



# Análisis de Series de Tiempo

Carrera de Especialización en Inteligencia Artificial

---

**Clase 7**

Ing. Magdalena Bouza, Ing. Carlos German Carreño Romano

# Contenido

---

- Modelos de regresión
  - Cómo incorporar información externa en el modelado de las series de tiempo
  - Cómo realizar tratamiento de outliers
- Heterocedasticidad

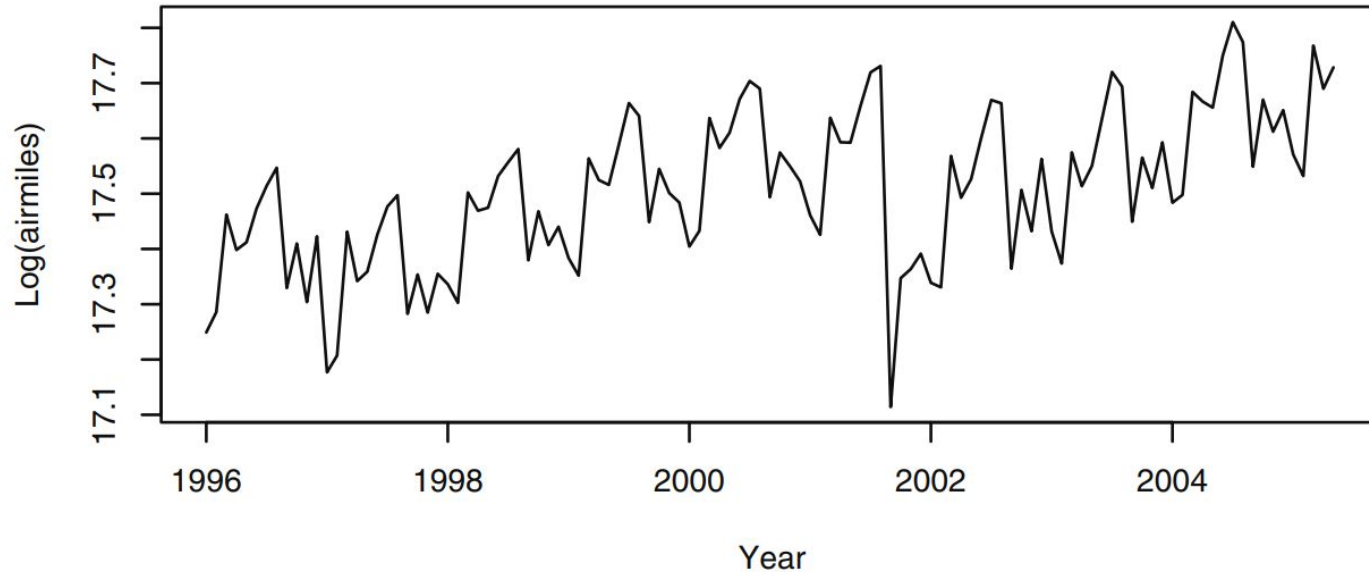
# Modelos de Regresión

# Análisis de intervenciones

# Análisis de intervenciones

---

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



# Análisis de intervenciones

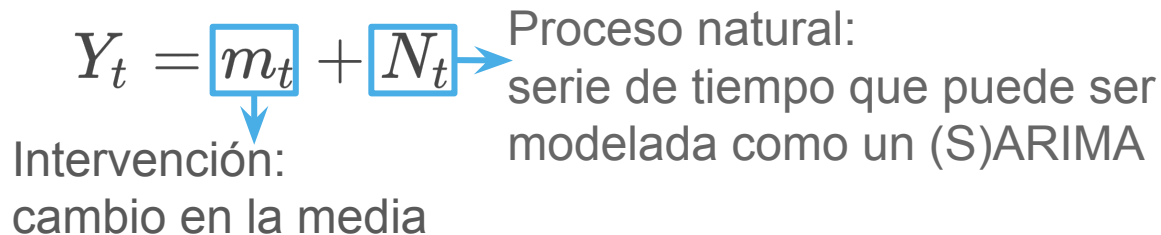
---

Consideremos el caso de una única intervención:

$$Y_t = \boxed{m_t} + \boxed{N_t}$$

Intervención:  
cambio en la media

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



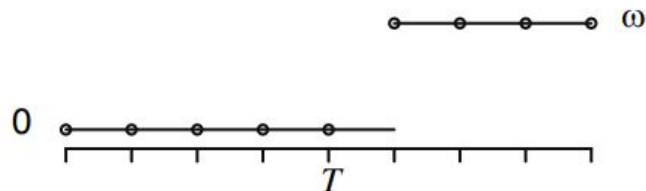
Si la intervención ocurre en un instante  $T$ , se asume que  $m_t = 0$  para  $t < T$  y a  $\{Y_t\}_{t < T}$  se la conoce como *datos preintervención* y puede ser usada para modelar  $N_t$ .

