



Análisis de Series de Tiempo

Carrera de Especialización en Inteligencia Artificial

Clase ⁶

Ing. Magdalena Bouza, Ing. Carlos German Carreño Romano

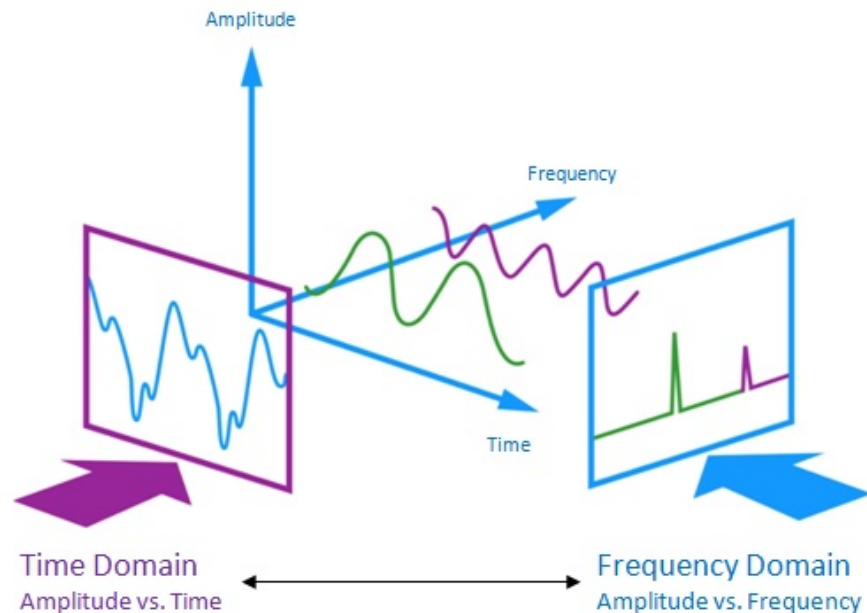
Contenido

- Análisis espectral
- Análisis de intervenciones
- Análisis de outliers
- Heteroscedasticidad
- Revisión bibliográfica

Introducción al análisis espectral

Introducción al análisis espectral

- Series periódicas
- Análisis en el dominio del tiempo
- Densidad espectral de potencia
- Periodograma
- Análisis en el dominio de frecuencias



Ref.: [1] *Time Series Analysis*, J.D. Cryer ; [2] *Introduction to Time Series Modeling*, G. Kitagawa

Series periódicas

Consideremos un proceso dado por la expresión

$$Y_t = R \cos(2\pi ft + \Phi)$$

Conceptos clave en este tipo de series son Amplitud, Frecuencia y Fase. Usando trigonometría podemos pasar a una expresión distinta:

$$Y_t = R \cos(2\pi ft + \Phi) = A \cos(2\pi ft) + B \sin(2\pi ft)$$

$$A = R \cos(\Phi) \quad B = -R \sin(\Phi)$$

$$R = \sqrt{A^2 + B^2} \quad \Phi = \arctan(-B/A)$$

Si tenemos la frecuencia, entonces podemos ajustar los parámetros A y B a partir de, por ejemplo, cuadrados mínimos.

Una generalización de lo anterior nos permite definir un proceso 'Y' a partir de una combinación lineal de senos y cosenos de distintas frecuencias, a partir de la expresión:

$$Y_t = A_0 + \sum_{j=1}^m (A_j \cos(2\pi f_j t) + B_j \sin(2\pi f_j t))$$

Nuevamente, podemos hacer estimaciones por cuadrados mínimos para encontrar los coeficientes A_j y B_j .

Fourier

Suponiendo que n es impar, las frecuencias de la forma $1/n, 2/n, \dots, k/n$ se denominan frecuencias de Fourier. La estimación por cuadrados mínimos resulta:

$$\hat{A}_0 = \bar{Y}$$
$$\hat{A}_j = \frac{2}{n} \sum_{t=1}^n Y_t \cos(2\pi t j / n) \quad \hat{B}_j = \frac{2}{n} \sum_{t=1}^n Y_t \sin(2\pi t j / n)$$

El resultado de la serie de suma de cosenos y senos es general. Esto es, cualquier serie de cualquier longitud n , sea determinística o estocástica y con o sin periodicidades verdaderas puede ajustarse perfectamente por este modelo. Luego habrá n parámetros a estimar para ajustar la serie de largo n .

Periodograma

Para muestras de tamaño impar ($n=2k+1$), el periodograma se define como:

$$I(j/n) = \frac{n}{2} (\hat{A}_j^2 + \hat{B}_j^2)$$

donde j/n es la frecuencia para $j=1, 2, \dots, k$.

Si la muestra tiene tamaño par ($n=2k$), en la frecuencia de corte $f=k/n=1/2$ resulta:

$$I(1/2) = n\hat{A}_k^2$$

Interpretación: Como el periodograma es proporcional a la suma de cuadrados de los coeficientes p/c/frecuencia, la altura muestra la intensidad relativa del par sin-cos a distintas frecs. en el comportamiento general de la serie.

Introducción al análisis espectral

example 1a

N=1000

f=50

T=1/f

t=np.arange(N)

Phi=np.random.normal(0,1,N)

X=sin(2*pi*f*t/N)

Y=sin(2*pi*f*t/N + Phi)

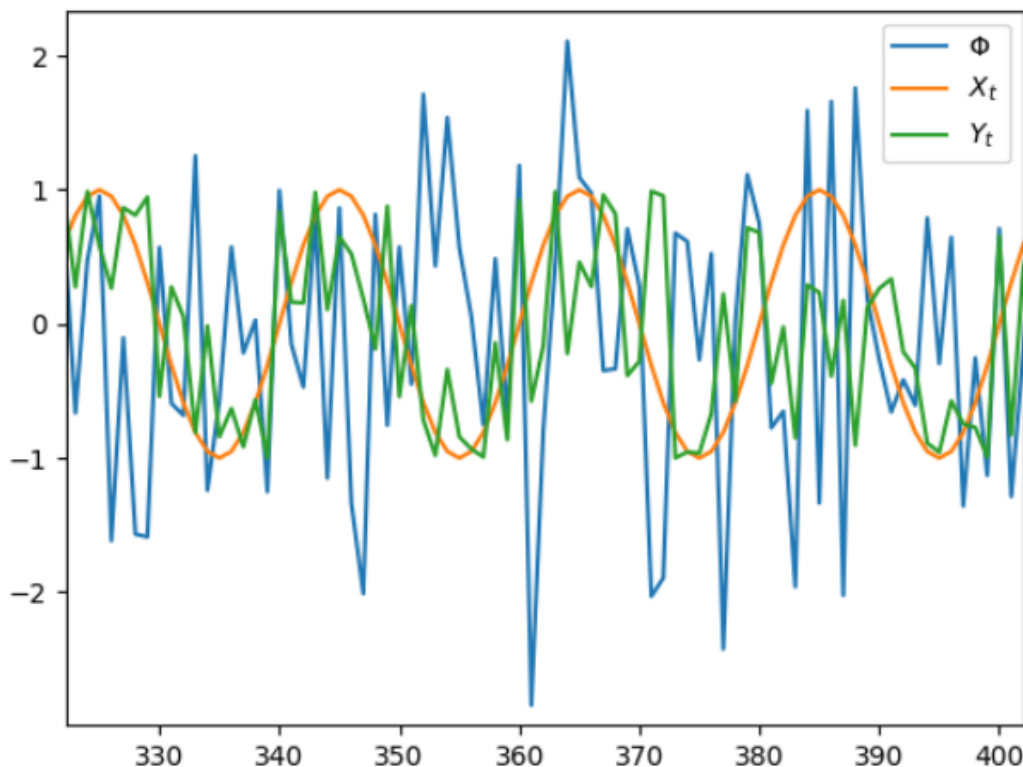
plt.plot(Phi)

plt.plot(X)

plt.plot(Y)

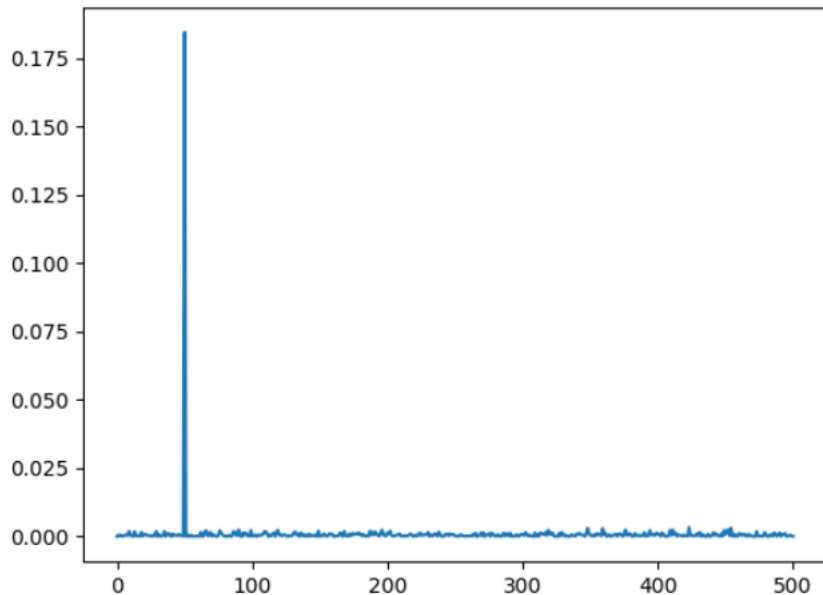
plt.legend([' Φ ', ' X_t ',

Y_t '])



Periodograma

```
G, Pyy_den = signal.periodogram(Y,N)  
plt.plot(G, Pyy_den)
```



Se puede observar el piso de ruido y una componente marcada en la frecuencia 50.

