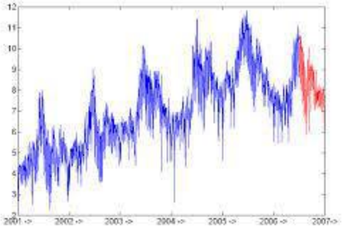
Series temporales y minería de flujos de datos

Series temporales



UNIVERSIDAD
DE GRANADA

DECSAI

Contenido

- Predicción
- Herramientas básicas
- Regresión
- Descomposición de series temporales
- Modelos ARIMA
- Modelos de predicción avanzados
- Minería de series temporales

Bibliografía

- R. Hyndman, G. Athanasopoulus, «Forecasting: principles and practice» 2021, 3rd Edition, www.otexts.org/fpp3; www.otexts.org/fppsp
- D. Peña, "Análisis de series temporales", Alianza Editorial, 2010
- C. Chatfield, «The analysis of time series: An Introduction», CRC Press, 2003
- P.J. Brockwell, R.A. Davis, «Introduction to Time Series and Forecasting», 2nd ed., Springer, 2002

Bibliografía (2)

- S.G. Makridakis, S.C. Wheelwright, R.J. Hyndman, «Forecasting», 3rd Ed., Wiley & Sons, 1998
- A.K.Palit, D. Popovic, «Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications», Springer, 2005
- R.H. Shumway, D.S. Stoffer, «Time Series Analysis and Its Applications», Springer, 2nd Ed., 2006

Software

- Python (pypi.org)
  - pandas (time series)
  - statsmodels
  - tslearn
  - numpy
  - matplotlib
  - Time series in Python (GitHub)

- *Entornos virtuales*
- *Contenedores*

Software

- R
  - CRAN Task View: [“Time Series Analysis”]
  - Paquetes de referencia:
    - forecast, fable (tidyverse)
    - ts, xts, zoo
    - fpp

Predicción

Predicción

- La predicción ha fascinado a los seres humanos durante miles de años
- Algunas referencias históricas:
  - Los profetas de la Biblia
  - Sacerdotes de Babilonia
  - Los griegos consultando el oráculo de Delfos
  - Adivinos ¿Rappel?

¿Qué se puede predecir?

- La predecibilidad de un evento o magnitud depende de varios factores:
  - Cómo de bien entendemos los factores que influyen
  - Cuántos datos están disponibles
  - Si la predicción puede afectar lo que se predice

Ejemplos

- Predicción de la demanda de energía eléctrica: muy precisa
- Predicción del tipo de cambio de divisas: más o menos
- Combinación de euromillones: no

Factores que afectan la predicción

- Horizonte temporal
- Tipo de patrones en los datos

Elección del método

- Depende de la disponibilidad de datos y de la predicibilidad de la cantidad a predecir
- Cada método tiene sus propias características: precisiones, restricciones, costes

Definición

- **Predicción**: Averiguar el futuro tan precisamente cómo sea posible, a partir de la información disponible incluyendo datos históricos y conocimiento de cualquier evento futuro que pueda influir en la predicción
- Habitualmente, es *parte de un proceso* de toma de decisiones

Forecasting vs. Prediction

- Conviene distinguir estas dos palabras del inglés:
- **Forecasting**: predicción sobre el futuro
Weather forecast, economic forecast
- **Prediction**: Averiguar un valor desconocido a partir de otros, no necesariamente relacionados por momentos de tiempo distintos
predictive maintenance

predecir prever pronosticar

Clasificación basada en el horizonte

- A corto plazo
- A medio plazo
- A largo plazo

Determinar qué predecir

- ¿Qué predecir?
  - Una cantidad
  - Un resumen
- Horizonte de predicción: 1 paso, varios
- Frecuencia de predicción: una vez solo, varias, periódicamente

Predicción cuantitativa

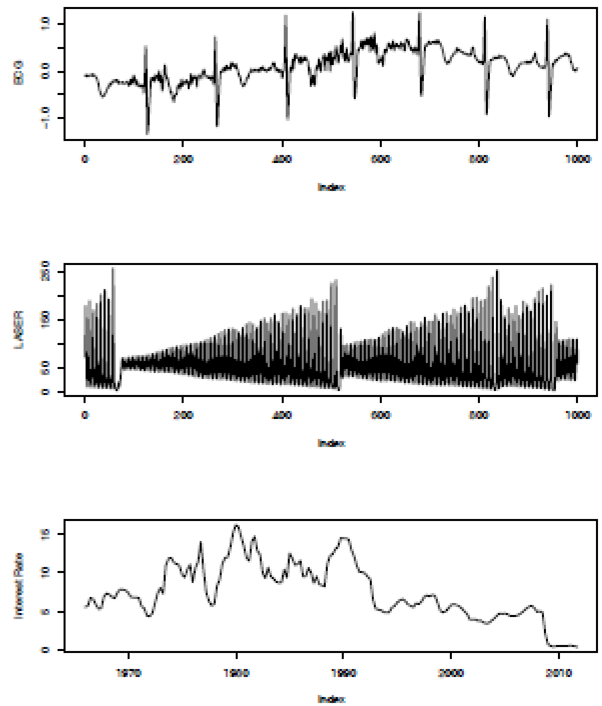
- Se puede aplicar cuando:
  - Hay datos numéricos del pasado disponibles
  - Es razonable *asumir* que algunos aspectos de patrones que ocurren en el pasado seguirán en el futuro

Datos transversales

- Datos recogidos en un mismo instante
- Predicción de datos transversales pretende predecir el valor de algo no observado en base a valores observados (en el mismo periodo temporal):
  - El precio de todas las casas vendidas en 2020 en un área concreta
  - Consumo de combustible para los vehículos de 2030

Series temporales

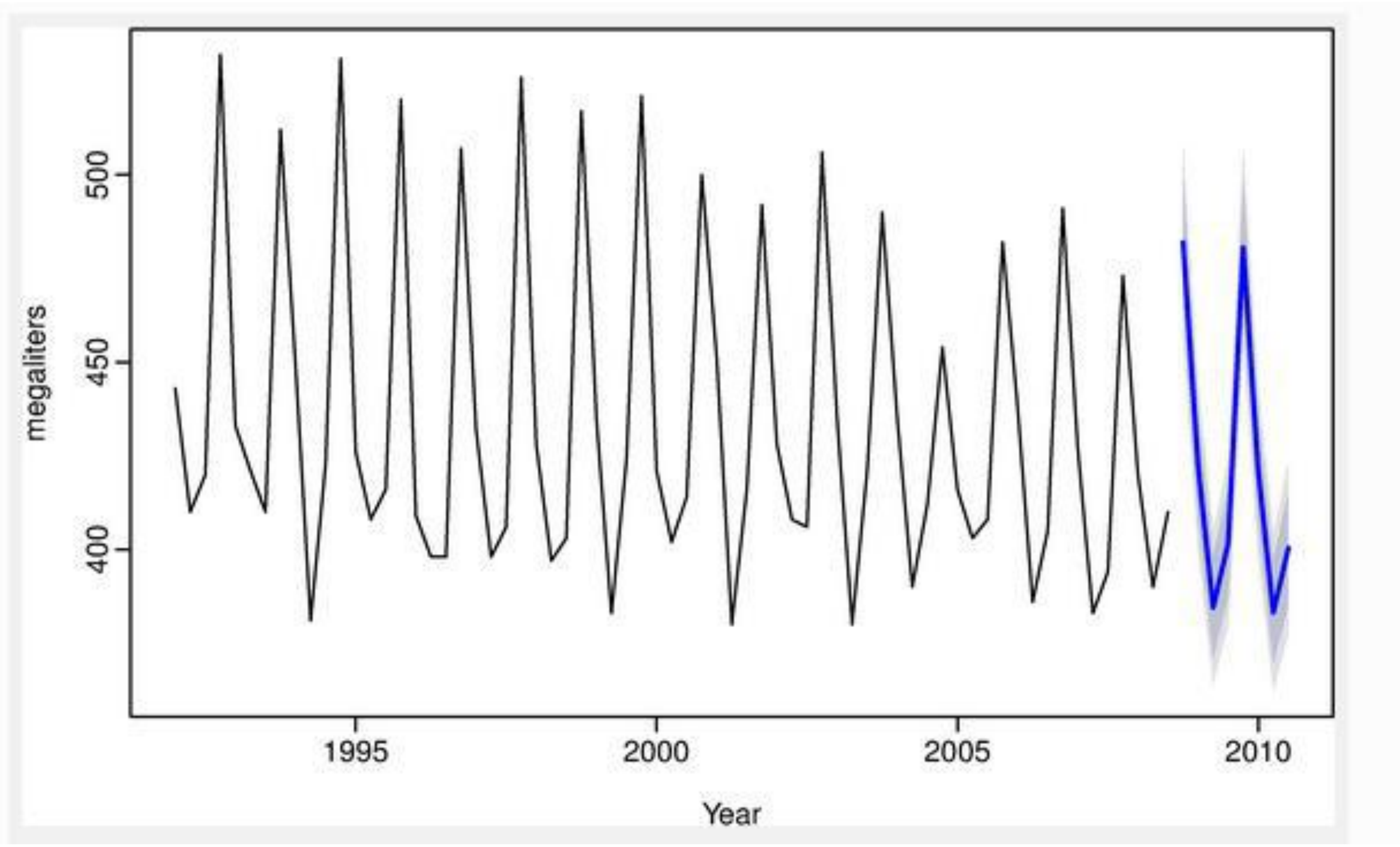
- Cualquier magnitud observada a lo largo del tiempo es una **serie temporal**
  (valores continuos o discretos)
- Series temporales observadas en intervalos regulares (cada minuto, hora, día, semana, ...)



Predicción de series temporales

- Los datos de las series temporales son útiles cuando se está prediciendo algo que cambia con el tiempo (p.ej. precios de acciones, ventas, beneficios, ...)
- La predicción de series temporales pretende calcular cómo continúa la serie temporal en el futuro

Predicción de producción de cerveza



Variables predictoras

- Predecir la demanda eléctrica (DE)
- DE = *f(temperatura actual*, *nivel de la economía*, *población*, *hora del día*, *día de la semana*, error)

Pasos básicos en predicción

1. Definición del problema
2. Reunir datos e información
3. Análisis preliminar (exploratorio)
4. Elección del modelo de ajuste
5. Usar y evaluar los modelos

Perspectiva estadística de la predicción

- La cantidad a predecir se puede considerar una variable aleatoria
- Cuanto más alejado sea el horizonte de predicción, más incertidumbre hay
- Predicción: aproximar la media del rango de los valores posibles, o en general, la esperanza de la variable aleatoria

Herramientas básicas

Las herramientas del analista

- Gráficos
- Resúmenes numéricos
- Transformaciones
- Evaluación de la precisión del modelo
- Diagnóstico de residuos
- Intervalos de predicción

Gráficos

- Gráfico temporal
- Gráfico de puntos

Gráfico temporal

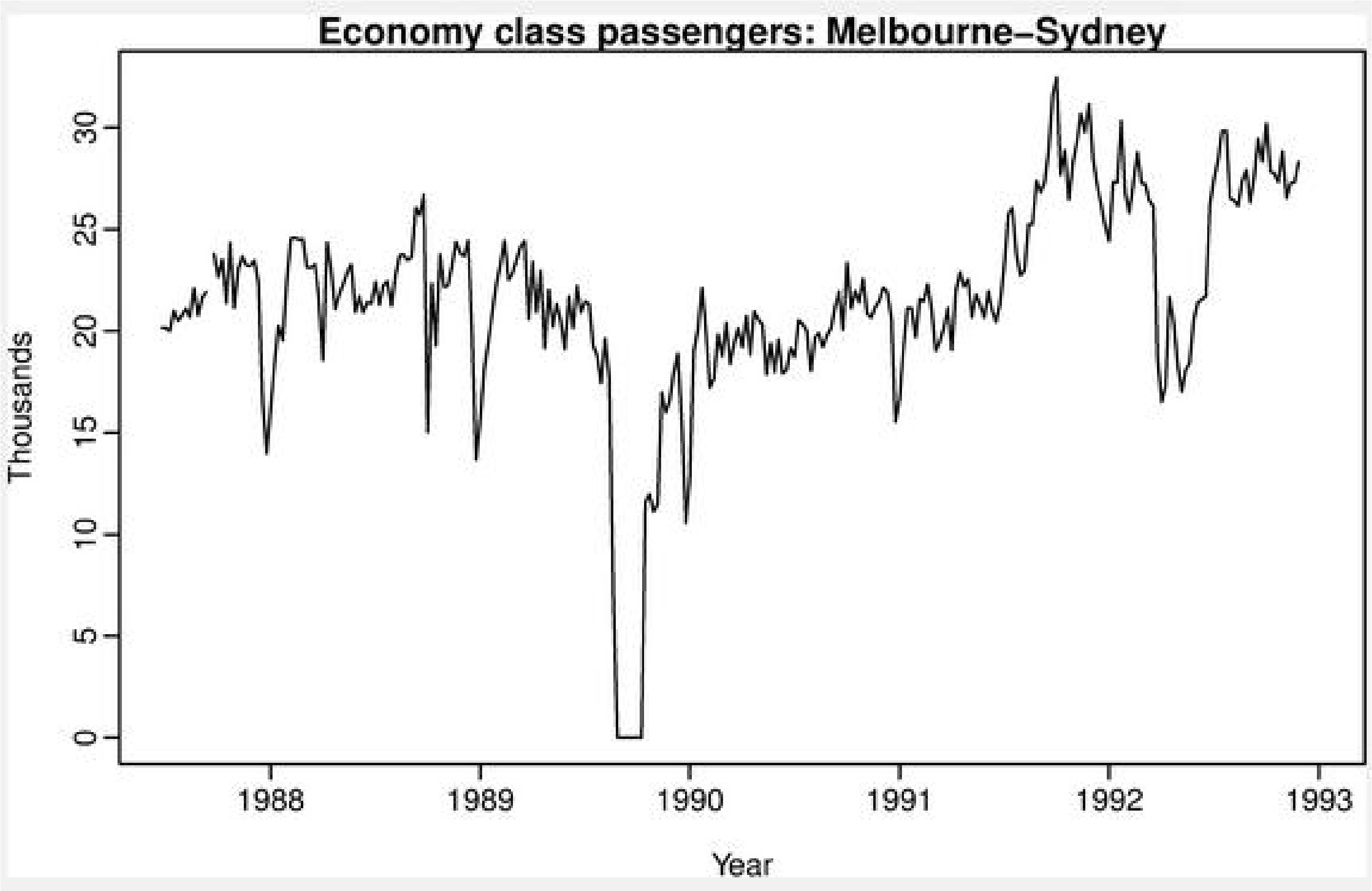


Gráfico temporal

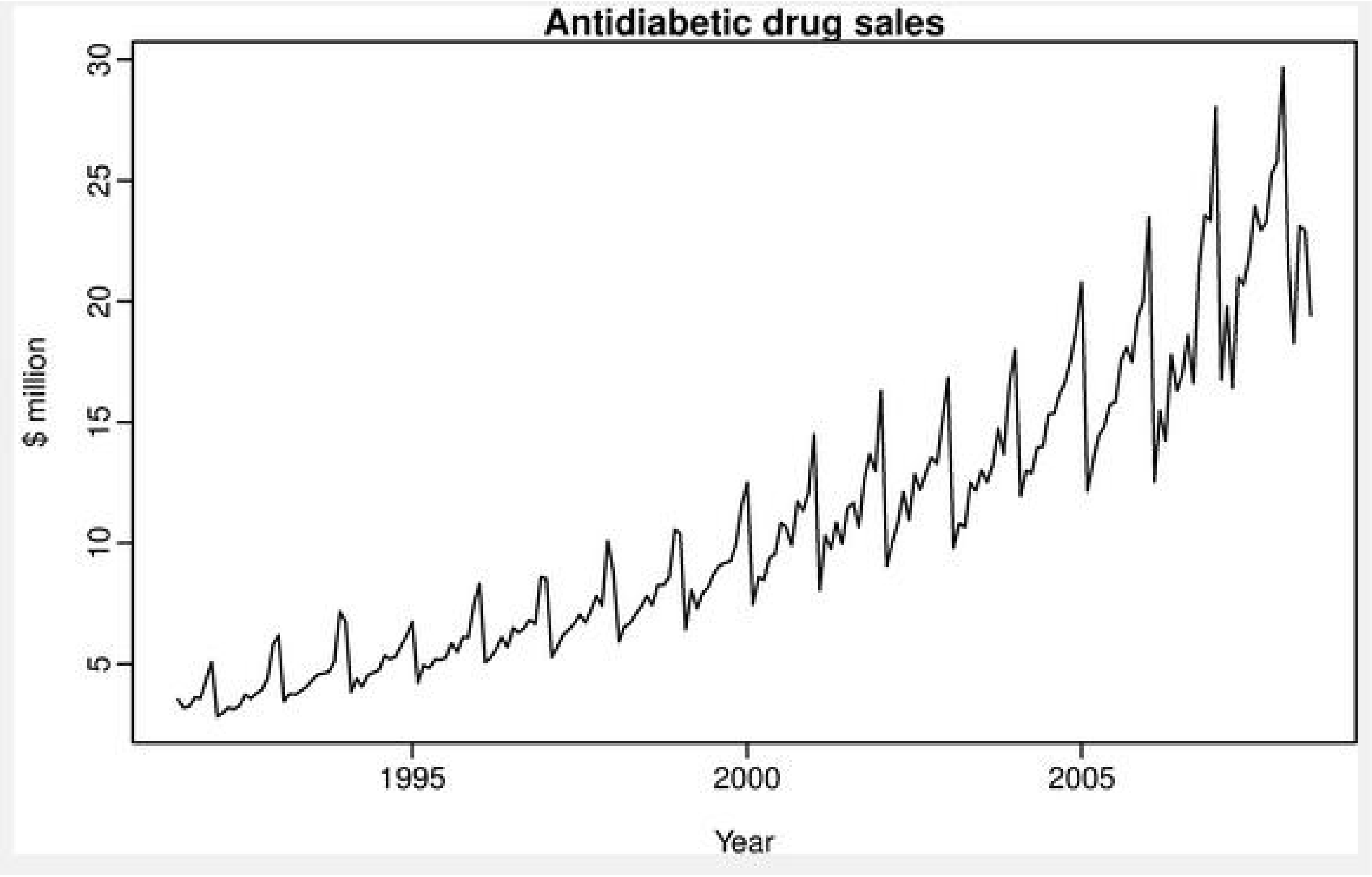
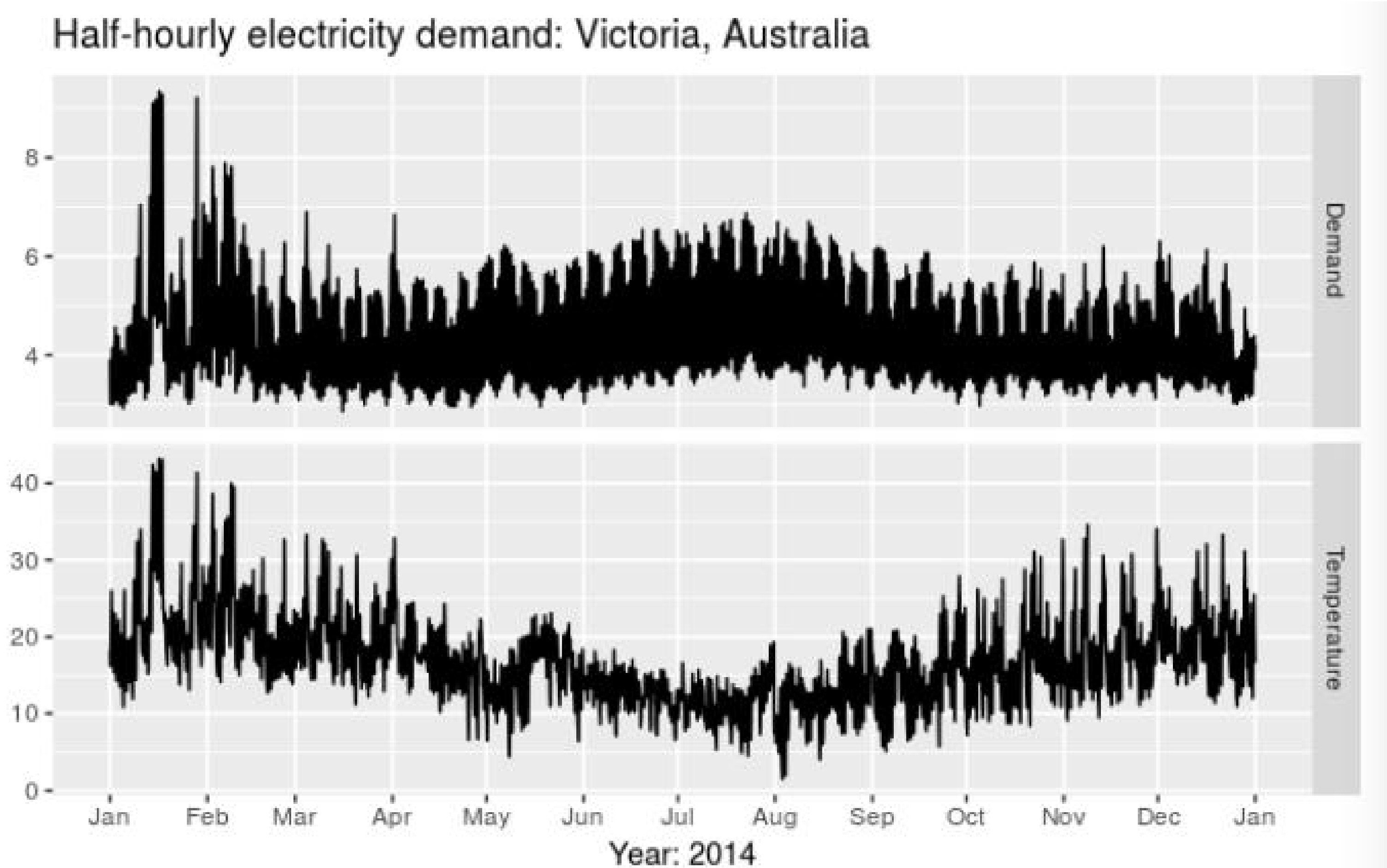
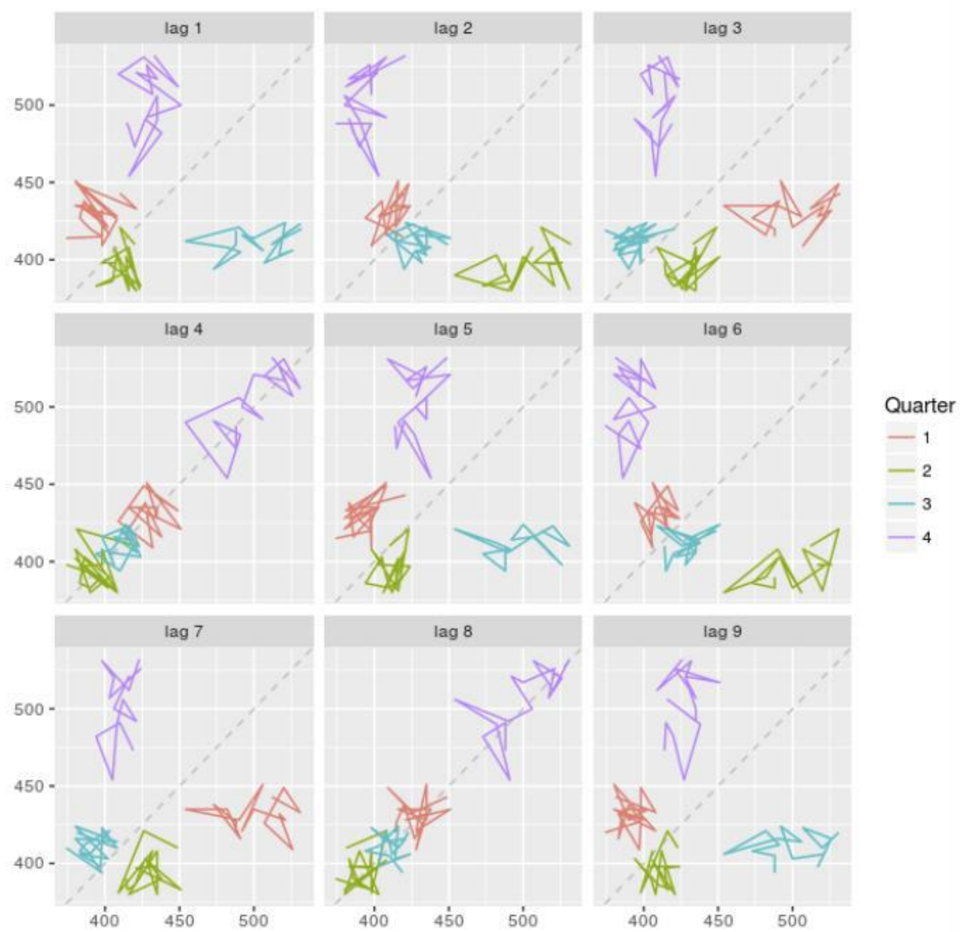


