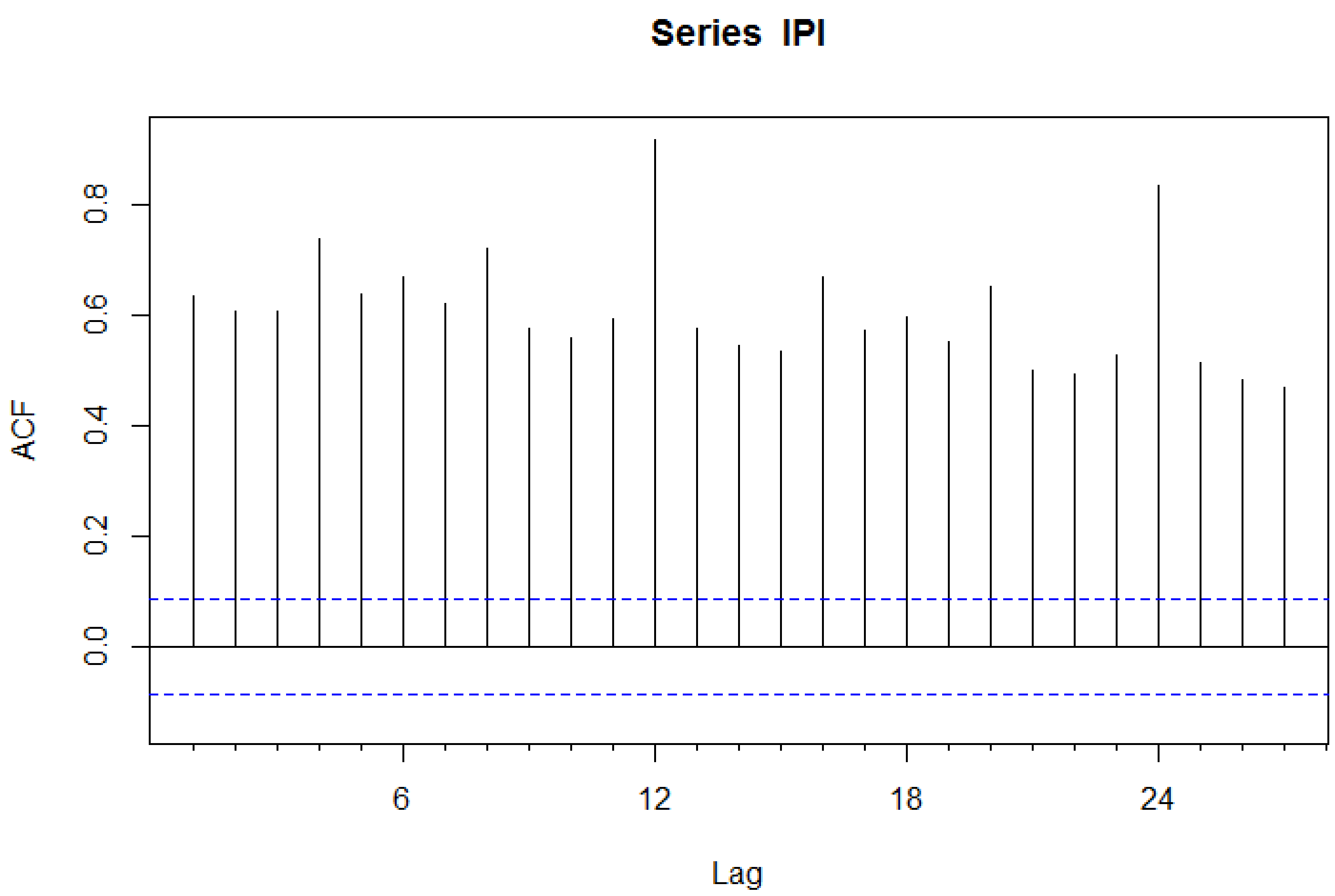
Autocorrelación

- Autocorrelaciones muy parecidas a realizar las siguientes autorregresiones:

$$y_t = c_0^{1)} + c_1 y_{t-1}$$
$$y_t = c_0^{2)} + c_2 y_{t-2}$$
$$\vdots$$
$$y_k = c_0^{k)} + c_k y_{t-k}$$

- **Función de autocorrelación simple** es la representación gráfica de los coeficientes r_k frente a los retardos.

Autocorrelación



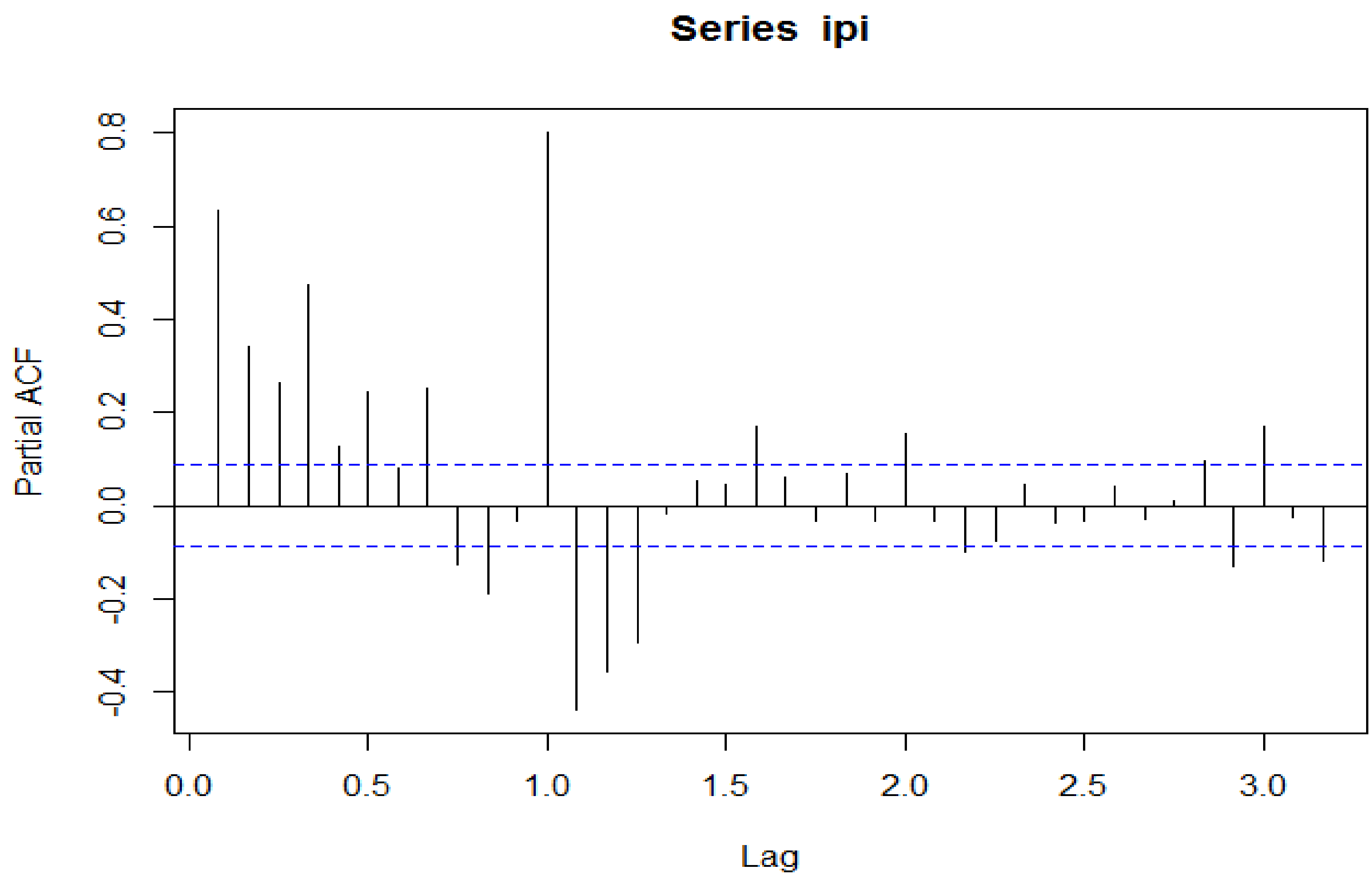
Autocorrelación

- Autocorrelaciones parciales:

$$y_t = p_0^{1)} + p_1 y_{t-1}$$
$$y_t = p_0^{2)} + p_1^{2)} y_{t-1} + p_2 y_{t-2}$$
$$\vdots$$
$$y_k = p_0^{k)} + p_1^{k)} y_{t-1} + \cdots + p_k y_{t-k}$$

- **Función de autocorrelación parcial** es la representación gráfica de los coeficientes p_k frente a los retardos.

