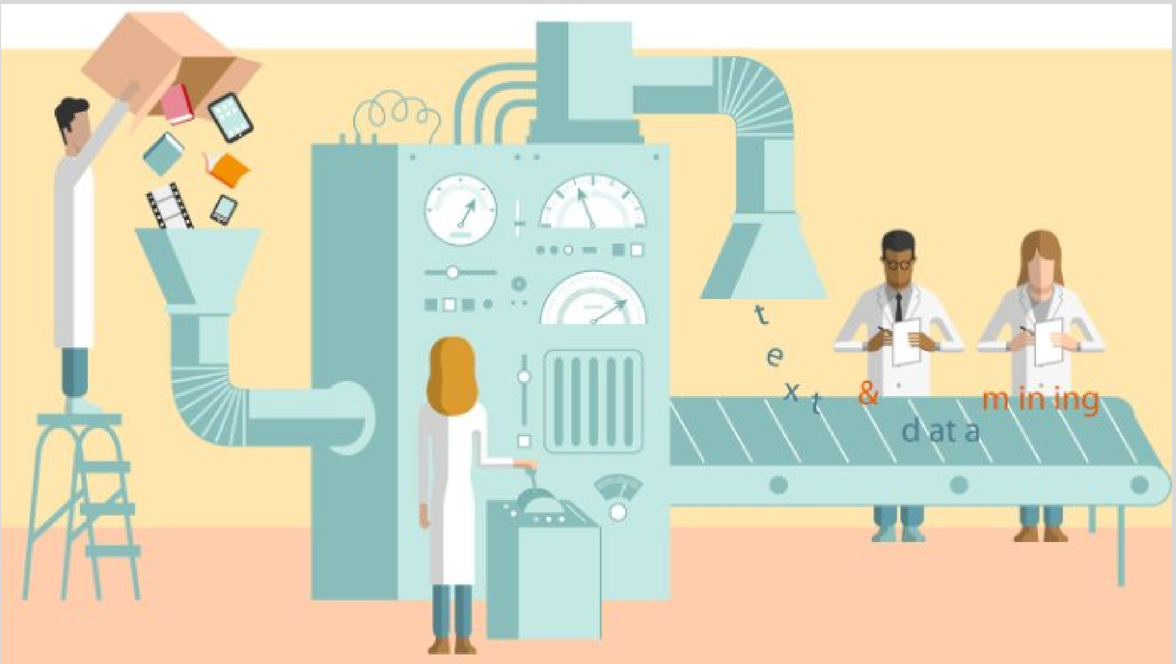


DATA SCIENCE

MODULO 5

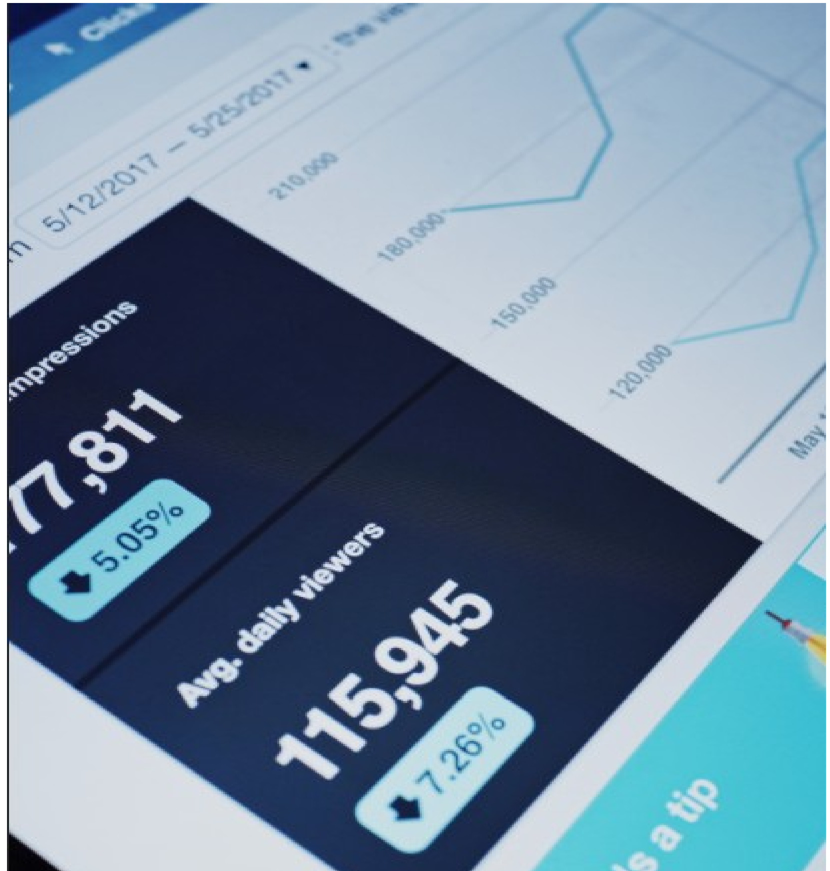
Series de Tiempo I

Series de Tiempo

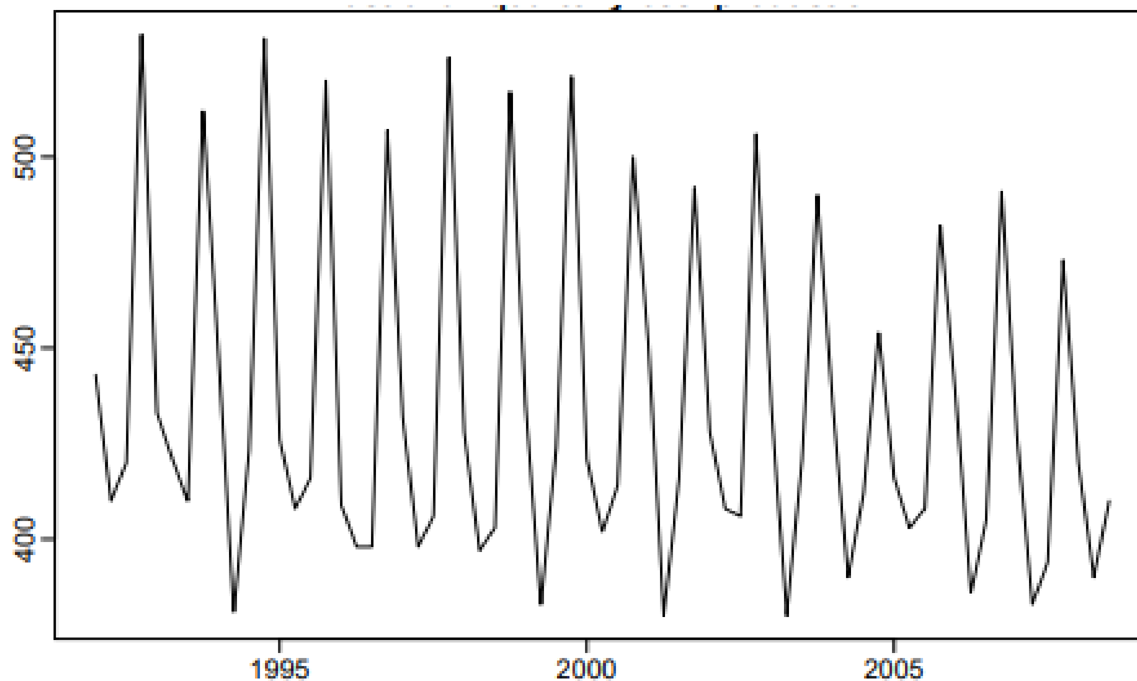
1. **Introducir los datos de series temporales, entender el concepto de estacionariedad.**
2. **Presentar modelos simples para datos de series temporales**
3. **Comprender el esquema de validación cruzada a aplicar con datos de series temporales**

Introducción. Series de Tiempo

DigitalHouse > Coding School

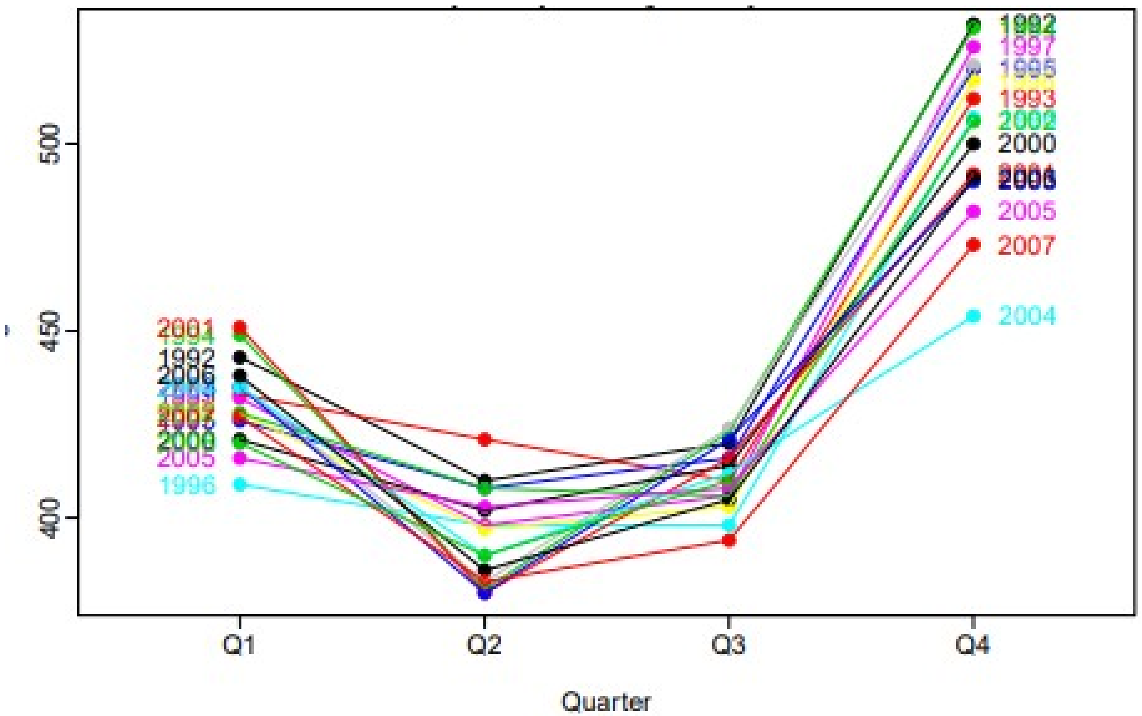


Introducción. Series de Tiempo



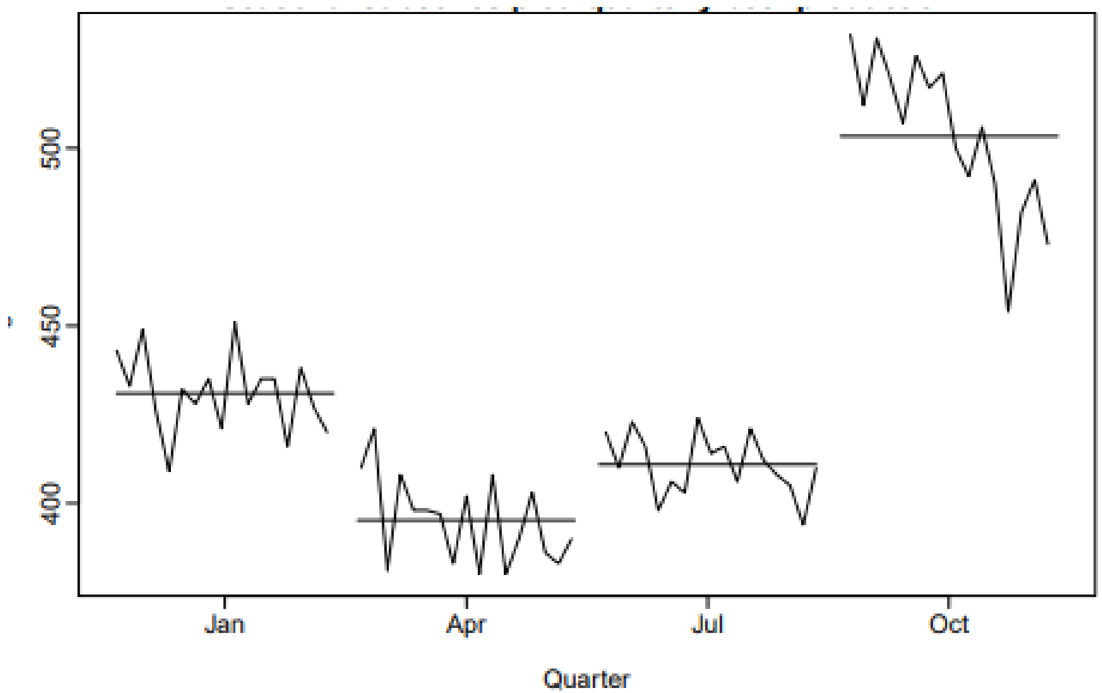
Introducción. Series de Tiempo

Seasonal plots



Introducción. Series de Tiempo

Seasonal subseries plots



Introducción. Series de Tiempo

Patrones en series de tiempo

- El **patrón de tendencia** existe cuando a largo plazo hay un Incremento o disminución de los datos.
- El **patrón estacional** existe cuando una serie es influenciada por factores estacionales (por ejemplo, el trimestre del año, mes o día de la semana).
- Existe un **patrón cíclico** cuando los datos exhiben aumentos y caídas que no son de periodo fijo (duración normalmente de al menos 2 años).

Introducción. Series de Tiempo

Diferencias entre patrones en series de tiempo

- El patrón estacional tiene longitud constante mientras que el patrón cíclico tiene longitud variable
- La duración media del ciclo es más larga que la duración del patrón estacional.
- La magnitud de ciclo es más variable que la magnitud del patrón estacional
- El tiempo de los picos y valles es predecible con datos estacionales, pero impredecibles a largo plazo en el patrón cíclico.