Ejemplos hands-on
(ver spectral_intro.py)

Densidad espectral de potencia

Si la autocovarianza decrece rápido a medida que aumentan los retardos y se satisface que tienen energía finita, esto es:

$$\sum_{k=-\infty}^{\infty} |C_k| < \infty$$

se puede definir la transformada de Fourier de la autocovarianza. La función definida entre $-\frac{1}{2} < f < \frac{1}{2}$ como sigue se denomina ***densidad espectral de potencia (PSD)***:


$$p(f) = \sum_{k=-\infty}^{\infty} C_k e^{-2\pi j k f}$$

Densidad espectral de potencia


Se puede ver que:

$$p(f) = \sum_{k=-\infty}^{\infty} C_k e^{-2\pi h k f}$$

$$= \sum_{k=-\infty}^{\infty} C_k \cos(2\pi k f) - j \sum_{k=-\infty}^{\infty} C_k \sin(2\pi k f)$$

$p(f)$ es par! 

$$= \sum_{k=-\infty}^{\infty} C_k \cos(2\pi k f)$$

cos es par! 

$$= C_0 + 2 \sum_{k=0}^{\infty} C_k \cos(2\pi k f)$$

Densidad espectral de potencia

Entonces una serie de tiempo se puede expresar en función de componentes trigonométricas y la PSD es una representación en esos términos.

También podemos partir de una PSDm y querer encontrar la función de autocovarianza, que se puede calcular a partir de la transformada inversa de Fourier:

$$C_k = \int_{-1/2}^{1/2} p(f) e^{2\pi j k f} df = \int_{-1/2}^{1/2} p(f) \cos(2\pi k f) df$$

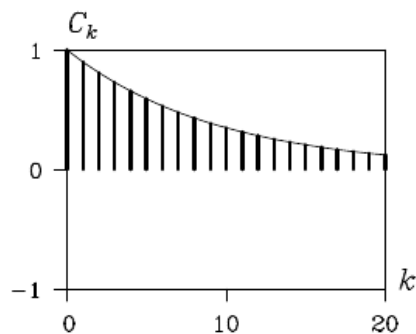
Periodograma y Densidad espectral de Potencia

Las expresiones que vimos recién corresponden a expresiones **teóricas**, que se calculan a partir del conocimiento del modelo.

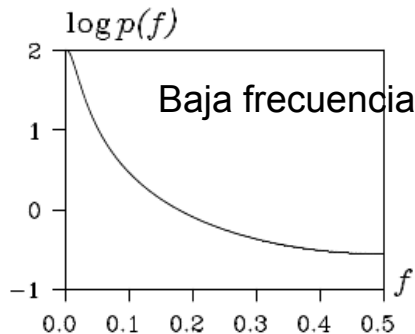
Observar que la PSD es la transformada de fourier de la función de covarianza, y por lo tanto podemos **estimarla** haciendo un periodograma de la misma

En python existe la función **scipy.signal.welch**, que devuelve la estimación de la PSD

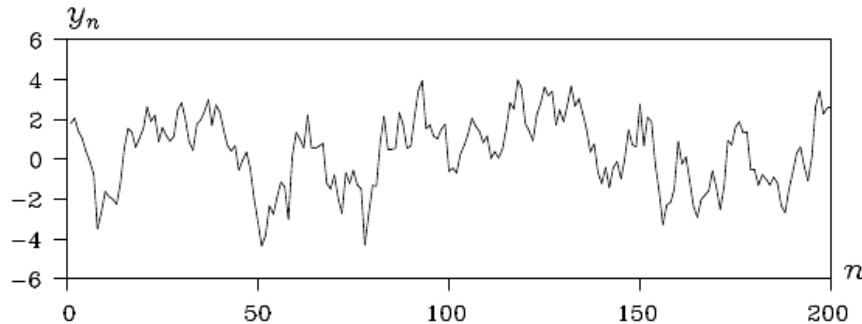
Ejemplo 1: densidad espectral de un proceso AR(1)



(a) Autocorrelation function



(b) Power spectrum



(c) Simulated time series

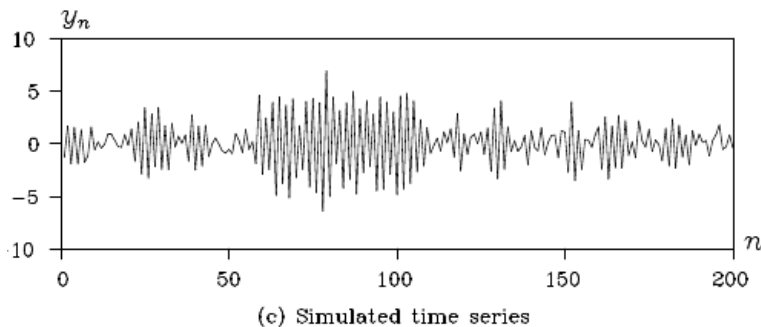
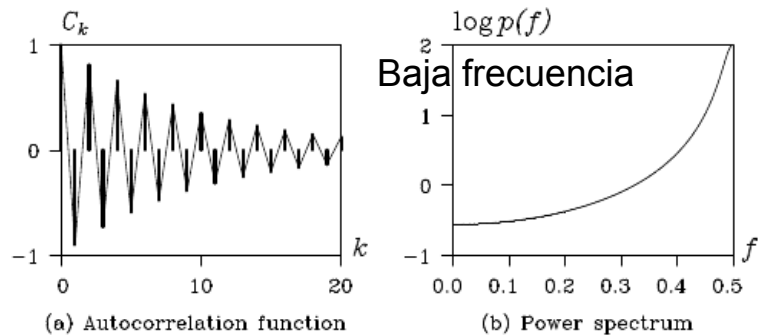
AR(1), $a = 0.9$

$$Y_t = a_1 Y_{t-1} + e_t$$

$$C_k = \sigma^2 \frac{a^{|k|}}{1 - a^2}$$

$$\begin{aligned} p(f) &= \frac{\sigma^2}{|1 - ae^{2\pi jf}|^2} \\ &= \frac{\sigma^2}{1 - 2a \cos 2\pi f + a^2} \end{aligned}$$

Ejemplo 2: densidad espectral de un proceso AR(1)



AR(1), $a = -0.9$

$$Y_t = a_1 Y_{t-1} + e_t$$

$$C_k = \sigma^2 \frac{a^{|k|}}{1 - a^2}$$

$$\begin{aligned} p(f) &= \frac{\sigma^2}{|1 - ae^{2\pi jf}|^2} \\ &= \frac{\sigma^2}{1 - 2a \cos 2\pi f + a^2} \end{aligned}$$


Análisis en el dominio espectral

Muchas veces la información en el dominio del tiempo para distintas series parece ser muy distinta y sin embargo en el dominio de frecuencias son similares. Cuando la representación espectral es continua, surge la necesidad de estimar el espectro.

Algunas técnicas de estimación espectral se llaman estimaciones no-paramétricas y asumen muy poco acerca de la densidad espectral “verdadera”.

La estimación paramétrica asume que un modelo autorregresivo de orden adecuado sirve para ajustar un modelo a la serie de tiempo. Luego la densidad espectral estimada va a tener las mismas propiedades que las del autorregresivo.

Análisis en el dominio espectral

- densidad espectral suavizada  <https://hakanyurdakul.com/smoothing-power-spectral-density/>
- sesgo
- varianza
- fugas
- ancho de banda
- ensanchamiento

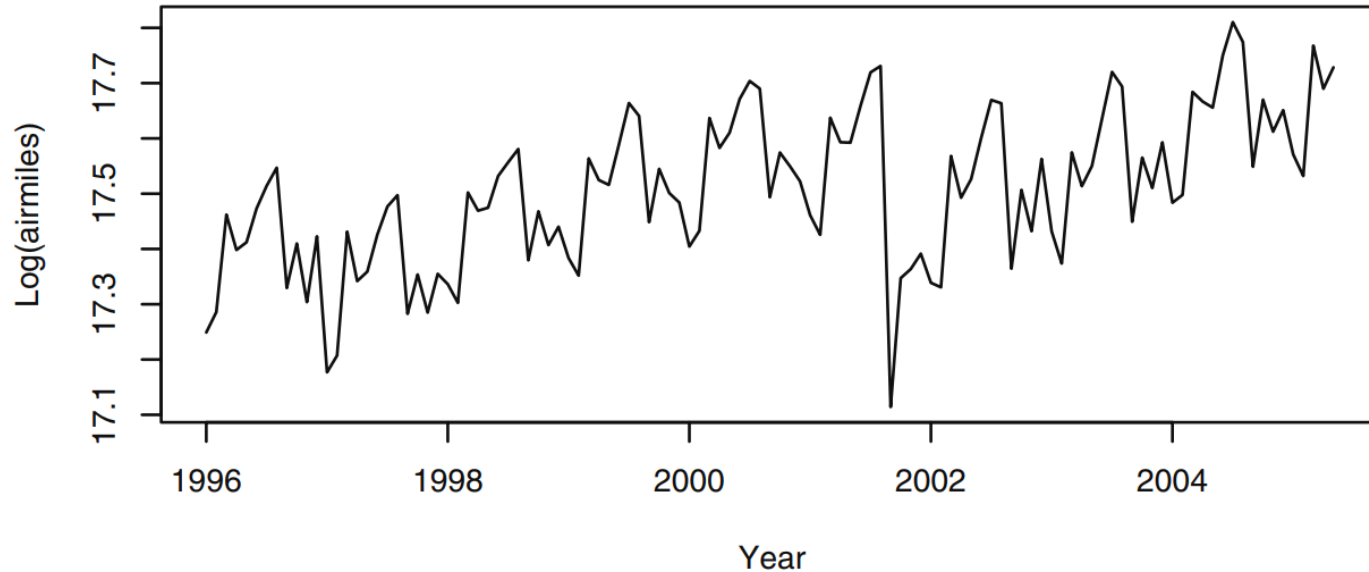
Resumen

- Modelos de series de tiempo como combinaciones lineales de senos y cosenos, introducción al análisis espectral
- El Periodograma es una herramienta para encontrar la contribución de varias frecuencias en la representación espectral
- Se puede extender el modelado a rangos continuos de frecuencias
- Se exploran las densidades espectrales de procesos ARMA
- Algunas propiedades de la densidad espectral muestral sugieren que la densidad espectral no es un estimador consistente, por lo que hay que buscar más para estimar espectros...