Autocorrelación



Autocorrelación

- **Ruido blanco**: $y_t = a_t$. Serie temporal sin autocorrelación con media cero y varianza constante (normalmente se requiere independencia). Es una serie puramente aleatoria y por tanto impredecible.
  - ¿Cómo serán las funciones de autocorrelación simple y parcial?
  - ¿Cuál será la predicción óptima de este modelo?
- **Paseo aleatorio**: $y_t = y_{t-1} + a_t$
  - ¿Cómo serán las funciones de autocorrelación simple y parcial de $y_t - y_{t-1} = a_t$?
  - ¿Cuál será la predicción óptima de este modelo?

Autocorrelación

- Pormateau test de autocorrelación: Ljung-Box test. Se realiza para *m* valores de la función de autocorrelación simple.

$$Q_m = T(T+2)\sum_{k=1}^{m}\frac{r_k^2}{T-k} \approx \chi^2_{m-número\ parámetros}$$

PRÁCTICA #2.3

Estacionariedad

- Las herramientas estadísticas se pueden desarrollar para series temporales que son estacionarias en media y varianza. Es decir, que tienen la media y varianza constantes a lo largo del tiempo.
- Sin embargo, la mayor parte de las series temporales **NO** son estacionarias.
- Existen varias soluciones para este problema, pero la que utilizaremos en este curso es la de aplicar transformaciones a los datos para reducirlas a estaconariedad.

Transformación de varianza Box-Cox

$$z_t = \begin{cases} ln(y_t), & \text{con } \lambda = 0 \\ \dfrac{\left(y_t^{\lambda} - 1\right)}{\lambda}, & \text{con } \lambda \neq 0 \end{cases}$$

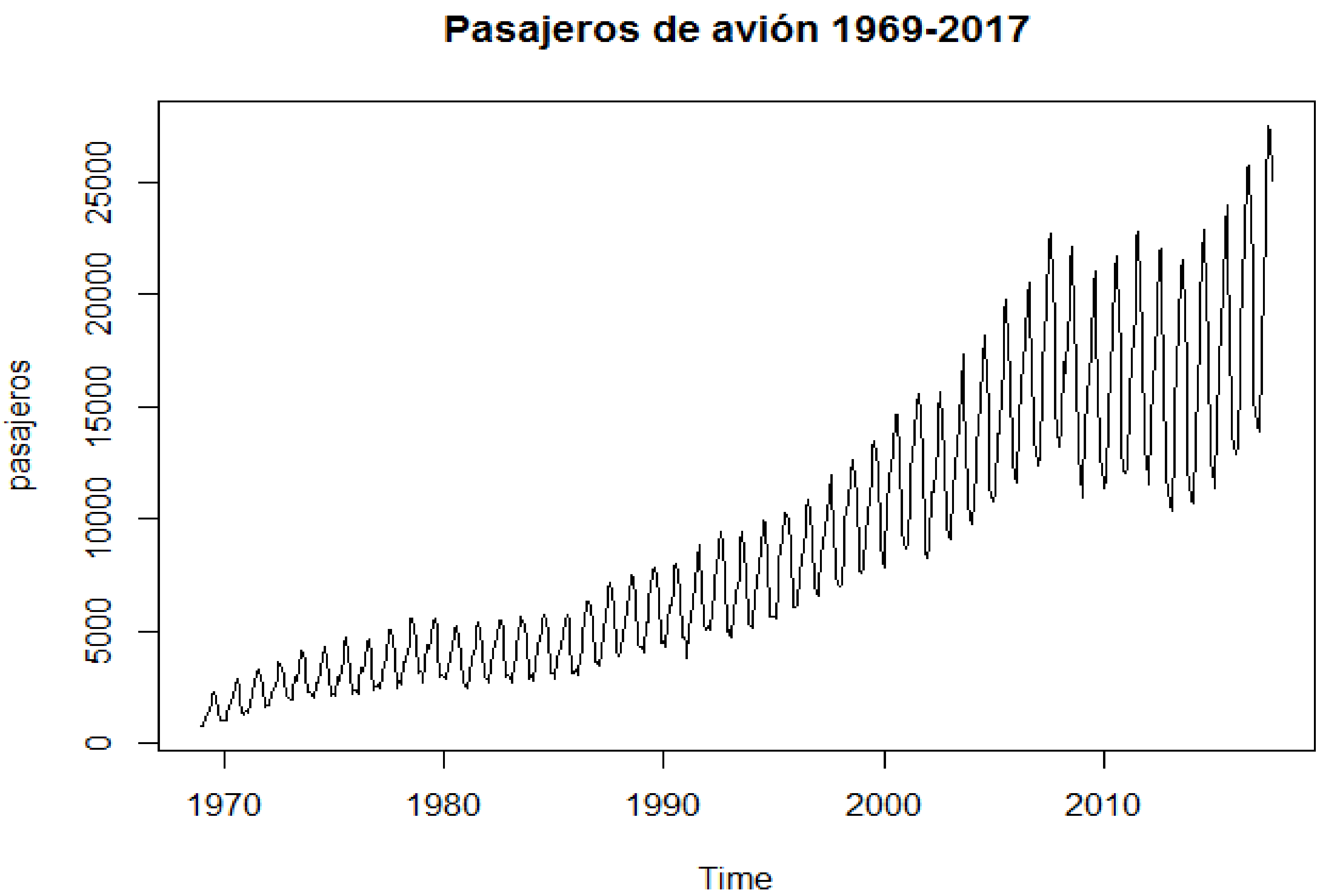
- $\lambda = 1$, no hay transformación
- $\lambda = 0{,}5$, raíz cuadrada
- $\lambda = 0$, logaritmo neperiano
- $\lambda = -1$, inversa

- A veces la transformación no soluciona el problema
- A menudo los resultados no son muy sensibles al valor de λ
- $\lambda = 0$ fuerza a que las predicciones sean positivas