Diagrama de dispersión



Gráficos de retardos



Patrones en las series temporales

- *Tendencia*: incrementos o decrementos a largo plazo en los datos
- *Patrón estacional*: datos afectados por un patrón estacional tales como el día del año o el día de la semana
- *Ciclo*: los datos muestran subidas o bajadas pero no asociadas a un periodo fijo; la longitud suele ser variable y desconocida

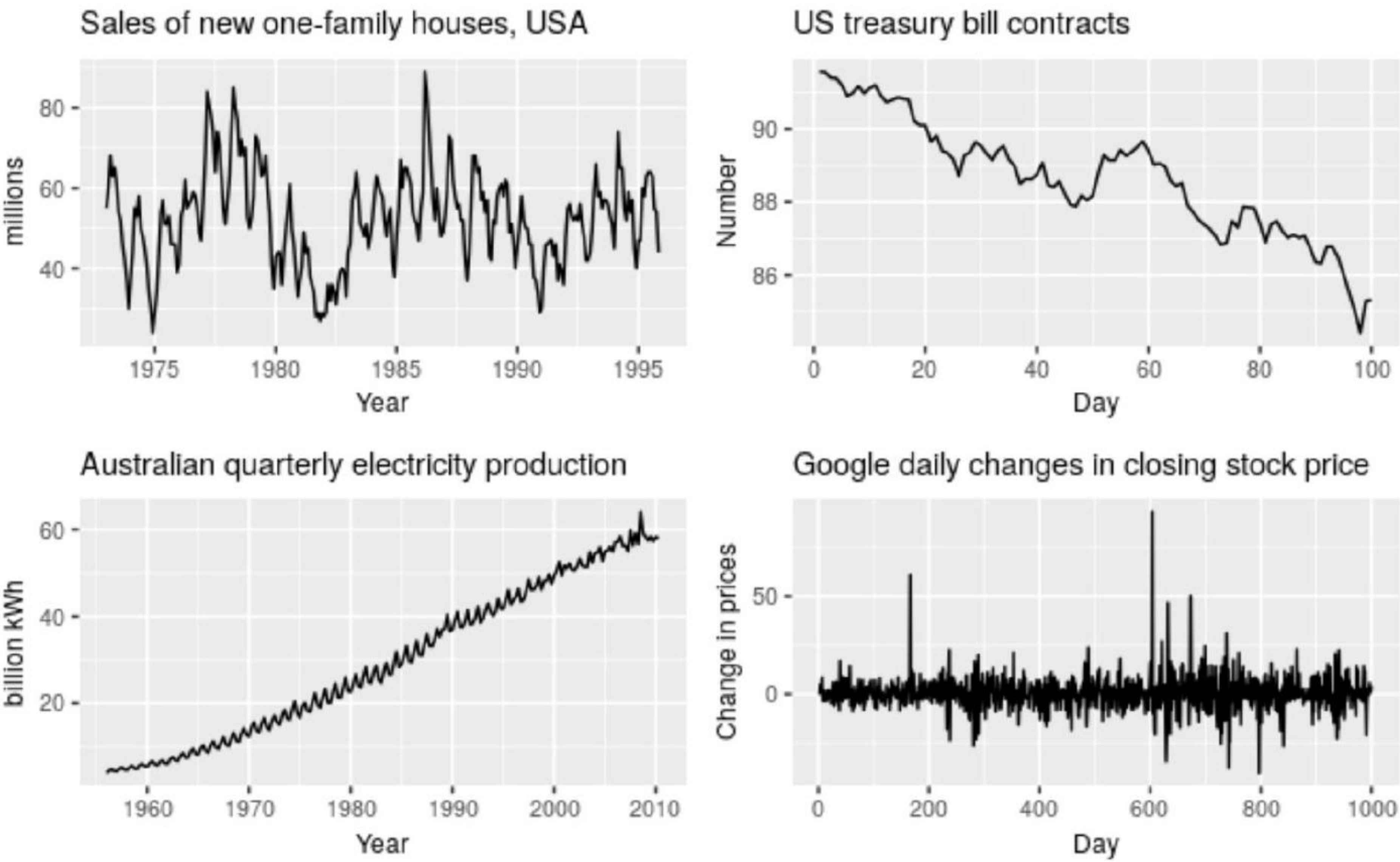
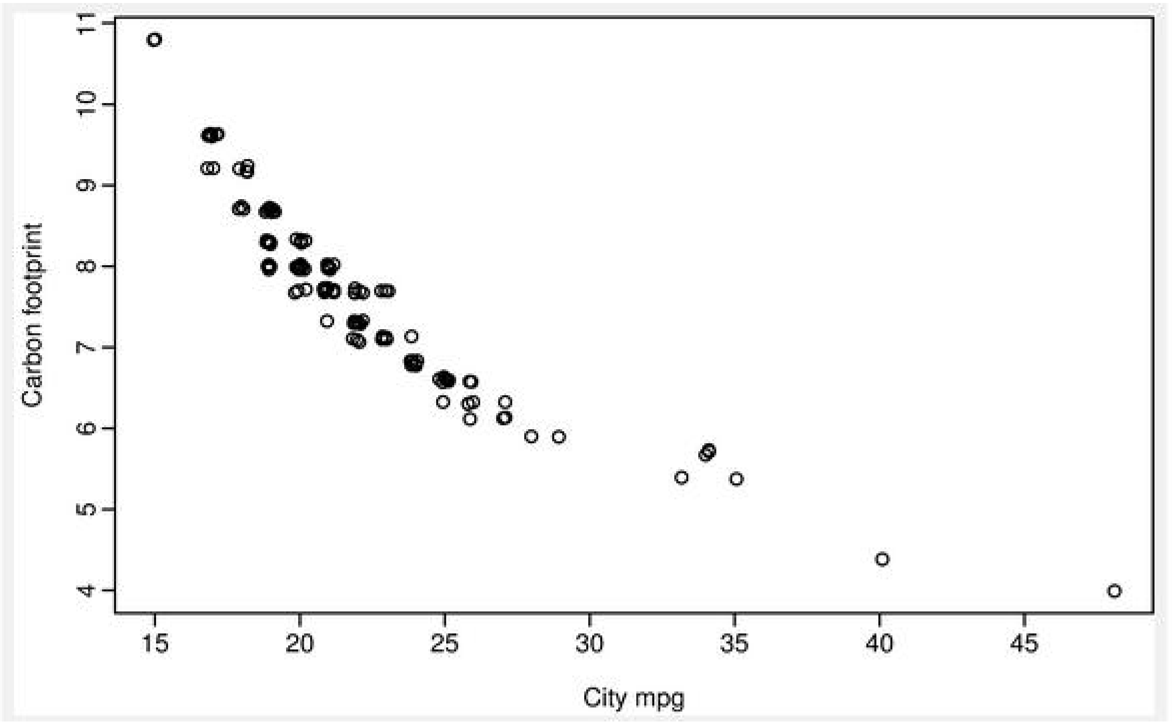
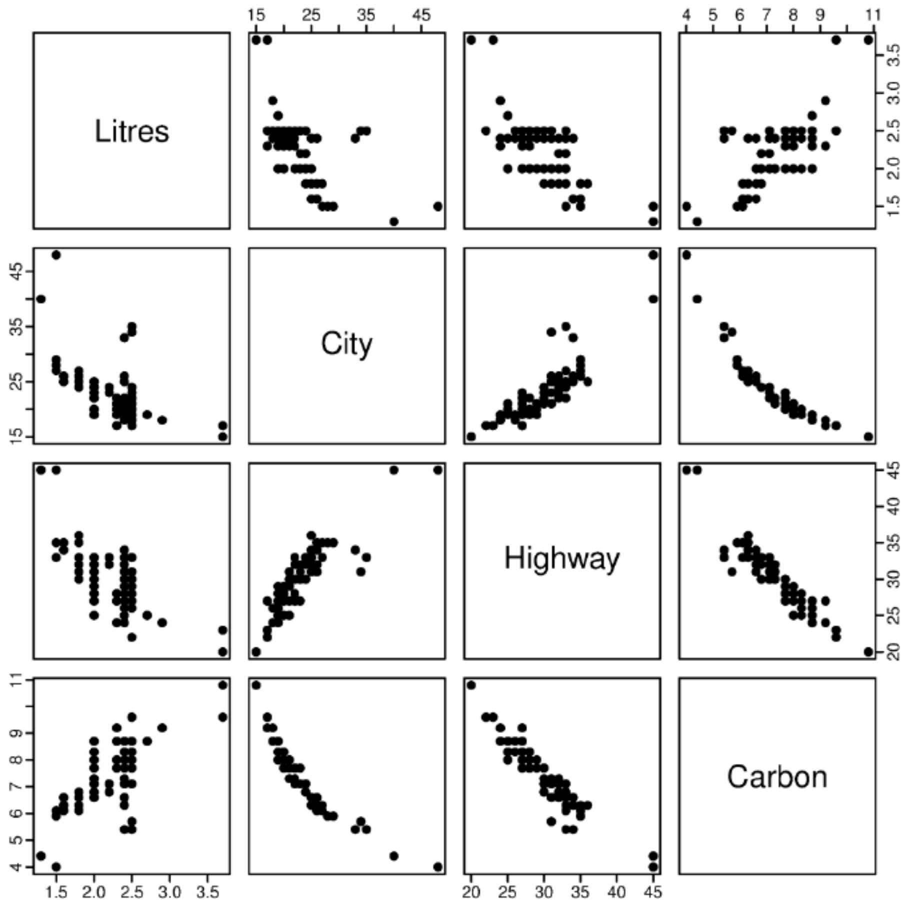

Sales of new one-family houses, USA
millions
80
60
40
1975
1980
1985
1990
1995
Year
US treasury bill contracts
Number
90
88
86
0
20
40
60
80
100
Day
Australian quarterly electricity production
billion kWh
60
40
20
1960
1970
1980
1990
2000
2010
Year
Google daily changes in closing stock price
Change in prices
50
0
0
200
400
600
800
1000
Day

Gráfico de puntos



Matriz de gráficos de puntos



Resúmenes numéricos

- Estadísticas univariables
  - Media

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i = (x_1 + x_2 + x_3 + \cdots + x_N)/N.$$

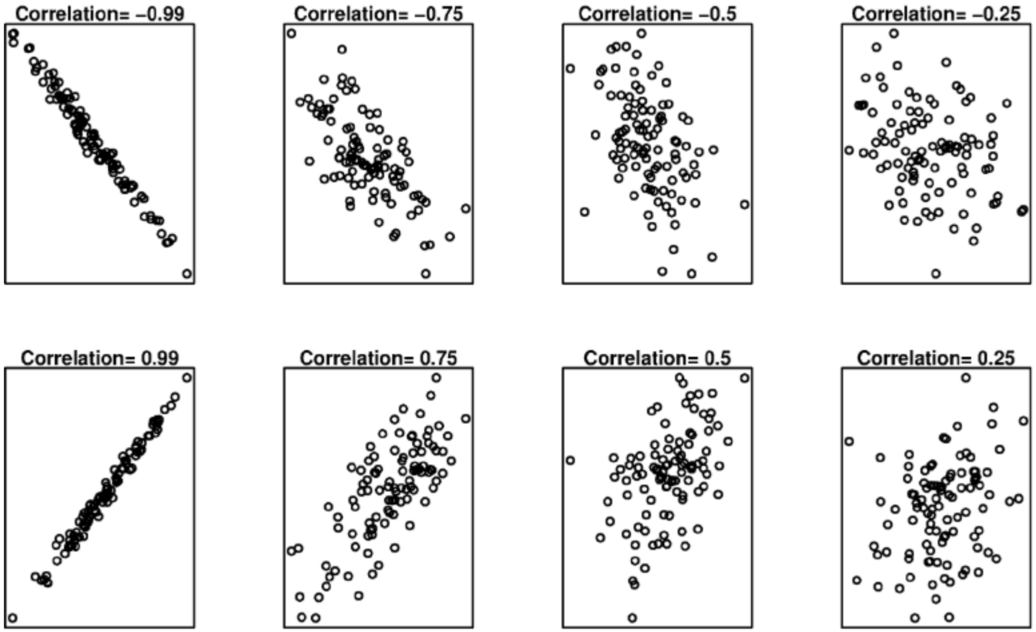
  - Mediana
  - Percentiles
  - Rango Intercuartílico (IQR)
  - Desviación estándar

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}.$$

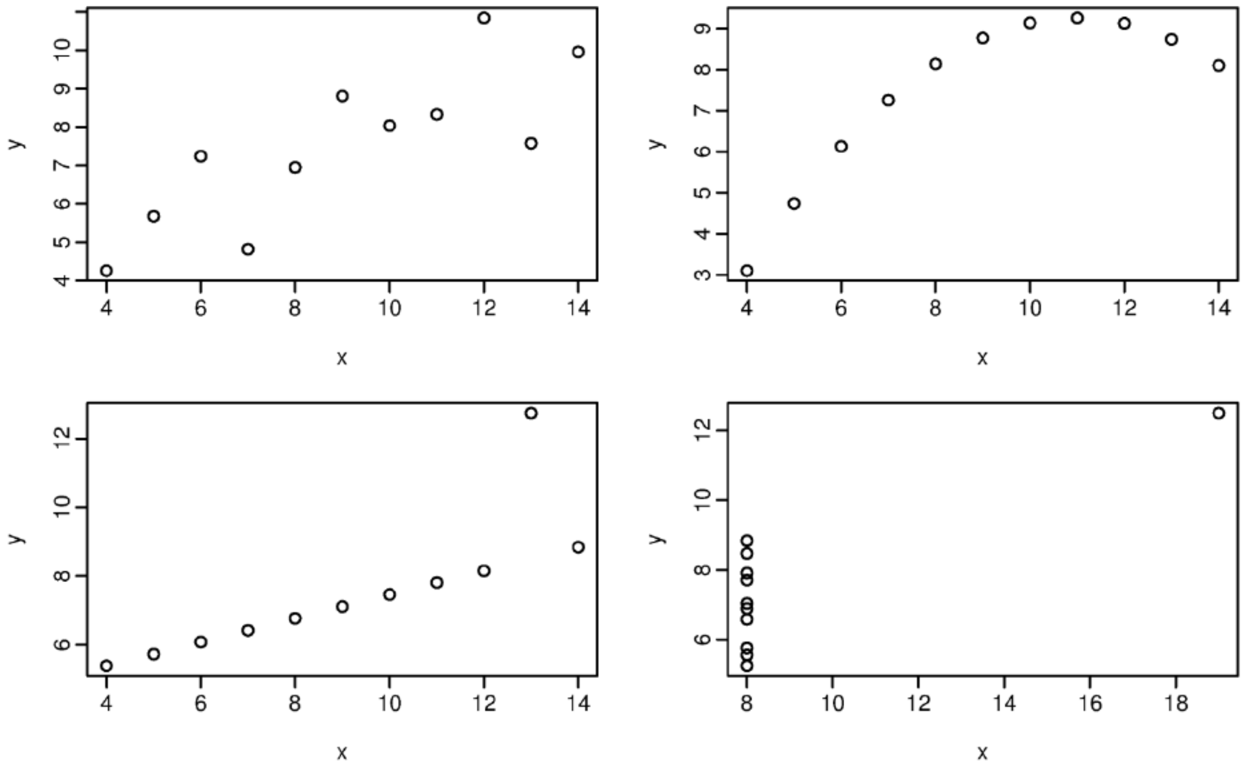
Coeficiente de correlación

Fuerza de la relación lineal

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2}\sqrt{\sum (y_i - \bar{y})^2}},$$



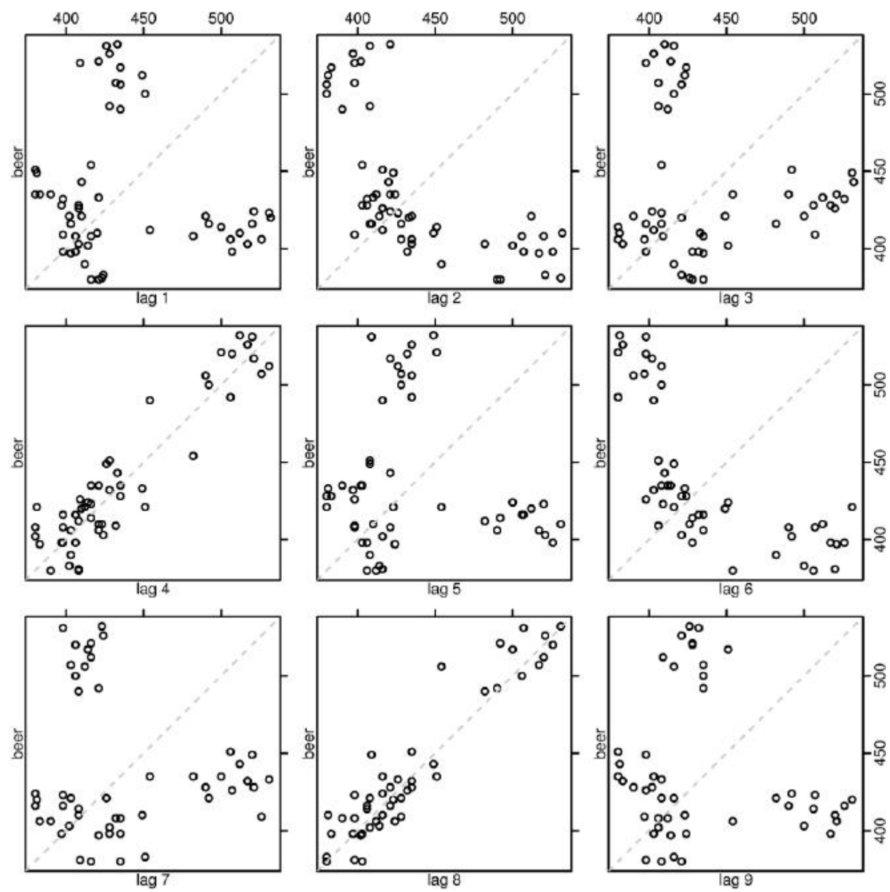
Coeficiente de correlación 0,82



Autocorrelación

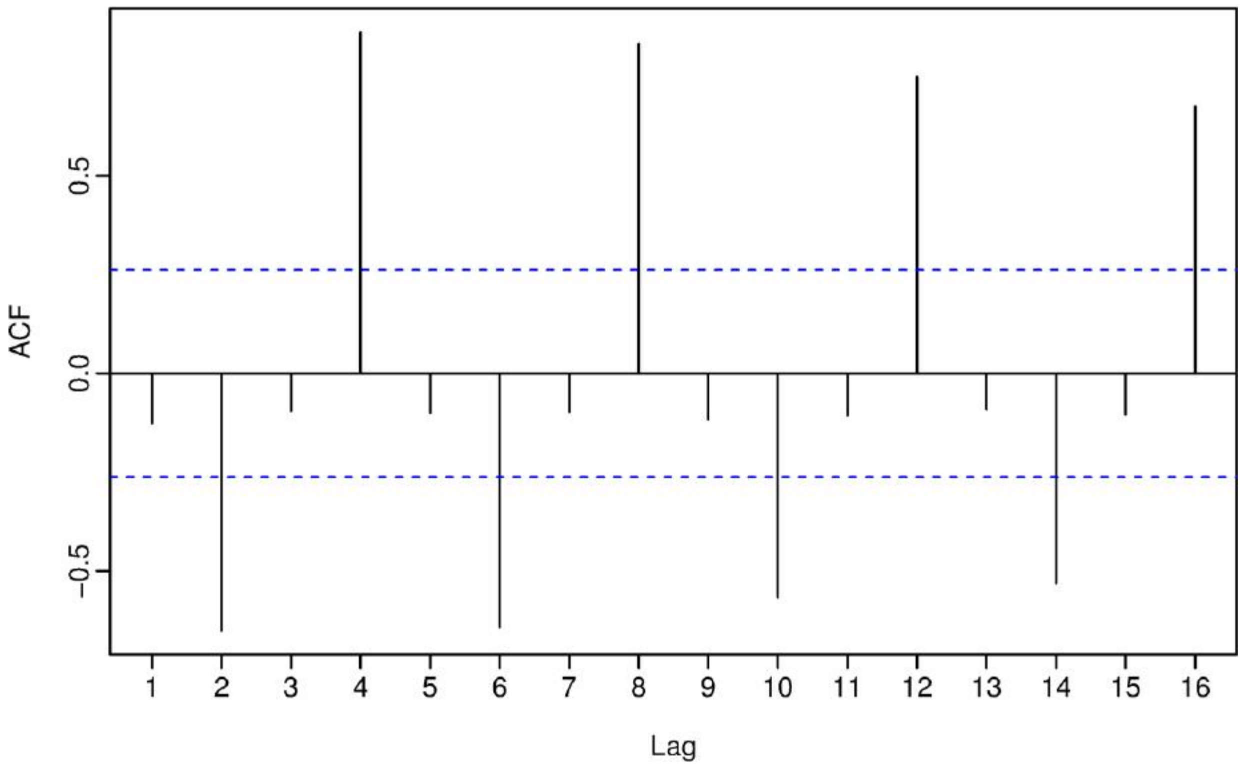
- Relación entre valores retrasados de una serie temporal

$$r_k = \frac{\sum_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T} (y_t - \bar{y})^2},$$