Introducción. Series de Tiempo

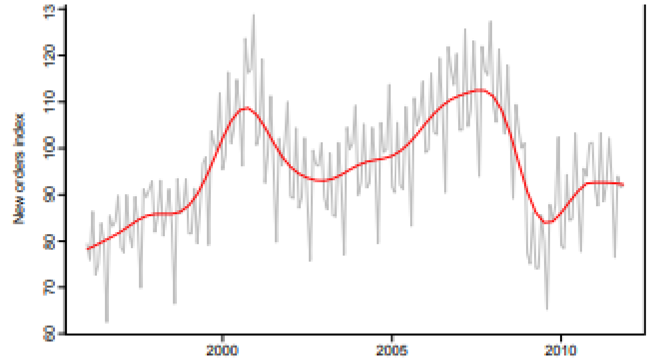
DigitalHouse> Coding School

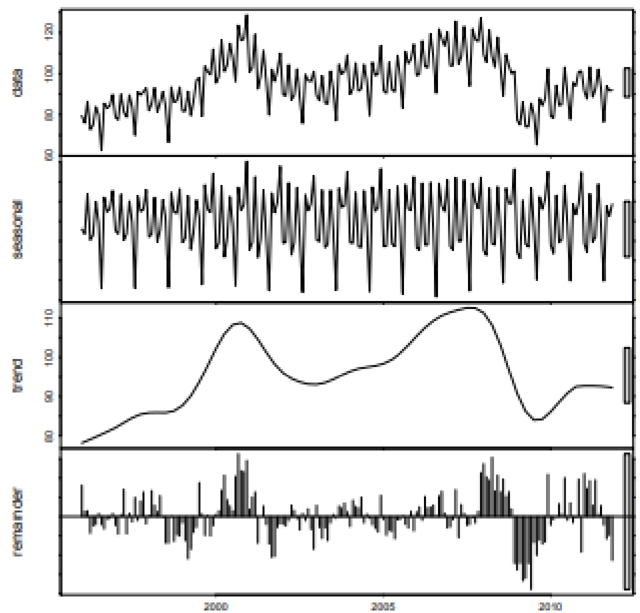
$$Y_t = f(S_t, T_t, E_t)$$

$Y_t =$ data at period t
$S_t =$ seasonal component at period t
$T_t =$ trend component at period t
$E_t =$ remainder (or irregular or error) component at period t

Additive decomposition: $Y_t = S_t + T_t + E_t$.

Introducción. Series de Tiempo





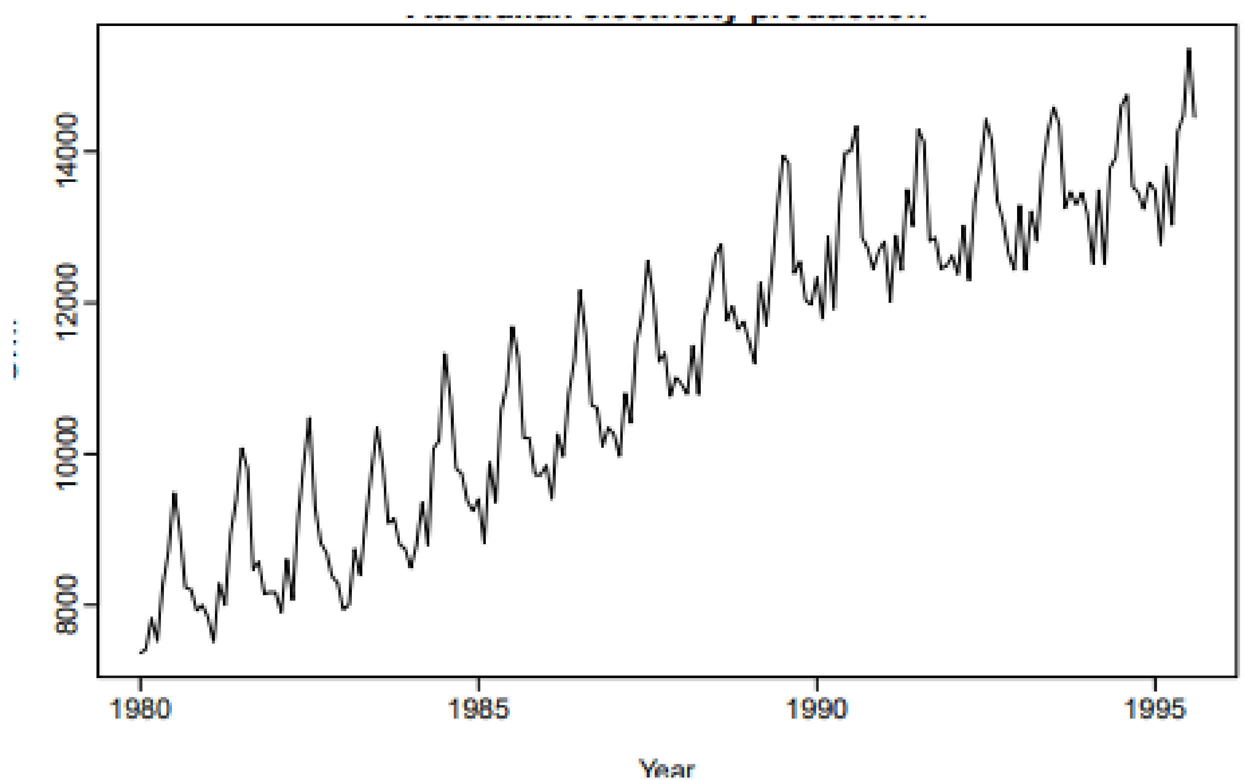
Introducción. Series de Tiempo

DigitalHouse> Coding School

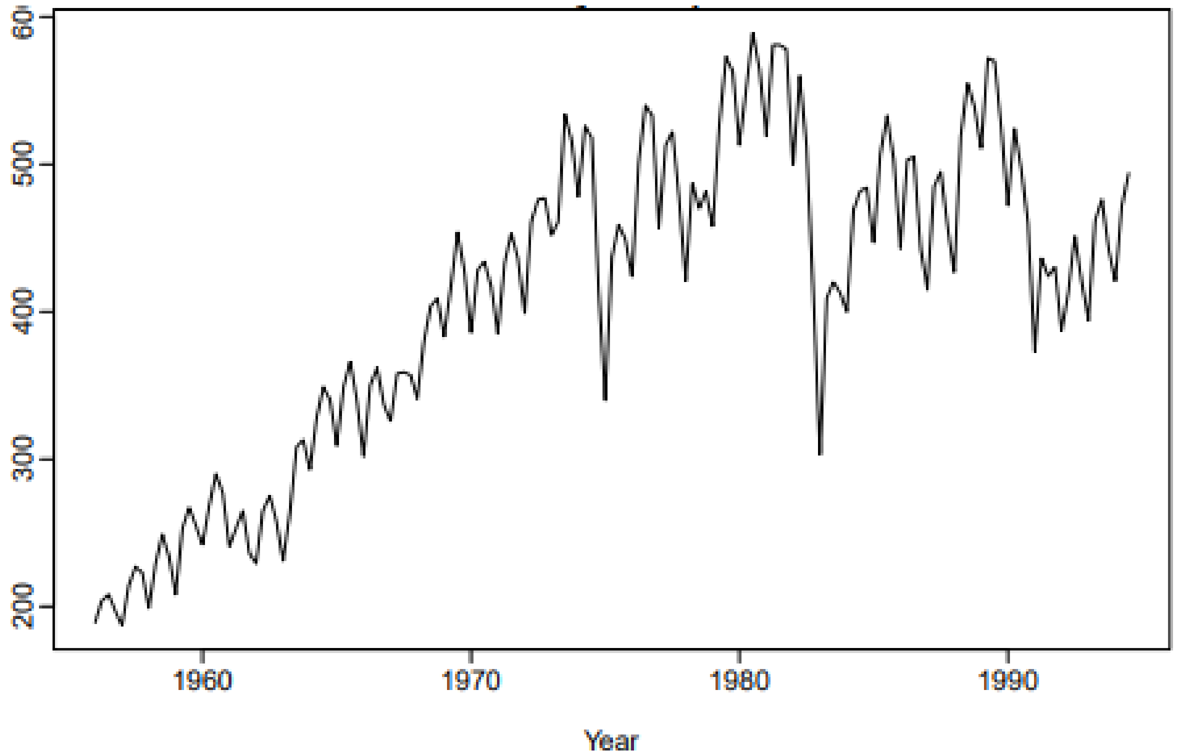
STL decomposition

- Muy versátil y robusto.
- El componente estacional puede cambiar con el tiempo y la tasa de cambio es controlada por el usuario.
- Suavidad de tendencia-ciclo también controlado por el usuario.
- Robusto a los outliers
- Sólo aditivo.