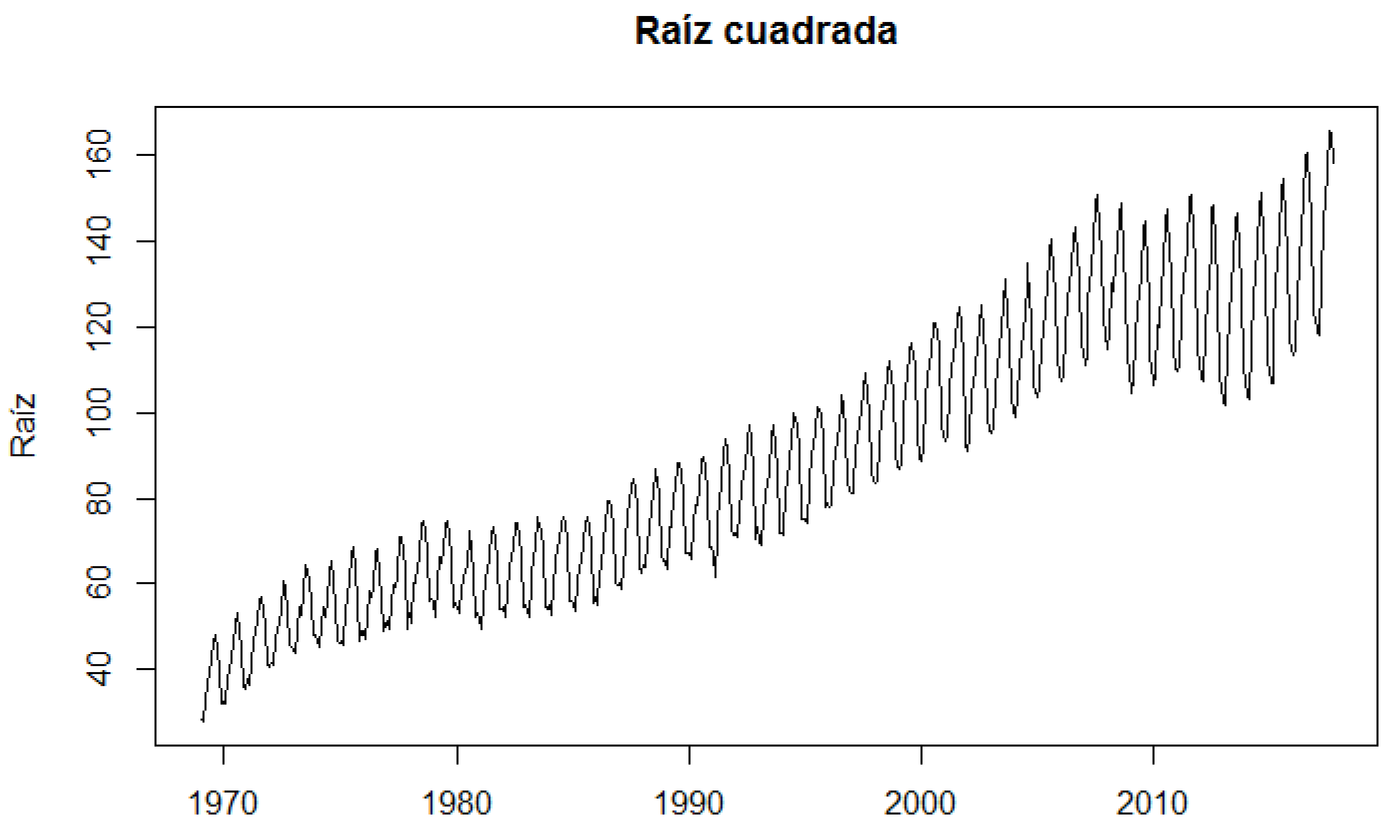
Transformación de varianza Box-Cox

- La transformación hay que invertirla para volver a las unidades iniciales de la variable

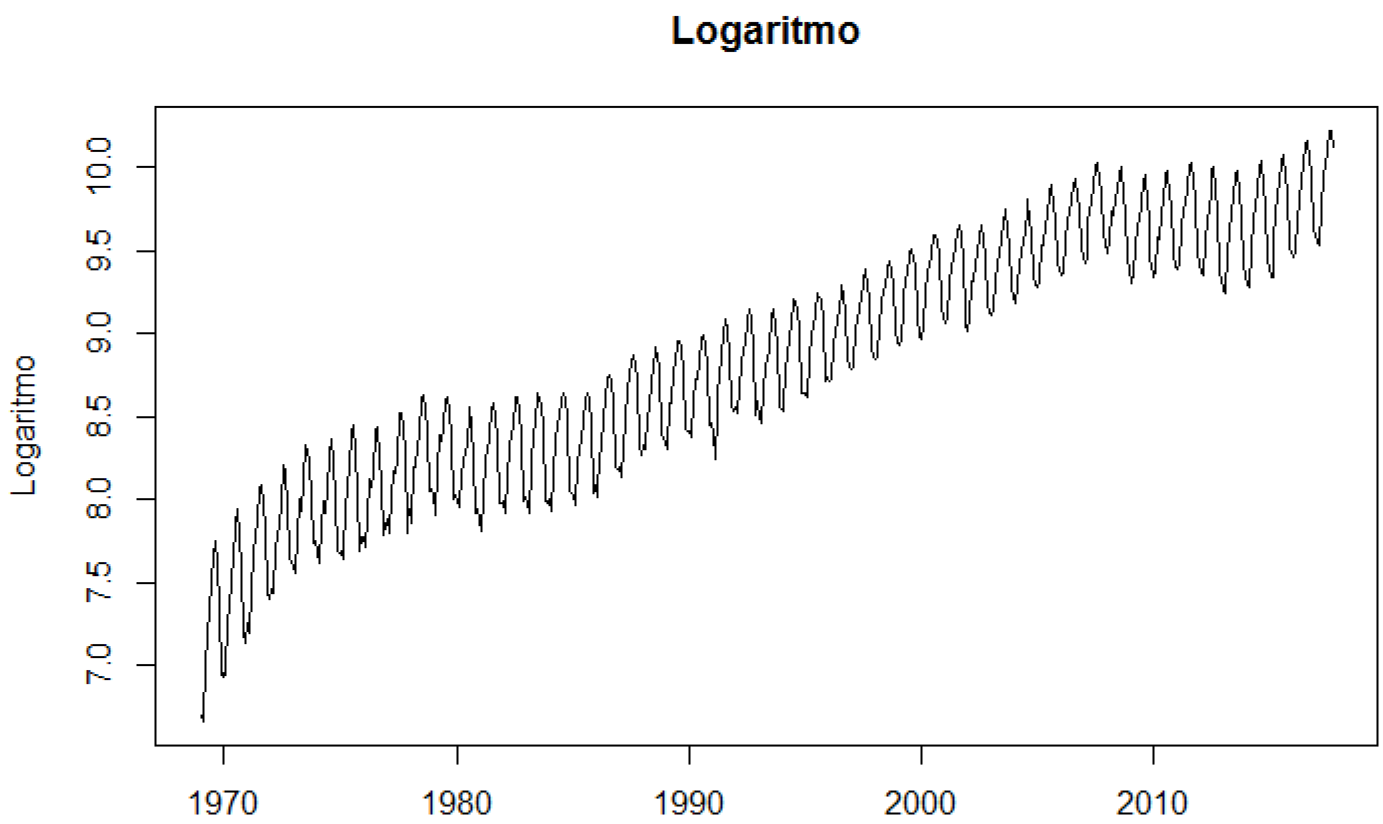
$$y_t = \begin{cases} exp(z_t), & \text{con } \lambda = 0 \\ (\lambda z_t + 1)^{1/\lambda}, & \text{con } \lambda \neq 0 \end{cases}$$


Pasajeros de avión 1969-2017
pasajeros
0
5000
10000
15000
20000
25000
1970
1980
1990
2000
2010
Time