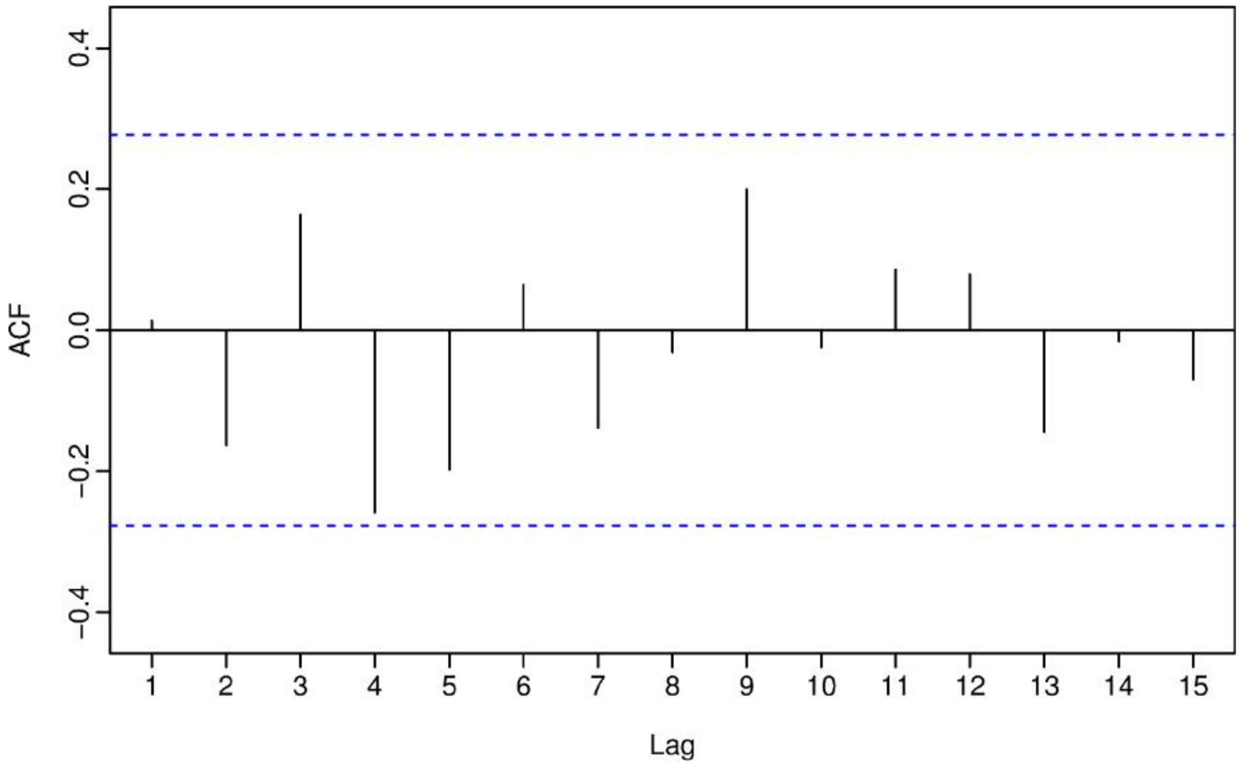
Función de autocorrelación (ACF)

r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9
-0.126	-0.650	-0.094	0.863	-0.099	-0.642	-0.098	0.834	-0.116



Ruido blanco

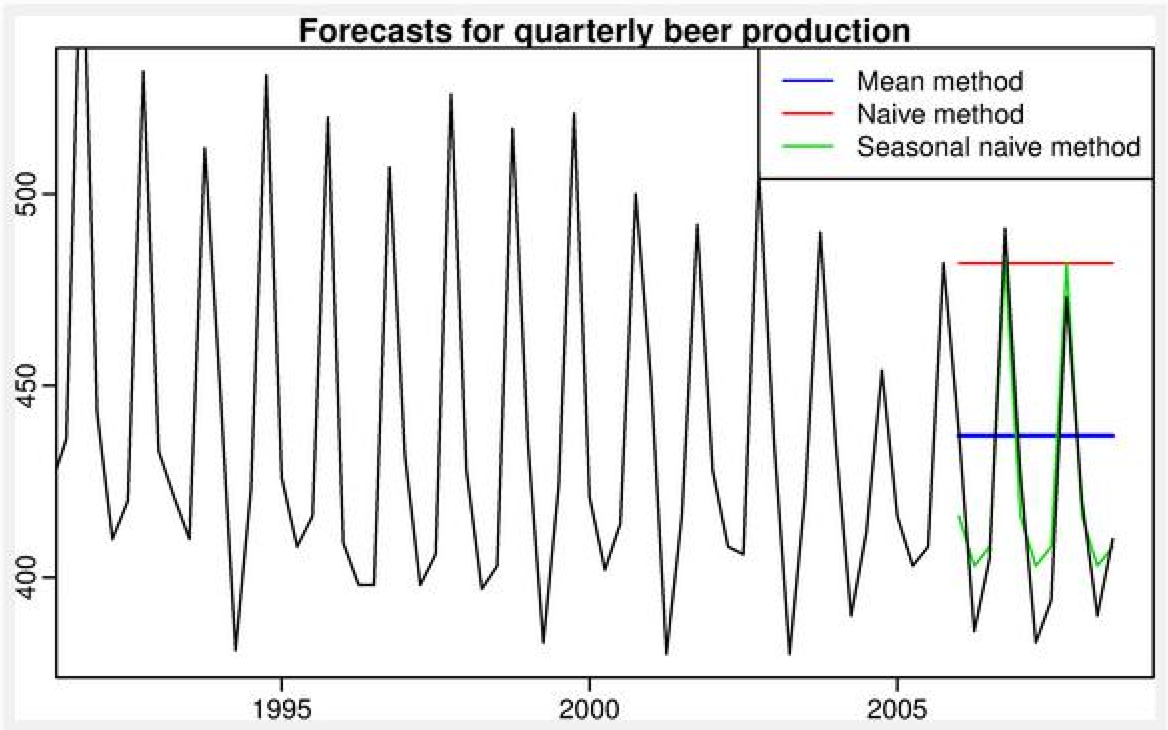
- Series que no muestran autocorrelación



Transformaciones

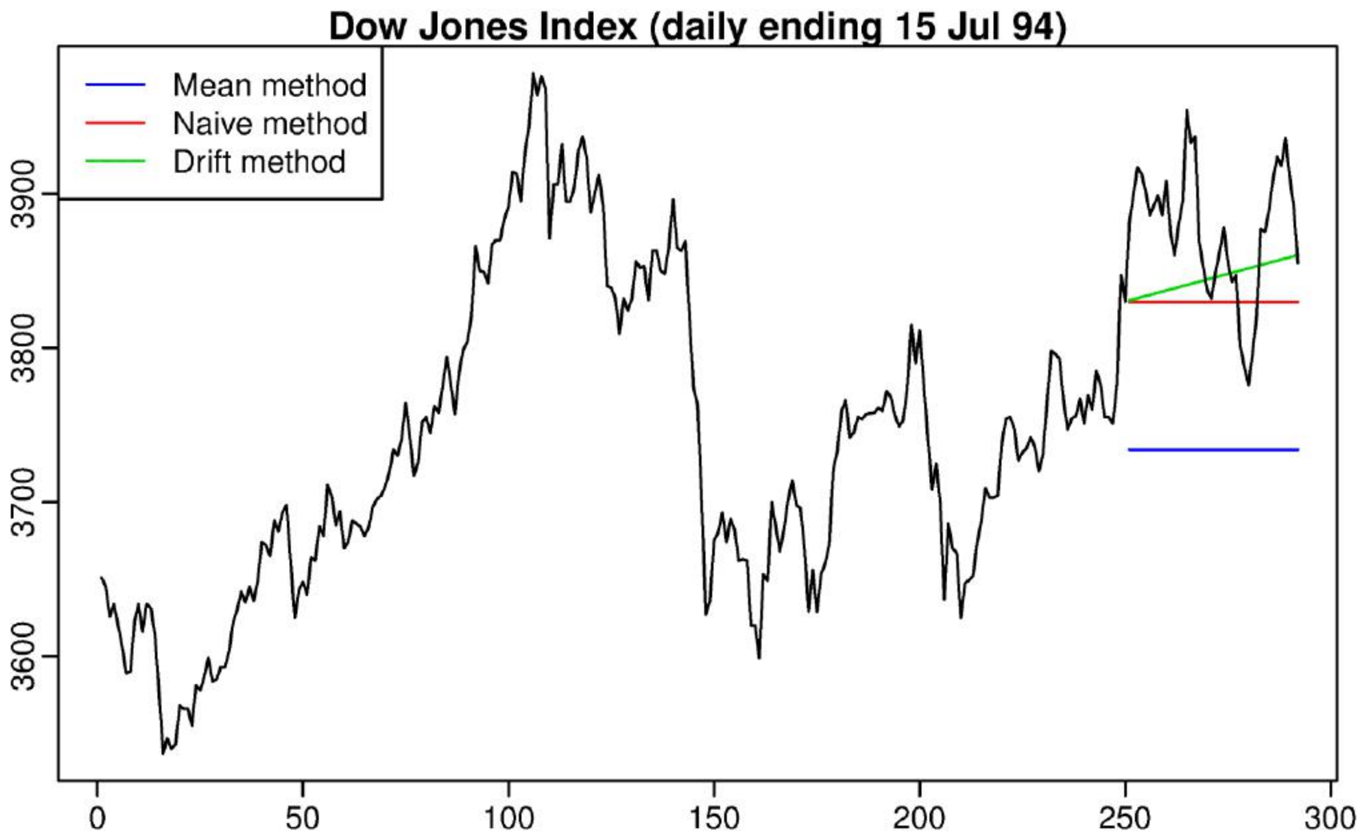
- Ajuste de datos históricos puede llevar a un modelo de predicción más sencillo
- Transformación matemática
  - Logaritmos, exponenciales
  - Transformada Box-Cox
  - Fourier, Wavelet, Laplace
- Ajuste de calendario (promedio por día)
- Ajuste de población (indicadores per cápita)
- Ajuste de inflación

Series temporales estacionales



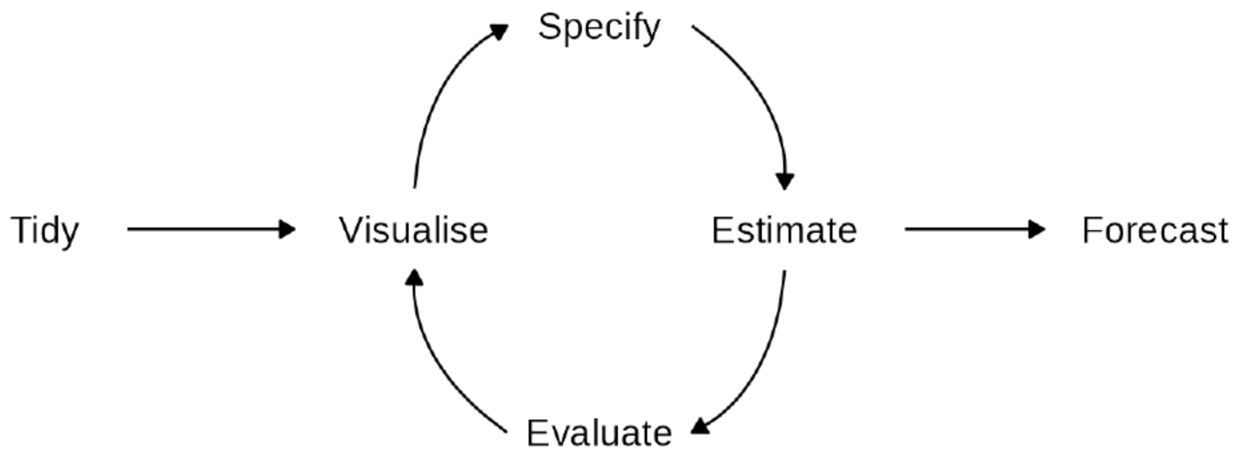
Method	RMSE	MAE	MAPE	MASE
Mean method	38.01	33.78	8.17	0.61
Naïve method	70.91	63.91	15.88	1.15
Seasonal naïve method	12.97	11.27	2.73	0.20

Series temporales no estacionales



Method	RMSE	MAE	MAPE	MASE
Mean method	148.24	142.42	3.66	8.70
Naïve method	62.03	54.44	1.40	3.32
Drift method	53.70	45.73	1.18	2.79

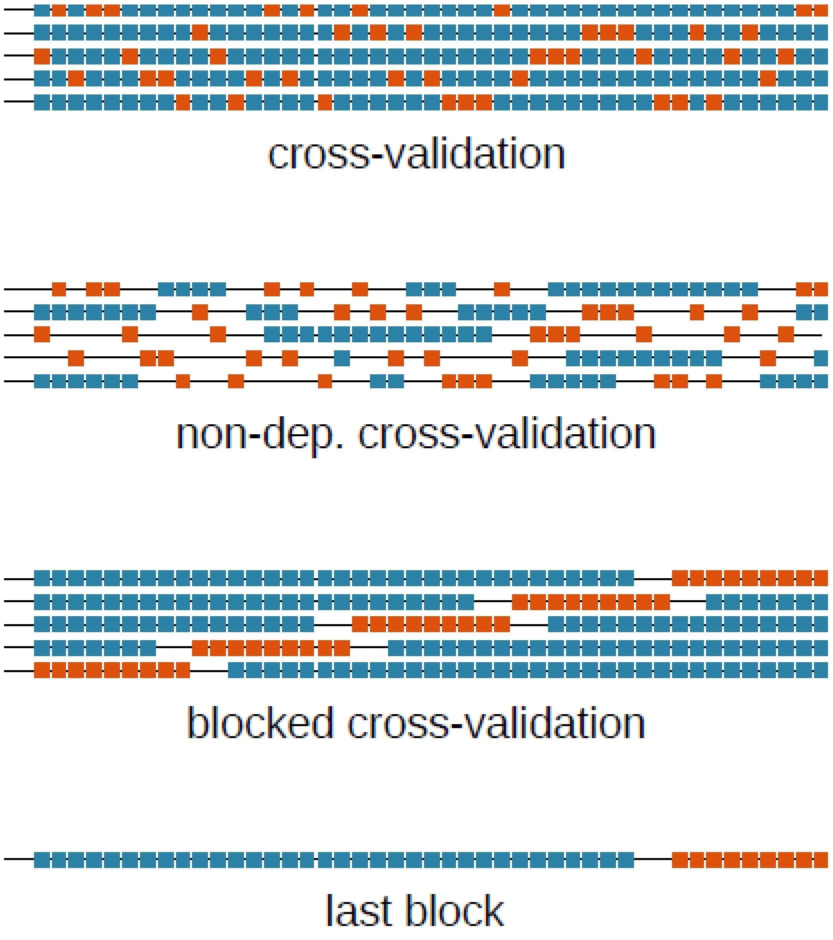
Flujo de trabajo en predicción



Metodología

- Como en cualquier otra tarea de modelado es esencial realizar una evaluación correcta
- Dividir datos en: entrenamiento y prueba
- Validación cruzada mejorada
- Blocked Cross-Validation

Procedimientos de selección de modelos



Diagnóstico de residuos

- **Residuo**: diferencia entre el valor observado y la predicción
- Un buen método de predicción debe producir residuos que cumplan:
  - No están correlados
  - La media es cero
- Cualquier método que no verifique las propiedades previas puede ser mejorado
- Además, es interesante si:
  - Los residuos tienen varianza constante
  - Los residuos se distribuyen según una normal

Evaluando la calidad de la predicción

- Error: $e_i = y_i - \hat{y}_i$
- Error dependiente de escala

$$\text{MAE} = \text{mean}(|e_i|), \qquad \text{RMSE} = \sqrt{\text{mean}(e_i^2)}.$$

- Error en %: $p_i = 100 e_i / y_i$
- Errores escalados

$$\text{MAPE} = \text{mean}(|p_i|). \qquad \text{sMAPE} = \text{mean}\left(200|y_i - \hat{y}_i| / (y_i + \hat{y}_i)\right)$$

$$q_j = \frac{e_j}{\frac{1}{T-1}\sum_{t=2}^{T}|y_t - y_{t-1}|}. \qquad \text{MASE} = \text{mean}(|q_j|).$$

Intervalos de predicción

- Un intervalo de predicción es un intervalo dentro del cual está el valor esperado con una probabilidad especificada
- Cuando la predicción es de un paso, la desviación estándar del predictor es casi la misma que la desviación estándar de los residuos

Métodos de predicción sencillos (1)

- Media

$$\hat{y}_{T+h|T} = \bar{y} = (y_1 + \cdots + y_T)/T.$$

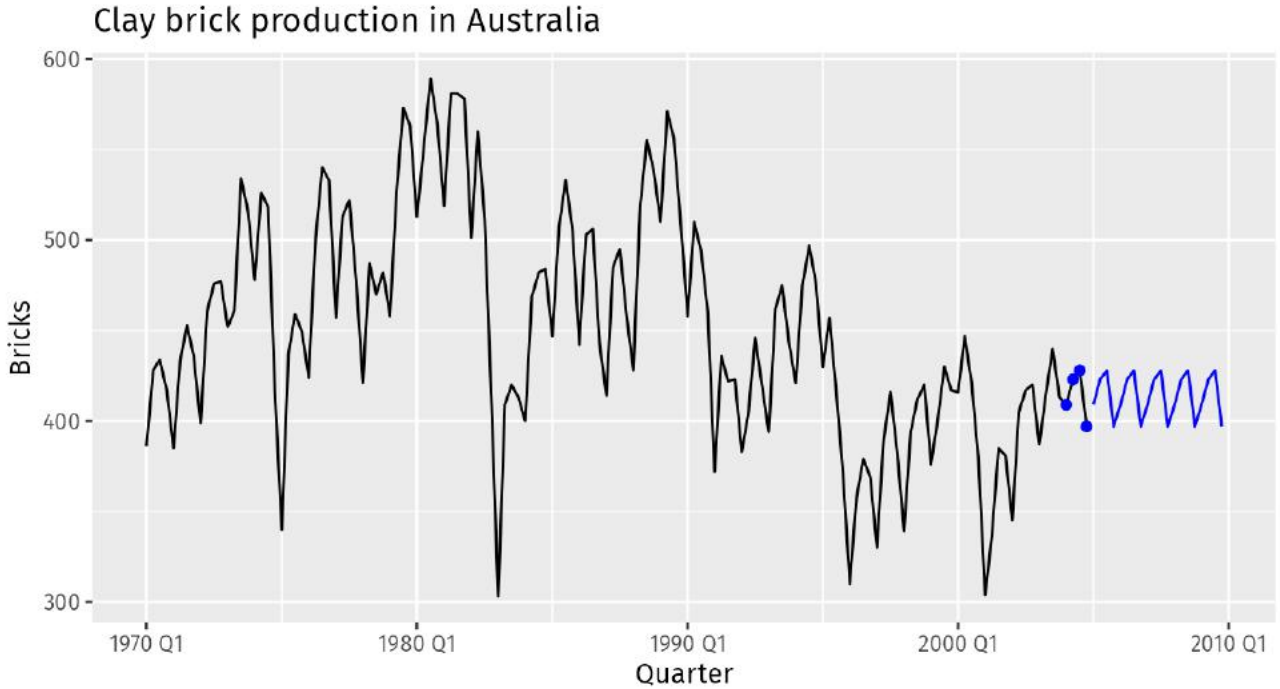
- Método de deriva (*drift*)

La predicción crece o decrete a lo largo del tiempo, en base a la media histórica

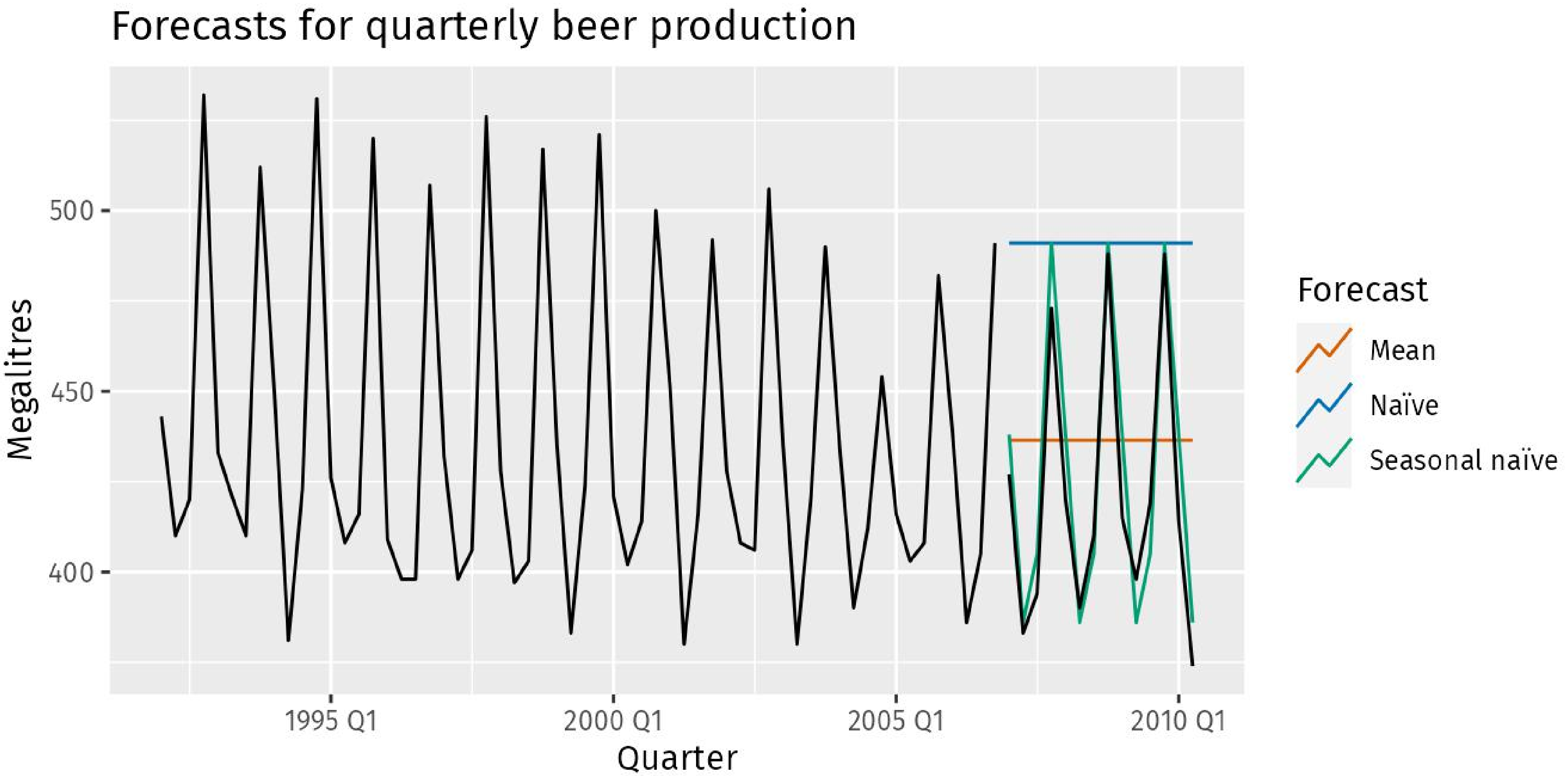
$$\hat{y}_{T+h|T} = y_T + \frac{h}{T-1}\sum_{t=2}^{T}(y_t - y_{t-1}) = y_T + h\left(\frac{y_T - y_1}{T-1}\right).$$