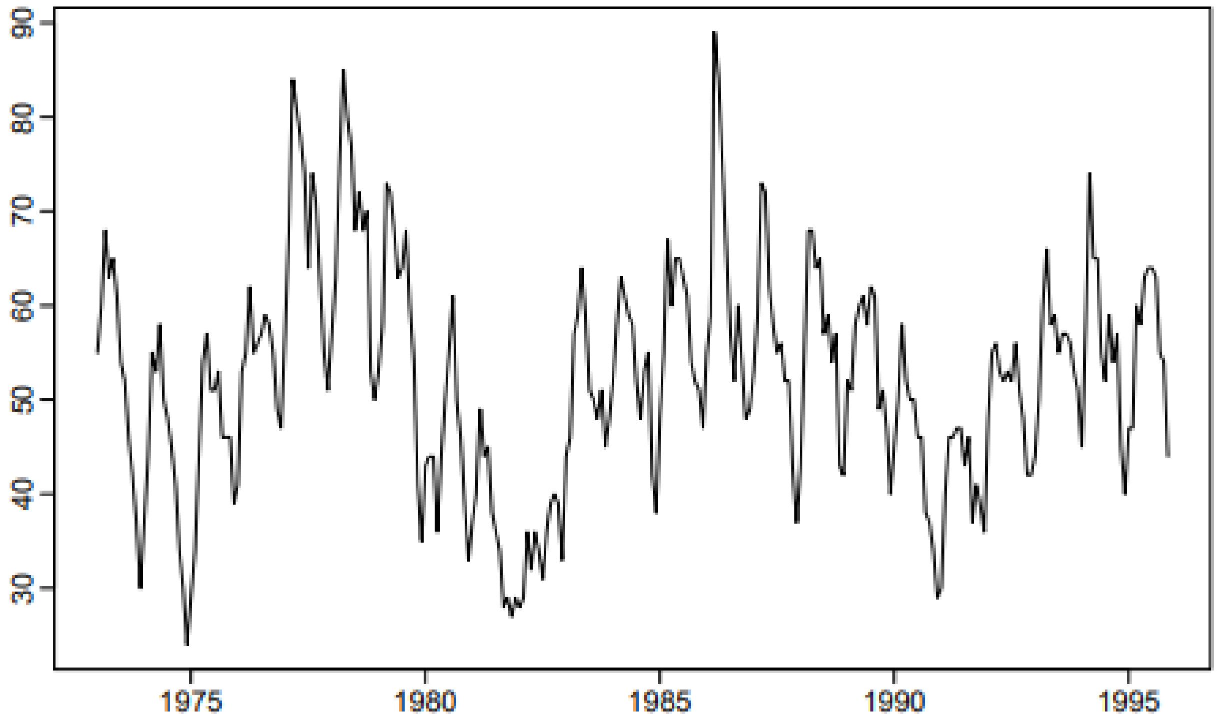
Introducción. Series de Tiempo



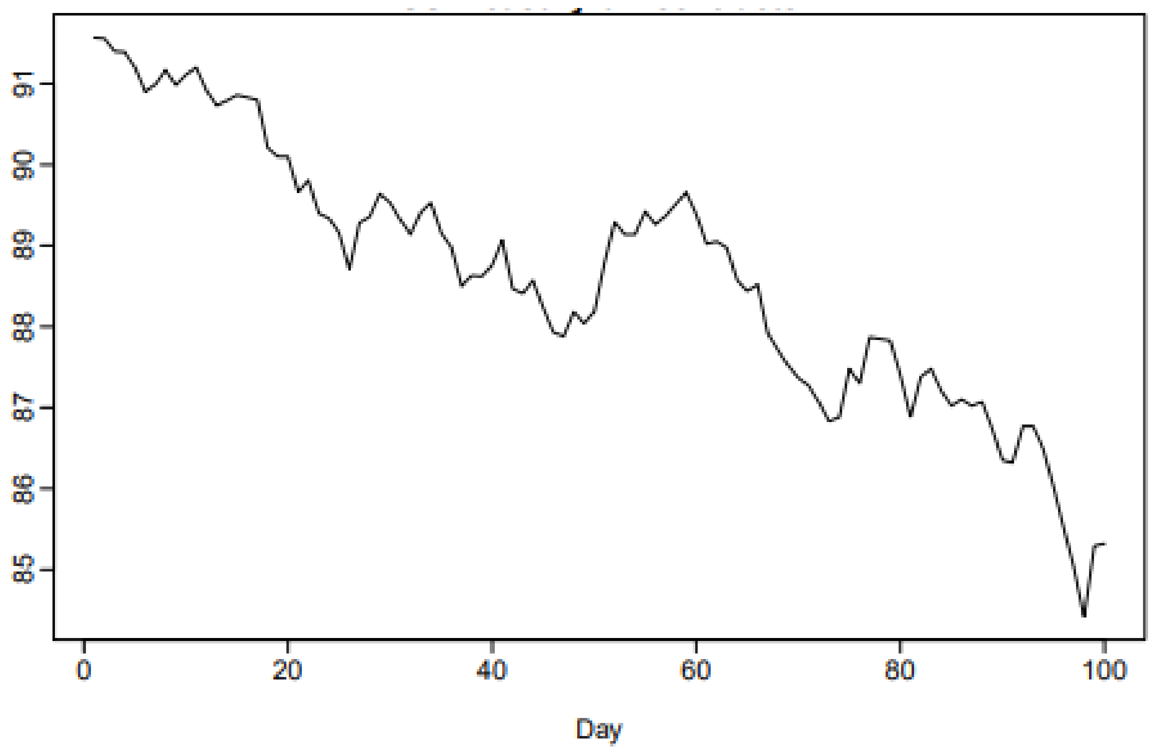
Introducción. Series de Tiempo



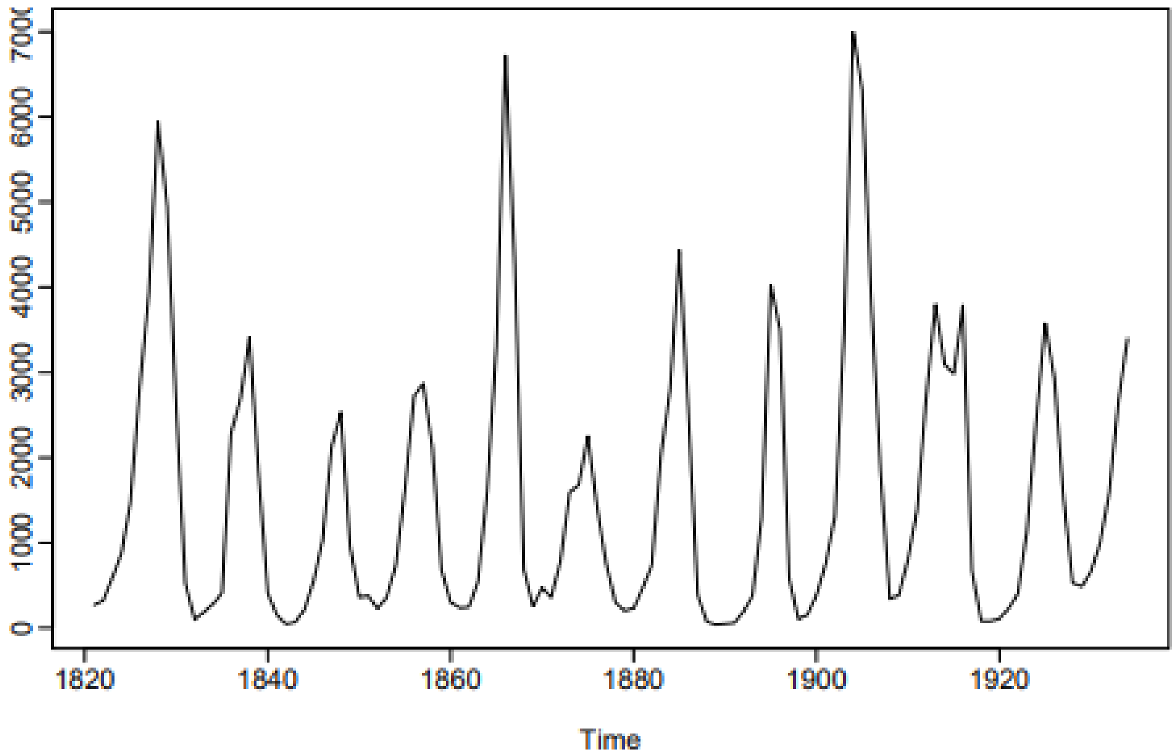
Introducción. Series de Tiempo



Introducción. Series de Tiempo



Introducción. Series de Tiempo



Introducción. Series de Tiempo

- Las series temporales económicas a menudo son analizadas después de calcular sus logaritmos o las variaciones en sus logaritmos.
- Una razón para ello es que muchas series económicas, tales como el producto interior bruto (PIB), presentan un crecimiento que es aproximadamente exponencial, es decir, a largo plazo la serie tiende a crecer a un determinado porcentaje medio anual; si es así, el logaritmo de la serie crece de forma aproximadamente lineal.

Introducción a los datos de series temporales y correlación serial

- Otra razón es que la desviación típica de muchas series temporales económicas es aproximadamente proporcional a su nivel, es decir, la desviación típica se puede expresar correctamente un porcentaje del nivel de las series; si es así, entonces la desviación típica del logaritmo de la serie es aproximadamente constante.
- En cualquier caso, resulta útil transformar las series para que las variaciones en las series transformadas sean variaciones proporcionales (o porcentuales) de la serie original, y esto se logra tomando el logaritmo de las series.

Introducción. Series de Tiempo.

- Covarianza y correlación: medida medida de la relación lineal entre dos variables (Y e X).

- Autocovarianza y autocorrelación: medida de la relación lineal entre los valores rezagados de una series de tiempo y.

Introducción. Series de Tiempo

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

ck es la autocovarianza. Si dividimos por la varianza obtenemos las autocorrelaciones.

$$r_k = c_k / c_0$$

- Las autocovarianzas y autocorrelaciones j-ésimas poblacionales del pueden ser estimadas mediante las autocovarianzas y autocorrelaciones j-ésimas muestrales

Introducción. Series de Tiempo

- r1 indica cómo los valores sucesivos de y se relacionan con cada uno

  otro

- r2 indica cómo se relacionan los valores de y con dos periodos separados

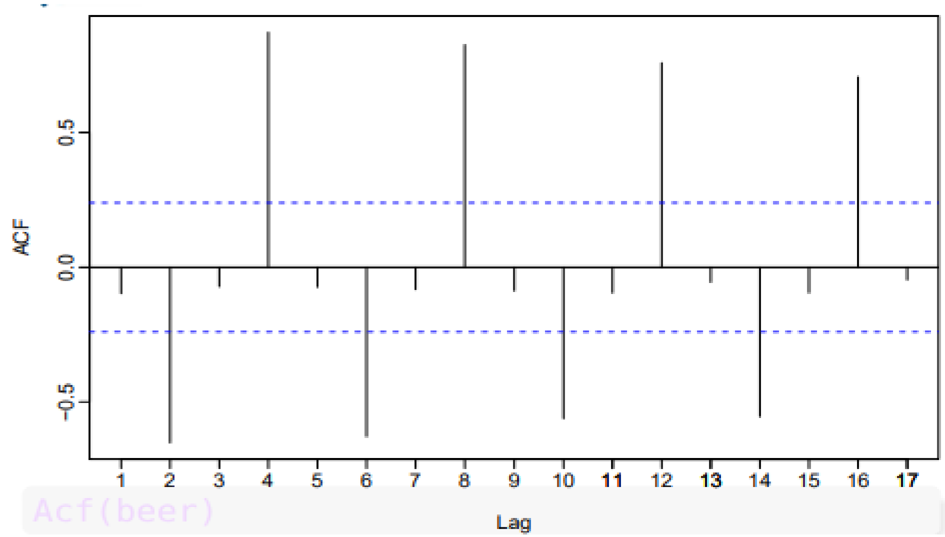
  el uno del otro

- rk es casi lo mismo que la correlación muestral entre yt y yt − k.

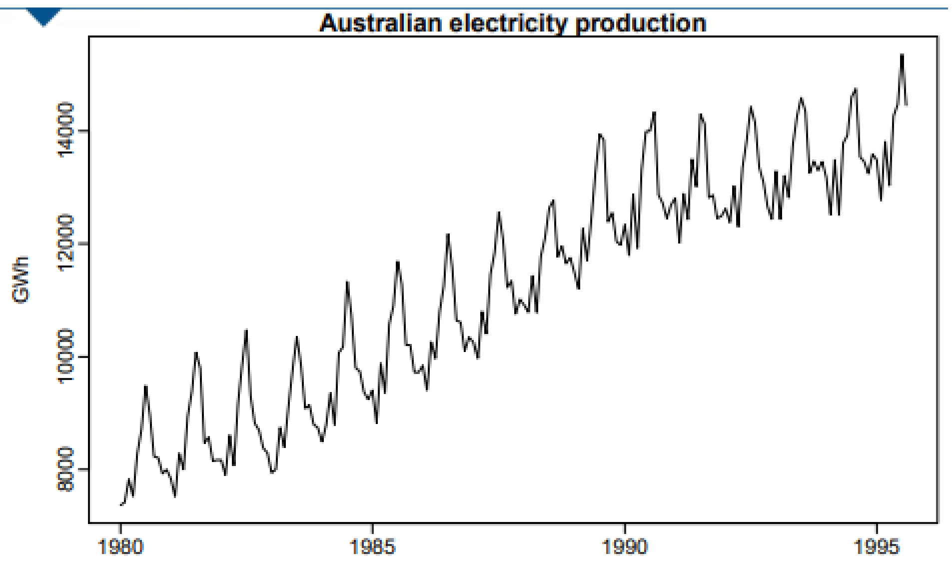
- Las autocorrelaciones en los lags 1,2.. forman la función de autocorrelación.

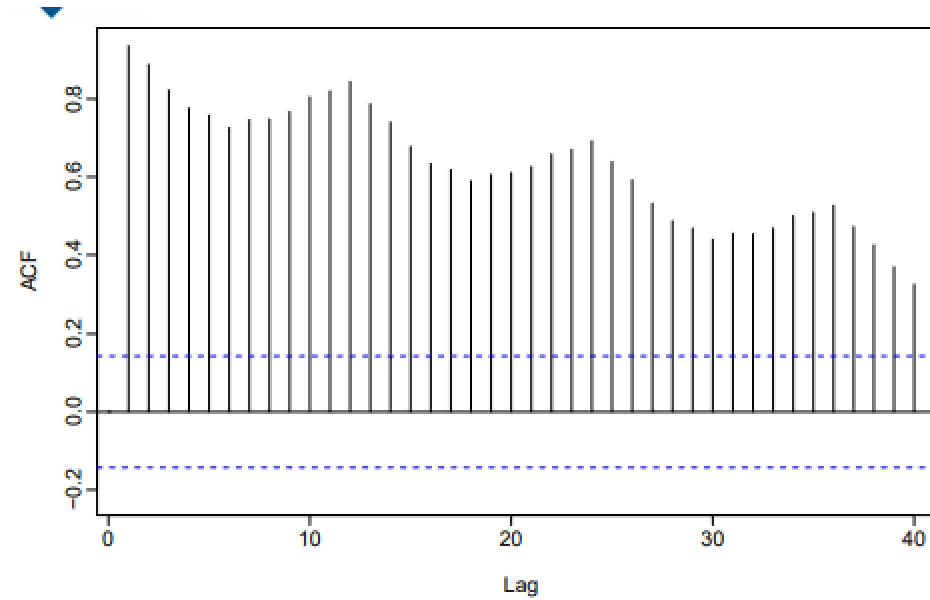
  o ACF.

- El gráfico se conoce como correlograma.



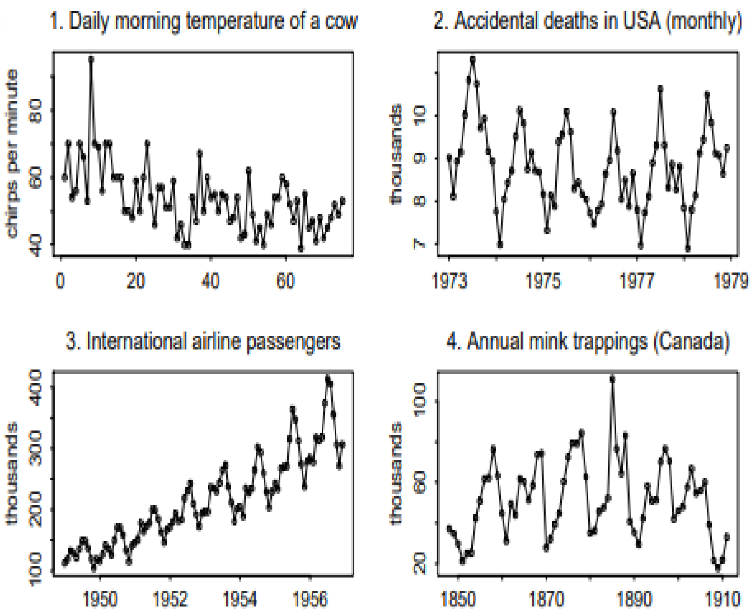
Introducción a los datos de series temporales y correlación serial

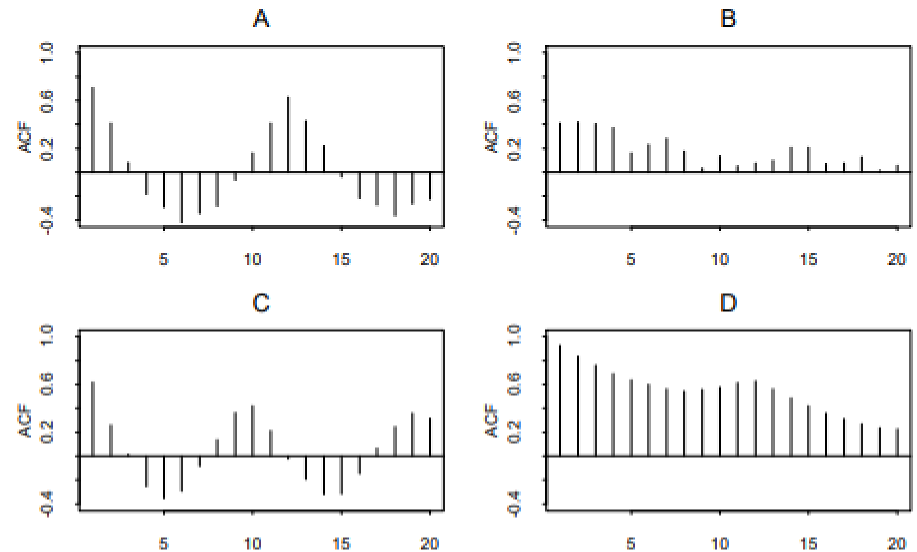




Introducción a los datos de series temporales y correlación serial

¿Cuál es cual?