Análisis de intervenciones

Análisis de intervenciones

Exhibit 11.1 Monthly U.S. Airline Miles: January 1996 through May 2005



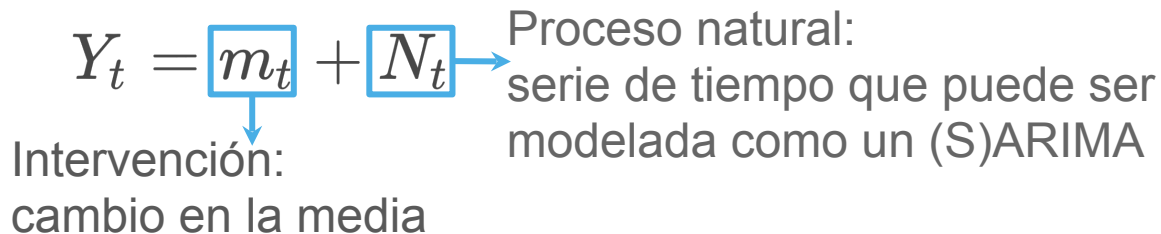
Análisis de intervenciones

Consideremos el caso de una única intervención:

$$Y_t = m_t + N_t$$

Intervención:
cambio en la media

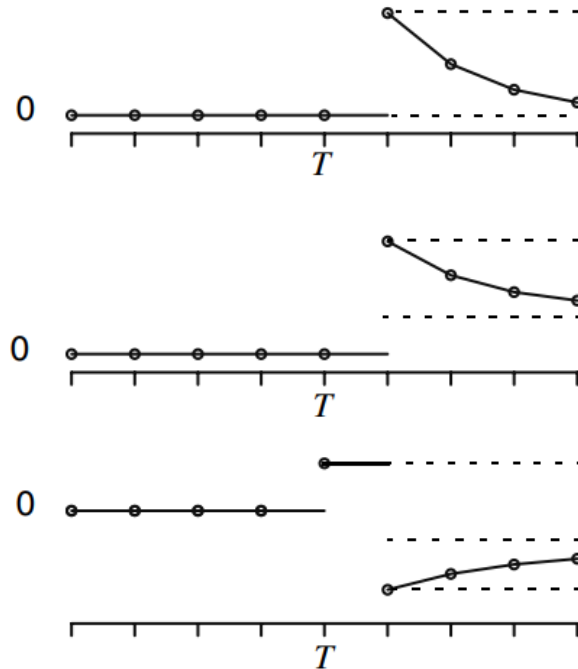
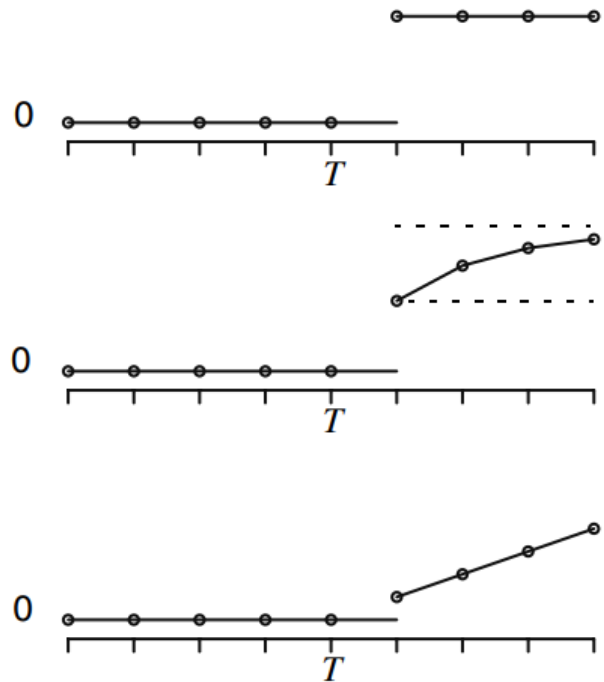
Proceso natural:
serie de tiempo que puede ser
modelada como un (S)ARIMA



Si la intervención ocurre en un instante T , se asume que $m_t = 0$ para $t < T$ y a $\{Y_t\}_{t < T}$ se la conoce como *datos preintervención* y puede ser usada para modelar N_t .

La estimación de los parámetros de m_t se puede realizar por MV o el enfoque de modelo de estados ya presentados.

Análisis de intervenciones

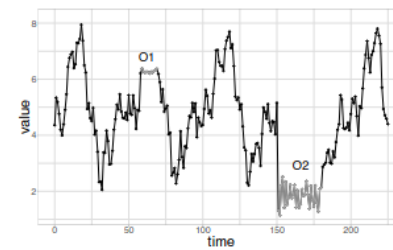
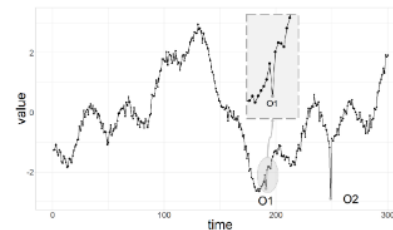
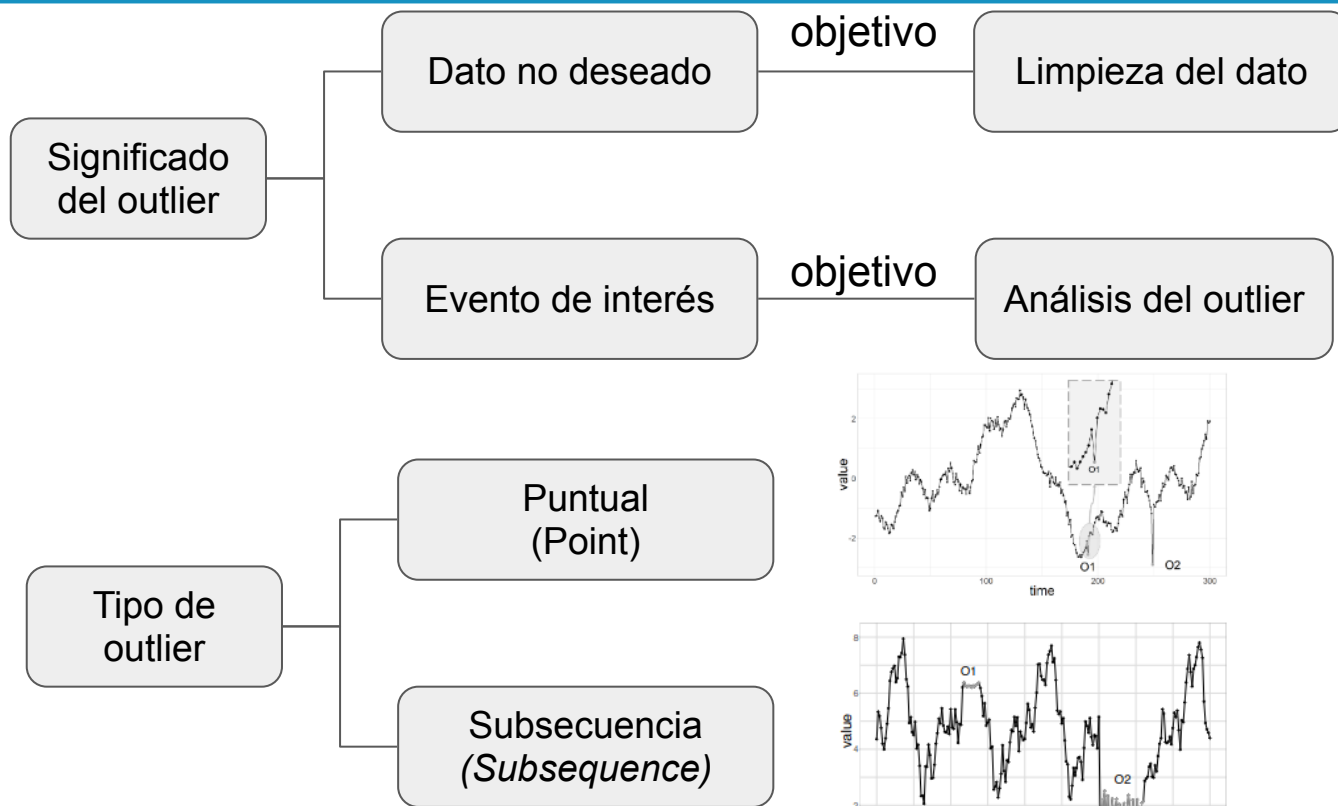


Ejemplo

Ver notebook: [intervention_analysis.ipynb](#)