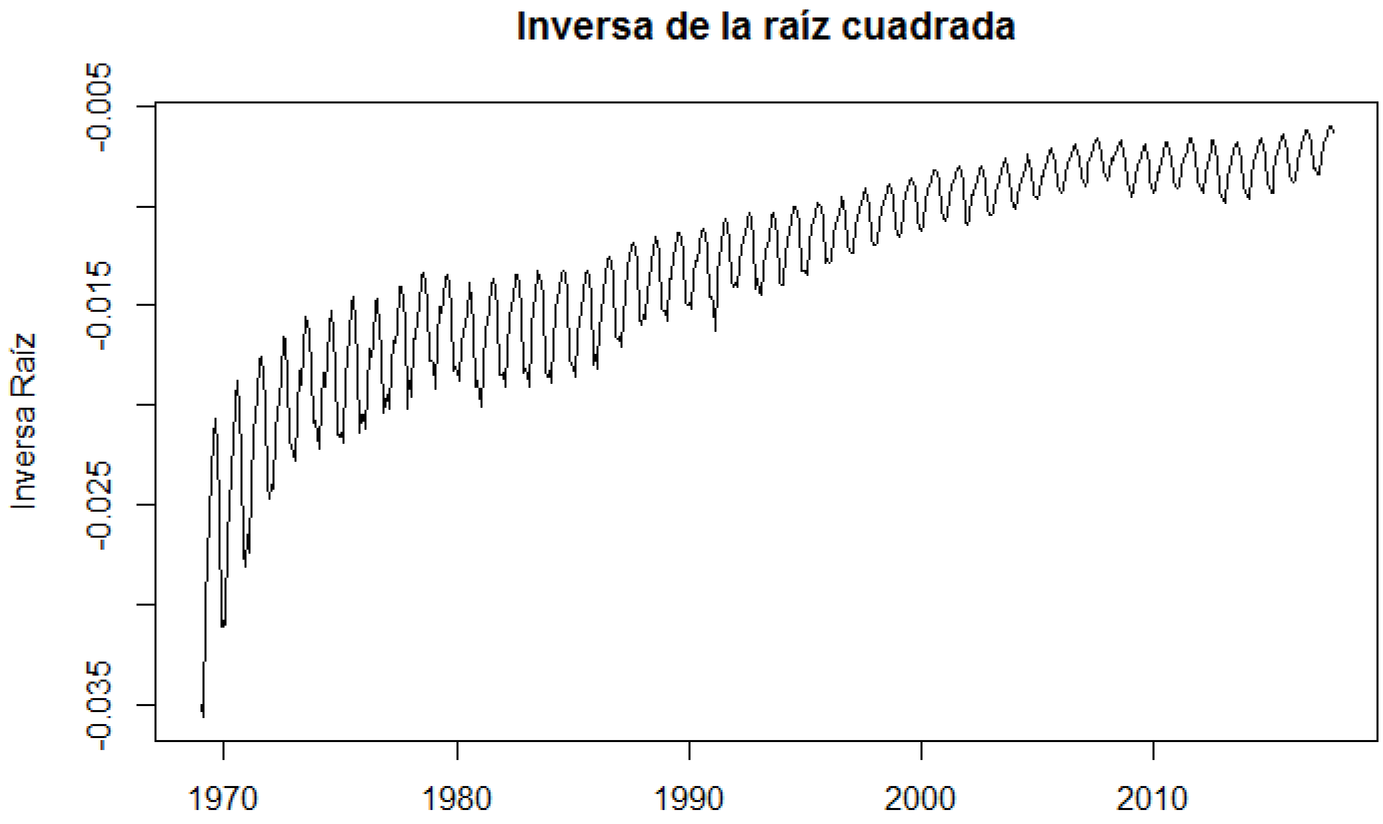
Transformación de varianza Box-Cox



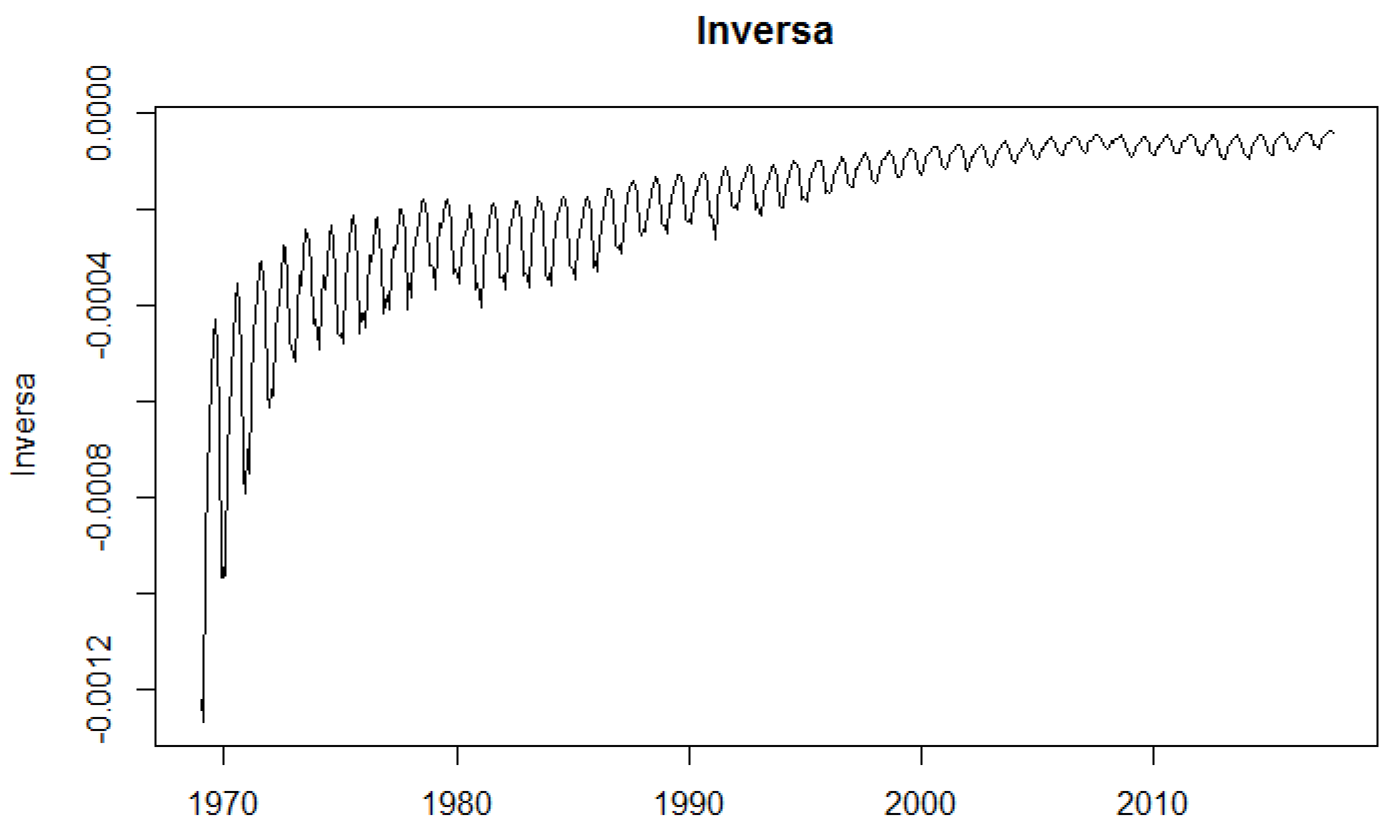
Transformación de varianza Box-Cox



Transformación de varianza Box-Cox



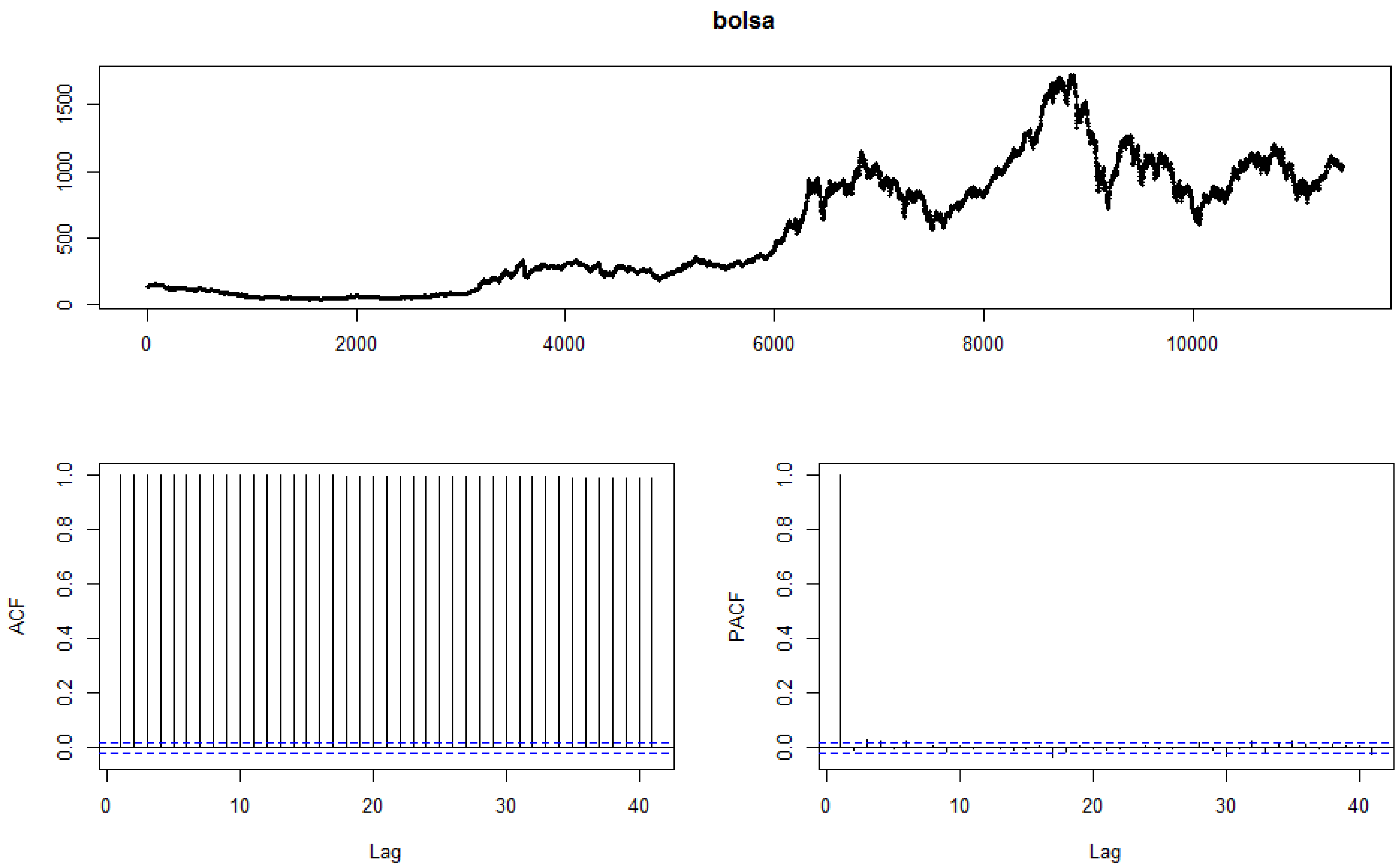
Transformación de varianza Box-Cox



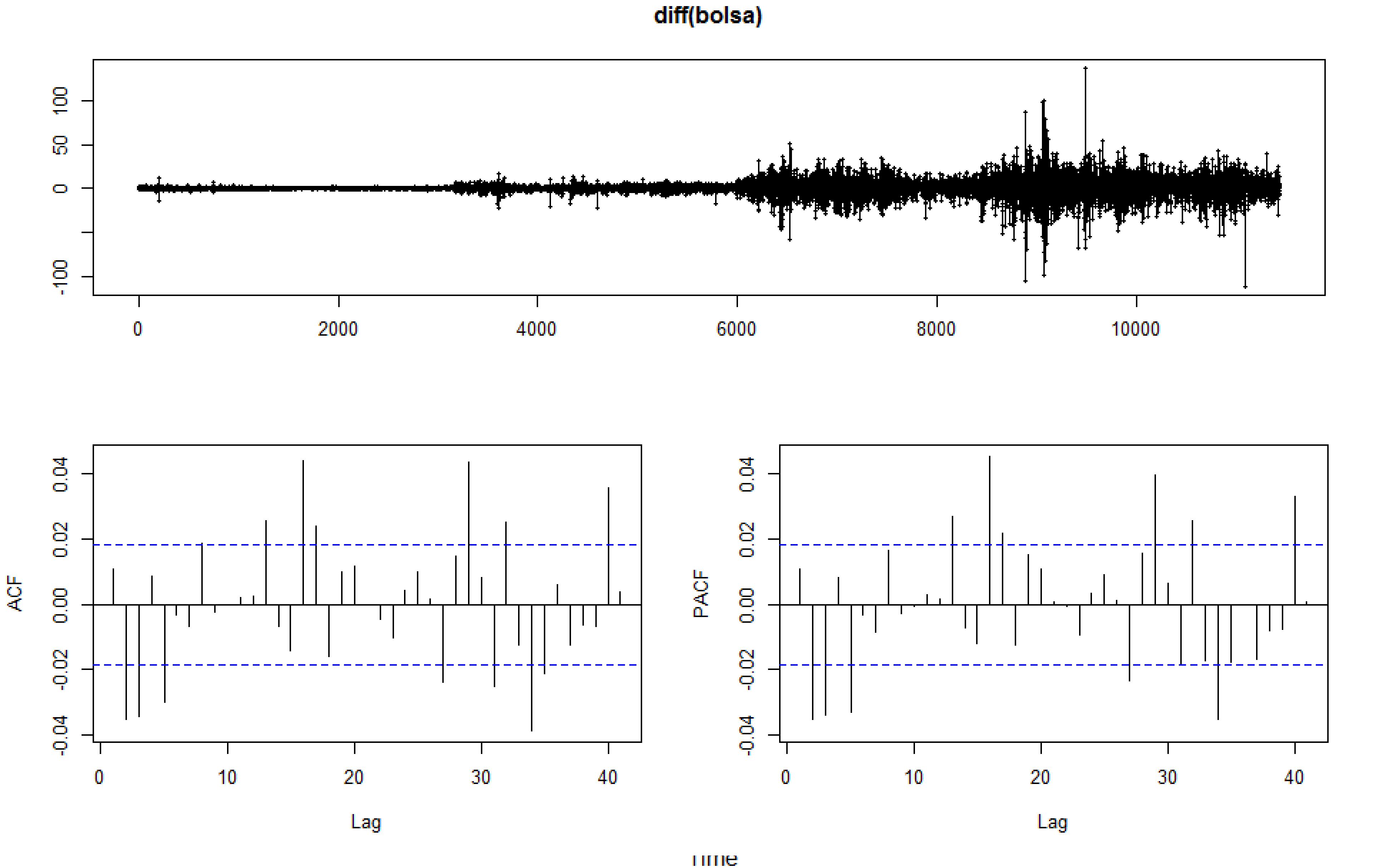
Transformación de media (diferenciación)

- Muchas series no son estacionarias en media, pero se puede lograr que lo sean mediante la diferenciación