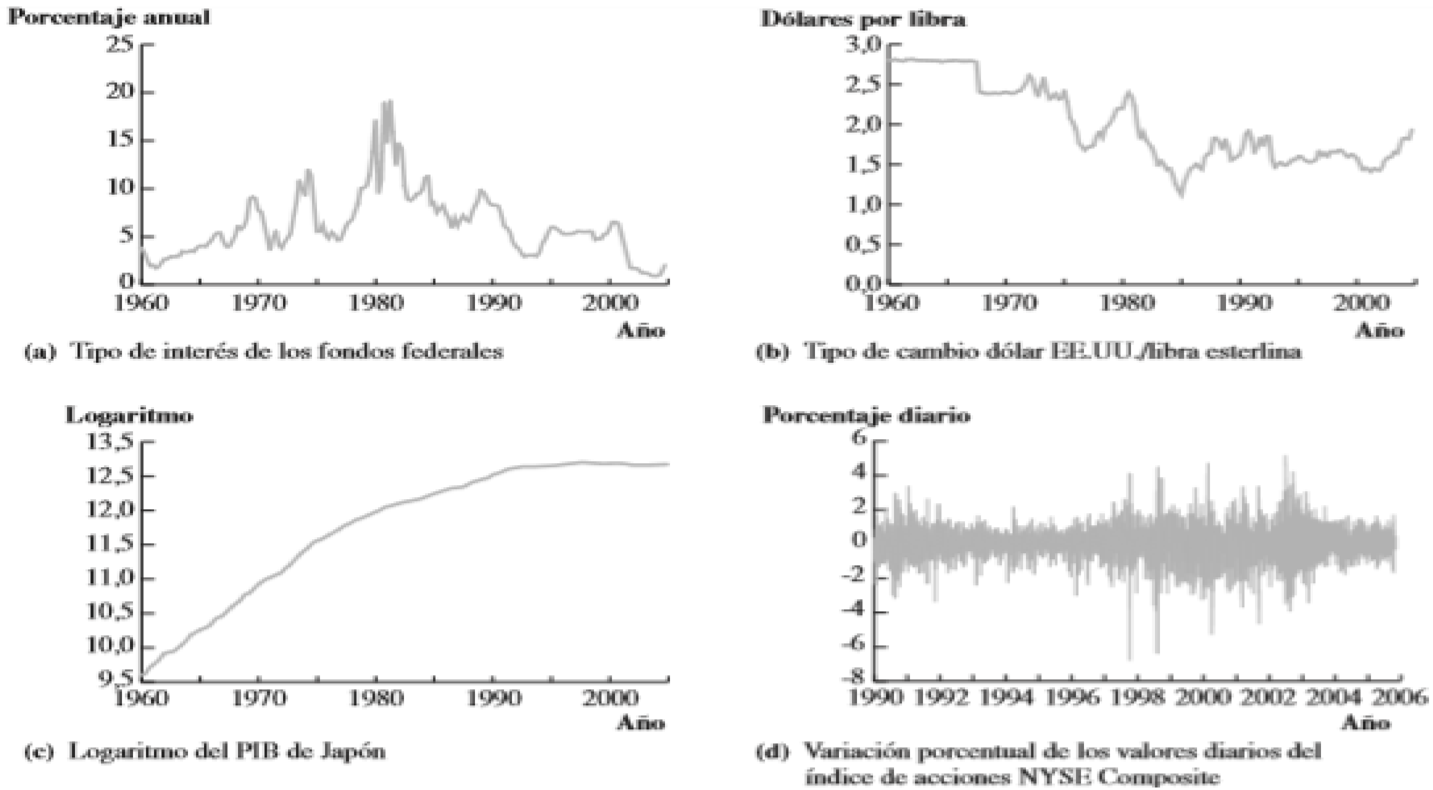


Introducción a los datos de series temporales y correlación serial.

- Otros ejemplos de series temporales económicas
  - el tipo de interés de los fondos federales en EE.UU.
  - el tipo de cambio entre el dólar y la libra esterlina
  - el logaritmo del producto interior bruto de Japón
  - la rentabilidad diaria en el índice del mercado de acciones Standard and Poor’s 500 (S&P 500).

Introducción a los datos de series temporales y correlación serial.

DigitalHouse>
Coding School



(a) Tipo de interés de los fondos federales

(b) Tipo de cambio dólar EE.UU./libra esterlina

(c) Logaritmo del PIB de Japón

(d) Variación porcentual de los valores diarios del índice de acciones NYSE Composite

Forecasting

- Algunos métodos simples de forecasting
  - Average

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

  - Naive

$$\hat{y}_{T+h|T} = y_T$$

  - Seasonal naive

Forecasts: $\hat{y}_{T+h|T} = y_{T+h-km}$ where $m =$ seasonal period and $k = \lfloor (h-1)/m \rfloor + 1$.

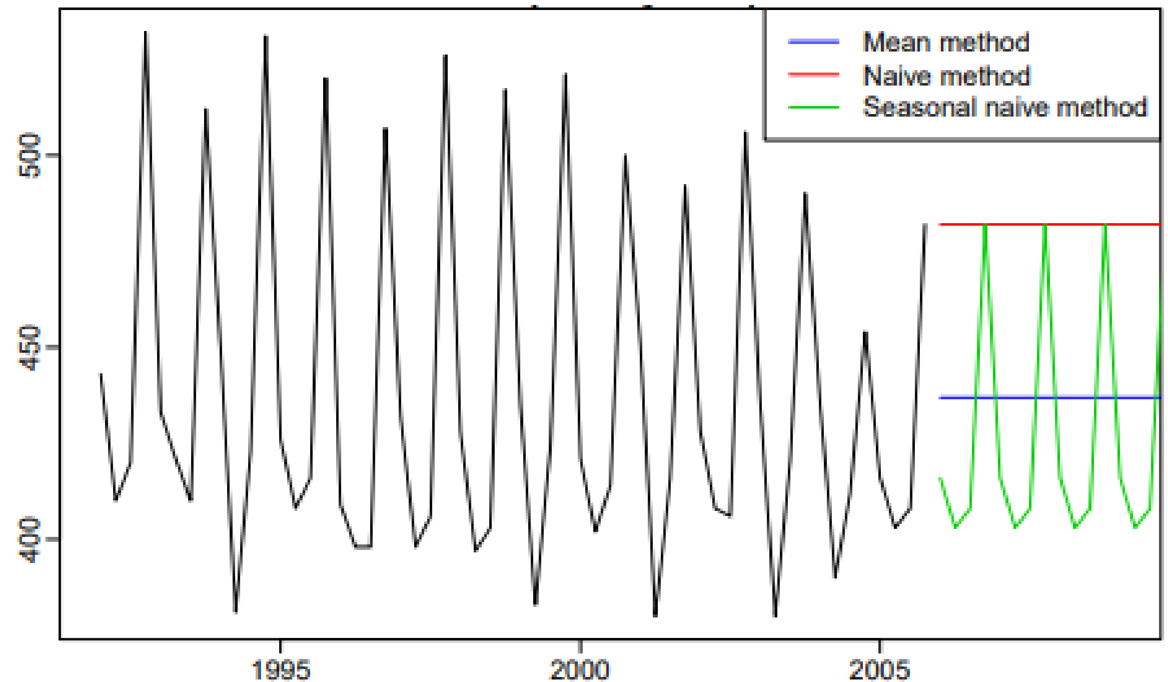
Forecasting

- Algunos métodos simples de forecasting
  - Drift method

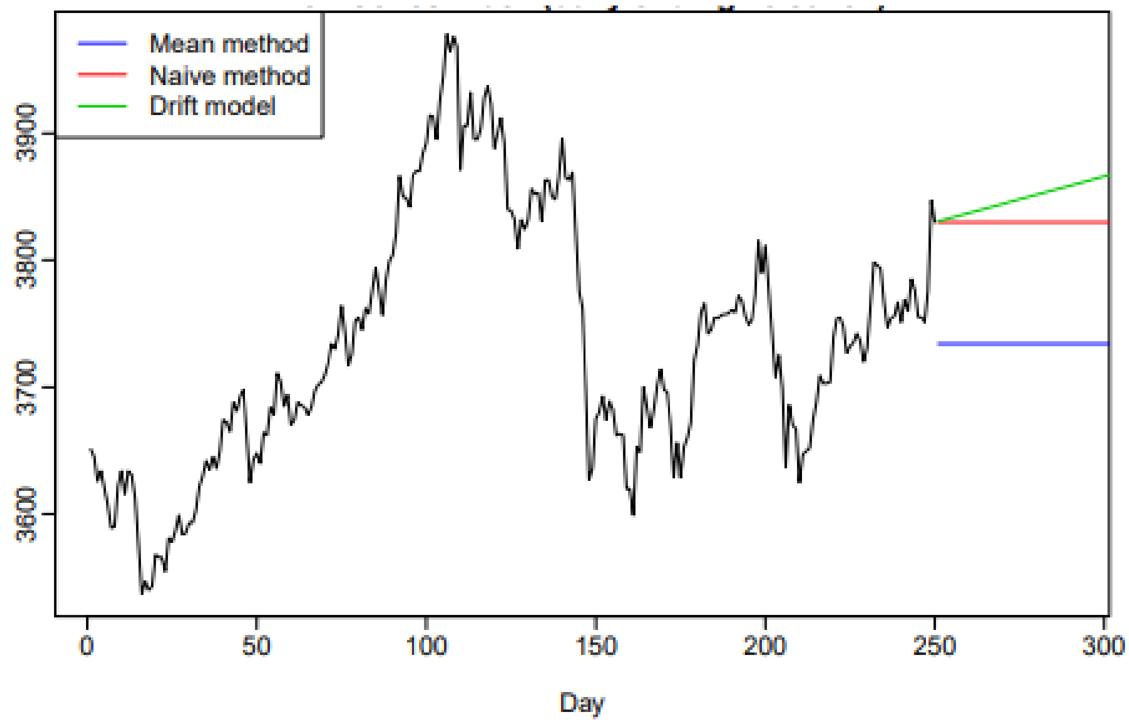
$$\hat{y}_{T+h|T} = y_T + \frac{h}{T-1}\sum_{t=2}^{T}(y_t - y_{t-1})$$

$$= y_T + \frac{h}{T-1}(y_T - y_1).$$

Forecasting



Forecasting



Forecasting

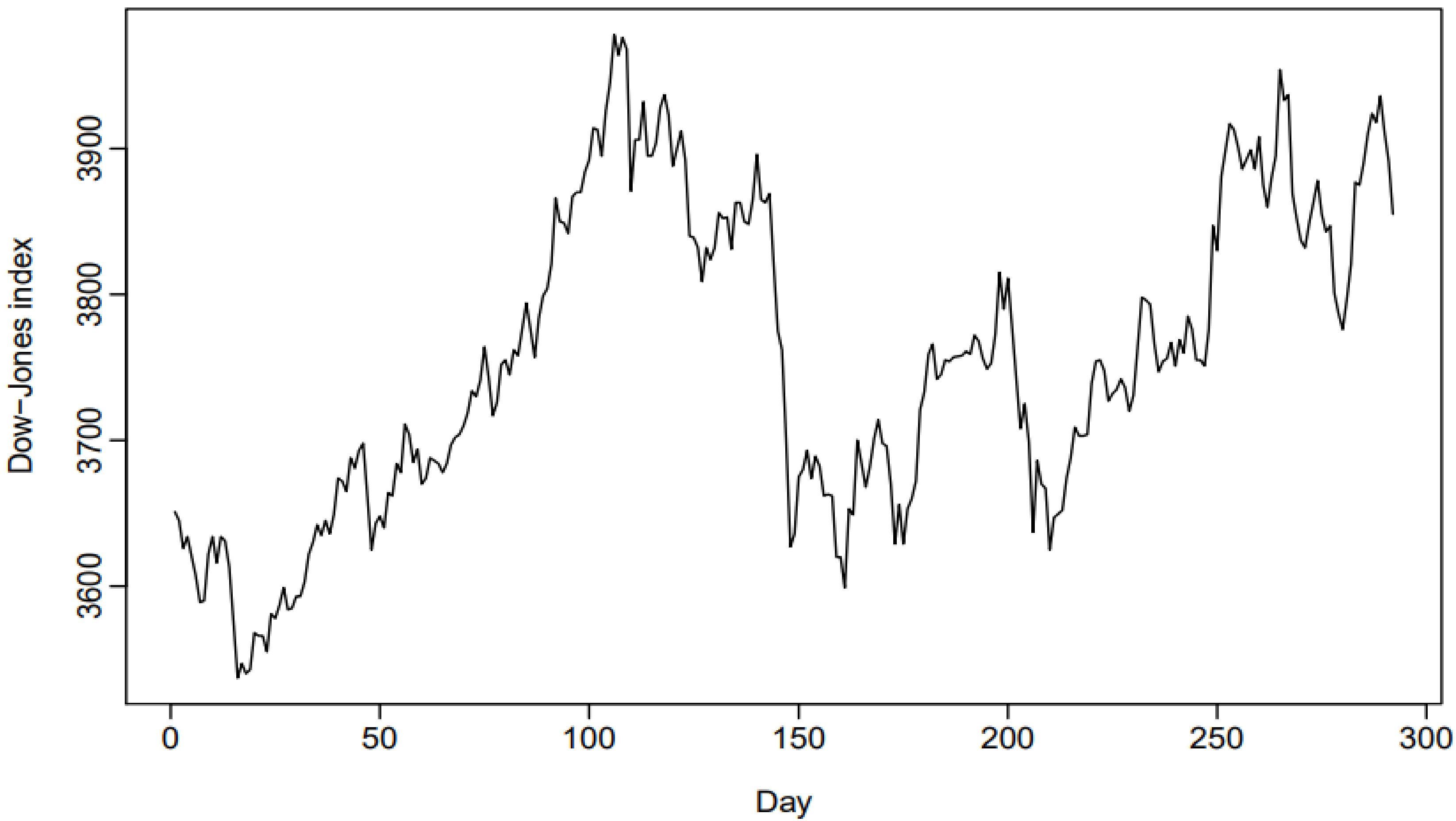
- ¿Qué características nos gustaría ver en los residuos?
- Usemos un modelo ingenuo para el Dow Jones