Métodos de predicción sencillos (2)

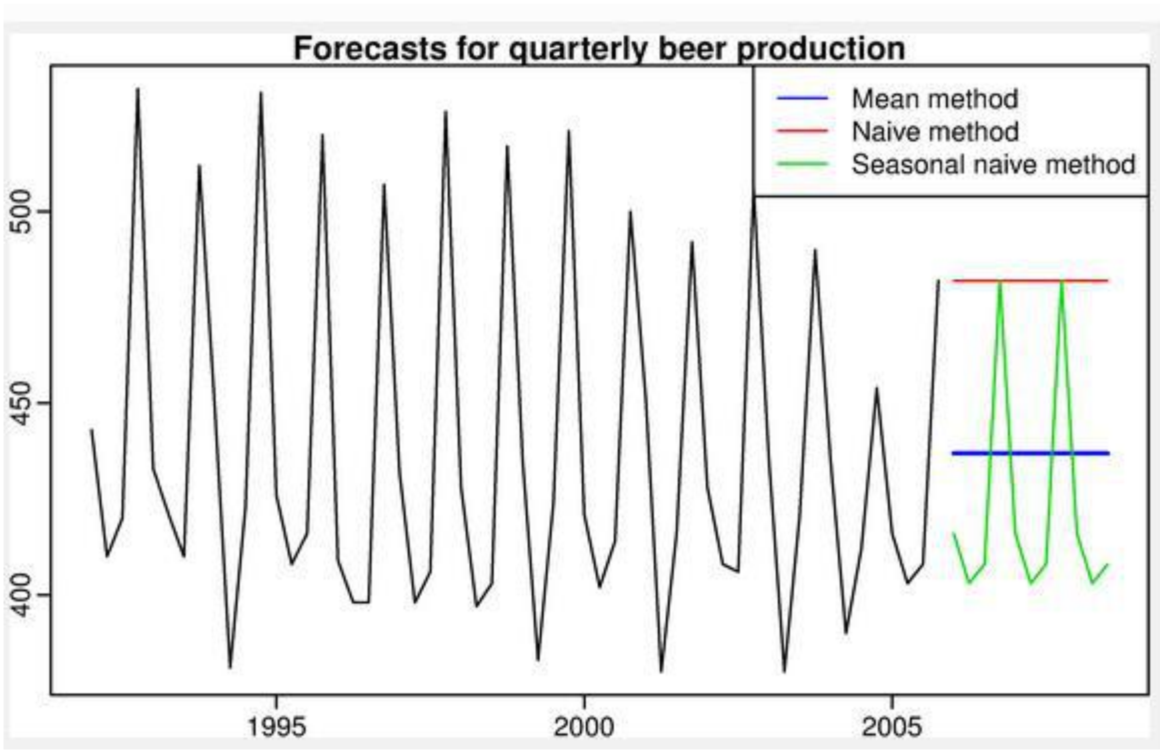
- Predictor ingenuo (*naive*)
  - Predicción: último valor
- Predictor ingenuo estacional



Métodos de predicción sencillos (3)



Métodos de predicción sencillos (4)



Regresión

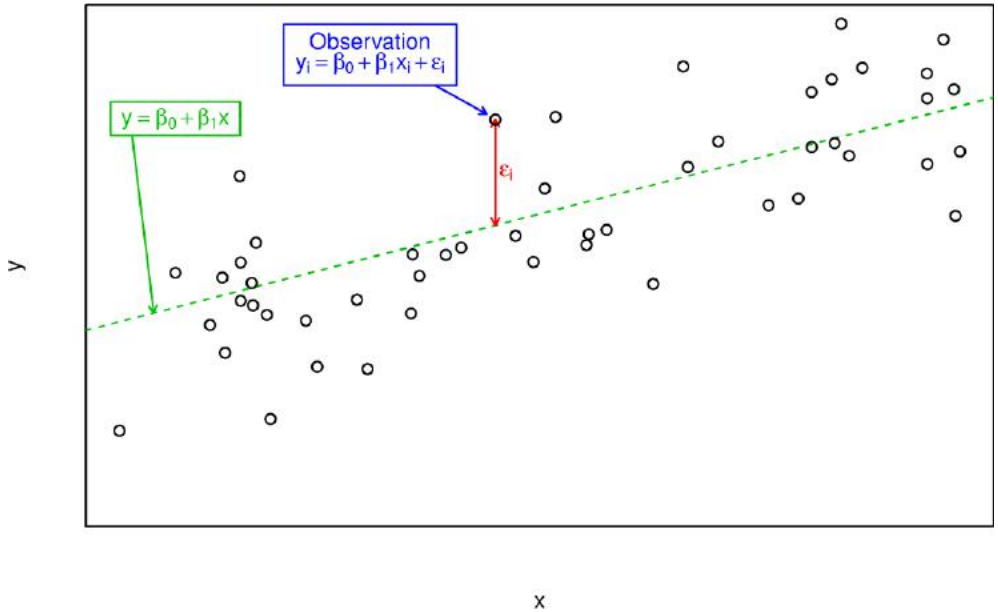
Predicción por regresión

Concepto básico: se asume una relación lineal entre la serie de interés y otras series.

Modelo lineal sencillo

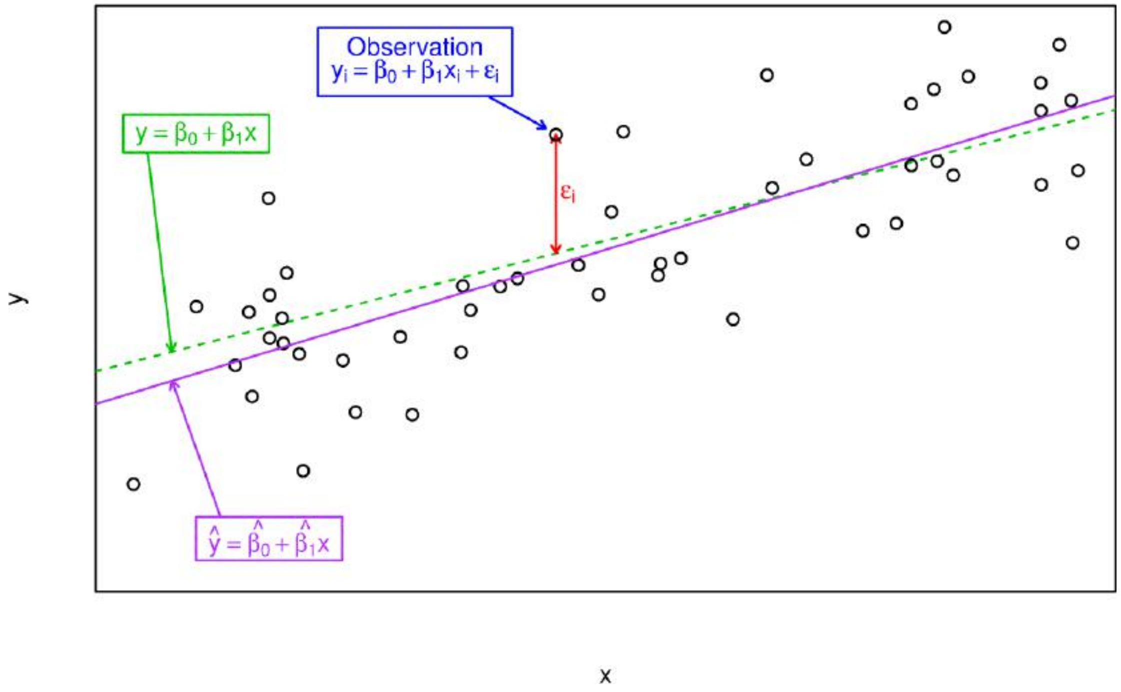
- Asumamos una relación lineal entre el predictor y la variable de salida

$$y = \beta_0 + \beta_1 x + \varepsilon$$



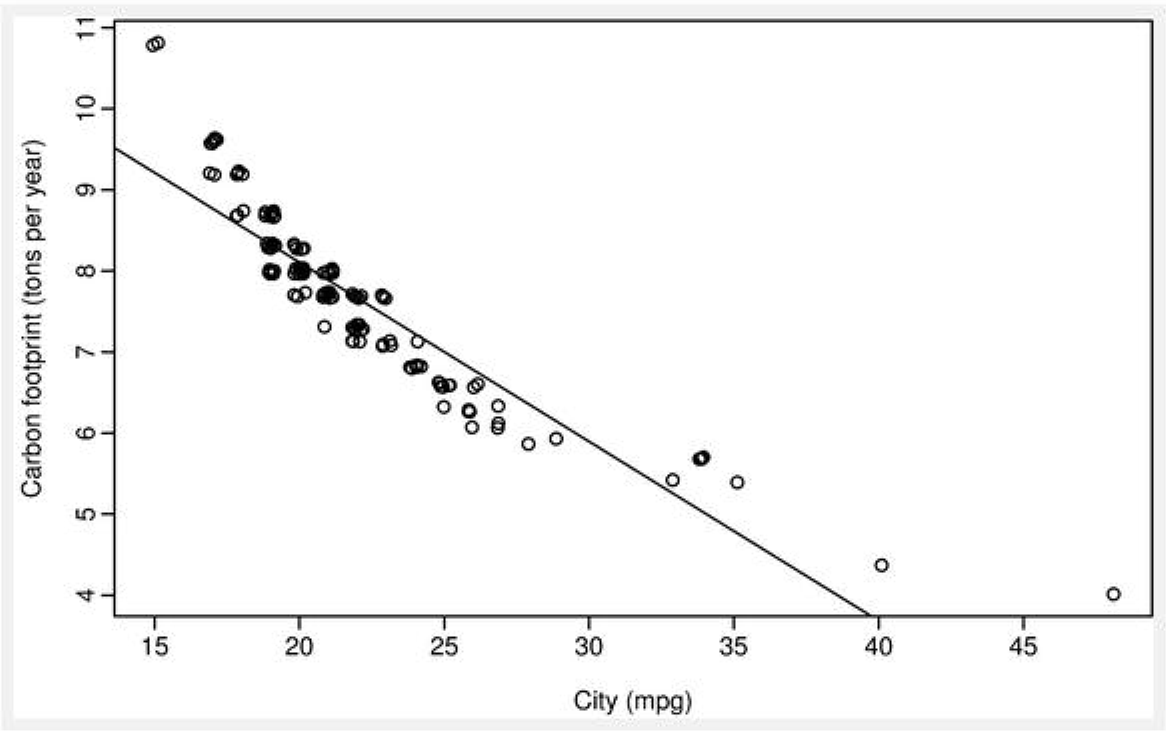
Aproximación por mínimos cuadrados

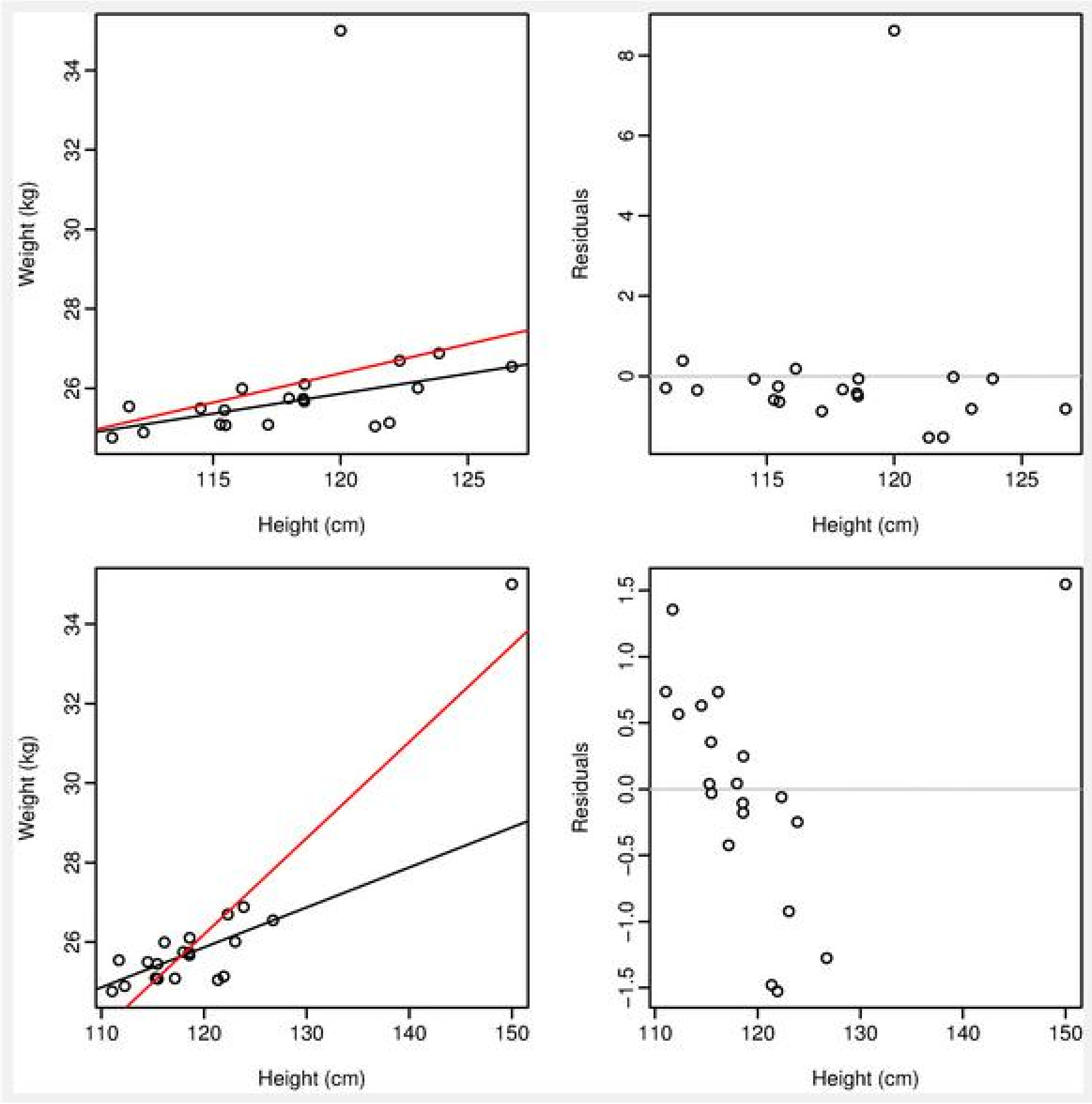
$$\sum_{i=1}^{N} \epsilon_i^2 = \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x_i)^2$$



Correlación y regresión

$$\hat{\beta}_1 = r\frac{s_y}{s_x}$$





Weight (kg)
Height (cm)
Residuals
Height (cm)
Weight (kg)
Height (cm)
Residuals
Height (cm)

Calidad del ajuste

- Coeficiente de ajuste, R^2:

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2},$$

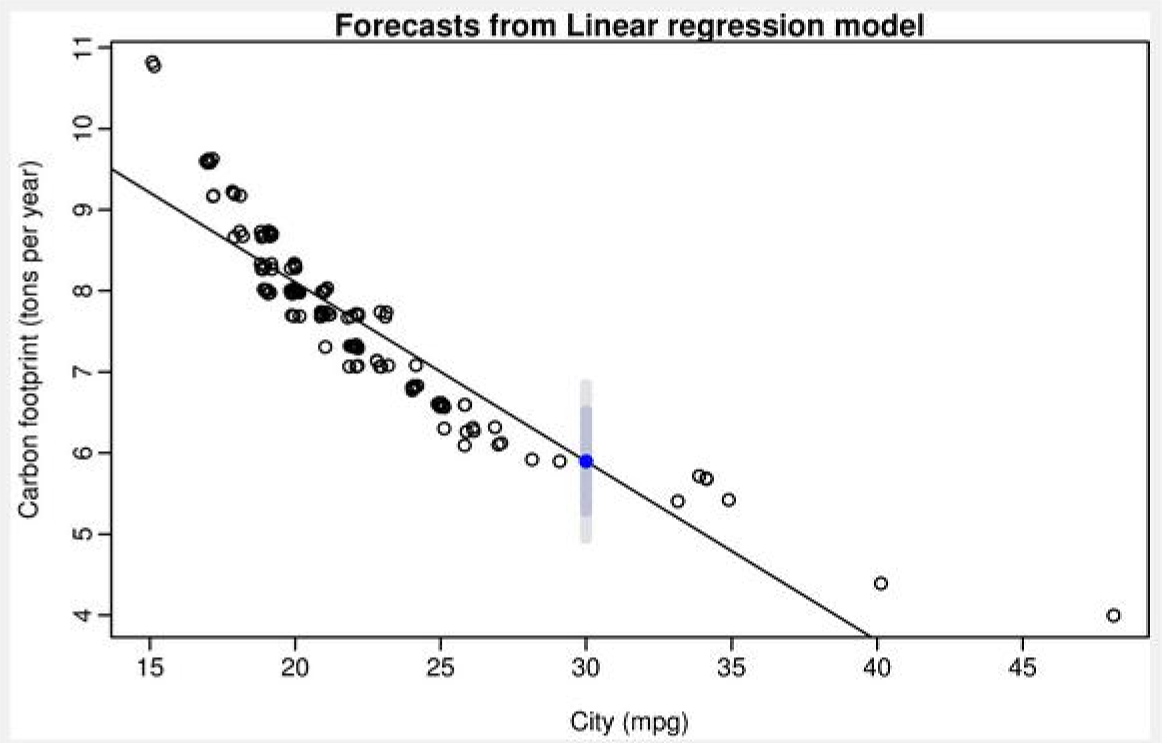
- Error estándar de la regresión

$$s_e = \sqrt{\frac{1}{N-2}\sum_{i=1}^{N} e_i^2}.$$

Predicción con regresión

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

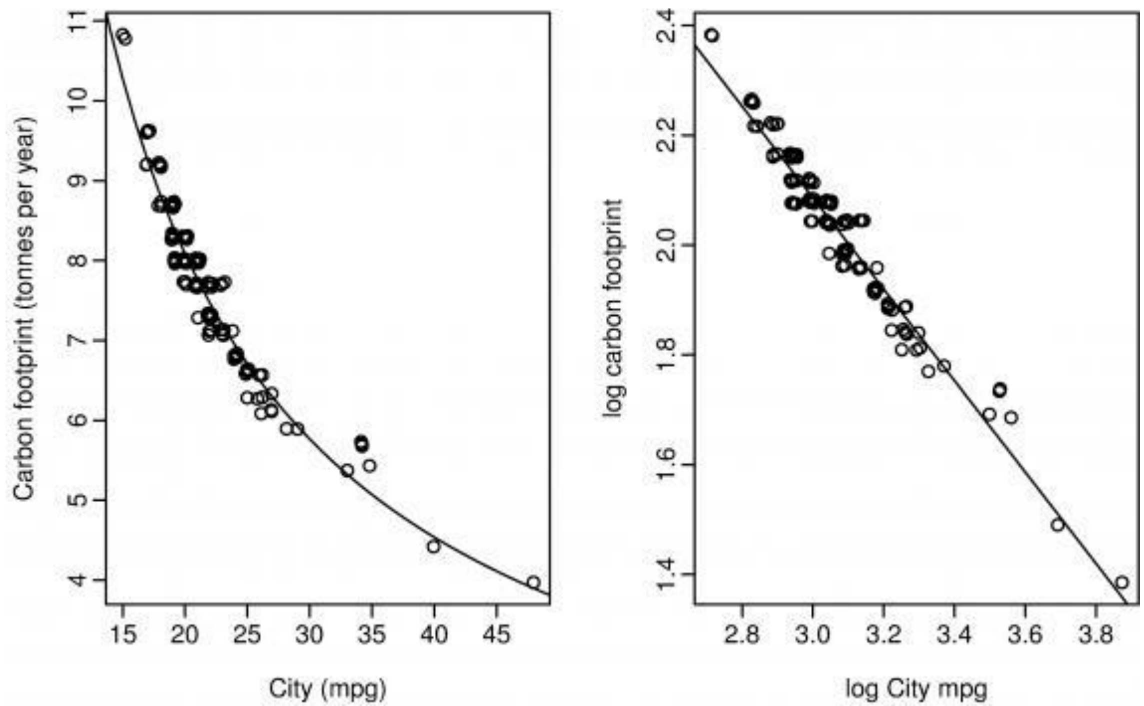
$$\hat{y} \pm 1.96 s_e \sqrt{1 + \frac{1}{N} + \frac{(x - \bar{x})^2}{(N-1)s_x^2}}$$