- La no estacionariedad en media se puede identificar mediante:
  - El gráfico de la serie, que presenta una tendencia
  - La autocorrelación simple decrece lentamente
  - La primera autocorrelación es grande y positiva
- Tomar una diferencia regular es restar a cada observación el valor anterior: $d_t = y_t - y_{t-1}$. Se pierde una observación

Transformación de media (diferenciación)



Transformación de media (diferenciación)



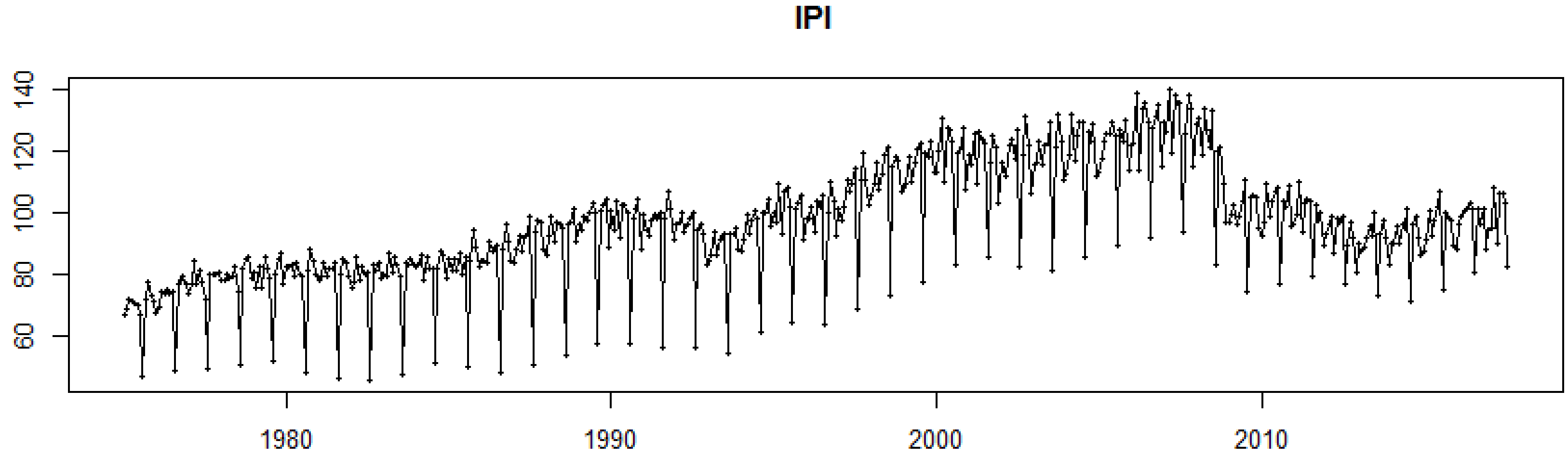
Transformación de media (diferenciación)