$$\hat{y}_{t|t-1} = y_{t-1}$$

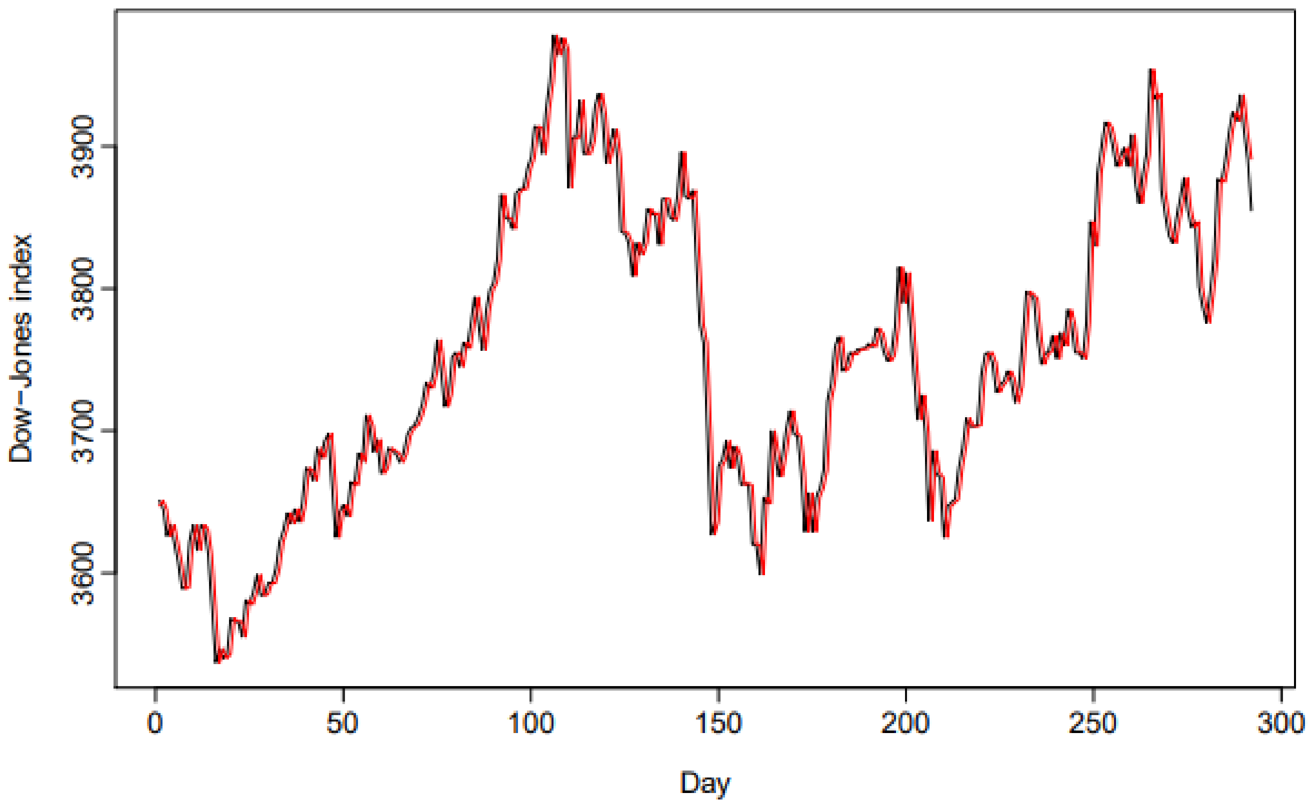
$$e_t = y_t - y_{t-1}$$

Forecasting

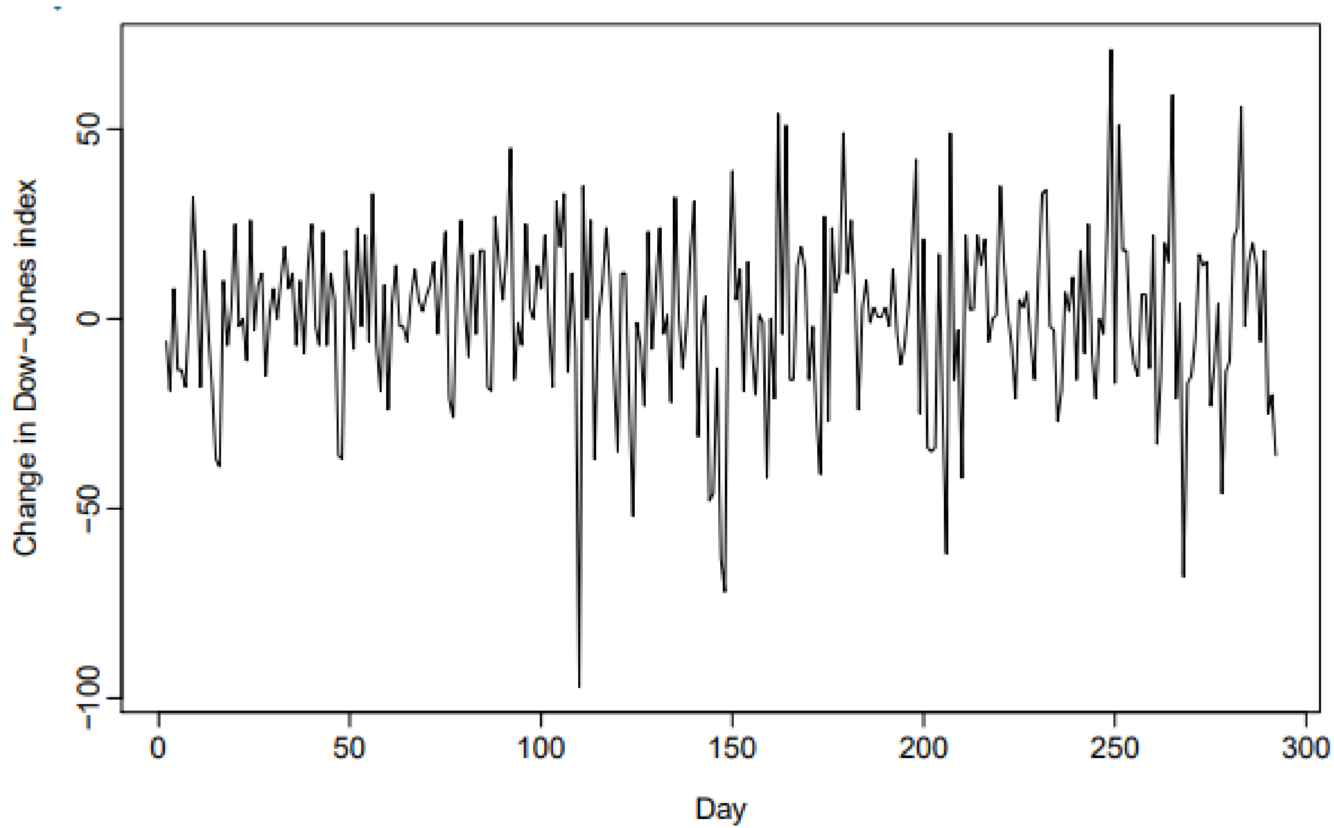
DigitalHouse>
Coding School



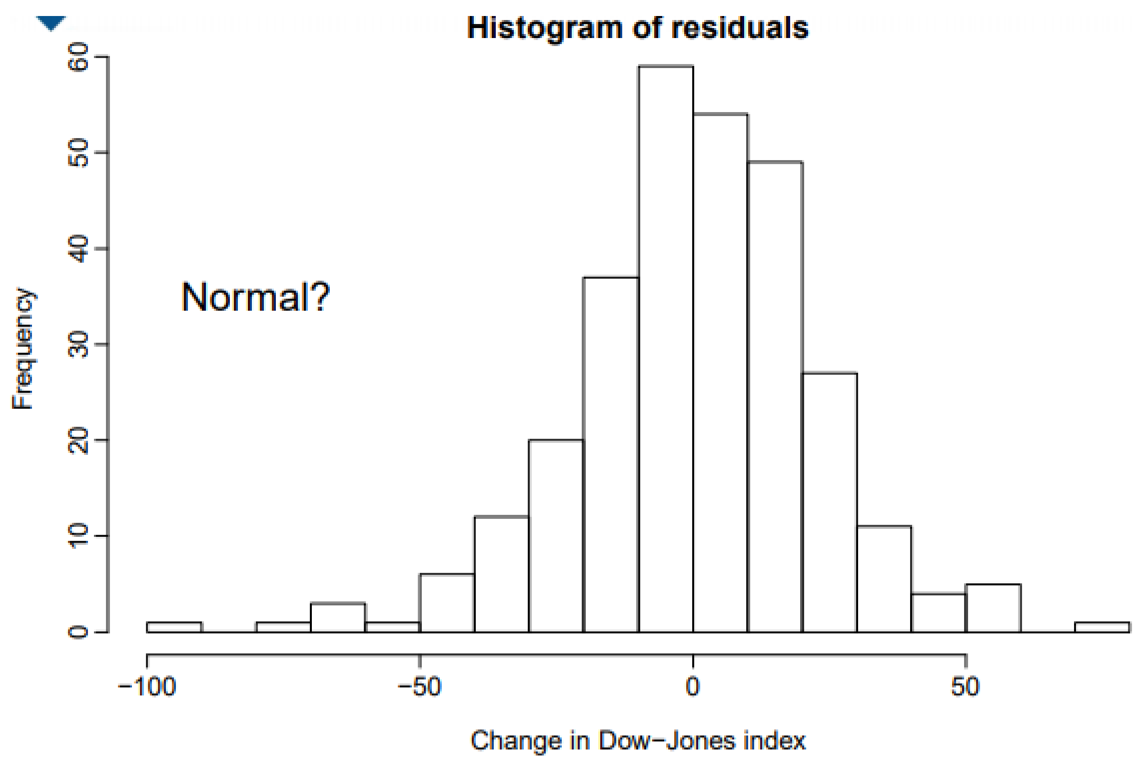
Forecasting



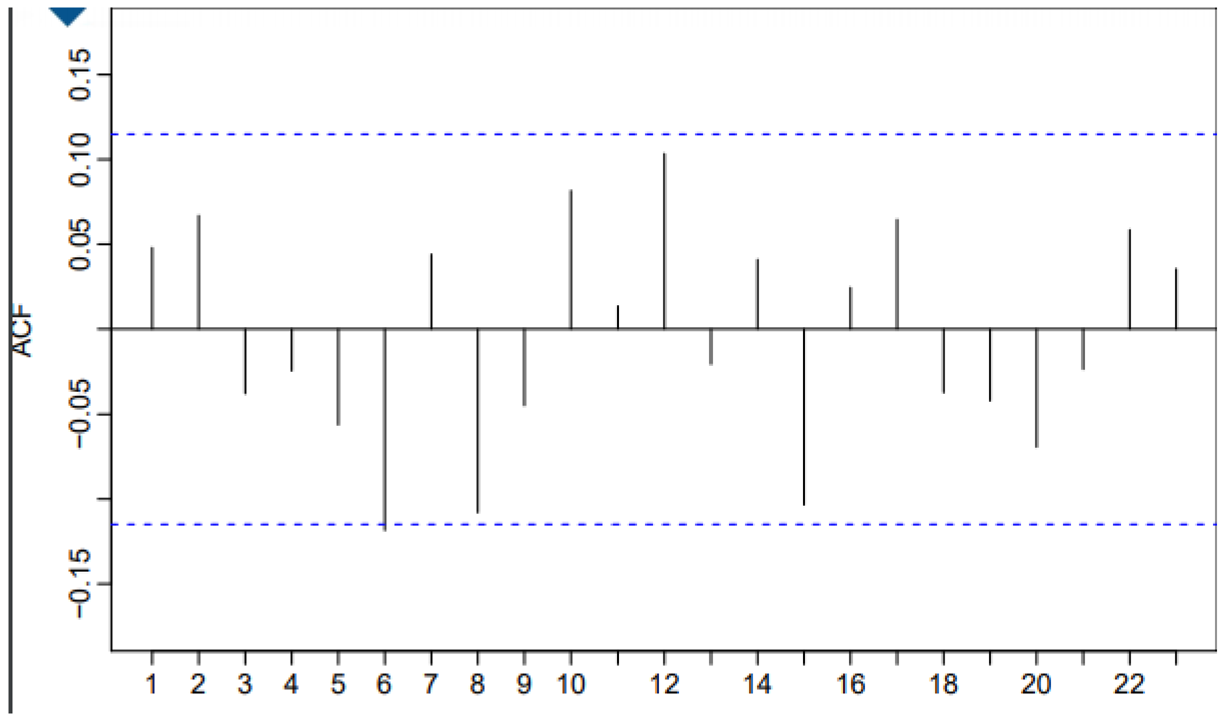
Forecasting



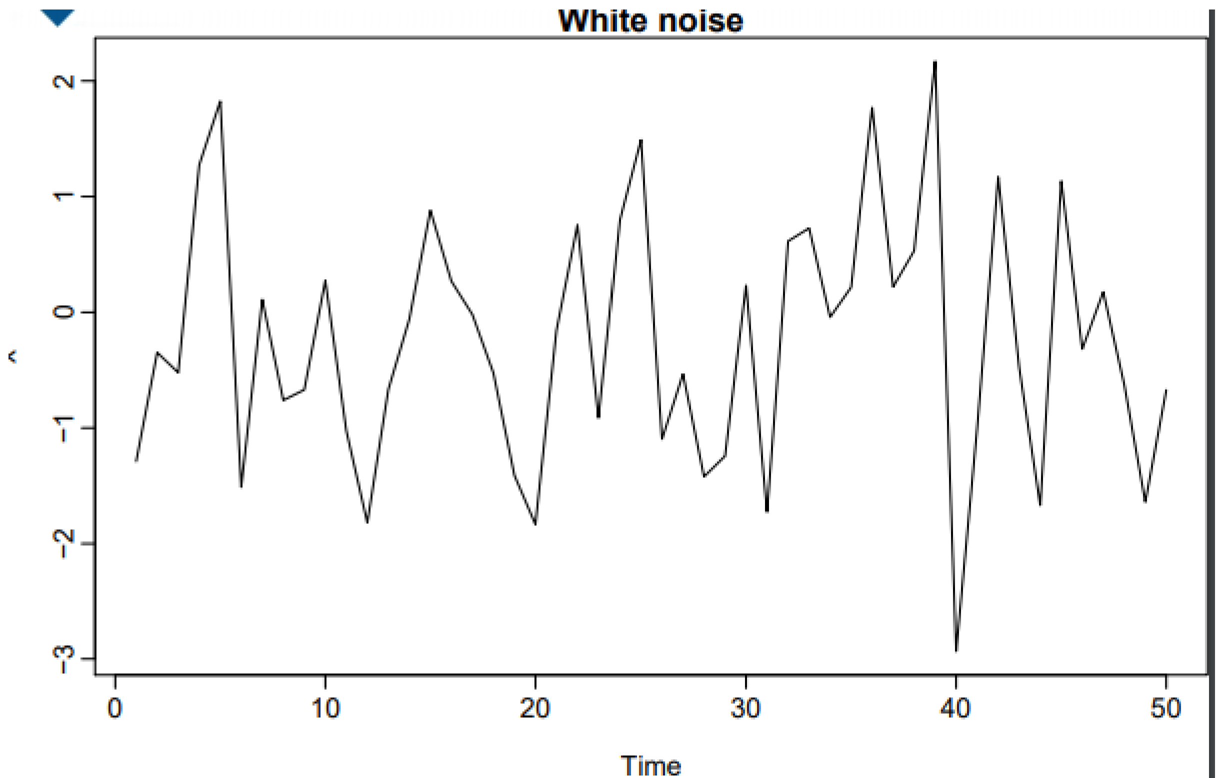
Forecasting



Forecasting

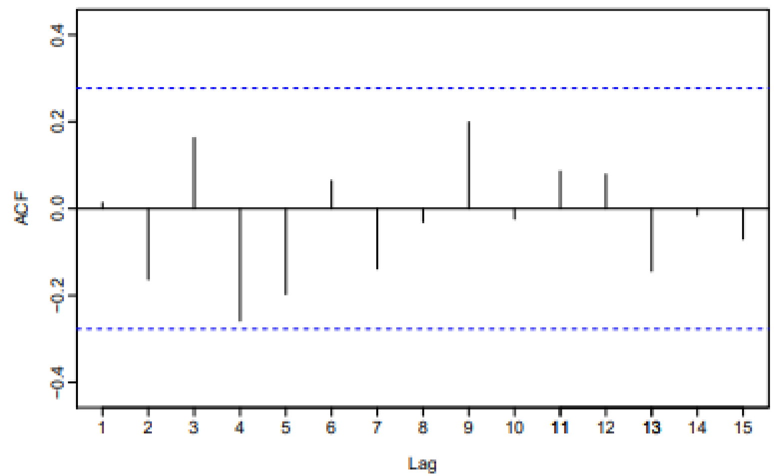


Ruido blanco

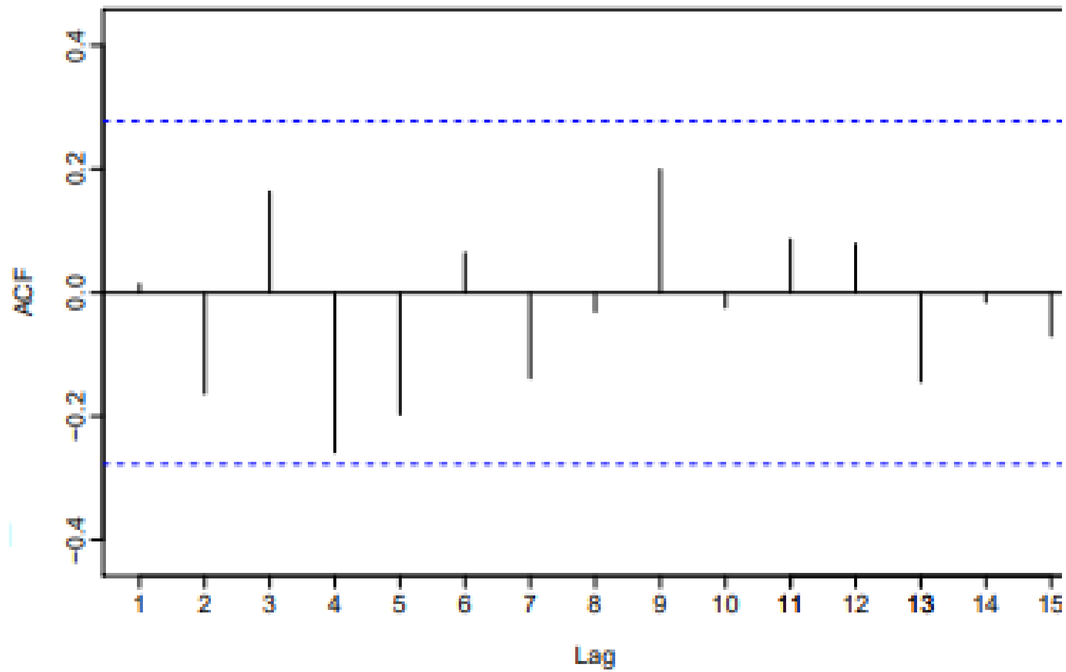


Ruido blanco

$$
\begin{aligned}
r_1 &= 0.013\\
r_2 &= -0.163\\
r_3 &= 0.163\\
r_4 &= -0.259\\
r_5 &= -0.198\\
r_6 &= 0.064\\
r_7 &= -0.139\\
r_8 &= -0.032\\
r_9 &= 0.199\\
r_{10} &= -0.240
\end{aligned}
$$



Ruido blanco

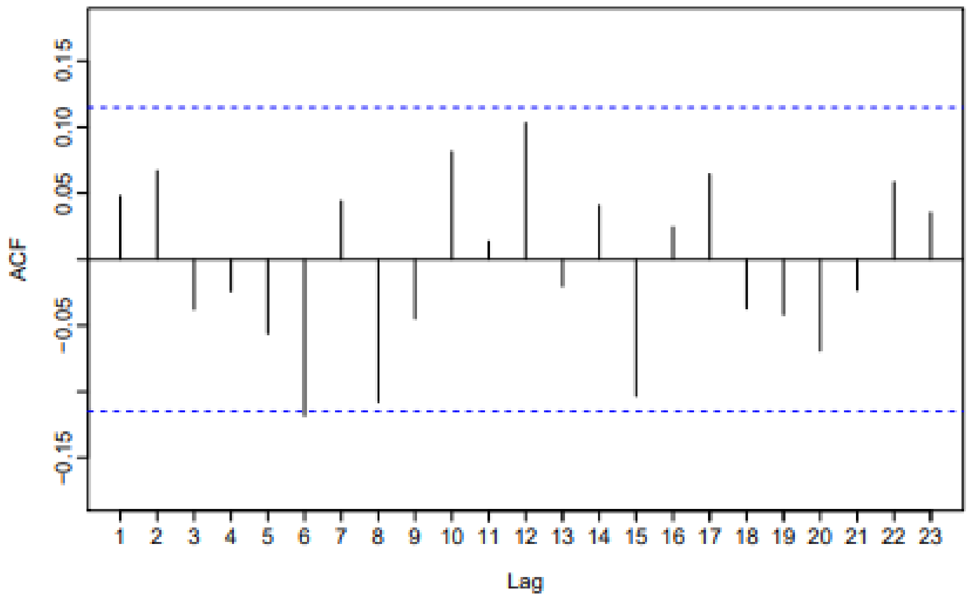


Ruido blanco

- Suponemos que los residuos son ruido blanco (no correlacionado, media cero, varianza constante).
- Si no lo son, entonces hay información que quede en los residuos que deben usarse en el forecast.
- En el ACF de los residuos de un método de forecast esperamos que estos se vean como ruido blanco.

Ruido blanco

Residuos Dow Jones



Parecen ruido blanco pero recuerde que ACF es un problema de multiple testing.

Ruido blanco

Test de Box Pierce

$$Q = T \sum_{k=1}^{h} r_k^2$$

donde h es el máximo rezago considerado y T es el tamaño muestral

Si los datos son ruido blanco entonces Q tiene distribución Chi Cuadrado con (h-K) grados de libertad, donde K es el número de parámetros del modelo. .

Forecast accuracy

$$\text{MAE} = T^{-1} \sum_{t=1}^{T} |y_t - \hat{y}_{t|t-1}|$$

$$\text{MSE} = T^{-1} \sum_{t=1}^{T} (y_t - \hat{y}_{t|t-1})^2 \qquad \text{RMSE} = \sqrt{T^{-1} \sum_{t=1}^{T} (y_t - \hat{y}_{t|t-1})^2}$$

$$\text{MAPE} = 100T^{-1} \sum_{t=1}^{T} |y_t - \hat{y}_{t|t-1}|/|y_t|$$

Forecast accuracy

DigitalHouse> Coding School

- MAE, MSE, RMSE son todos dependientes de la escala.
- MAPE es scale independent pero sólo es razonable si yt es positivo para todo t y si la serie tiene un cero natural.

Mean Absolute Scaled Error

$$\text{MASE} = T^{-1} \sum_{t=1}^{T} |y_t - \hat{y}_{t|t-1}| / Q$$