Análisis de outliers

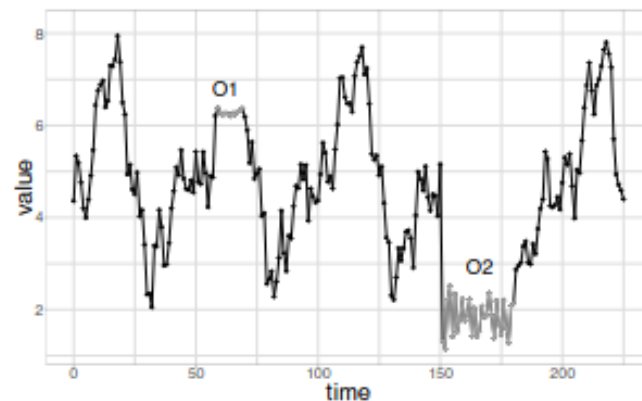
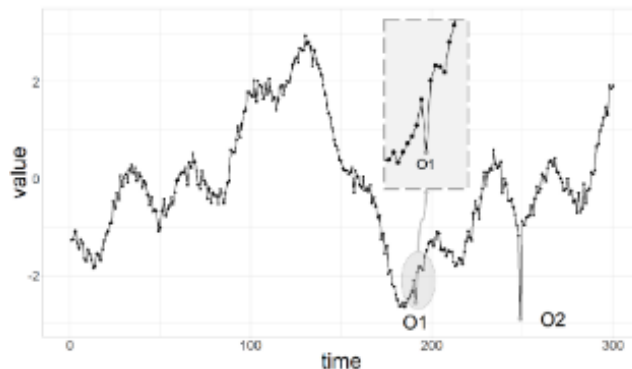
Tipos de outliers



Tipos puntuales

Outliers puntuales: Dato que se comporta de manera inusual en un instante de tiempo específico, comparado con la serie total, o sus puntos vecinos. Son los más comunes.

Outliers de subsecuencia: corresponde a puntos que en conjunto se comportan de forma peculiar respecto del resto de la serie, aunque cada observación de forma individual no resulta necesariamente un outlier.

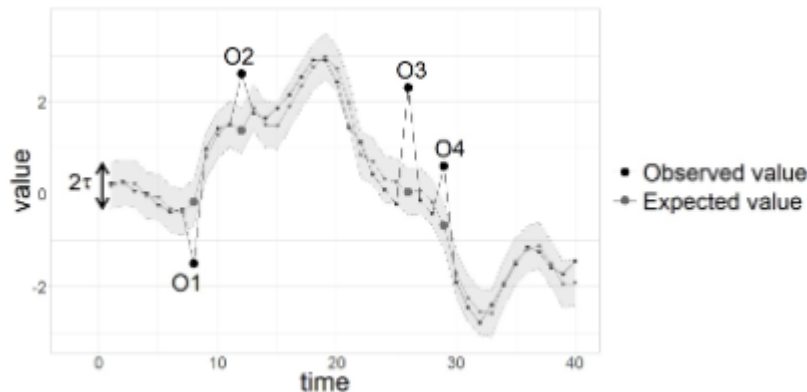


Análisis de Outliers puntuales:

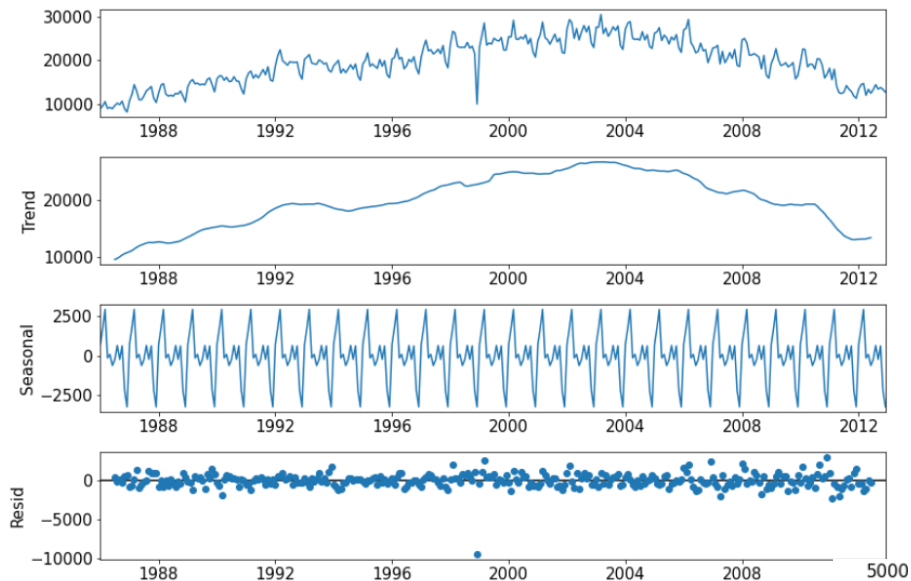
Podemos definir un outlier como un punto que se desvía significativamente de su valor esperado, es decir $|x_t - \hat{x}_t| > \tau$

Hay distintas formas de determinar ese valor esperado y el umbral.

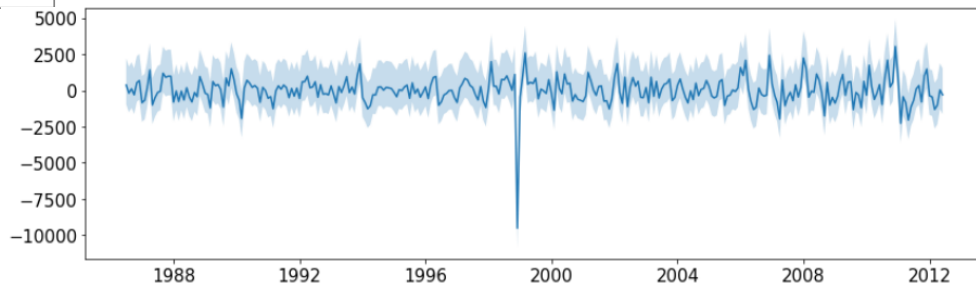
- Modelos constantes a tramos, donde se toman estadísticas como mediana o MAD (mean absolute deviation)
- Identificación de puntos poco probables bajo un cierto modelo o distribución
- Analizando residuos (ej. usando la descomposición STL, o directamente analizando residuos del ARIMA)
- Etc.



Detección de outliers usando la descomposición STL



```
fig = plt.figure(figsize=(14,4))
q975 = residuals.quantile(.975)
q025 = residuals.quantile(.025)
plt.plot(residuals)
plt.fill_between(x=residuals.index,
                 y1=residuals + q025,
                 y2=residuals + q975,
                 alpha=0.25
                )
```



Cómo tratar los outliers

- Usar la media en una ventana para suavizar el outlier
- Modelar el dato basándose en el problema de estudio
- Quitar el punto (mucho cuidado!)

Ejemplo

Ver notebook: [outlier_detection.ipynb](#)

Heterocedasticidad

Motivación

Hasta ahora: $\mathbb{E}[Y_t | Y_{t-1}, Y_{t-2}, \dots, Y_1]$

También podemos estudiar $\mathbb{V}[Y_t | Y_{t-1}, Y_{t-2}, \dots, Y_1]$, que mide la incerteza de la desviación de Y_t de su esperanza condicional.

Muy relevante cuando queremos modelar **series financieras**.

Si Y_t sigue un modelo ARMA, la varianza condicional es constante para cualquier paso fijo en el tiempo (a un paso coincide con la varianza del ruido).

En la práctica, la var. condicional es un proceso aleatorio, y por lo tanto puede variar con valores presentes y pasados de la serie.

Se desarrollan entonces **modelos para la varianza condicional**.

NOTA: la varianza condicional es una medida del riesgo

Medidas de Riesgo

- **Desvío estándar:** mide la variabilidad de los data points alrededor del valor medio de la serie de tiempo.
- **Valor en Riesgo (VaR):** medida estadística que estima la máxima pérdida que una inversión puede sufrir con un determinado intervalo de confianza sobre un período específico de tiempo.
- **VaR Condicional (CVaR):** provee un estimador de la expectativa de pérdida más allá del valor de VaR.
- **ARCH/GARCH:** modelos estadísticos para estimar riesgo de una serie de tiempo basado en las variabilidad pasada.
- **Monte Carlo:** simulaciones que generan miles de potenciales escenarios para estimar el rango de potenciales resultados y sus probabilidades.

Ejemplo

Sea $\{p_t\}$ una serie de tiempo correspondiente al precio diario de un activo financiero.

El **rendimiento** del día t se define como

$$r_t = \log\left(\frac{p_t}{p_{t-1}}\right) \times 100 = [\log(p_t) - \log(p_{t-1})] \times 100$$

Se denomina **volatilidad** al fenómeno donde la varianza condicional de la serie varía a lo largo del tiempo.