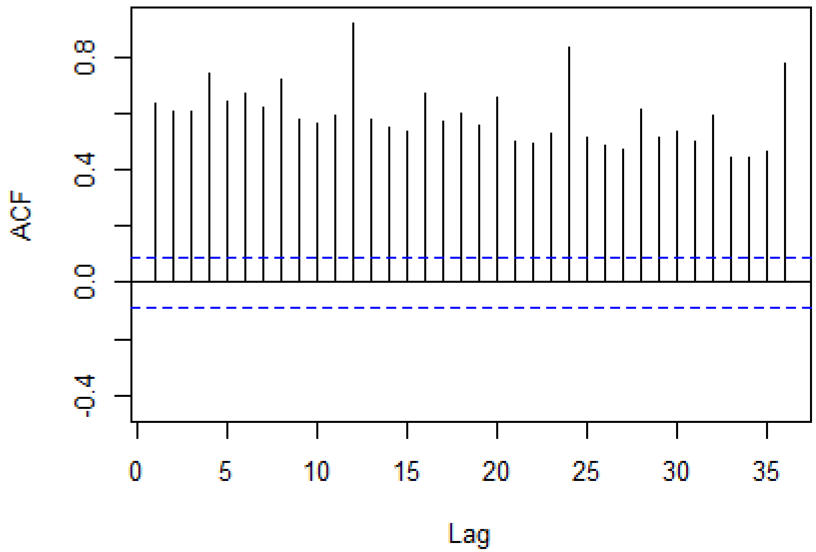
- Si con una diferencia regular no se consigue estacionariedad en media se debe tomar otra diferencia:

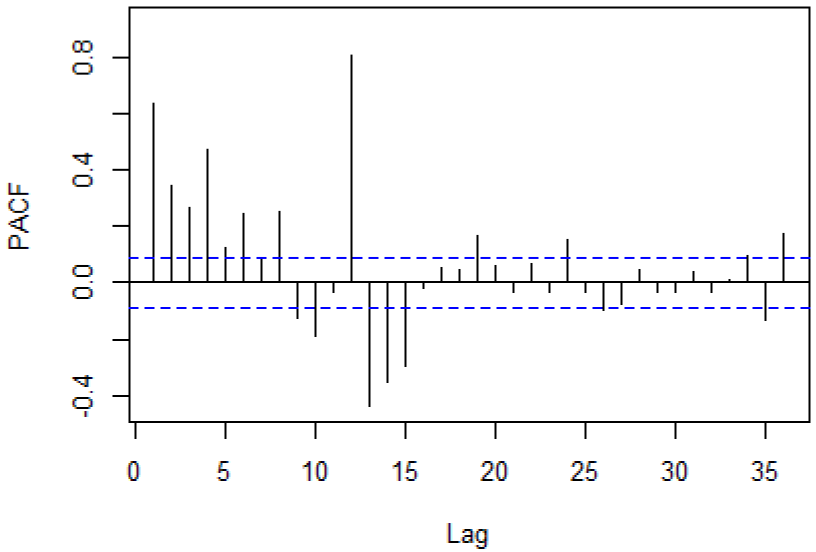
$$z_t = d_t - d_{t-1} =$$
$$(y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$$

- Se pierden dos observaciones
- Es muy raro tener que diferenciar más de dos veces
- Cuando la estacionalidad se detecta visualmente es porque es necesario tomar una diferencia estacional, es decir, $d_t = y_t - y_{t-s}$, siendo *s* el número de observaciones por año (12 para observaciones mensuales, 4 para trimestrales, etc.)

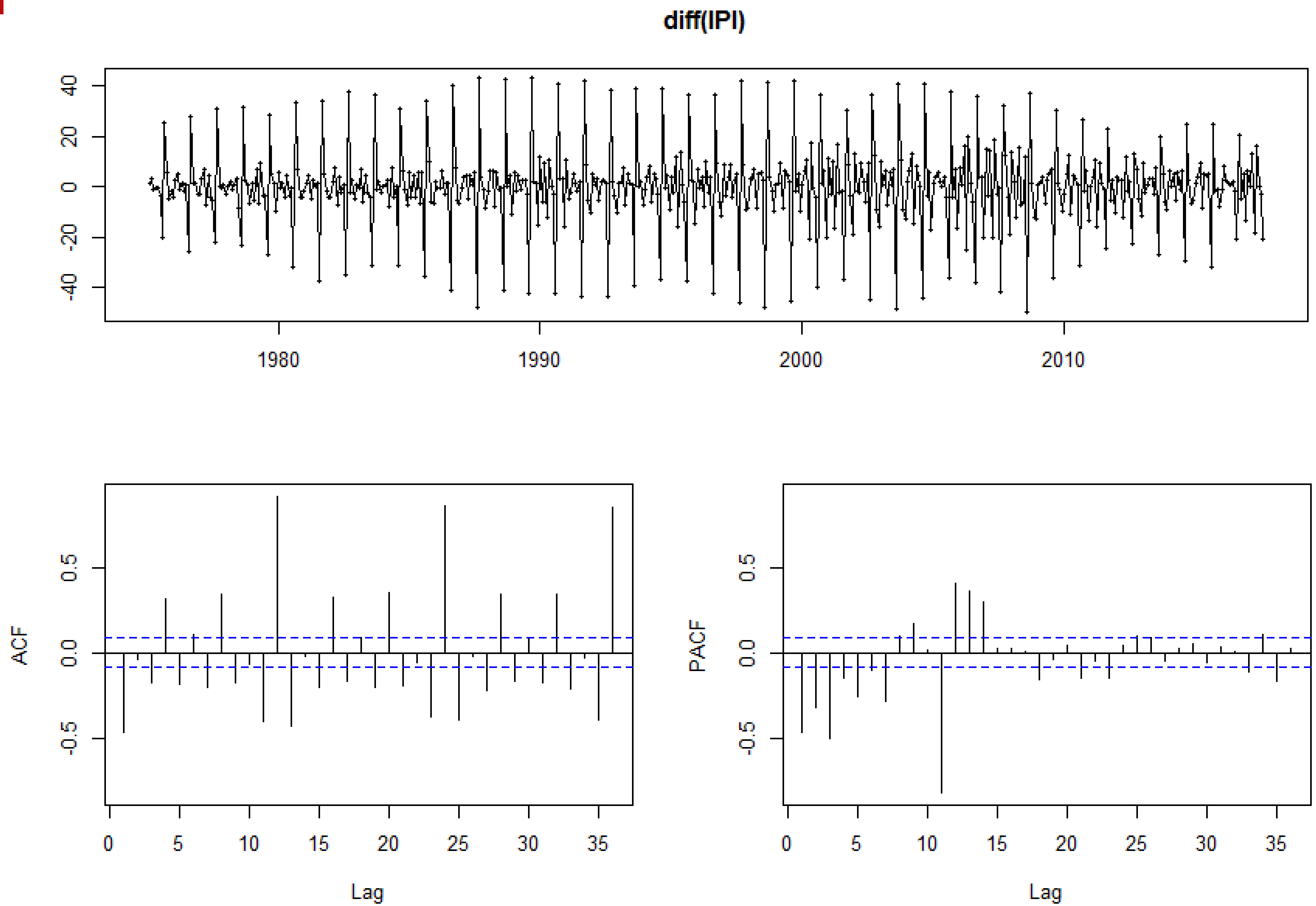
Transformación de media (diferenciación)