Donde Q es una medida de la escala de la serie

Para series no estacionales

$$Q = (T-1)^{-1} \sum_{t=2}^{T} |y_t - y_{t-1}|$$

En series con

$$Q = (T-m)^{-1} \sum_{t=m+1}^{T} |y_t - y_{t-m}|$$

Exponential Smoothing

- State space perspective

- Observed data: $y_1, \ldots, y_T$.
- Unobserved state: $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T$.

$$\text{Forecast } \hat{y}_{T+h|T} = \mathrm{E}(y_{T+h}|\boldsymbol{x}_T)$$

La varianza del forecast e $\mathrm{Var}(y_{T+h}|\boldsymbol{x}_T)$

Un intervalo de predicción o “interval forecast” es un rango de valores para y en T+h que tiene alta probabilidad

Simple Exponential Smoothing

Forecast equation	$\hat{y}_{t+h\|t} = \ell_t$
Smoothing equation	$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$

$$\ell_1 = \alpha y_1 + (1-\alpha)\ell_0$$

$$\ell_2 = \alpha y_2 + (1-\alpha)\ell_1 = \alpha y_2 + \alpha(1-\alpha)y_1 + (1-\alpha)^2\ell_0$$

$$\ell_3 = \alpha y_3 + (1-\alpha)\ell_2 = \sum_{j=0}^{2}\alpha(1-\alpha)^j y_{3-j} + (1-\alpha)^3\ell_0$$

$$\vdots$$

$$\ell_t = \sum_{j=0}^{t-1}\alpha(1-\alpha)^j y_{t-j} + (1-\alpha)^t\ell_0$$

Simple Exponential Smoothing

Forecast equation

$$\hat{y}_{t+h|t} = \sum_{j=1}^{t} \alpha(1-\alpha)^{t-j} y_j + (1-\alpha)^t \ell_0, \qquad (0 \le \alpha \le 1)$$

	Weights assigned to observations for:			
Observation	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.8$
y_t	0.2	0.4	0.6	0.8
y_{t-1}	0.16	0.24	0.24	0.16
y_{t-2}	0.128	0.144	0.096	0.032
y_{t-3}	0.1024	0.0864	0.0384	0.0064
y_{t-4}	$(0.2)(0.8)^4$	$(0.4)(0.6)^4$	$(0.6)(0.4)^4$	$(0.8)(0.2)^4$
y_{t-5}	$(0.2)(0.8)^5$	$(0.4)(0.6)^5$	$(0.6)(0.4)^5$	$(0.8)(0.2)^5$

Exponential Smoothing

- Component form

Forecast equation	$\hat{y}_{t+h\|t} = \ell_t$
Smoothing equation	$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$

- State Space form

Observation equation	$y_t = \ell_{t-1} + e_t$
State equation	$\ell_t = \ell_{t-1} + \alpha e_t$

Exponential Smoothing

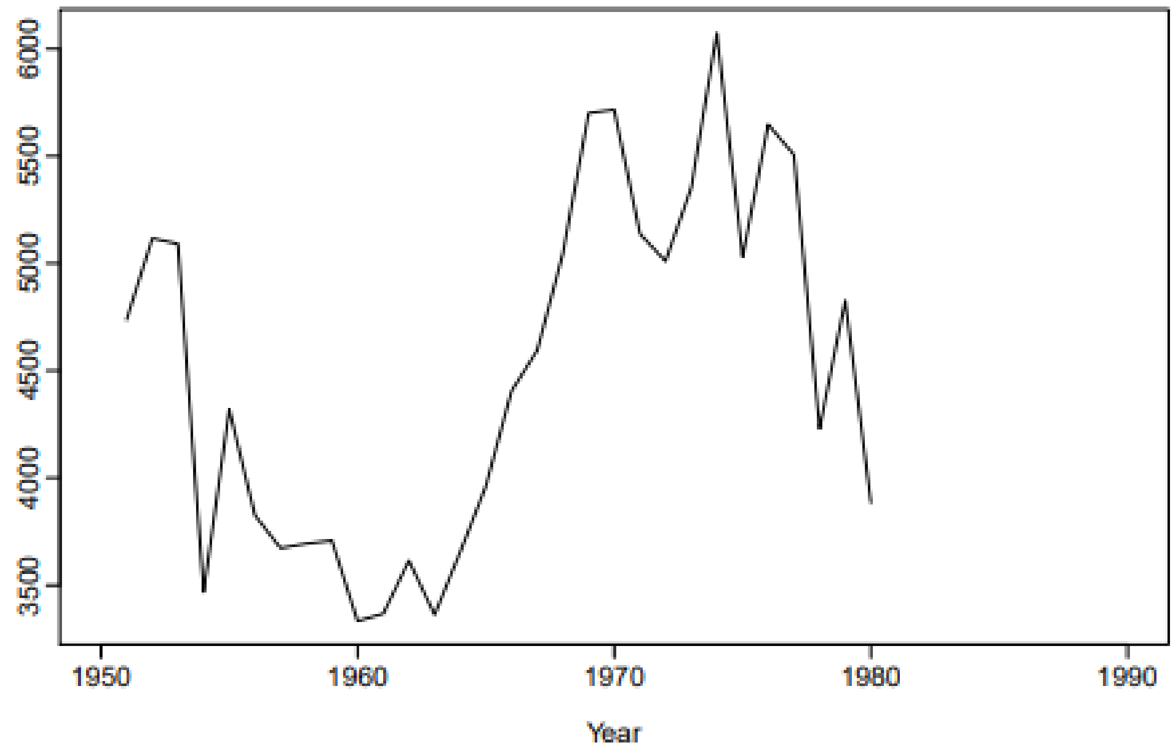
$$e_t = y_t - \ell_{t-1} = y_t - \hat{y}_{t|t-}$$