Regresión no lineal

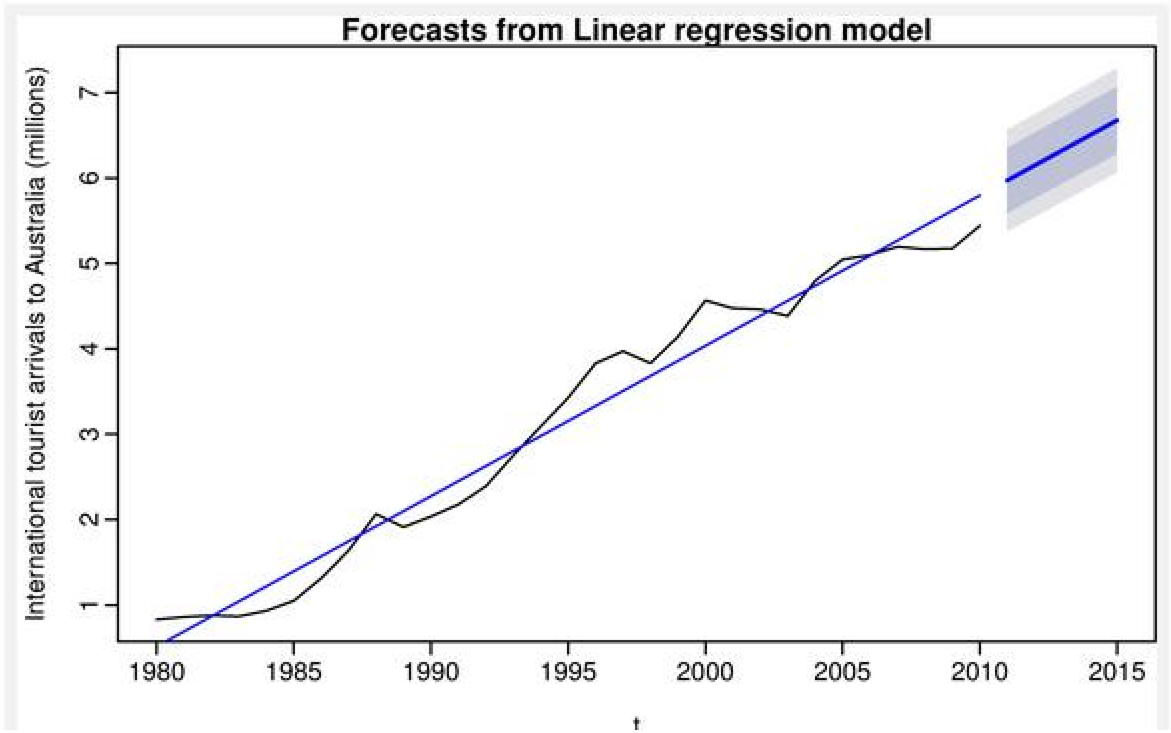
- Existen problemas para los que una función no lineal puede ser más adecuada que una lineal
- Se puede obtener mediante la transformación de x o y

$$\log y_i = \beta_0 + \beta_1 \log x_i + \varepsilon_i.$$

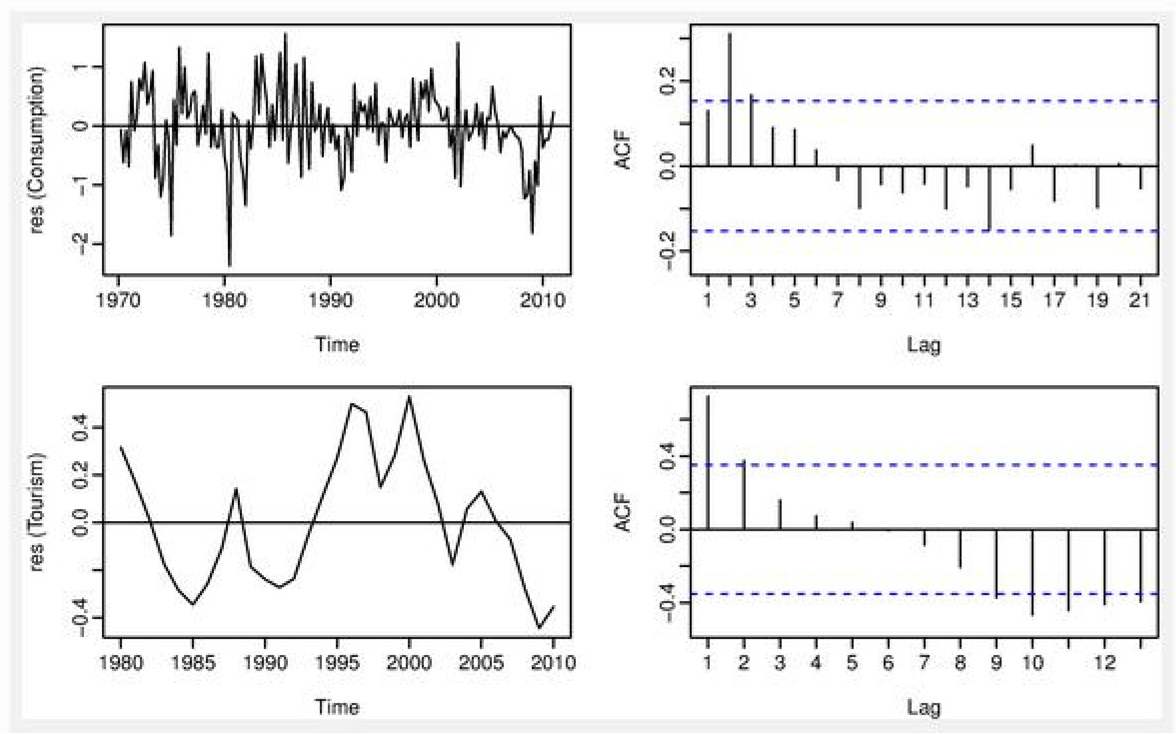


Regresión con datos de series temporales

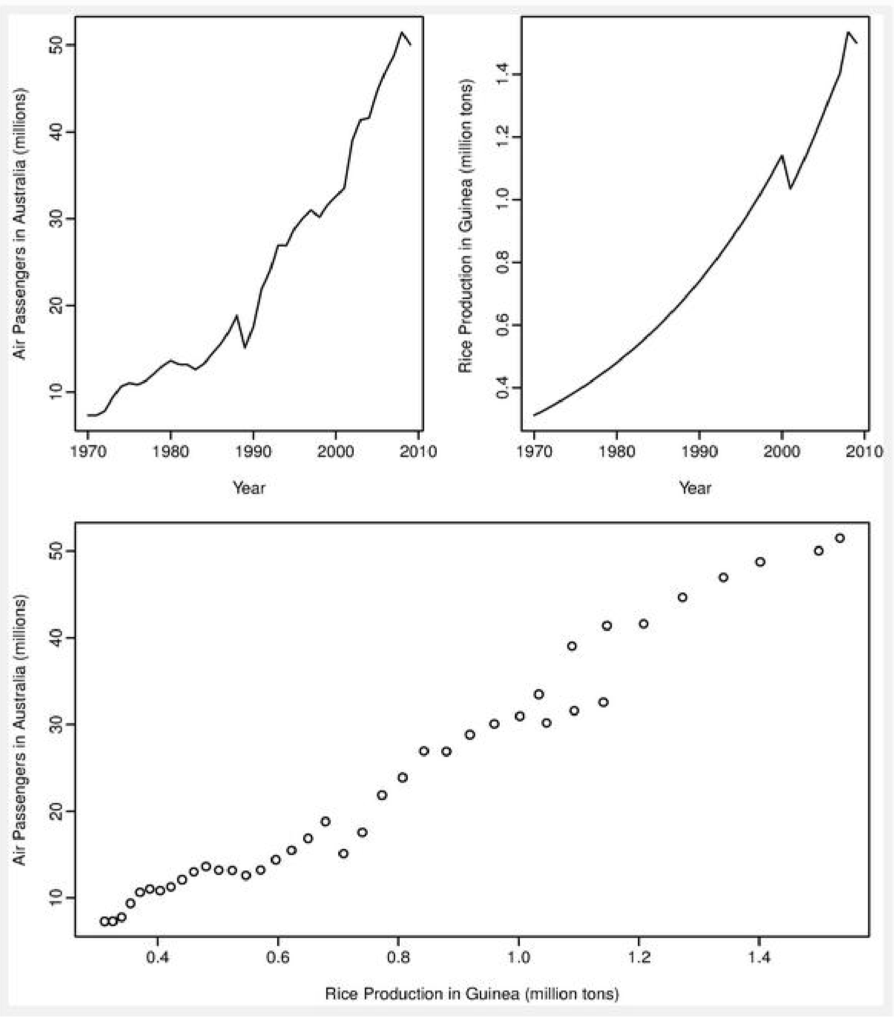
$$y_t = \beta_0 + \beta_1 t + \varepsilon_t$$



Autocorrelación de los residuos



Regresión espúrea



Regresión múltiple

Una variable para predecir y varias independientes

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_k x_{k,i} + e_i.$$

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1,i} - \cdots - \beta_k x_{k,i})^2.$$

Selección de modelos

- R^2 ajustado
- Validación cruzada
- Criterio de información de Akaike (AIC)
- Criterio de información de Akaike Corregido (AIC)
- Mejor subconjunto de regresión
- Regresión por pasos

Diagnóstico de residuos: gráficos

- Gráfico de puntos de residuos frente a variables independientes
- Gráfico de puntos de residuos frente a valores ajustados
- Autocorrelación de los residuos
  - Tests portmanteau: Box-Pierce, Ljung-Box
- Histograma de los residuos

Correlación no es causa

- Una variable x puede ser útil para predecir la variable y, pero eso no significa que x cause y
- Las correlaciones son útiles para la predicción, aunque no haya relaciones causales entre dos variables

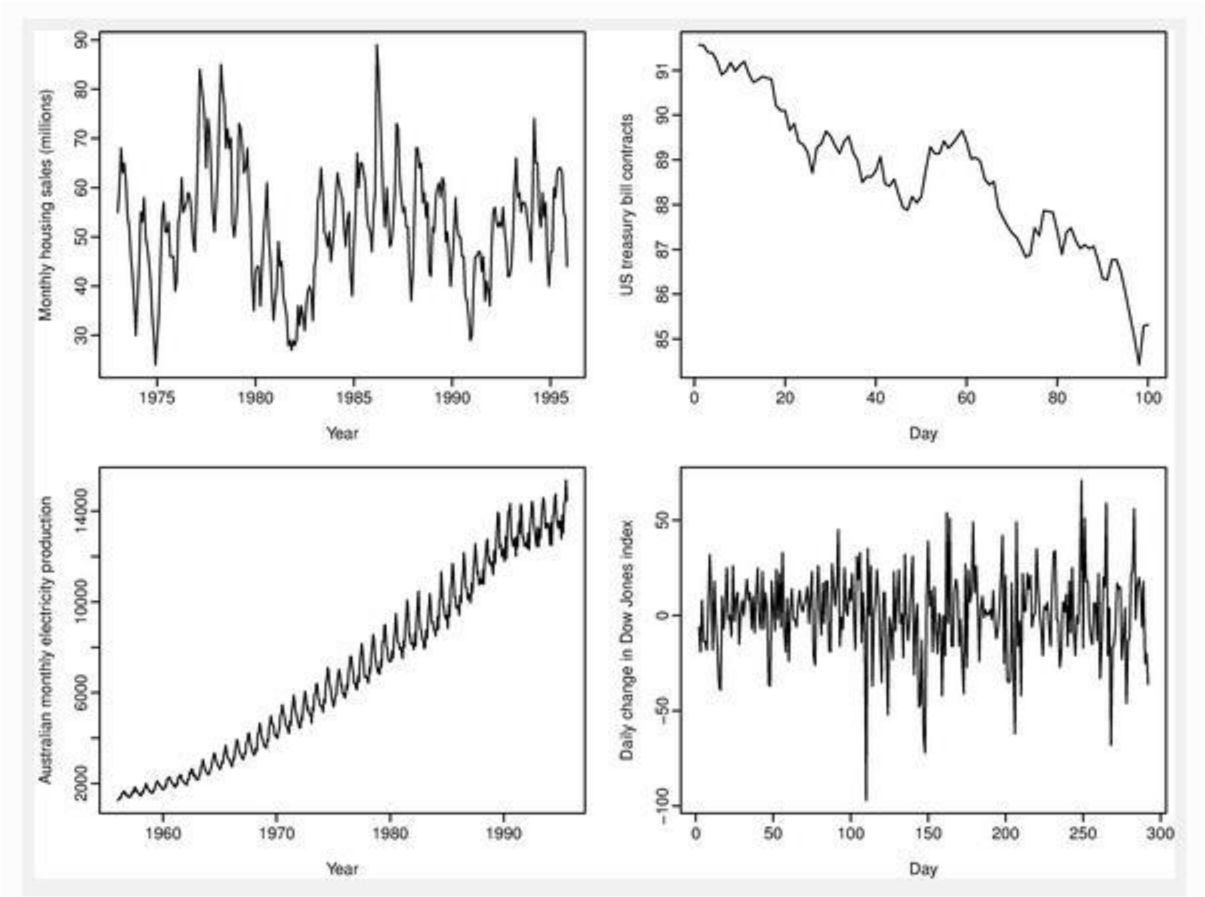
Descomposición de series temporales

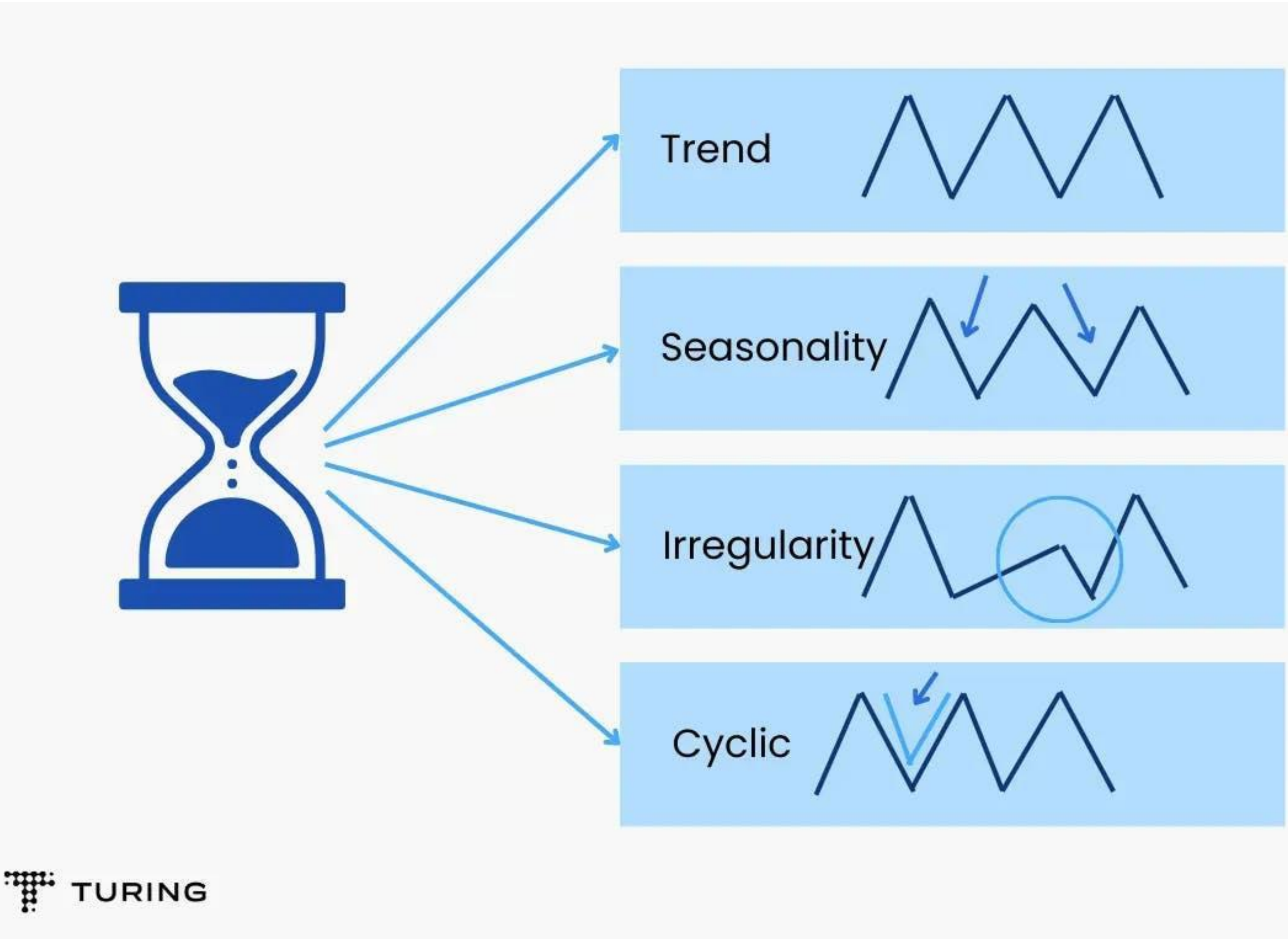
Descomposición de series temporales

- Las series temporales pueden mostrar una gran variedad de patrones y es útil categorizarlos y los comportamientos que se pueden visualizar
- A veces es útil descomponer una serie en componentes diferenciados, cada uno de los cuales representa un aspecto

Componentes de las series

Tendencia, estacionalidad, ciclo




Trend
Seasonality
Irregularity
Cyclic
TURING

Descomposición

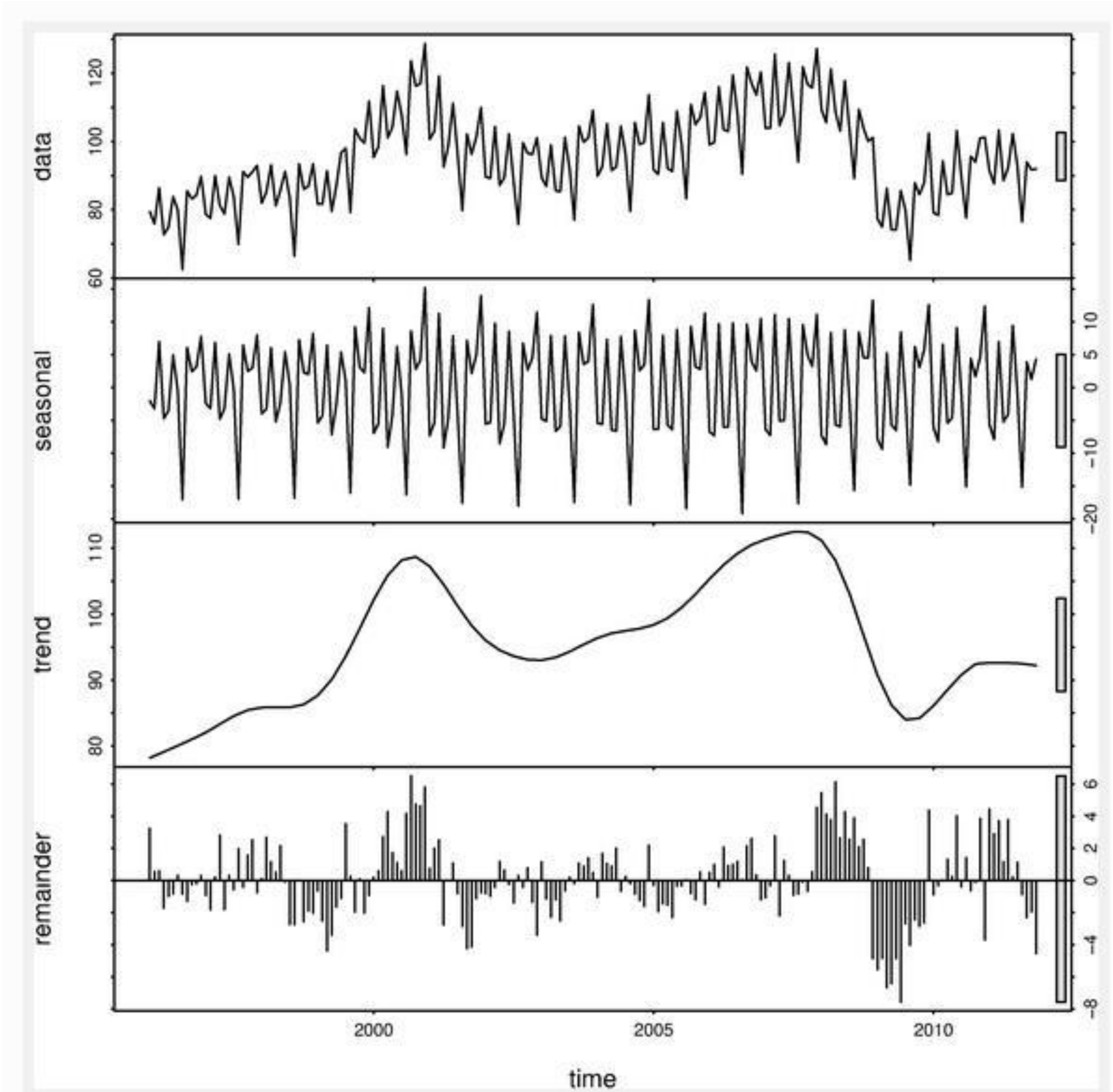
Descomposición aditiva

$$y_t = S_t + T_t + E_t$$

Indicada cuando la magnitud de las fluctuaciones estacionales por las variaciones entorno a la tendencia no varían con el valor de la serie temporal