En la gráfica de rendimiento se ven clusters de volatilidad, donde se ve un patrón de alternancia entre períodos “tranquilos” y períodos con volatilidad

Exhibit 12.1 Daily CREF Stock Values: August 26, 2004 to August 15, 2006

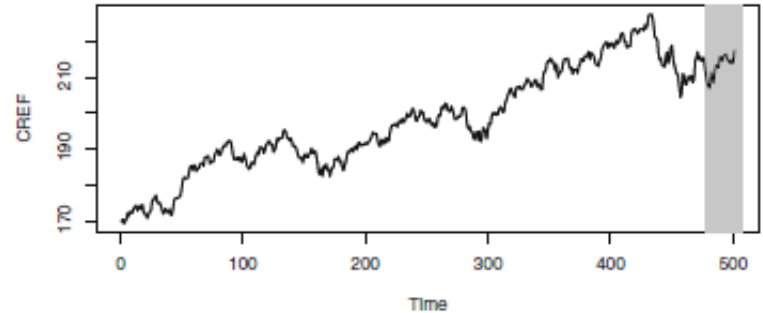
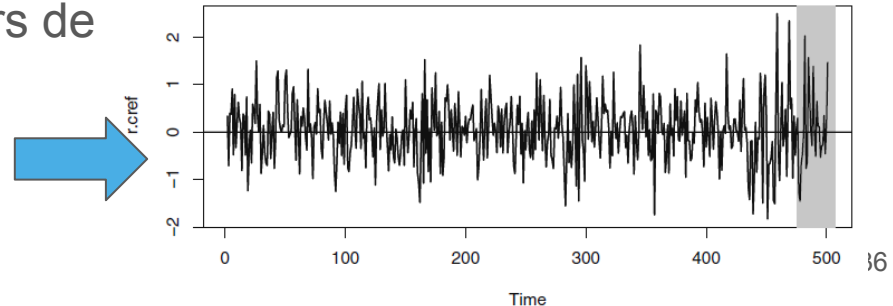


Exhibit 12.2 Daily CREF Stock Returns: August 26, 2004 to August 15, 2006



Ejemplo

Exhibit 12.3 Sample ACF of Daily CREF Returns: 8/26/04 to 8/15/06

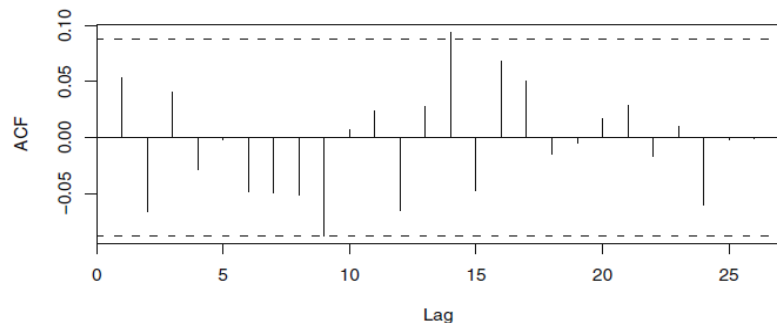
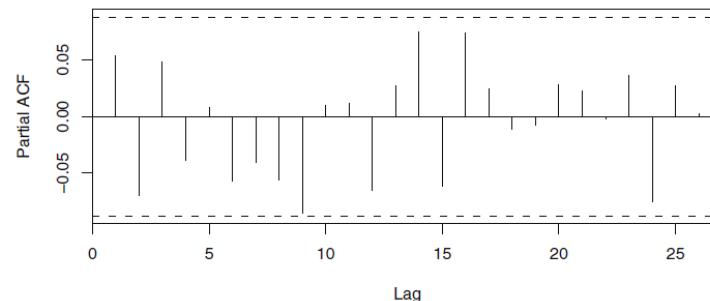


Exhibit 12.4 Sample PACF of Daily CREF Returns: 8/26/04 to 8/15/06



La ACF y PACF parecen indicar que la serie de retornos está **descorrelacionada**. **Pero** los clusters de volatilidad indican que **no hay independencia**, sino la varianza sería constante a lo largo de la serie de tiempo.

independencia \neq descorrelación.

Si la serie es independiente, la independencia se mantiene frente a transformaciones. No es cierto para la descorrelación.

Algunas formas de identificar falta de independencia

- **Métodos visuales:** Podemos observar la ACF del retorno absoluto o bien del retorno cuadrático. Si estas ACF presentan autocorrelaciones significativas, tenemos evidencia de que la serie original de retorno no era independiente.

Exhibit 12.7 Sample ACF of the Squared Daily CREF Returns

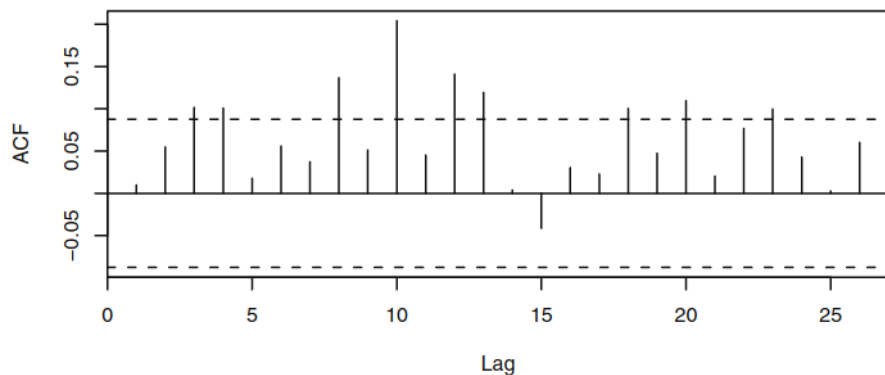
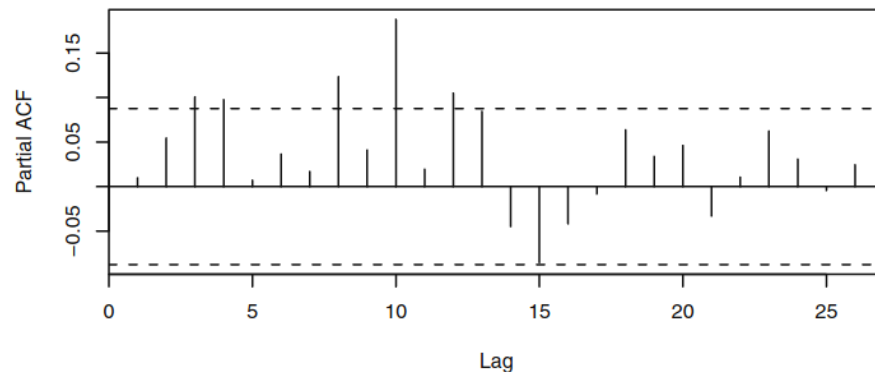


Exhibit 12.8 Sample PACF of the Squared Daily CREF Returns



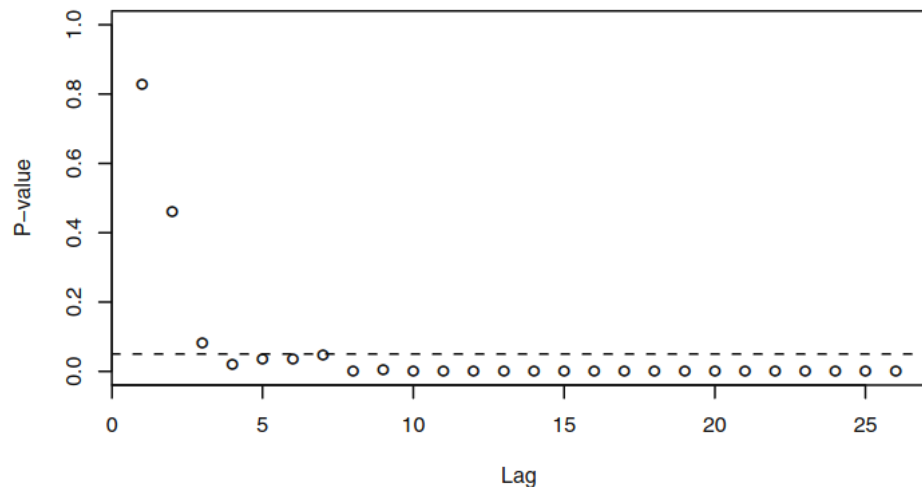
Algunas formas de identificar falta de independencia

Tests estadísticos: Podemos aplicar el test de Box-Ljung para determinar si el cuadrado de los datos está autocorrelacionada.

Se puede extender este test al análisis del cuadrado de los residuos de una serie temporal donde un modelo ARMA sirve para describir la estructura de la autocorrelación de los datos.

Al test de Box-Ljung sobre el cuadrado de los residuos o datos se lo conoce como **test de McLeod-Li**

Exhibit 12.9 McLeod-Li Test Statistics for Daily CREF Returns



Modelos

Modelo ARCH

El modelo ARCH (AutoRegressive Conditional Heteroskedasticity) resulta una **regresión** sobre la varianza condicional (o **volatilidad** condicional), donde la varianza condicional es la variable a predecir y los valores r_1^2, \dots, r_{t-1}^2 son los regresores.

Denotaremos $\sigma_t^2 | t-1 = \mathbb{V}(r_t | r_{t-1}, \dots, r_1)$

El modelo ARCH(1) considera que la serie $\{r_t\}$ es generado según

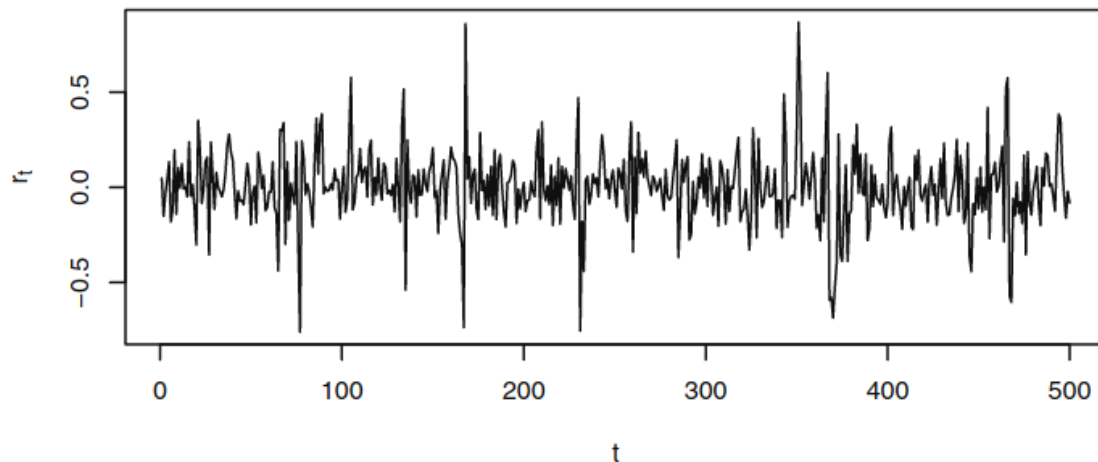
$$r_t = \sigma_{t|t-1} \varepsilon_t \longrightarrow \varepsilon_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$$

con ε_t indep. de r_{t-j} , $j \geq 1$; α y ω constantes desconocidas.