Es el within sample one step forecast error

ℓ_t es el estado inobservable

Necesitamos elegir valores par α y ℓ_0

Exponential Smoothing



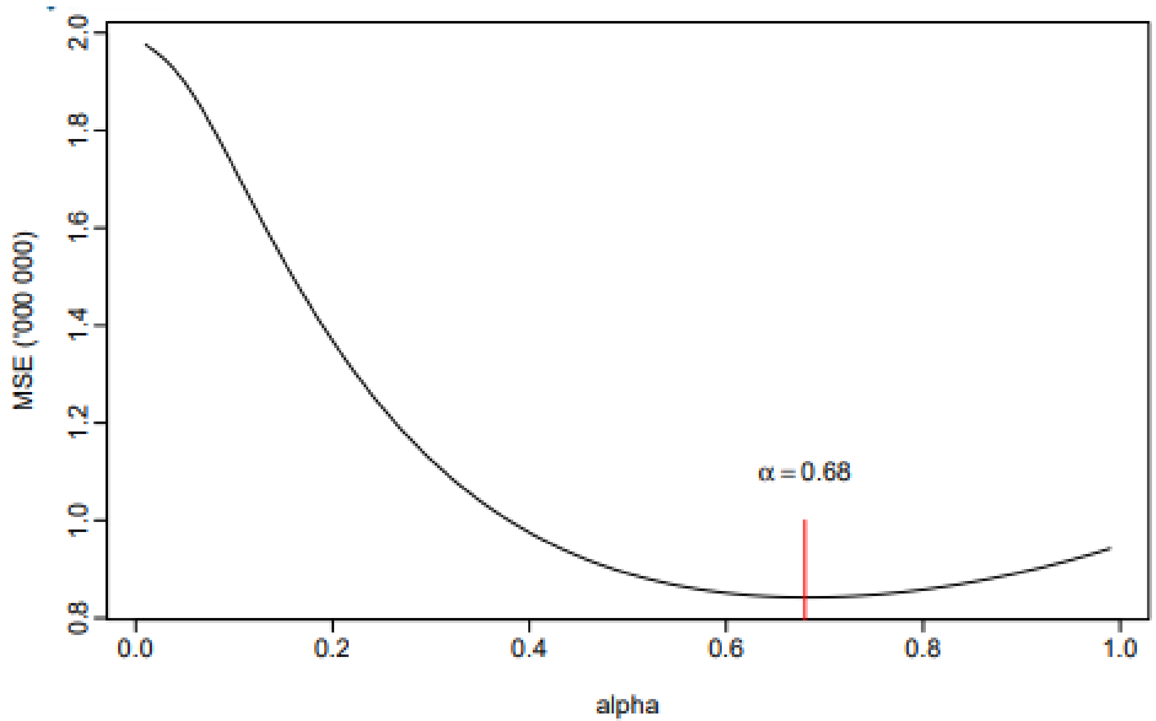
Exponential Smoothing

- Busquemos esos valores minimizando MSE

$$\text{MSE} = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_{t|t-1})^2 = \frac{1}{T}\sum_{t=1}^{T} e_t^2.$$

- A diferencia de regresión no hay solución de forma cerrada usamos optimización numérica.

Exponential Smoothing



Exponential Smoothing

- Multistep forecast

$$\hat{y}_{T+h|T} = \hat{y}_{T+1|T}, \qquad h = 2, 3, \ldots$$

- Una función de pronóstico "plana".

- Recuerde, un pronóstico es una media estimada de un valor futuro así que sin tendencia, sin estacionalidad, y sin otros patrones los pronósticos son constantes.

Exponential Smoothing

- Holt extiende SES para forecasting de series con tendencias.
- Dos parámetros de smoothing alpha y beta, ambos entre 0 y 1.

$$
\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t + hb_t \\
\ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}
\end{aligned}
$$

El componente lt estima el nivel de la serie en el momento t

El componente bt estima la pendiente de la serie en el momento t

Exponential Smoothing

Holt´s linear trend

- Component form

Forecast	$\hat{y}_{t+h\|t} = \ell_t + hb_t$
Level	$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$
Trend	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-}$

- State Space form

Observation equation	$y_t = \ell_{t-1} + b_{t-1} + e_t$
State equations	$\ell_t = \ell_{t-1} + b_{t-1} + \alpha e_t$
	$b_t = b_{t-1} + \beta e_t$

Exponential Smoothing

Holt´s linear trend

$$\beta = \alpha\beta^*$$
$$e_t = y_t - (\ell_{t-1} + b_{t-1}) = y_t - \hat{y}_{t|t-1}$$

Necesito estimar $\alpha, \beta, \ell_0, b_0$

Exponential Smoothing

Multiplicative versión del método de Holt

- State Space form

Forecast equation	$\hat{y}_{t+h\|t} = \ell_t b_t^h$
Observation equation	$y_t = (\ell_{t-1} b_{t-1}) + e_t$
State equations	$\ell_t = \ell_{t-1} b_{t-1} + \alpha e_t$
	$b_t = b_{t-1} + \beta e_t / \ell_{t-1}$

Exponential Smoothing

Multiplicative versión del método de Holt

- Gardner y McKenzie (1985) sugirieron que las tendencias deben ser "amortiguadas" para ser más conservadoras para horizontes de pronóstico más largos.
- Parámetro de amortiguamien $0 < \phi < 1$.
- State Space form

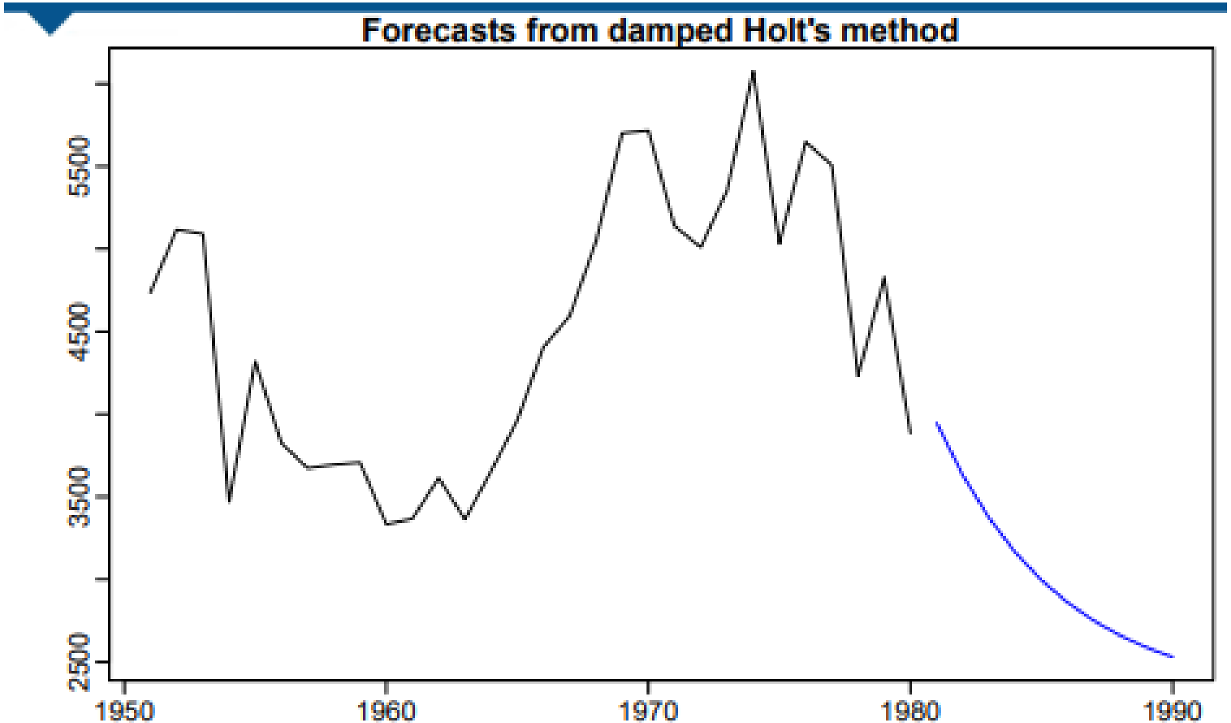
Forecast equation	$\hat{y}_{t+h\|t} = \ell_t + (\phi + \phi^2 + \cdots + \phi^h) b_t$
Observation equation	$y_t = \ell_{t-1} + \phi b_{t-1} + e_t$
State equations	$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha e_t$
	$b_t = \phi b_{t-1} + \beta e_t$

Exponential Smoothing

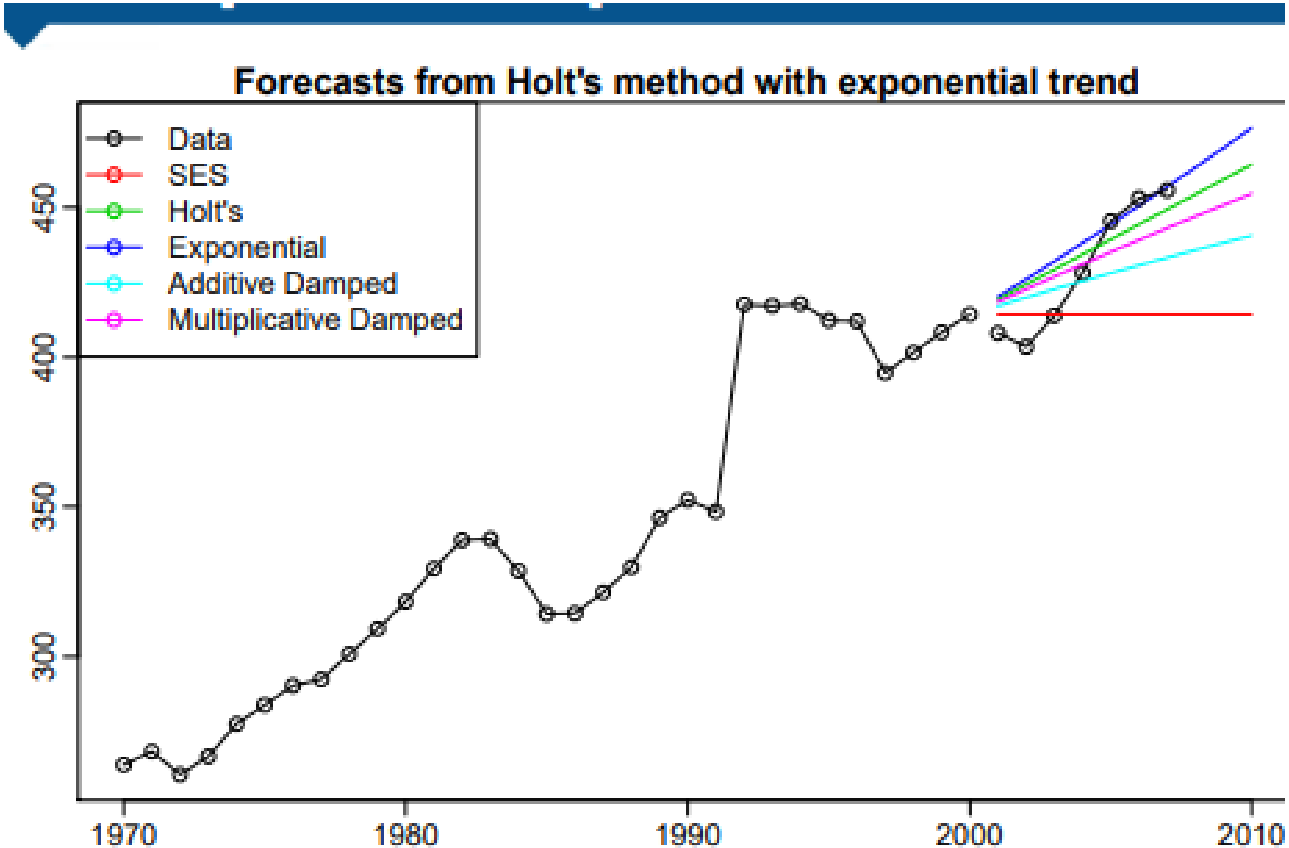
- Si $\phi = 1$ idéntico al Holt´s linear trend
- Short run forecasts con tendencia, a largo plazo constantes

$$h \to \infty, \hat{y}_{T+h|T} \to \ell_T + \phi b_T/(1-\phi)$$

Exponential Smoothing



Exponential Smoothing



Exponential Smoothing

Holt-Winter additive method

- Holt y Winters extendieron el método de Holt para capturar estacionalidad
- Tres ecuaciones de suavizado: una para el nivel, una por tendencia, y una por estacionalidad.
- Parámetros : m es el período de la estacionalidad

$$0 \leq \alpha \leq 1, \ 0 \leq \beta^* \leq 1, 0 \leq \gamma \leq 1 - \alpha$$

Exponential Smoothing

Holt-Winter additive method

– State Space form

$$\hat{y}_{t+h|t} = \ell_t + hb_t + s_{t-m+h_m^+} \qquad h_m^+ = \lfloor (h-1) \mod m \rfloor + 1$$

$$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + e_t$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha e_t$$

$$b_t = b_{t-1} + \beta e_t$$

$$s_t = s_{t-m} + \gamma e_t.$$

Exponential Smoothing

Holt-Winter damped method

- State Space form

$$
\begin{aligned}
y_t &= (\ell_{t-1} + \phi b_{t-1}) s_{t-m} + e_t \\
\ell_t &= \ell_{t-1} + \phi b_{t-1} + \alpha e_t / s_{t-m} \\
b_t &= \phi b_{t-1} + \beta e_t / s_{t-m} \\
s_t &= s_{t-m} + \gamma e_t / (\ell_{t-1} + \phi b_{t-1}).
\end{aligned}
$$

Model selections

- Criterio de Información de Akaike

$$\mathrm{AIC} = -2\log(\mathrm{Likelihood}) + 2p$$

donde p es el número de parámetros estimados

  - Buscamos minimizar el AIC.

Model selections

- Criterio de Información de Akaike corregido por small sample bias

$$\mathrm{AIC_C} = \mathrm{AIC} + \frac{2(p+1)(p+2)}{n-p}$$

donde p es el número de parámetros estimados

  - Buscamos minimizar el AIC.