Descomposición multiplicativa

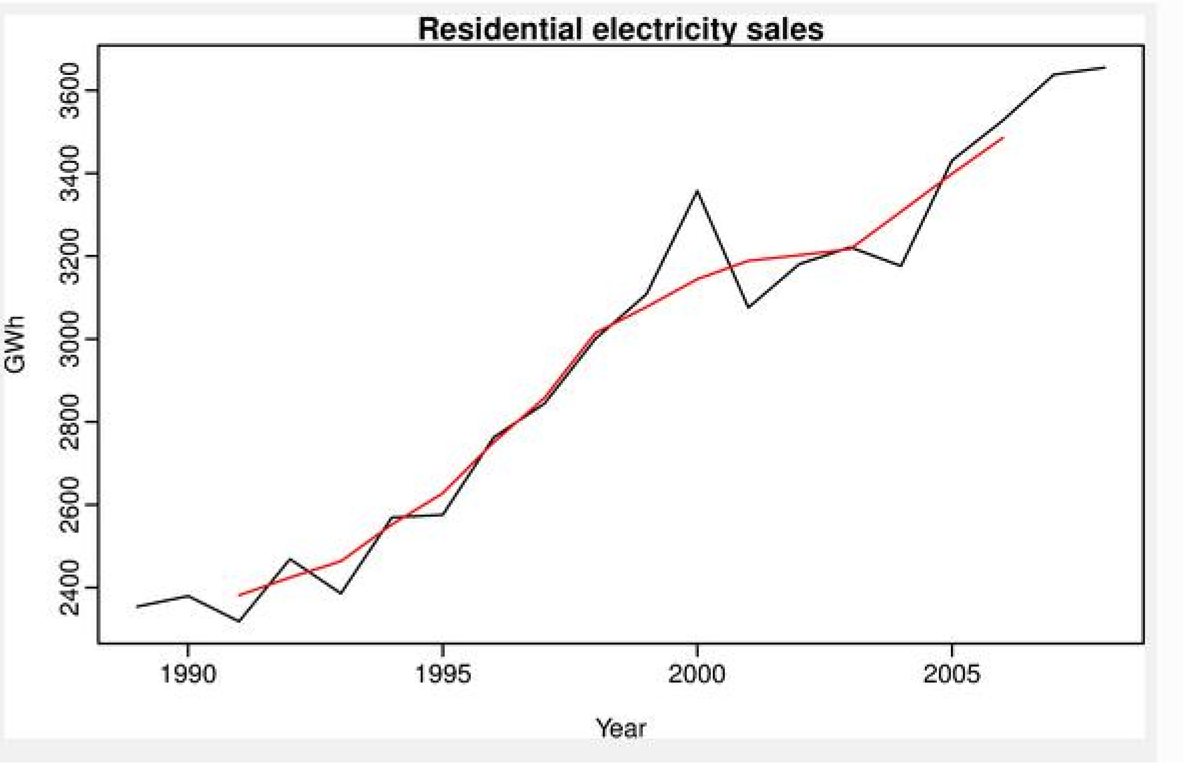
$$y_t = S_t \times T_t \times E_t$$

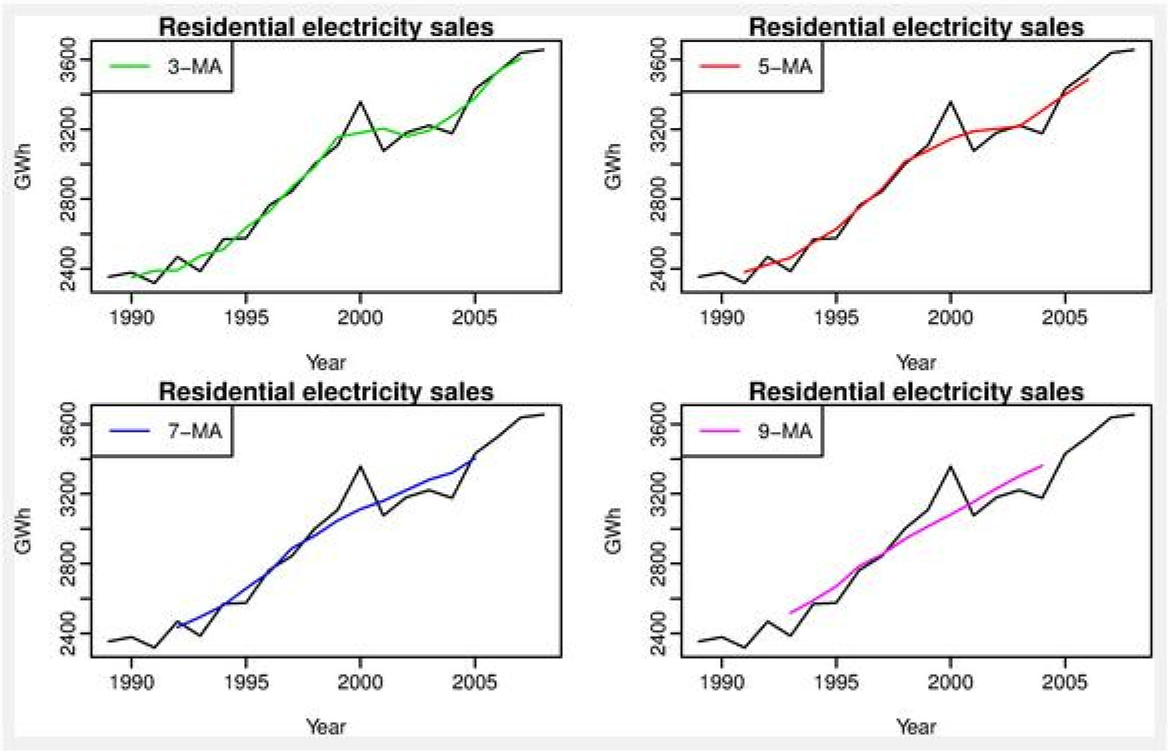


data
seasonal
trend
remainder
60
80
100
120
-20
-10
0
5
10
80
90
100
110
-8
-4
0
2
4
6
2000
2005
2010
time

Medias móviles

$$\hat{T}_t = \frac{1}{m} \sum_{j=-k}^{k} y_{t+j}$$





Utilizado frecuentemente para detectar tendencia y ciclo en datos estacionales

Medias móviles ponderadas

$$\hat{T}_t = \sum_{j=-k}^{k} a_j y_{t+j}$$

Su principal ventaja es que proporcionan una aproximación más “suave” de la tendencia-ciclo

Descomposición aditiva clásica

Step 1

If m is an even number, compute the trend-cycle component using a $2 \times m$ $-$MA to obtain $\hat{T}_t$. If m is an odd number, compute the trend-cycle component using an $m-$MA to obtain $\hat{T}_t$.

Step 2

Calculate the detrended series: $y_t - \hat{T}_t$.

Step 3

To estimate the seasonal component for each month, simply average the detrended values for that month. For example, the seasonal index for March is the average of all the detrended March values in the data. These seasonal indexes are then adjusted to ensure that they add to zero. The seasonal component is obtained by stringing together all the seasonal indices for each year of data. This gives $\hat{S}_t$.

Step 4

The remainder component is calculated by subtracting the estimated seasonal and trend-cycle components: $\hat{E}_t = y_t - \hat{T}_t - \hat{S}_t$.

Descomposición multiplicativa clásica

Step 1

If m is an even number, compute the trend-cycle component using a $2 \times m$ $-$MA to obtain $\hat{T}_t$. If m is an odd number, compute the trend-cycle component using an $m-$MA to obtain $\hat{T}_t$.

Step 2

Calculate the detrended series: $y_t/\hat{T}_t$.

Step 3

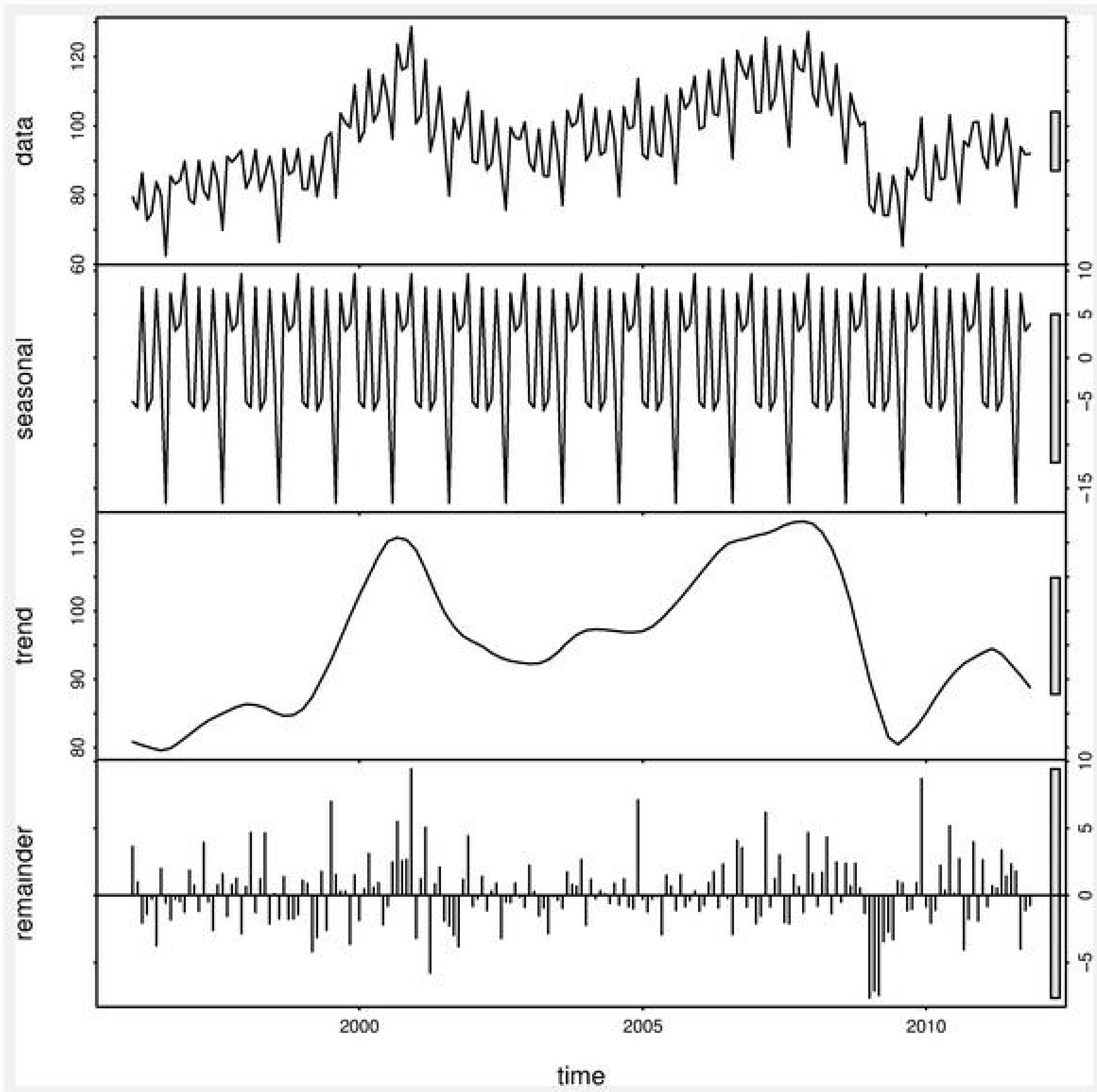
To estimate the seasonal component for each month, simply average the detrended values for that month. For example, the seasonal index for March is the average of all the detrended March values in the data. These seasonal indexes are then adjusted to ensure that they add to m. The seasonal component is obtained by stringing together all the seasonal indices for each year of data. This gives $\hat{S}_t$.

Step 4

The remainder component is calculated by subtracting the estimated seasonal and trend-cycle components: $\hat{E}_t = y_t/(\hat{T}_t\hat{S}_t)$.

Descomposición STL

- STL: método robusto y versátil de descomposición
  - It can handle any type of seasonality
  - The seasonal component is allowed to change over time, within a range controllable by the user
  - The smoothness of the trend-cycle can also be controlled by the user
  - It is robust to outliers


data
seasonal
trend
remainder
60
80
100
120
10
5
0
-5
-15
80
90
100
110
10
5
0
-5
2000
2005
2010
time

Predicción con descomposición

- Simple: predecir los componentes individuales y agregarlos

$$y_t = \hat{S}_t + \hat{A}_t$$

$$y_t = \hat{S}_t \hat{A}_t,$$

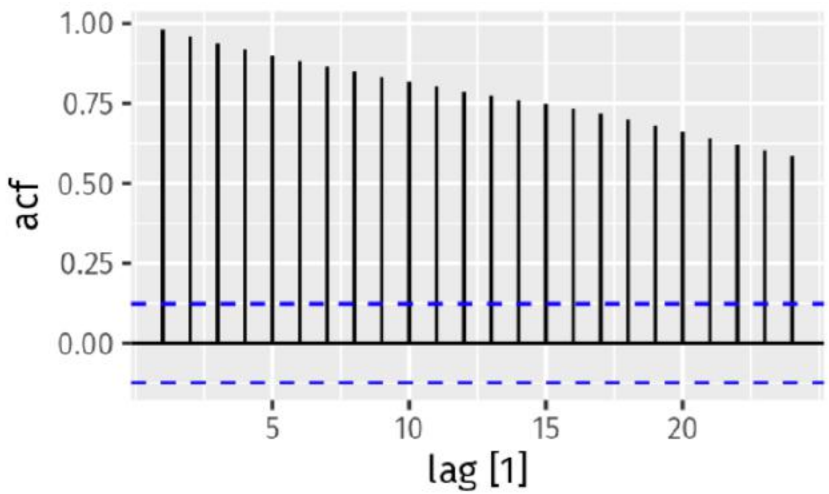
SEATS: Seasonal Extraction in ARIMA Time Series

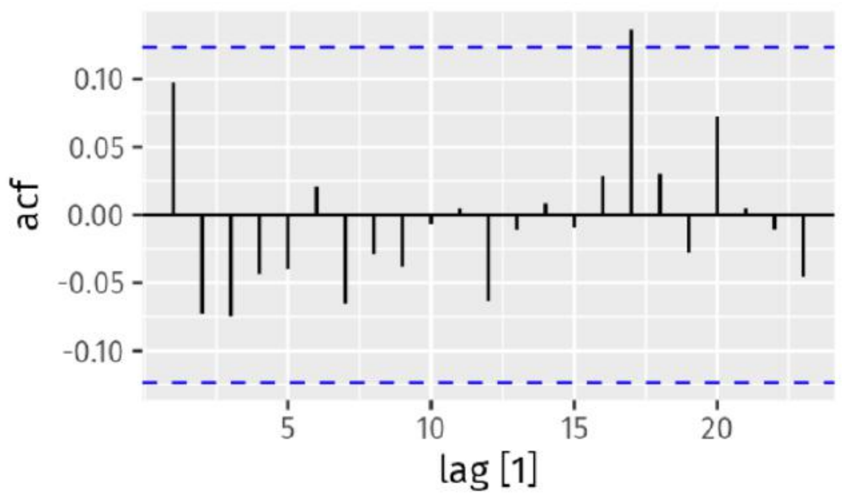
Modelos ARIMA

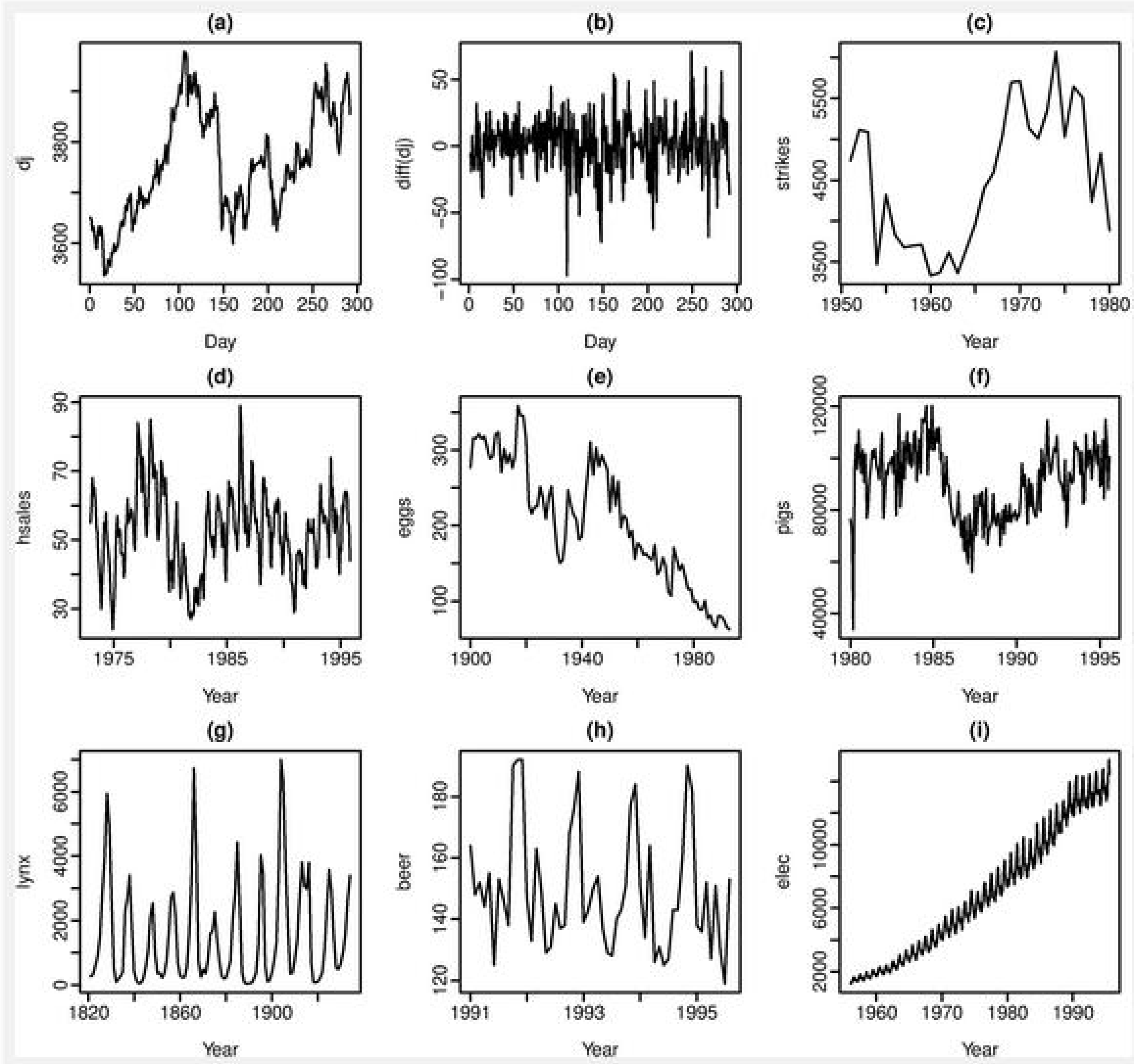
Estacionareidad

- Una serie estacionaria es aquélla cuyas propiedades no dependen del momento en que se observa la serie
- El gráfico ACF es útil para identificar series no estacionarias. Para series estacionarias se aproxima a cero rápidamente, mientras que para las series no estacionarias lo hace más lentamente

Detección de estacionareidad de series en el ACF






(a)
dj
3600
3800
0 50 100 150 200 250 300
Day
(b)
diff(dj)
-100 -50 0 50
0 50 100 150 200 250 300
Day
(c)
strikes
3500 4500 5500
1950 1960 1970 1980
Year
(d)
hsales
30 50 70 90
1975 1985 1995
Year
(e)
eggs
100 200 300
1900 1940 1980
Year
(f)
pigs
40000 80000 120000
1980 1985 1990 1995
Year
(g)
lynx
0 2000 4000 6000
1820 1860 1900
Year
(h)
beer
120 140 160 180
1991 1993 1995
Year
(i)
elec
2000 6000 10000
1960 1970 1980 1990
Year

Calcular diferencias

- Calcular las diferencias entre observaciones sucesivas: $z_t = y_t - y_{t-1}$
- Las transformaciones como el logaritmo pueden ayudar a estabilizar la varianza de las series temporales
- Las diferencias pueden ayudar a estabilizar la media de la serie temporal eliminando cambios en el valor de la serie, y por tanto, eliminando la tendencia y la estacionalidad

Modelo de paseo aleatorio

- Es una serie construida añadiendo el término de error a cada nuevo término de la serie:

$$y_t = y_{t-1} + e_t$$

donde la media de e_t es cero y su desviación típica es constante

- Los paseos aleatorios típicamente tienen:
  - Periodos largos de tendencias aparentes alcistas o bajistas
  - Cambios de dirección repentinos e impredecibles

Pruebas de raíz unidad

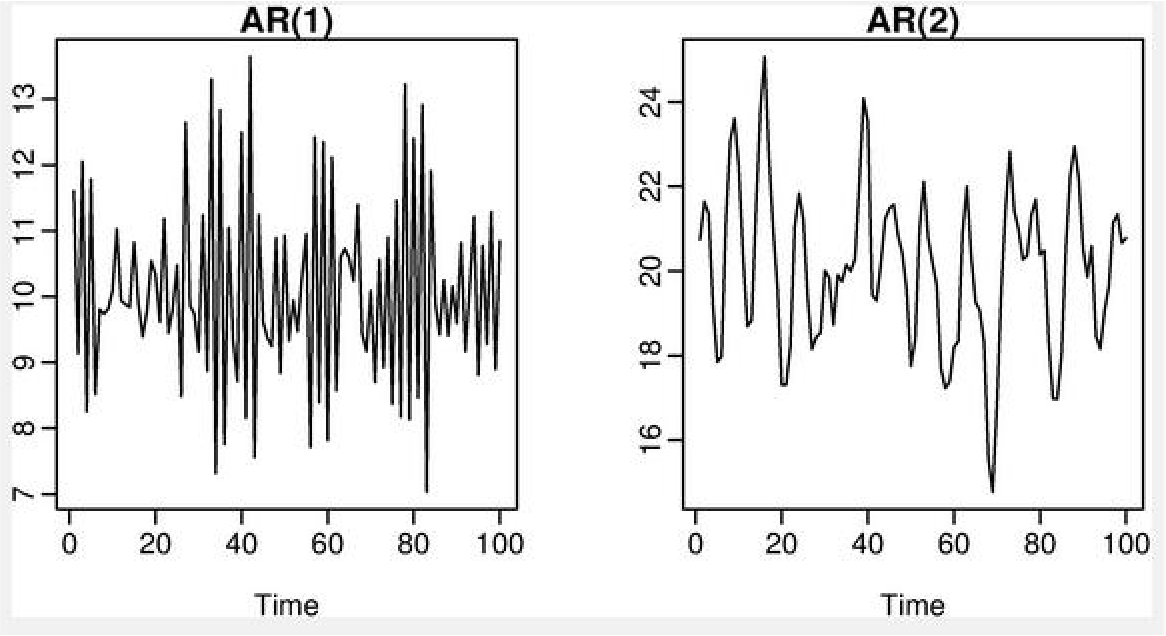
- Test de hipótesis estadístico de estacionareidad diseñados para determinar si es necesario hacer diferencias
- Test Dickey-Fuller ampliado

$$y'_t = \phi y_{t-1} + \beta_1 y'_{t-1} + \beta_2 y'_{t-2} + \cdots + \beta_k y'_{t-k}$$

Modelos autorregresivos

- Predicción usando una combinación lineal de valores pasados de la serie:

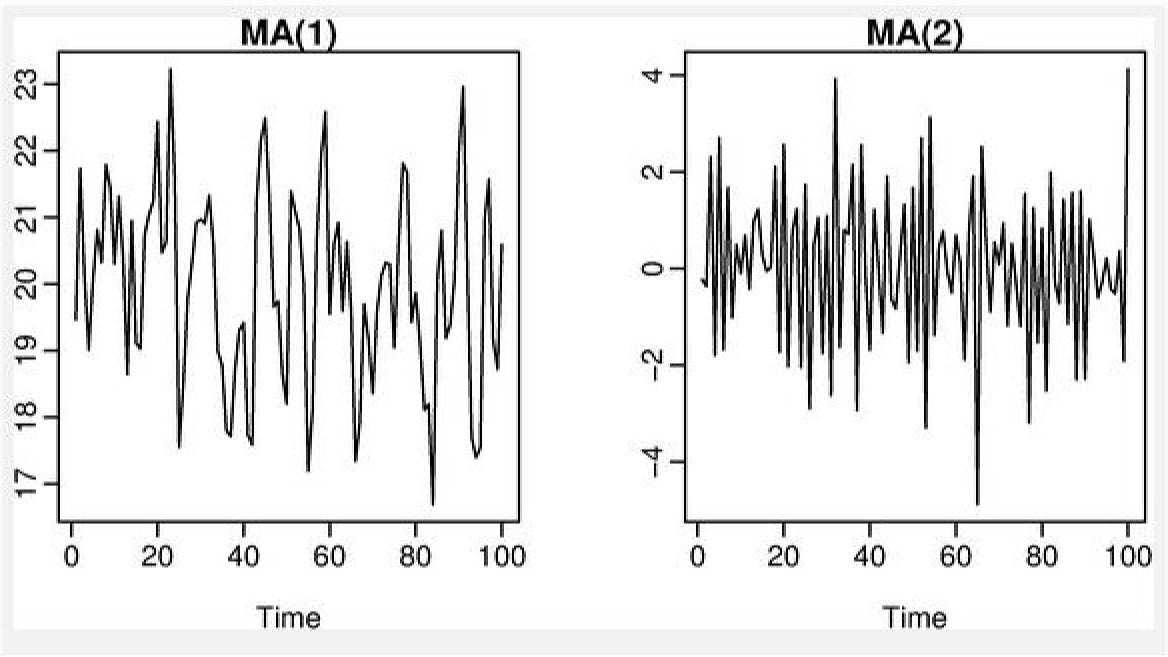
$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + e_t$$