$$\sigma_{t|t-1}^2 = \omega + \alpha r_{t-1}^2$$

Modelo ARCH(1)

Exhibit 12.11 Simulated ARCH(1) Model with $\omega = 0.01$ and $\alpha_1 = 0.9$



```
> set.seed(1235678); library(tseries)
> garch01.sim=garch.sim(alpha=c(.01,.9),n=500)
> plot(garch01.sim,type='l',ylab=expression(r[t]), xlab='t')
```

$$r_t = \sigma_{t|t-1} \varepsilon_t$$

$$\sigma_{t|t-1}^2 = \omega + \alpha r_{t-1}^2$$

Modelo ARCH

Si bien el modelo ARCH se parece a un AR, la variable $\sigma_{t|t-1}$ no es observable, y se la llame variable **latente** \Rightarrow modificaciones en el análisis de los datos

- No es directo cómo explorar la relación de la regresión visualmente.

Si reemplazamos la variable latente $\sigma_{t|t-1}$ por un observable, podemos ver que el **modelo ARCH(1) puede reescribirse como un AR(1)**:

Definiendo $\eta_t = r_t^2 - \sigma_{t|t-1}^2$, se puede ver que $\{\eta_t\}$ es una serie descorrelacionada de media nula, y además η_t se encuentra descorrelacionada de retornos previos. Luego,

$$r_t^2 = \omega + \alpha r_{t-1}^2 + \eta_t$$

Modelo ARCH(q)

Se puede generalizar el modelo ARCH(1) a una dependencia temporal más prolongada:

$$\sigma_{t|t-1}^2 = \omega + \alpha_1 r_{t-1}^2 + \alpha_2 r_{t-2}^2 + \cdots + \alpha_q r_{t-q}^2$$

donde $\alpha_1 > \alpha_2 > \dots > \alpha_q$.

En este caso, podremos reescribir el modelo ARCH(q) como un AR(q)

Modelo GARCH

El modelo **GARCH(p,q)** (**Generalized AutoRegressive Conditional Heteroskedasticity**), incorpora p lags de la varianza condicional al modelo:

$$\sigma_{t|t-1}^2 = \omega + \beta_1 \sigma_{t-1|t-2}^2 + \dots + \beta_p \sigma_{t-p|t-p-1}^2 + \alpha_1 r_{t-1}^2 + \alpha_2 r_{t-2}^2 + \dots + \alpha_q r_{t-q}^2$$

o en términos del operador de backshift:

$$(1 - \beta_1 B - \dots - \beta_p B^p) \sigma_{t|t-1}^2 = \omega + (\alpha_1 B + \dots + \alpha_q B^q) r_t^2$$

Nuevamente, es conveniente reescribir estas expresiones en términos de $\{\eta_t\}$, que sigue siendo descorrelacionada e independiente de retornos previos.

$$r_t^2 = \omega + (\beta_1 + \alpha_1) r_{t-1}^2 + \dots + (\beta_{\max(p,q)} + \alpha_{\max(p,q)}) r_{t-\max(p,q)}^2 + \eta_t - \beta_1 \eta_{t-1} - \dots - \beta_p \eta_{t-p}$$
$$\Rightarrow \{r_t^2\} \text{ sigue un modelo ARMA}(\max(p,q), p)$$

Exhibit 12.12 Simulated GARCH(1,1) Process

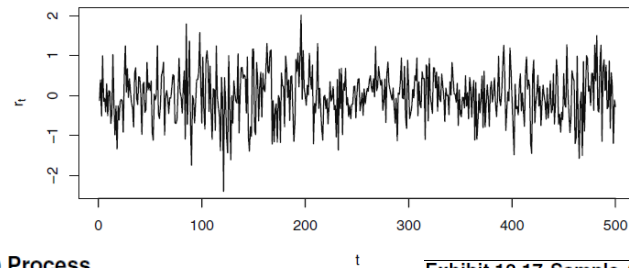


Exhibit 12.14 Sample PACF of Simulated GARCH(1,1) Process

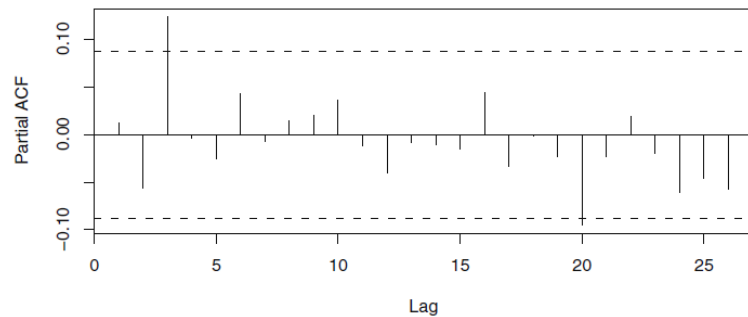


Exhibit 12.13 Sample ACF of Simulated GARCH(1,1) Process

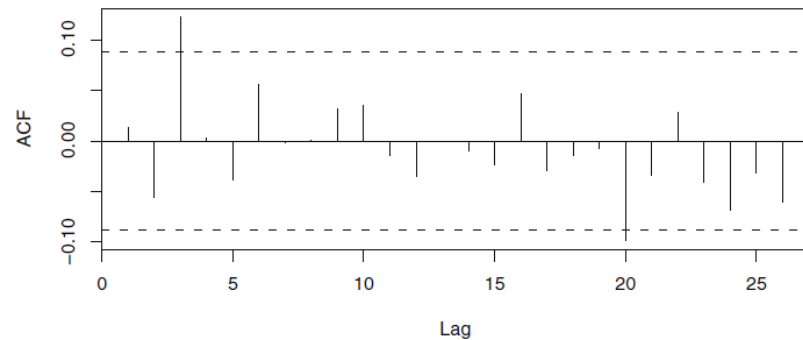


Exhibit 12.17 Sample ACF of the Squared Values of the Simulated GARCH(1,1) Process

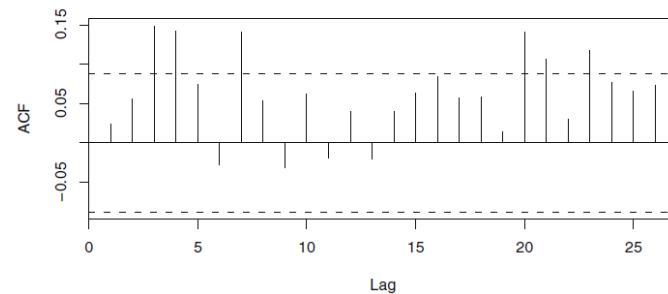
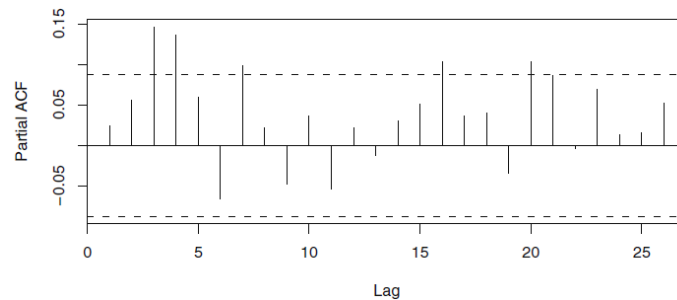


Exhibit 12.18 Sample PACF of the Squared Values of the Simulated GARCH(1,1) Process



Identificación del modelo.

Para estimar el orden del modelo GARCH(p, q), vamos a analizar las funciones de autocorrelación y autocorrelación parcial del cuadrado de los retornos (o residuos) para identificar los parámetros de ARMA($\max(p, q)$, p).

Observar que si $q < p$, el valor de p queda enmascarado en el modelo. En ese caso es conveniente ajustar primero un modelo GARCH(p, p), y luego estimar el valor de q mirando la significancia de los coeficientes (como $q < p$ estoy estimando parámetros de sobra, voy a ajustar el modelo quitando aquellos que tengan baja significancia).

Condiciones de estacionariedad del modelo GARCH

Se puede mostrar que para que un modelo GARCH(p,q) sea (débilmente) estacionario es condición necesaria y suficiente que:

$$\sum_{i=1}^{\max(p,q)} (\beta_i + \alpha_i) < 1$$

bajo el supuesto de que $\alpha_i, \beta_j \geq 0$ para garantizar que tanto r_i^2 como $\sigma_{t|t-h}$ resulten positivos.

Además, la distribución asintótica del modelo GARCH es de colas pesadas, incluso cuando las innovaciones tienen una distribución normal.

Predicción a h pasos

Armando el mismo esquema recursivo que venimos usando, podemos demostrar que

$$\sigma_{t+h|t}^2 = \omega + \sum_{i=1}^p (\alpha_i + \beta_i) \sigma_{t+h-i|t}^2, \quad h > p$$

(la expresión general para cualquier h está en la pag. 297 de Time Series Analysis)

Verificación del modelo

Definimos los residuos estandarizados como

$$\hat{\varepsilon}_t = r_t / \hat{\sigma}_{t|t-1}$$

Si el modelo está bien especificado los $\hat{\varepsilon}_t$ deberían ser i.i.d..