Model selections

- Criterio de Información Bayesiano (Schwartz)

$$\mathrm{BIC} = \mathrm{AIC} + p(\log(n) - 2)$$

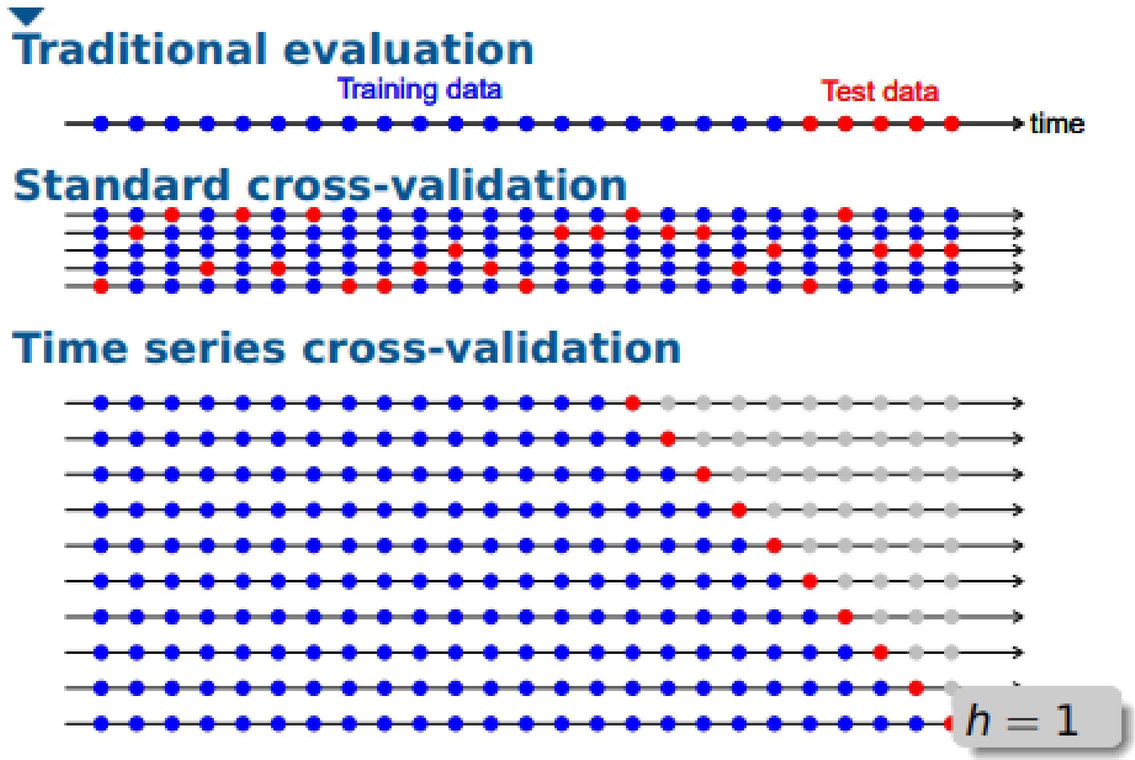
Model selections

- Aplique cada modelo que sea apropiado para el datos. Optimizar parámetros y valores iniciales utilizando MLE (o algún otro criterio).
- Seleccione el mejor método utilizando AICc:
- Producir pronósticos utilizando el mejor método.
- Obtener intervalos de predicción utilizando el modelo de espacio de estado subyacente.

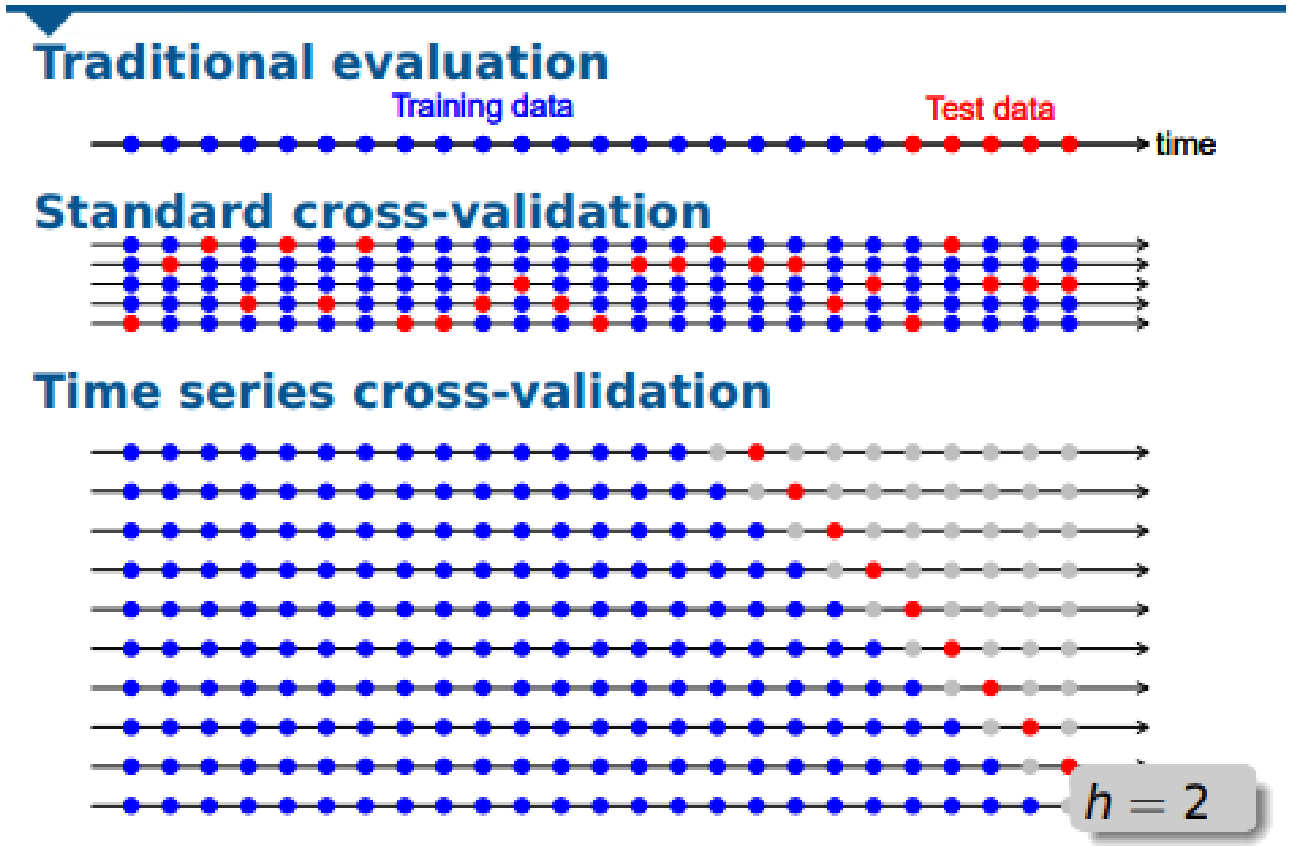
Validación del modelo

Cross Validation

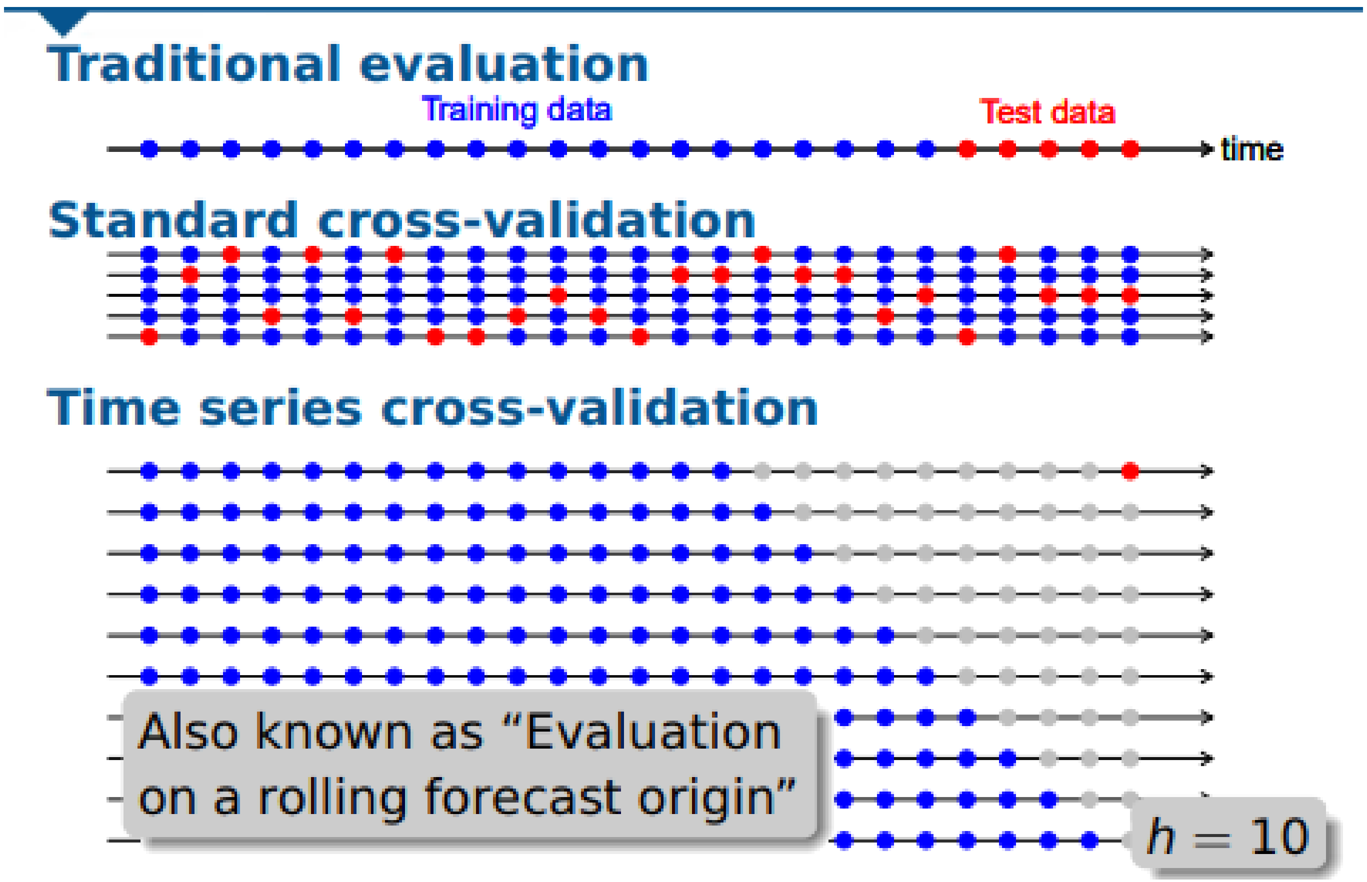
DigitalHouse>
Coding School



Cross Validation



Cross Validation



Anexo: más sobre smoothing

Exponential Smoothing

Taylor multiplicative dampening

$$\hat{y}_{t+h|t} = \ell_t b_t^{(\phi+\phi^2+\cdots+\phi^h)}$$
$$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} b_{t-1}^{\phi})$$
$$b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*) b_{t-1}^{\phi}$$

- Si $\phi = 1$ exponential trend method
- Los forecast convergen a $\ell_T + b_T^{\phi/(1-\phi)}$ cuando h se hace arbitrariamente grande.

Exponential Smoothing

Holt-Winter additive method

- Métodos de suavizado exponencial.
  - Algoritmos que devuelven las previsiones puntuales.
- Innovations State space models con innovaciones
  - Genera los mismos pronósticos puntuales pero también puede generar intervalos de forecast.
- Un proceso estocástico generador de datos que puede generar una distribución completa de forecasts.
- Permite una selección de modelo adecuada.

Exponential Smoothing

Trend Component		Seasonal Component N (None)	Seasonal Component A (Additive)	Seasonal Component M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

General notation E T S : **E**xponen**T**ial **S**moothing

Error **T**rend **S**easonal

Exponential Smoothing

Trend Component		Seasonal Component N (None)	A (Additive)	M (Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
A_d	(Additive damped)	A_d,N	A_d,A	A_d,M
M	(Multiplicative)	M,N	M,A	M,M
M_d	(Multiplicative damped)	M_d,N	M_d,A	M_d,M

N,N: Simple exponential smoothing
A,N: Holt's linear method
A_d,N: Additive damped trend method
M,N: Exponential trend method
M_d,N: Multiplicative damped trend method
A,A: Additive Holt-Winters' method
A,M: Multiplicative Holt-Winters' method

Exponential Smoothing

Holt-Winter additive method

- Todos los modelos ETS pueden ser escritos en innovations state space form.
- Las versiones aditivas y multiplicativas dan los mismos point forecasts pero diferentes intervalos de predicción.

Observation equation	$y_t = \ell_{t-1} + \varepsilon_t,$
State equation	$\ell_t = \ell_{t-1} + \alpha\varepsilon_t$

$$e_t = y_t - \hat{y}_{t|t-1} = \varepsilon_t$$

$$\varepsilon_t \sim \text{NID}(0, \sigma^2)$$

Innovations State Space models

$$\mathbf{x}_t = (\ell_t, b_t, s_t, s_{t-1}, \ldots, s_{t-m+1})$$

$$\varepsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$$

$$\begin{aligned} y_t &= \underbrace{h(\mathbf{x}_{t-1})}_{\mu_t} + \underbrace{k(\mathbf{x}_{t-1})\varepsilon_t}_{e_t} \\ \mathbf{x}_t &= f(\mathbf{x}_{t-1}) + g(\mathbf{x}_{t-1})\varepsilon_t \end{aligned}$$

Innovations State Space models

- Errores aditivos

$$k(x) = 1. \qquad y_t = \mu_t + \varepsilon_t$$

- Errores multiplicativos

$$k(\mathbf{x}_{t-1}) = \mu_t. \qquad y_t = \mu_t(1 + \varepsilon_t).$$

$\varepsilon_t = (y_t - \mu_t)/\mu_t$ is relative error.