Modelos de medias móviles

Regresión sobre **errores de predicción** previos

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$



Modelos ARIMA no estacionales

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \\ + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + e_t$$

ARIMA(p,d,q)

- p: orden de la parte autorregresiva
- d: grado de las diferencias
- q: orden de la parte de medias móviles

White noise	ARIMA(0,0,0)
Random walk	ARIMA(0,1,0) with no constant
Random walk with drift	ARIMA(0,1,0) with a constant
Autoregression	ARIMA(p,0,0)
Moving average	ARIMA(0,0,q)

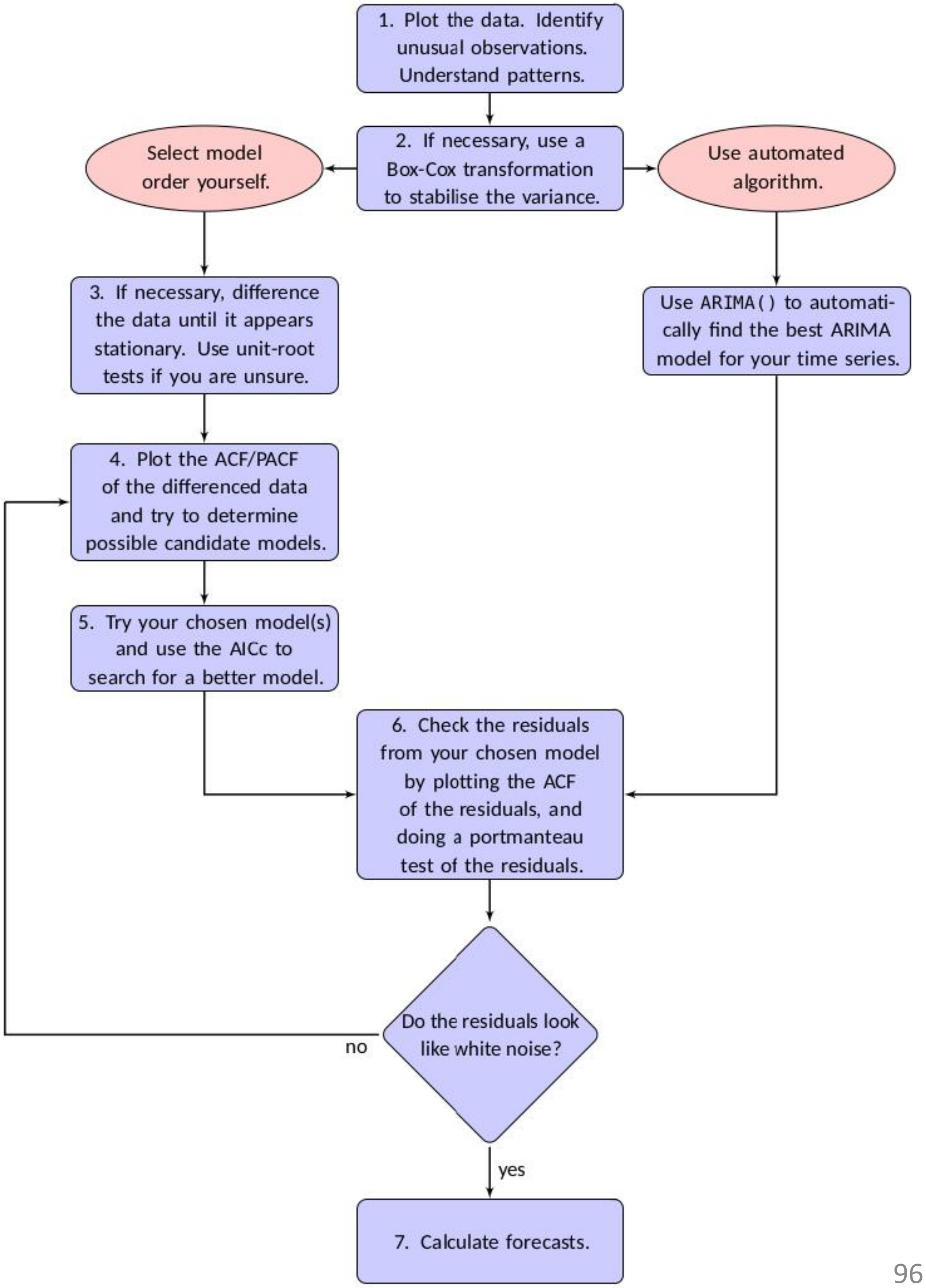
Consideraciones sobre modelos ARIMA

The constant c has an important effect on the long-term forecasts obtained from these models.

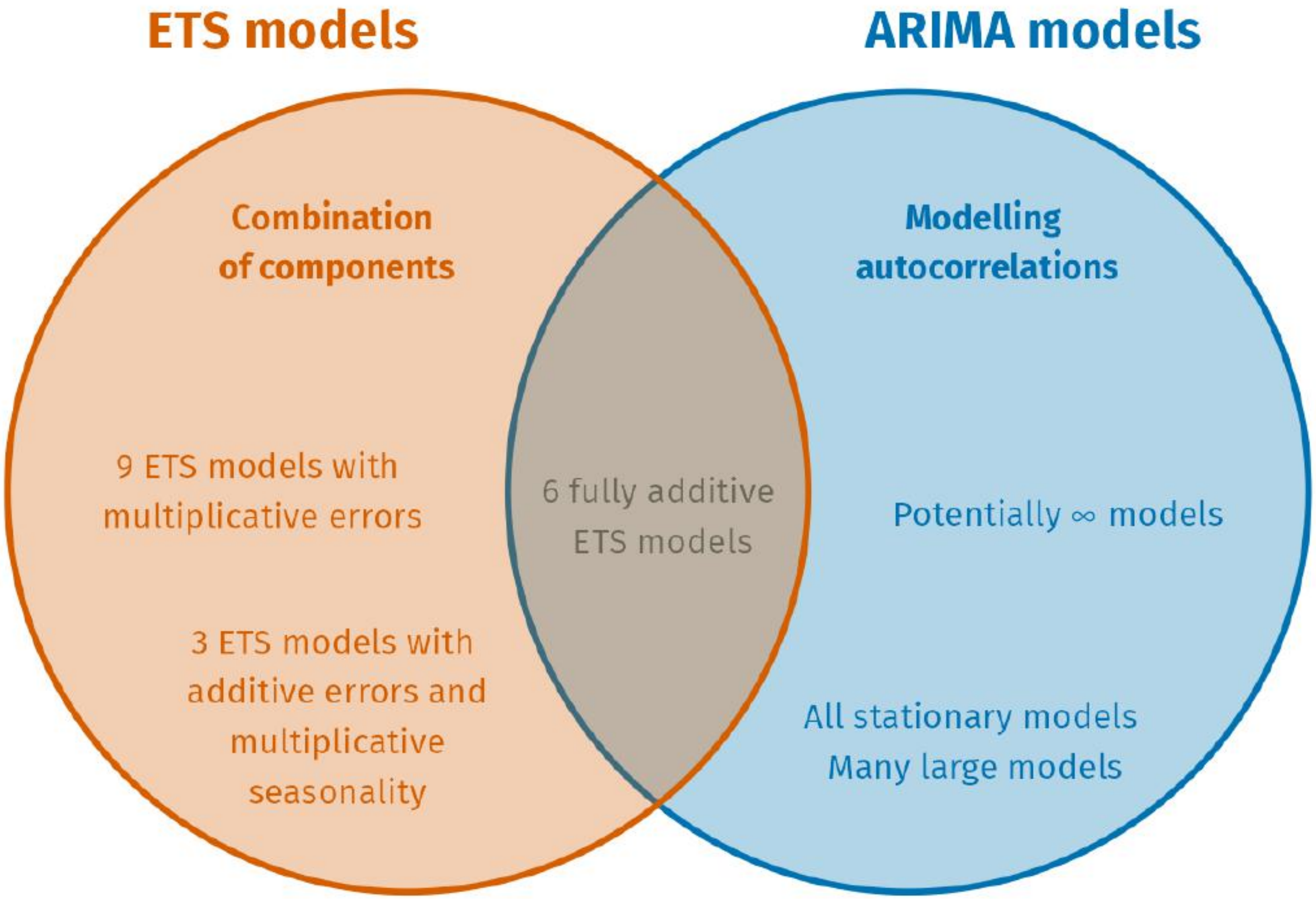
- If $c = 0$ and $d = 0$, the long-term forecasts will go to zero.
- If $c = 0$ and $d = 1$, the long-term forecasts will go to a non-zero constant.
- If $c = 0$ and $d = 2$, the long-term forecasts will follow a straight line.
- If $c \neq 0$ and $d = 0$, the long-term forecasts will go to the mean of the data.
- If $c \neq 0$ and $d = 1$, the long-term forecasts will follow a straight line.
- If $c \neq 0$ and $d = 2$, the long-term forecasts will follow a quadratic trend. (This is not recommended, and `fable` will not permit it.)

Función ARIMA() en el paquete fable

Aplicación de ARIMA



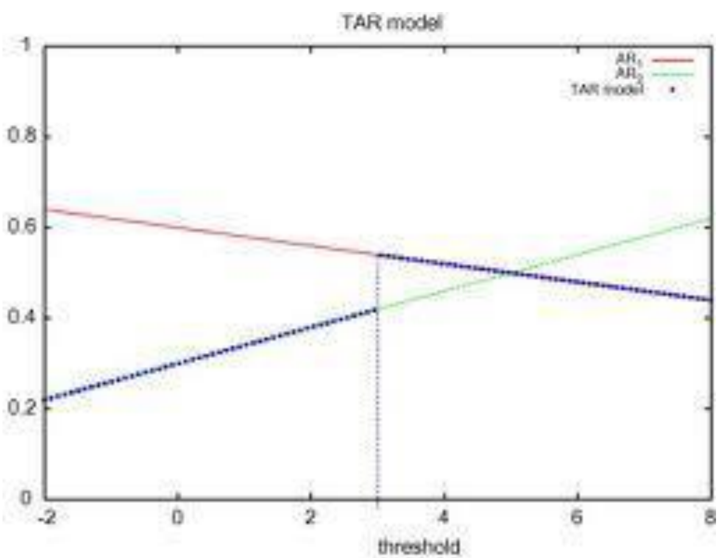
ARIMA vs. ETS

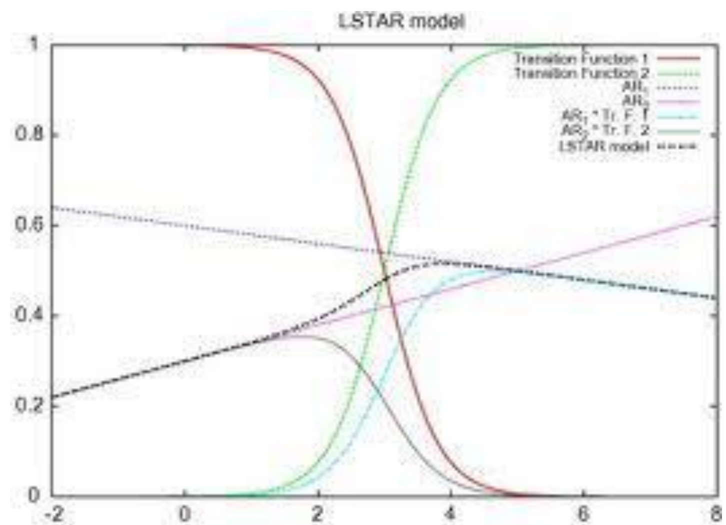

ETS models
ARIMA models
Combination of components
Modelling autocorrelations
9 ETS models with multiplicative errors
3 ETS models with additive errors and multiplicative seasonality
6 fully additive ETS models
Potentially ∞ models
All stationary models
Many large models

Métodos de predicción avanzados

Threshold Autoregressive Models

- Múltiples métodos autorregresivos, definidos por funciones indicador





Familia de modelos TAR

- TAR: Threshold AR
- STAR: Smooth TAR
- LSTAR: Logistic TAR
- NCSTAR: Neural Coefficient STAR

Conexiones entre las familias de modelos

- Muchas familias son realmente distintas caras de la misma idea básica
- Existen relaciones cercanas entre esas familias, lo que permite intercambiar propiedades y procedimientos
- ANN = FRBS (Fuzzy Rule-Based Systems)
- TAR = FRBS
- SVR = FRBS

Modelo Prophet

- Generado por Facebook para predicciones estacionales complejas
- Modelo de regresión no lineal

- g(t): tendencia lineal a trozos
- s(t): patrones estacionales
- h(t): efectos vacacionales

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t$$

Incrustamiento de series temporales

- (*Embedding*)

lag4	lag3	lag2	lag1	target
	...	...		
14	10	26	11	-13
10	26	11	-13	-15
26	**11**	**-13**	**-15**	**-8**
11	-13	-15	-8	35
-13	-15	-8	35	40
-15	**-8**	**35**	**40**	**-8**
-8	**35**	**40**	**-8**	**-16**
35	**40**	**-8**	**-16**	**7**
40	-8	-16	7	17
	...	...		

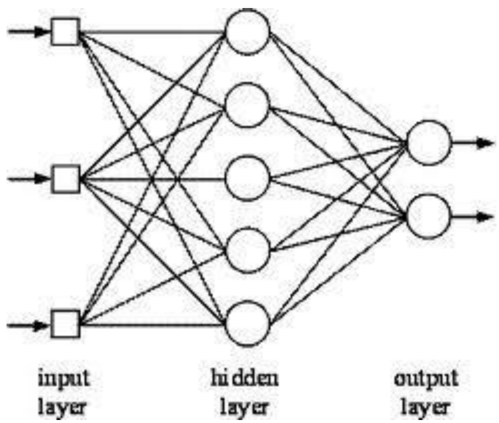
Redes Neurales Artificiales

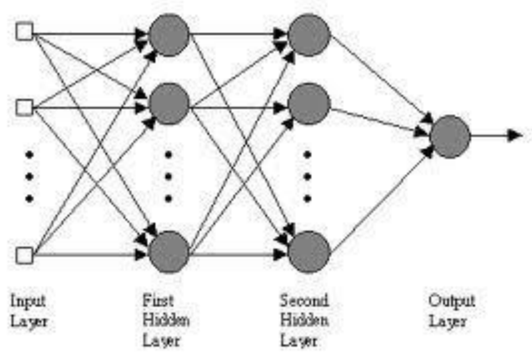
- Perceptrones multicapa
- RBF
- Redes Neuronales Recurrentes
  - Elman
  - Jordan
  - Recurrent MLP
- Deep Learning
  - LSTM
  - Transformers

MLP

- Los MLPS son los modelos más conocidos y usados
- Debido a su eficacia en problemas de regresión se aplican frecuente en predicción de series temporales
- Las mismas consideraciones que se hacen para problemas de regresión regular deben hacerse para la predicción de series temporales

Capas ocultas





Pasos para la aplicación de MLP

1. Definir el problema: entradas y salidas
2. Preprocesamiento: transformaciones sobre los datos
3. Definir la arquitectura de la red
   – Número de capas; número de unidades en cada capa
   – Función de activación
4. Definir algoritmos de aprendizaje y parámetros
5. Ajustar el modelo
6. Validar el modelo
7. Aplicarlo

Software para RNA

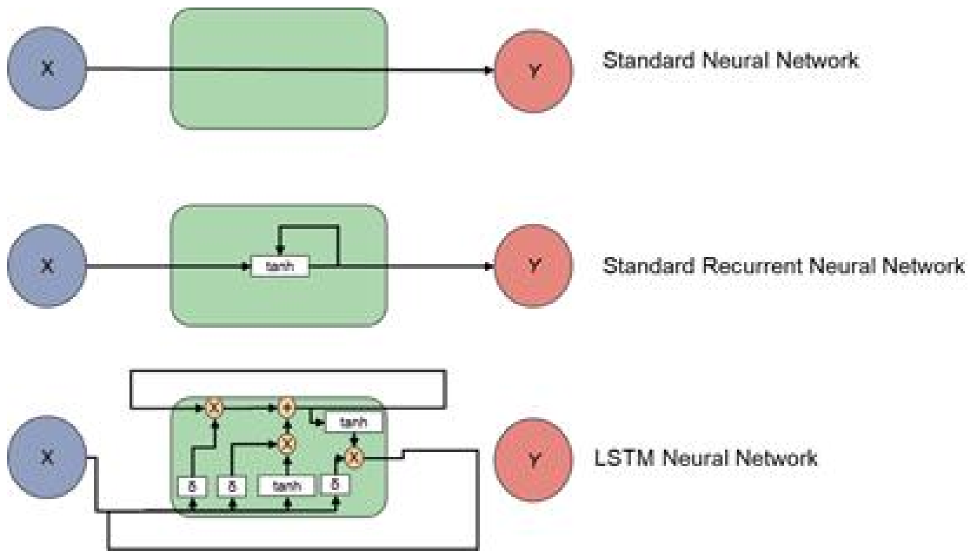
R

- nnet
- RSNNS
- caret

Python

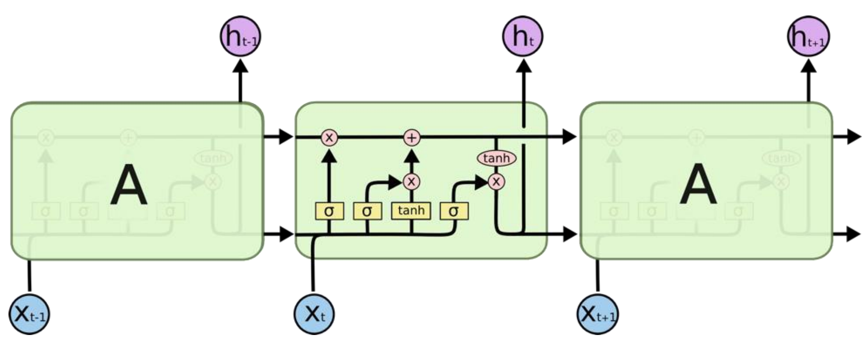
- sckit-learn
- Tensorflow
- Pytorch

LSTM



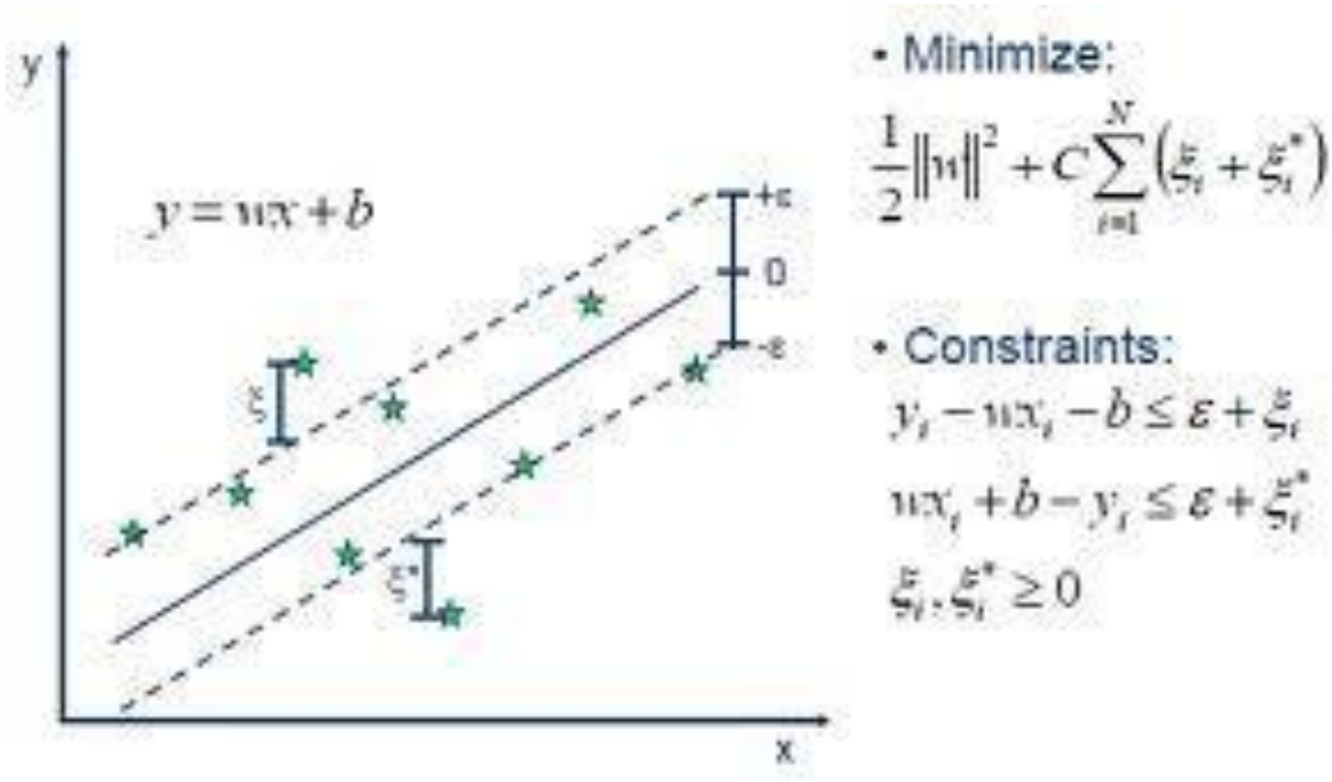
Memorias a:

a) Corto plazo
b) Largo plazo



Support Vector Regression

- Esta es la versión para regresión de las SVM

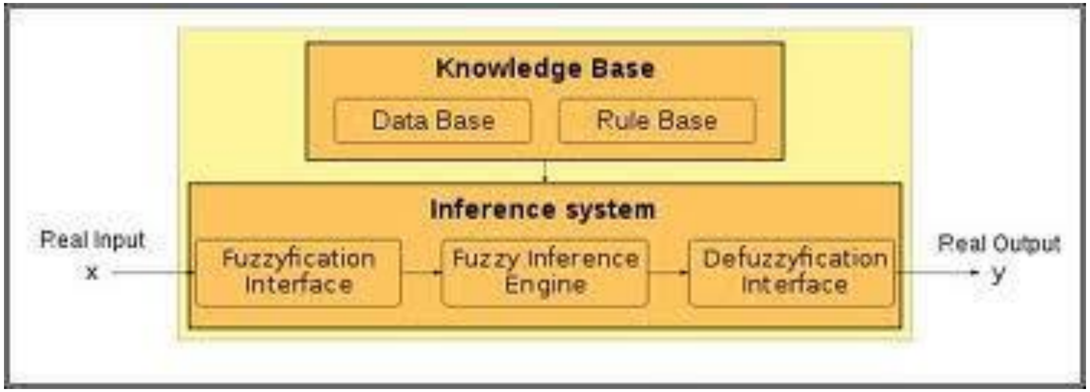


Pasos para la aplicación de SVR

- Definir el problema: entradas y salidas
- Preprocesamiento: transformaciones sobre los datos
- Definir los parámetros de la SVM
  - Tipo de la función kernel
  - Parámetros
- Ajustar el modelo
- Validar el modelo
- Aplicarlo