Para verificar el modelo podemos estudiar estos residuos estandarizados vamos a analizar

- Normalidad de $\hat{\varepsilon}_t$ (QQplot o Shapiro-Wilk)
- Independencia (análisis de la función de autocorrelación)

Exhibit 12.26 Standardized Residuals from the Fitted GARCH(1,1) Model of Daily CREF Returns

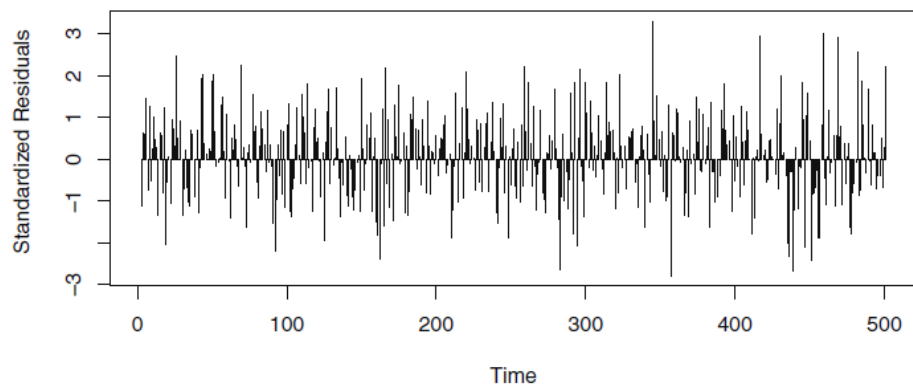


Exhibit 12.27 QQ Normal Scores Plot of Standardized Residuals from the Fitted GARCH(1,1) Model of Daily CREF Returns

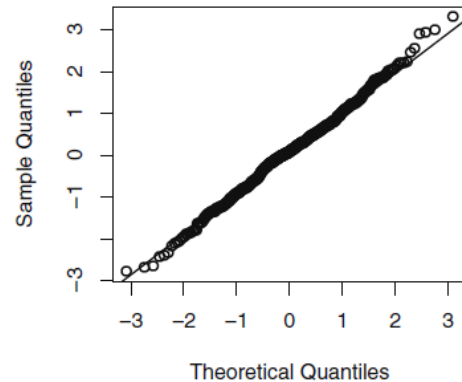
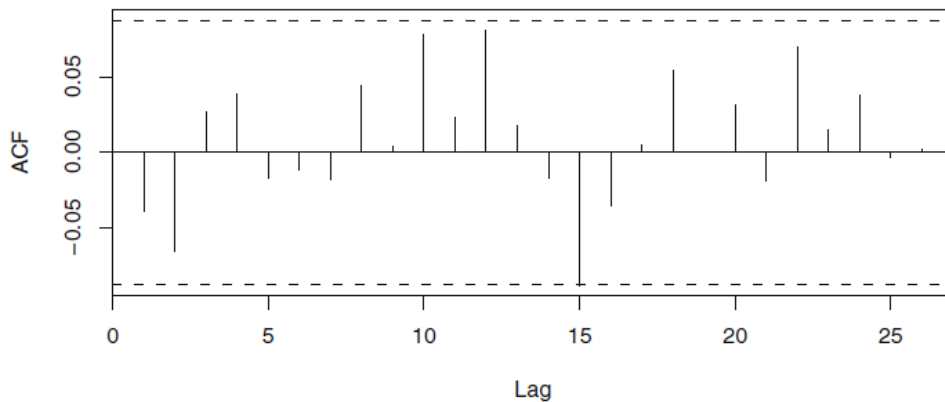


Exhibit 12.28 Sample ACF of Squared Standardized Residuals from the GARCH(1,1) Model of the Daily CREF Returns



Ejemplo

[Ver ejemplo](#)

Modelos multivariados

Modelos VAR

El modelo VAR (Vector AutoRegressive) es una extensión de los modelos AR para series multivariadas.

Sea $\underline{Y} = (Y_{1t}, Y_{2t}, \dots, Y_{nt})^T$. El modelo autorregresivo vectorial de orden p resulta

$$\underline{Y}_t = \underline{c} + \mathbf{\Pi}_1 \underline{Y}_{t-1} + \dots + \mathbf{\Pi}_p \underline{Y}_{t-p} + \underline{\varepsilon}_t, \quad \underline{\varepsilon}_t = (\varepsilon_{1,t}, \dots, \varepsilon_{n,t})^T, \quad \mathbf{\Pi}_i \in \mathbb{R}^{n \times n}$$

donde $\text{cov}(\varepsilon_{k,t}, \varepsilon_{j,s}) = \sigma_{k,j}$ si $t = s$, 0 si no.

Bibliografía

- [Documentación](#) del paquete ARCH
- [Modelos VAR](#)

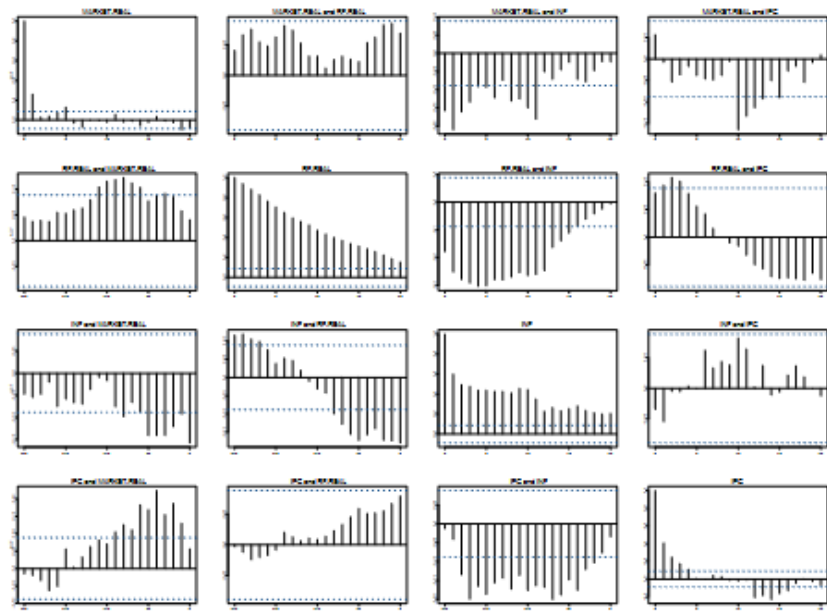


FIGURE 11.11. Autocorrelations and cross correlations at leads and lags of data in VAR model.

Revisión de bibliografía

Bibliografía

1. [Kitagawa] "Introduction to Time Series Modeling with applications in R", Genshiro Kitagawa, p.143-150
2. [Godfrey2017] Neural Decomposition of Time-Series Data
3. [Chrobok2005] Theory and Application of Advanced Traffic Forecast Methods
4. [Hochreiter1997] LONG SHORT-TERM MEMORY
5. [Gers2000] Learning to Forget: Continual Prediction with LSTM
6. Decompose: <https://towardsdatascience.com/different-types-of-time-series-decomposition-396c09f92693>
7. [Detección de outliers](#)
8. [Tutorial de detección de outliers](#)
9. RPIC2019, Exponential Smoothing: Ref.: <https://otexts.com/fpp2/expsmooth.html>

Appendix D: The Backshift Operator

Many other books and much of the time series literature use what is called the **backshift operator** to express and manipulate ARIMA models. The backshift operator, denoted B , operates on the time index of a series and shifts time back one time unit to form a new series.[†] In particular,

$$BY_t = Y_{t-1}$$

The backshift operator is linear since for any constants a , b , and c and series Y_t and X_t , it is easy to see that

$$B(aY_t + bX_t + c) = aBY_t + bBX_t + c$$

Consider now the MA(1) model. In terms of B , we can write

$$\begin{aligned} Y_t &= e_t - \theta e_{t-1} = e_t - \theta B e_t = (1 - \theta B) e_t \\ &= \theta(B) e_t \end{aligned}$$

where $\theta(B)$ is the MA characteristic polynomial “evaluated” at B .

Since BY_t is itself a time series, it is meaningful to consider BBY_t . But clearly $BBY_t = BY_{t-1} = Y_{t-2}$, and we can write

$$B^2 Y_t = Y_{t-2}$$

More generally, we have

$$B^m Y_t = Y_{t-m}$$

for any positive integer m . For a general MA(q) model, we can then write

$$\begin{aligned} Y_t &= e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \\ &= e_t - \theta_1 B e_t - \theta_2 B^2 e_t - \cdots - \theta_q B^q e_t \\ &= (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) e_t \end{aligned}$$

or

$$Y_t = \theta(B) e_t$$

where, again, $\theta(B)$ is the MA characteristic polynomial evaluated at B .

For autoregressive models AR(p), we first move all of the terms involving Y to the left-hand side

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \cdots - \phi_p Y_{t-p} = e_t$$

and then write

$$Y_t - \phi_1 B Y_t - \phi_2 B^2 Y_t - \cdots - \phi_p B^p Y_t = e_t$$

or

[†] Sometimes B is called a **Lag operator**.

Appendix D: The Backshift Operator

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) Y_t = e_t$$

which can be expressed as

$$\phi(B) Y_t = e_t$$

where $\phi(B)$ is the AR characteristic polynomial evaluated at B .

Combining the two, the general ARMA(p, q) model may be written compactly as

$$\phi(B) Y_t = \theta(B) e_t$$

Differencing can also be conveniently expressed in terms of B . We have

$$\begin{aligned} \nabla Y_t &= Y_t - Y_{t-1} = Y_t - B Y_t \\ &= (1 - B) Y_t \end{aligned}$$

with second differences given by

$$\nabla^2 Y_t = (1 - B)^2 Y_t$$

Effectively, $\nabla = 1 - B$ and $\nabla^2 = (1 - B)^2$.