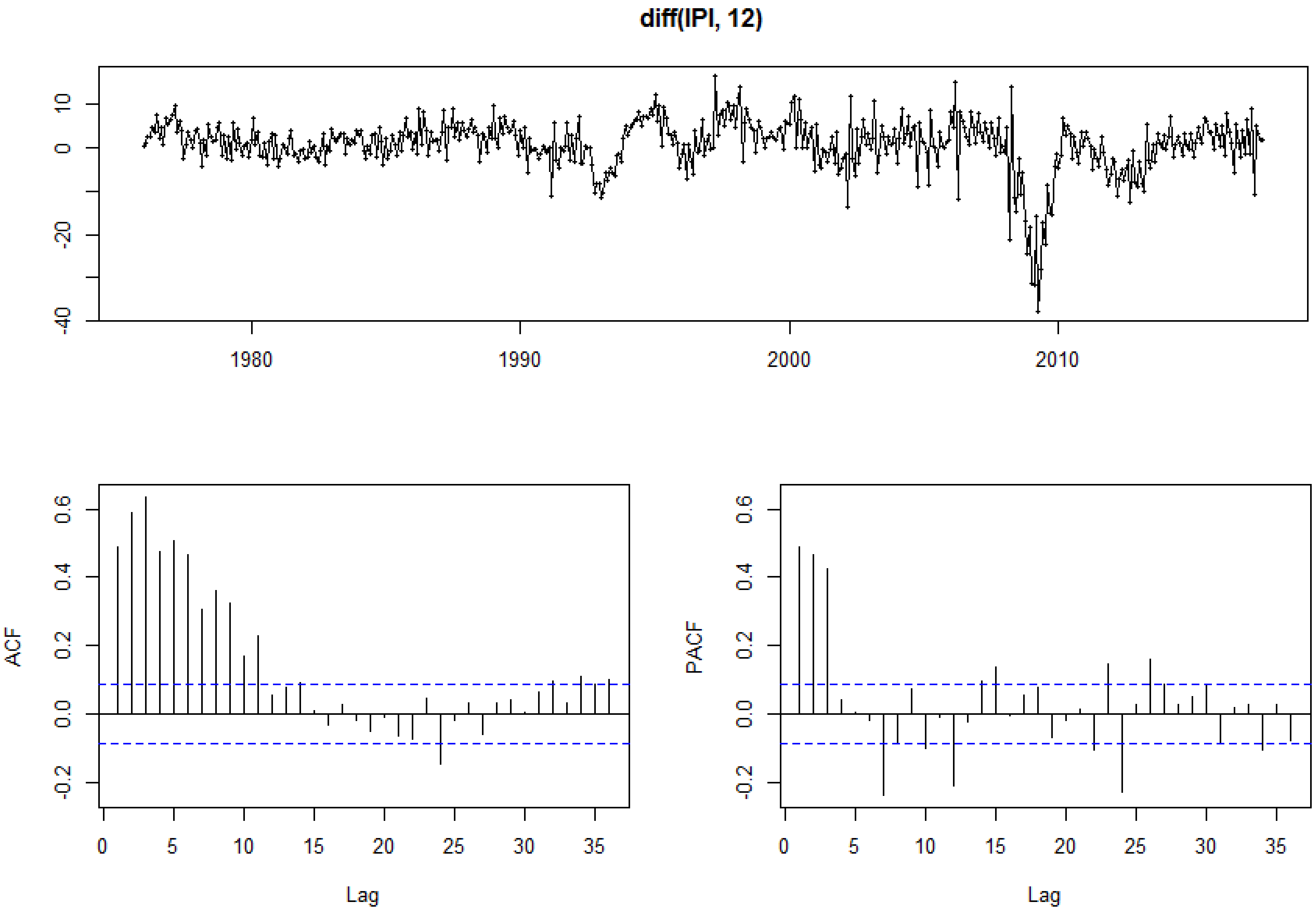




Transformación de media (diferenciación)



Transformación de media (diferenciación)



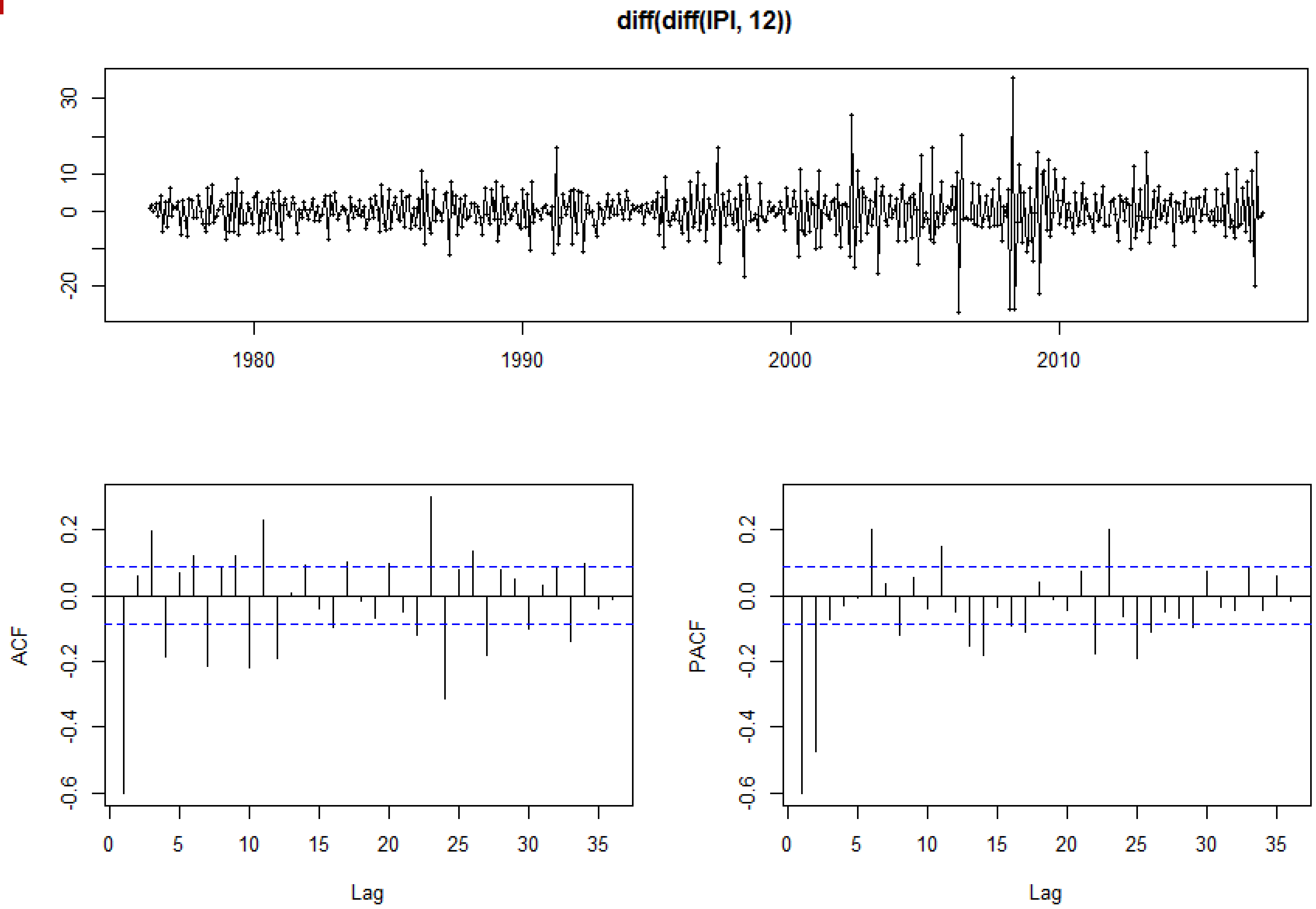
Transformación de media (diferenciación)