Sistemas basados en reglas difusas

- En particular, los TSK, más adecuados para regresión
- Pueden aportar una interpretación de los modelos



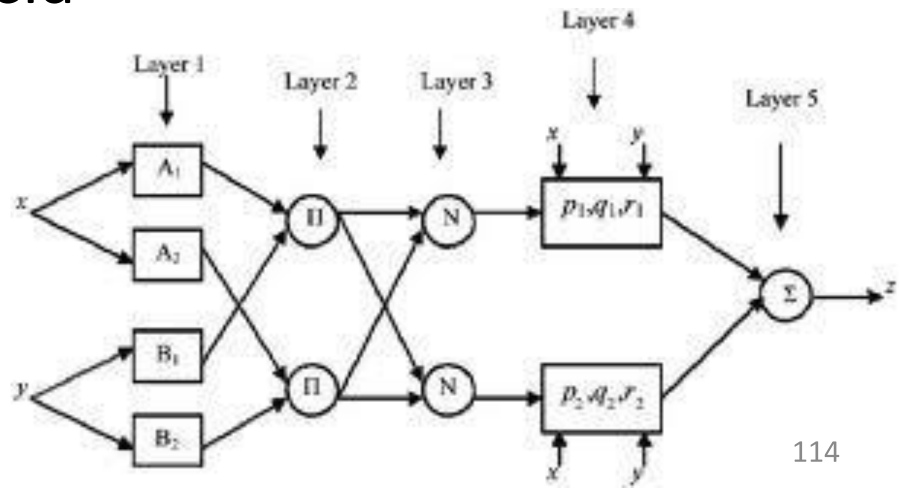
Paquete: frbs

Pasos para la aplicación de FRBS

1. Definir el problema: entradas y salidas
2. Aplicar posibles transformaciones a los datos
3. Definir la partición lingüística:
   - Granularidad: número de etiquetas
   - Semántica:
     - Tipo de funciones de pertenencia (guassianas, trapezoidales, triangulares,…)
     - Parámetros para cada etiqueta
   1. Estructura de los consecuentes
4. Crear base de reglas inicial:
   - A partir de expertos y conocimiento inicial
   - Derivado de datos (aprendizaje automático)
5. Definir el método de entrenamiento y ajuste y sus parámetros
   - Descenso en gradiente
   - Algoritmos evolutivos
   - Modelos híbridos (ANFIS)
6. Ajustar el modelo
7. Validar el modelo
8. Aplicarlo

ANFIS

- Modelo neuro-difuso con algoritmo de entrenamiento híbrido:
  - Optimización basada en mínimos cuadrados para los coeficientes de salida
  - Descenso en gradiente para los parámetros de las funciones de pertenencia



Otras técnicas

- Wavelets
- Spectral Analysis
- State-space models
- Hidden Markov models
- MARS
- Functional analysis
- …

Minería de datos en series temporales

Tareas

- Clasificación
- Clustering
- Identificación de patrones
- Detección de anomalías

Clasificación de series temporales

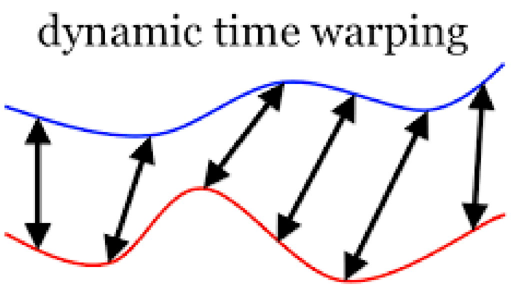
- Medidas de similitud
- Representación de series temporales
- Técnicas

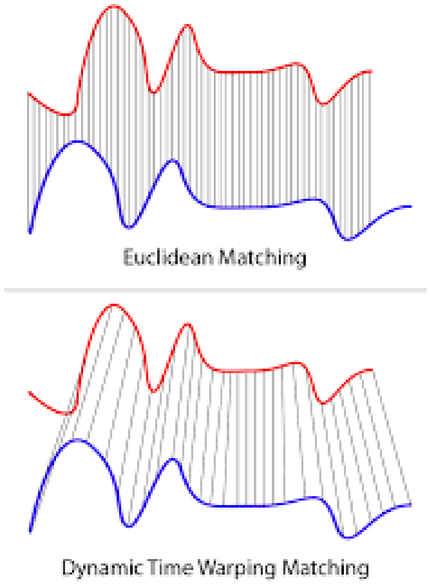
Medidas de similitud

- Métricas clásicas
  - Euclídea
  - Suma
  - Máximo
- Dynamic Time Warping (DTW)
- Subsecuencia común más larga

Dynamic Time Warping

- Método para calcular el emparejamiento óptimo entre dos secuencias bajo ciertas reglas. $O(n^2)$

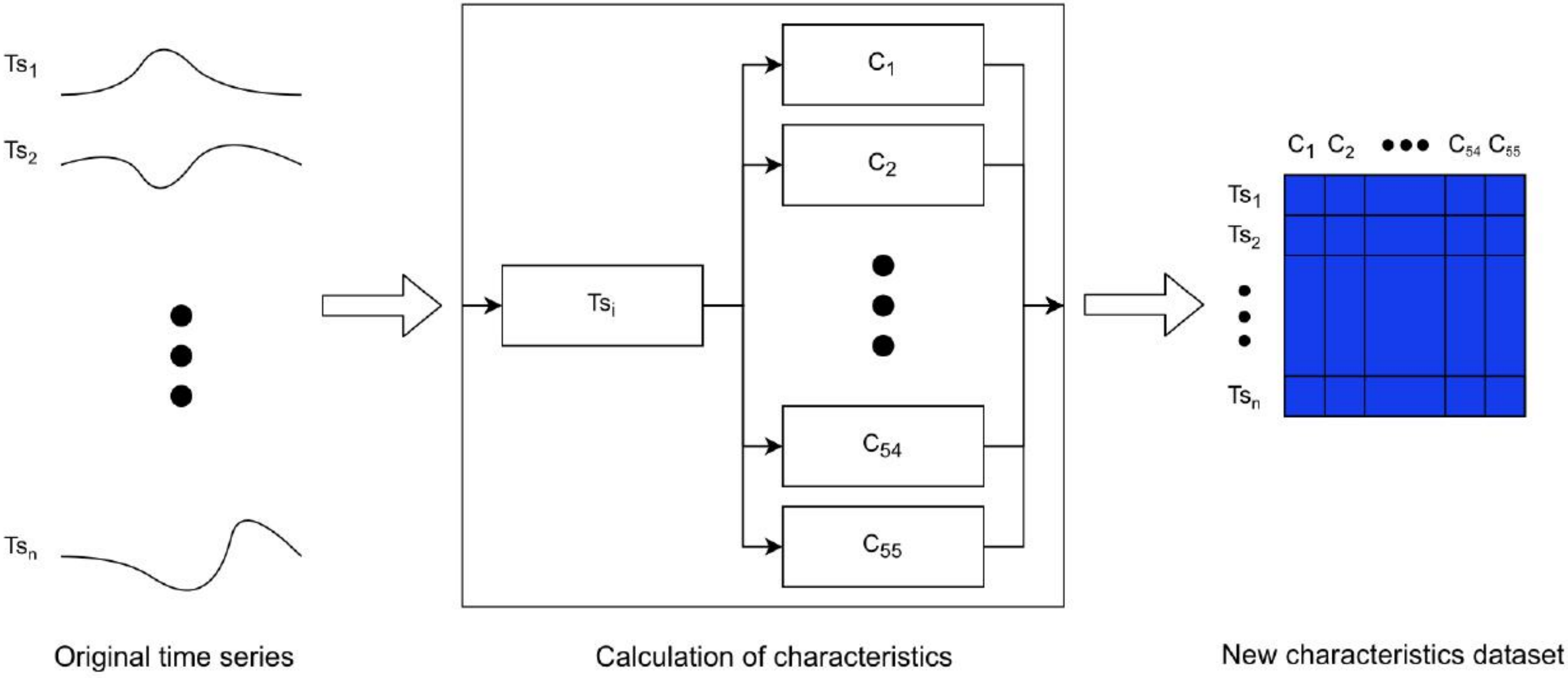


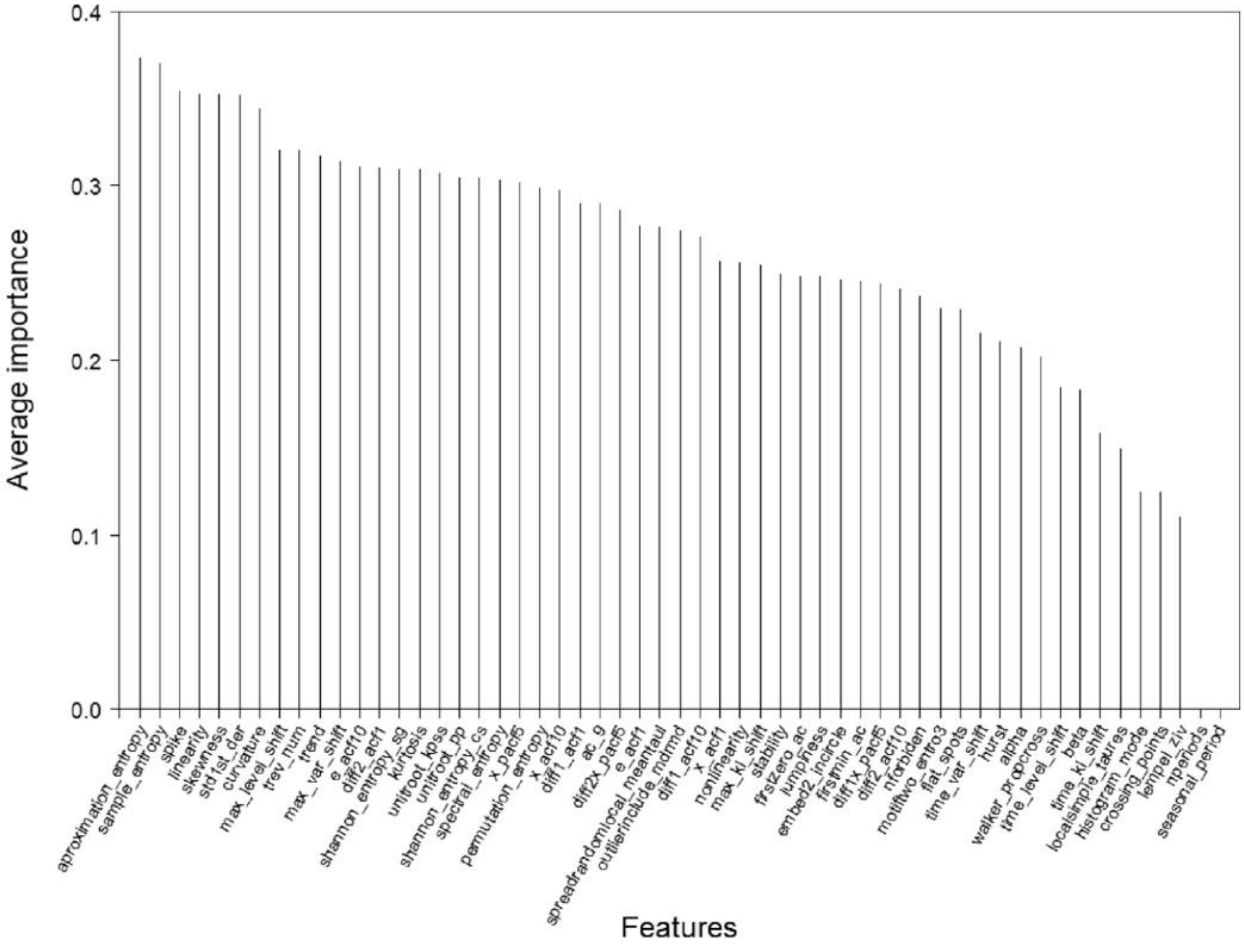


Representación de series temporales

- En el dominio del tiempo (vector de datos)
- No adaptativa:
  - Fourier
  - Laplace
  - Wavelet
- Adaptativa:
  - Symbolic Aggregate Approximation (SAX)
  - Piecewise Linear Representation (PLA)
- Basado en modelos: Modelos de Markov
- Características (Baldan et al.)

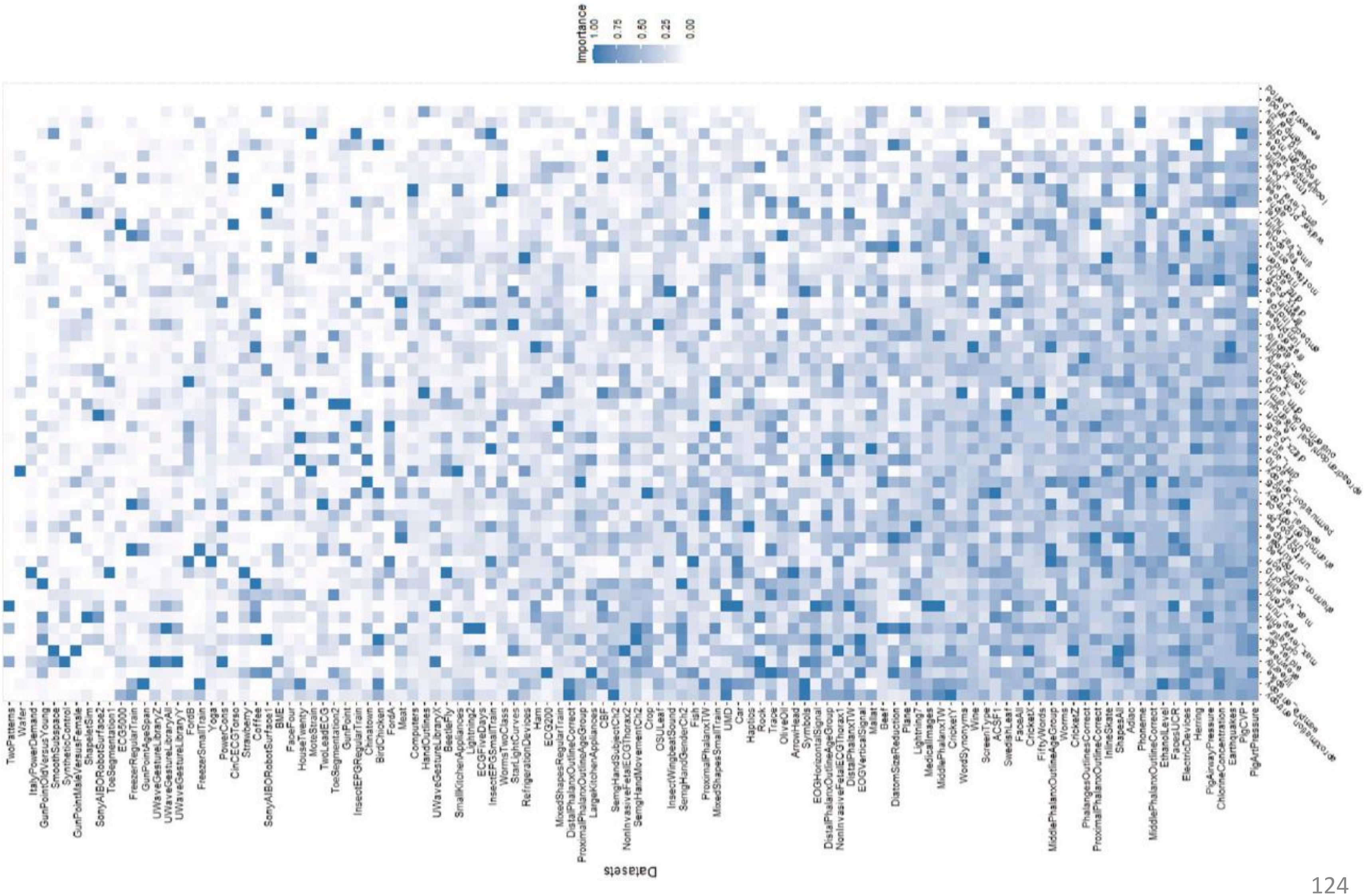
Representación con características





Average importance
0.0
0.1
0.2
0.3
0.4
Features
aproximation_entropy
sample_entropy
spike
linearity
skewness
std1st_der
curvature
max_level_shift
trev_num
trend
max_var_shift
e_acf10
diff2_acf1
shannon_entropy_sg
kurtosis
unitroot_kpss
unitroot_pp
shannon_entropy_cs
spectral_entropy
x_pacf5
permutation_entropy
x_acf10
diff1_acf1
ac_9
diff2x_pacf5
e_acf1
spreadrandomlocal_meantaul
outlierinclude_mdrmd
diff1_acf10
x_acf1
nonlinearity
max_kl_shift
stability
firstzero_ac
lumpiness
embed2_incircle
firstmin_ac
diff1x_pacf5
diff2_acf10
nforbiden
motiftwo_entro3
flat_spots
time_var_shift
hurst
alpha
walker_propcross
time_level_shift
beta
time_kl_shift
localsimple_taures
histogram_mode
crossing_points
lempel_ziv
nperiods
seasonal_period

Importancia por dataset



Técnicas

- Referencia habitual: 1-NN con DTW
- Con representación vectorial
  - Dominio de tiempo de igual longitud
  - Características
  - Todos los métodos habituales
- Repositorio de UCR

Clasificación en Big Data

Clasificación en Big Data: Distributed Shapelet

Sequential FS and DFST running time vs number of time series problem 1.

Number of time series	Sequential FS running time (s)	DFST running time (s)
100,000	2,088.06	12,190.91
200,000	4,119.02	12,310.29
300,000	6,699.75	9,748.17
400,000	7,909.84	11,222.16
500,000	11,643.08	9,025.84
600,000	13,407.48	12,057.48
700,000	16,317.37	8,965.22
800,000	18,454.49	9,798.00
900,000	20,764.48	12,226.55
1,000,000	22,513.41	11,855.29
1,100,000	24,622.18	9,651.16
1,200,000	26,634.61	11,536.39
1,300,000	30,618.57	13,686.43
1,400,000	30,524.07	13,741.24
1,500,000	33,556.26	13,075.48
1,600,000	36,532.96	14,814.28
1,700,000	38,695.64	14,998.85
1,800,000	41,606.93	15,583.09
1,900,000	35,099.40	16,625.55
2,000,000	49,435.64	15,128.27
2,100,000	48,421.24	15,692.12
2,200,000	53,036.16	15,667.97
3,000,000	NC	20,560.59
4,000,000	NC	26,823.02
5,000,000	NC	27,244.44
6,000,000	NC	33,295.01
7,000,000	NC	35,709.65
8,000,000	NC	42,389.48
9,000,000	NC	44,831.42
10,000,000	NC	52,685.68
11,000,000	NC	61,208.60
12,000,000	NC	67,723.17
13,000,000	NC	75,183.40
14,000,000	NC	75,973.21
15,000,000	NC	98,655.94
16,000,000	NC	103,155.57

Referencias

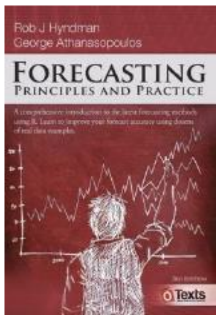
- C. Chatfield, «The analysis of time series: An Introduction», CRC Press, 2003
- J.D. Hamilton, «Time Series Analysis», Princeton University Press, 1994
- R. Hyndman, G. Athanasopoulus, «Forecasting and time series» 2013
- P.J. Brockwell, R.A. Davis, «Time Series: Theory and Methods», 2nd Ed., Springer, 1991
- J.S. Armstrong (ed), «Principles of Forecasting: A Handbook for Researchers and Practitioners», Springer, 2001

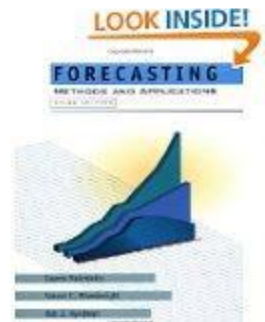
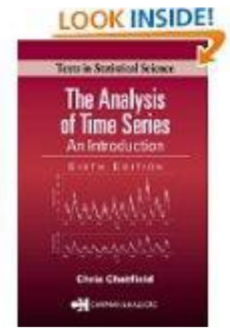
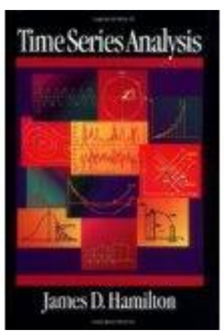
Referencias (2)

- S.G. Makridakis, S.C. Wheelwright, R.J. Hyndman, «Forecasting», 3rd Ed., Wiley & Sons, 1998
- P.J. Brockwell, R.A. Davis, «Introdution to Time Series and Forecasting», 2nd ed., Springer, 2002
- A.K.Palit, D. Popovic, «Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications», Springer, 2005
- R.H. Shumway, D.S. Stoffer, «Time Series Analysis and Its Applications», Springer, 2nd Ed., 2006

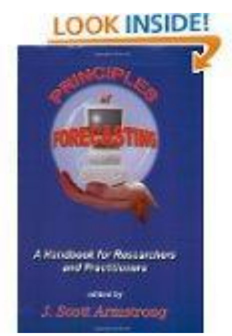
Referencias (3)

- T. Mitsa, “Temporal Data Mining”, CRC Press, 2010
- W. Hsu, M.L. Lee, J. Wang, “Temporal and Spatio-Temporal Data Mining”, IGI Publishing, 2008
- G. Dong, J. Pei, “Sequence Data Mining”, Springer Science, 2007


Rob J Hyndman
George Athanasopoulos
FORECASTING
PRINCIPLES AND PRACTICE
OTexts

LOOK INSIDE!
Introduction to Time Series and Forecasting

LOOK INSIDE!
FORECASTING
METHODS AND APPLICATIONS

LOOK INSIDE!
Texts in Statistical Science
The Analysis of Time Series
An Introduction
Chris Chatfield

Time Series Analysis
James D. Hamilton

LOOK INSIDE!
Computational Intelligence in Time Series Forecasting

LOOK INSIDE!

LOOK INSIDE!
PRINCIPLES of FORECASTING
A Handbook for Researchers and Practitioners
edited by
J. Scott Armstrong