The general ARIMA(p, d, q) model is expressed concisely as

$$\phi(B)(1 - B)^d Y_t = \theta(B) e_t$$

In the literature, one must carefully distinguish from the context the use of B as a backshift operator and its use as an ordinary real (or complex) variable. For example, the stationarity condition is frequently given by stating that the roots of $\phi(B) = 0$ must be greater than 1 in absolute value or, equivalently, must lie outside the unit circle in the complex plane. Here B is to be treated as a dummy variable in an equation rather than as the backshift operator.

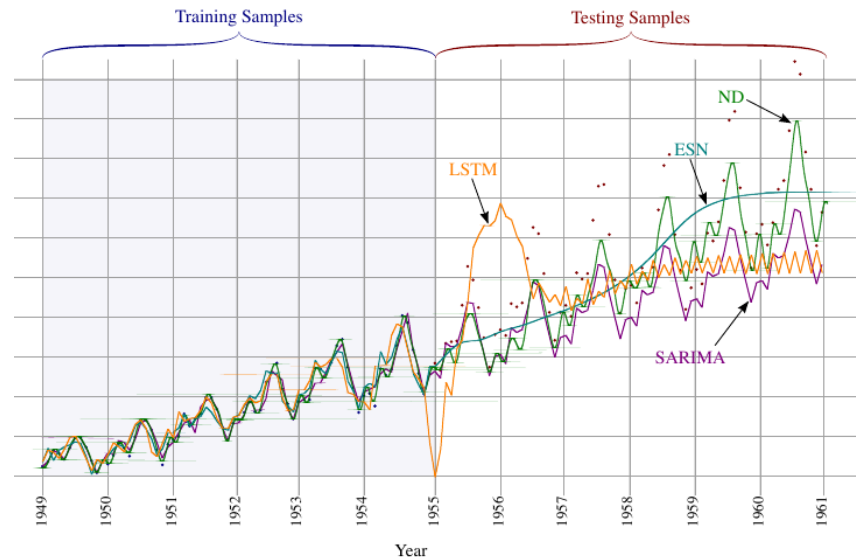
**Ref.: Introduction to Time Series Modeling,
Genshiro Kitagawa**

Neural decomposition

Our model is defined as follows. Let each a_k represent an amplitude, each w_k represent a frequency, and each ϕ_k represent a phase shift. Let t refer to time, and let $g(t)$ be an augmentation function that represents the nonperiodic components of the signal. Our model is defined as

$$x(t) = \sum_{k=1}^N (a_k \cdot \sin(w_k t + \phi_k)) + g(t). \quad (2)$$

Note that the lower index of the sum has changed from $k = 0$ in the DFT to $k = 1$ in our model. This is because ND can account for bias in the augmentation function $g(t)$. Therefore, only N sinusoids are required rather than $N + 2$.



Advanced Traffic Forecast Methods

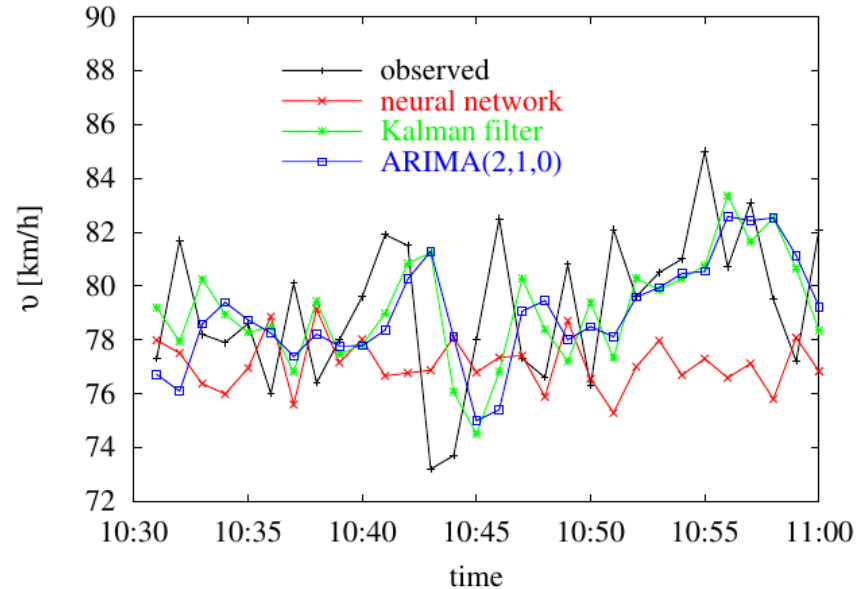
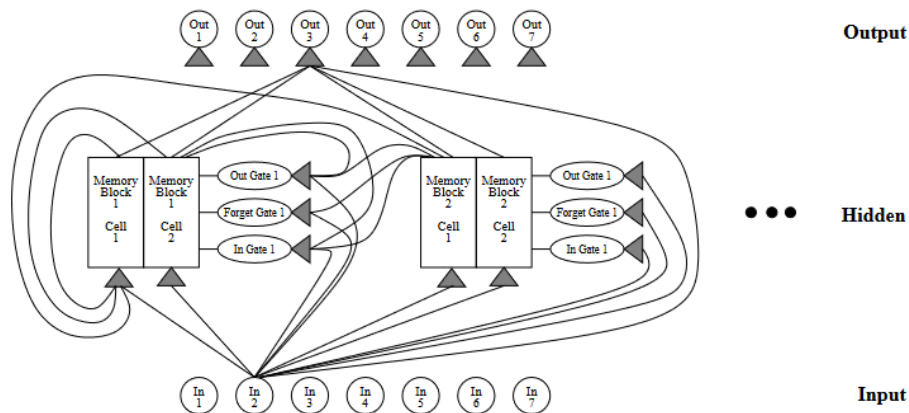
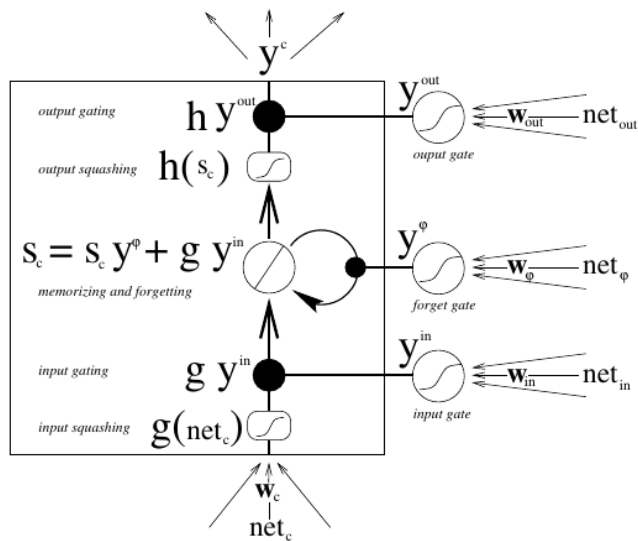


Fig. 3.9. Comparison of a four layer feedforward neural network, a Kalman filter, and an ARIMA(2,1,0) model as one-step-ahead velocity forecast methods by Lee and Choi [1998]. It seems that the velocity fluctuates around a mean value and that no model stands out in forecasting those fluctuations.

LSTM: Long Short Term Memory



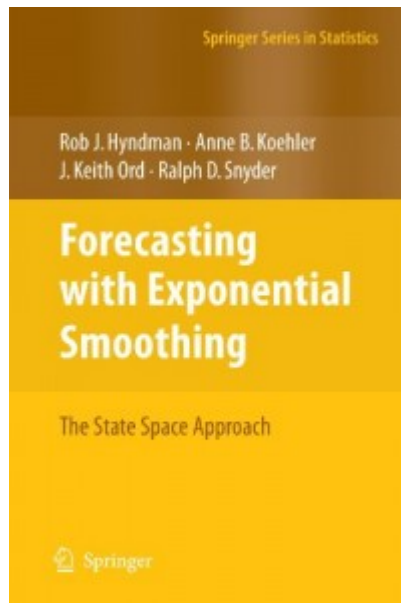


Table of contents

[Preface](#)

Part I: Introduction

1. Basic concepts [\[R code\]](#)
2. [Getting started \[R code\]](#) [\[Exercise solutions\]](#)

Part II: Essentials

3. Linear innovations state space models [\[R code\]](#) [\[Exercise solutions\]](#)
4. Non-linear and heteroscedastic innovations state space models [\[Exercise solutions\]](#)
5. Estimation of innovations state space models [\[R code\]](#) [\[Exercise solutions\]](#)
6. Prediction distributions and intervals [\[R code\]](#) [\[Exercise solutions\]](#)
7. Selection of models

Part III: Further topics

8. Normalizing seasonal components [\[R code\]](#)
9. Models with regressor variables [\[R code\]](#)
10. Some properties of linear models
11. Reduced forms and relationships with ARIMA models
12. Linear innovations state space models with random seed states
13. Conventional state space models
14. Time series with multiple seasonal patterns (with Phillip Gould) [\[R code\]](#)
15. Non-linear models for positive data (with Muhammad Akram) [\[R code\]](#)
16. Models for count data [\[R code\]](#)
17. Vector exponential smoothing (with Ashton de Silva)

Part IV: Applications

18. Inventory control application [\[R code\]](#)
19. Conditional heteroscedasticity and applications in finance [\[R code\]](#)
20. Economic applications: the Beveridge-Nelson decomposition (with Chin Nam Low and Heather Anderson)