- A menudo es necesario tomar los dos tipos de diferencias:

$$d_t = y_t - y_{t-1}$$
$$z_t = d_t - d_{t-s} = (y_t - y_{t-1}) - (y_{t-s} - y_{t-s-1})$$
$$z_t = y_t - y_{t-1} - y_{t-s} + y_{t-s-1}$$

- En general, es recomendable hacer primero la diferencia estacional porque puede resultar que es estacionaria, y entonces no es necesario tomar la diferencia regular
- Un procedimiento formal para ver las diferencias necesarias es utilizar tests de raíces unitarias

Transformación de media (diferenciación)



PRÁCTICA #2.4

Suavizado exponencial

Exponen**T**ial **S**mooting

Suavizado o alisado exponencial

- ETS parte de la idea intuitiva de que una predicción sensata del futuro es función del pasado descontado de forma que cuanto más nos alejamos en el tiempo menos influencia tiene. Por ejemplo:

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha) y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots$$

$$0 \leq \alpha \leq 1$$

Suavizado exponencial

- Basándose en esta idea algunos autores han desarrollado modelos en los que se permite combinar distintas versiones de modelos de tendencia, estacional y error, dentro de descomposiciones aditivas ($y_t = T_t + S_t + e_t$), multiplicativas ($y_t = T_t S_t e_t$) o mixtas ($y_t = T_t S_t + e_t$). En total son 30 combinaciones posibles:
  - Tendencia: Ninguna, Aditiva, Multiplicativa, (amortiguada o no)
  - Estacional: Ninguna, Aditiva, Multiplicativa.
  - Error: Aditivo o Multiplicativo.

Suavizado exponencial

- Se usa normalmente la nomenclatura ETS(Error, Tendencia, Estacional). Por ejemplo, un modelo ETS(A, A, A) responde a la siguiente formulación (α, β, γ son parámetros que hay que estimar y s es el período estacional, además es necesario estimar condiciones iniciales de los componentes):

Predicción: $\hat{y}_{T+1|T} = T_t + b_t + S_{t-s}$

Nivel: $T_t = \alpha(y_t - S_{t-s}) + (1-\alpha)(T_{t-1} + b_{t-1})$

Pendiente: $b_t = \beta(T_t - T_{t-1}) + (1-\beta)b_{t-1}$

Estacional: $S_t = \gamma(y_t - T_{t-1} - b_{t-1}) + (1-\gamma)S_{t-s}$

Además se pueden considerar errores autocorrelados, p.ej., siguiendo modelos ARMA.

Suavizado exponencial

- Se estiman los 30 modelos por máxima verosimilitud y con un criterio de información se selecciona el mejor de ellos. Se prima el uso del criterio de Akaike, aunque hay otros (L es la verosimilitud en el óptimo, T es el número de observaciones y k es el número de parámetros del modelo):
  - Criterio de Akaike: AIC$= -2\ln(L) + 2k$
  - Criterio de Akaike corregido: AICc $= \mathrm{AIC} + \frac{2k^2+2k}{T-k-1}$
  - Criterio de Schwartz: BIC $= -2\ln(L) + \ln(T)\,k$
  - …

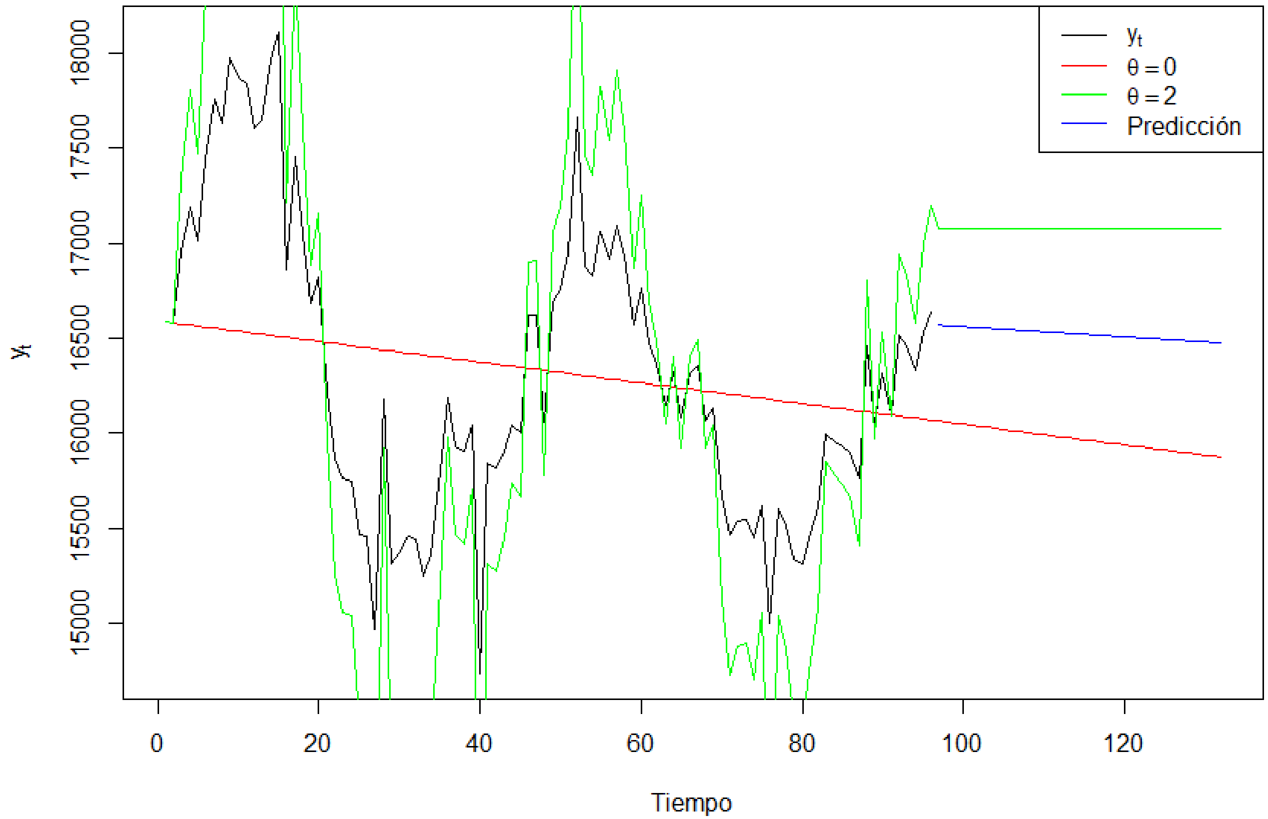
PRÁCTICA #2.5

Método Theta

- El método Theta es una forma ad-hoc de predecir series temporales con buenos resultados cuando se aplica ciegamente a bases de datos grandes.
- El coeficiente Theta se aplica a la segunda diferencia de la serie original. Es decir,
  - $z_t = \theta(y_t - 2y_{t-1} + y_{t-2})$
- El método consiste en los siguientes pasos:
  - Se desestacionaliza la serie por algún método ($\hat{y}_t = y_t - \hat{S}_t$ o $\hat{y}_t = y_t / \hat{S}_t$).
  - Se predicen independientemente dos transformaciones Theta, con parámetros $\theta = 0$ y $\theta = 2$. La primera es en realidad la predicción de una regresión lineal, la segunda se predice con alisado exponencial simple.

Método Theta

- Se combinan las dos predicciones.
- Se añade estacionalidad a la predicción.



PRÁCTICA #2.6