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

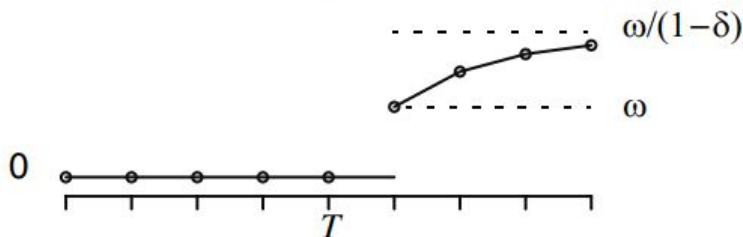
# Análisis de intervenciones

**Exhibit 11.3 Some Common Models for Step Response Interventions**  
(All are shown with a delay of 1 time unit)

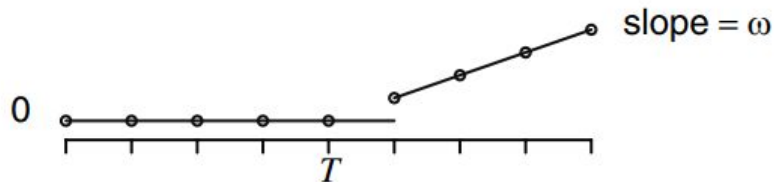
$$S_t^T \equiv \mathbf{I}\{t > T\} \quad \omega B S_t^{(T)}$$



$$(b) \quad \frac{\omega B}{1 - \delta B} S_t^{(T)}$$



$$(c) \quad \frac{\omega B}{1 - B} S_t^{(T)}$$



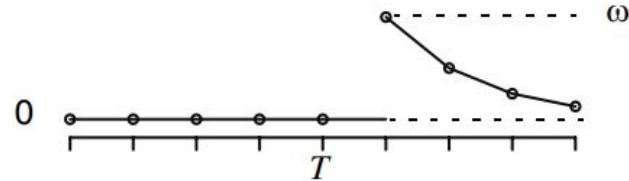
# Análisis de intervenciones

**Exhibit 11.4 Some Common Models for Pulse Response Interventions**  
(All are shown with a delay of 1 time unit)

$$[P_t^{(T)} \equiv \delta_T(t)]$$

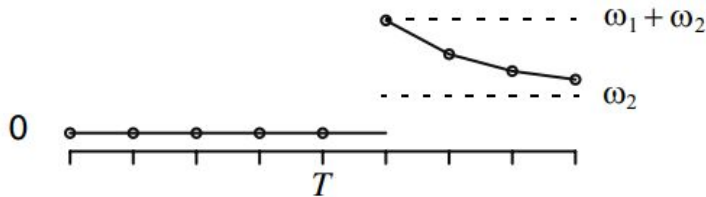
(a)

$$\frac{\omega B}{1 - \delta B} P_t^{(T)}$$



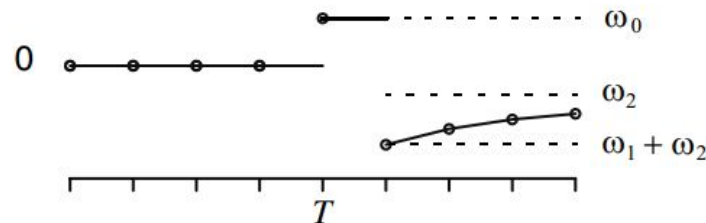
(b)

$$\left[ \frac{\omega_1 B}{1 - \delta B} + \frac{\omega_2 B}{1 - B} \right] P_t^{(T)}$$



(c)

$$\left[ \omega_0 + \frac{\omega_1 B}{1 - \delta B} + \frac{\omega_2 B}{1 - B} \right] P_t^{(T)}$$





# Análisis de outliers

# Análisis de outliers

---

Se corresponden con mediciones atípicas que aparecen por errores de medición o por cambios abruptos de corta duración en el proceso subyacente.

- **Outliers aditivos (AO):**  $Y'_t = Y_t + \omega_A P_t^{(T)}$ ,  $[P_t^{(T)} \equiv \delta_T(t)]$
- **Outliers innovativos (AI):** ocurre si el error (o innovación)  $e_t$  se ve perturbado, de forma que  $e'_t = e_t + \omega_I P_t^T$ . Al perturbarse el término de innovación, se ven perturbados también los valores de la serie en tiempos posteriores.

Usando la representación de proc. lineal general  $Y_t = e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \dots$  vemos que

$$\begin{aligned} Y'_t &= e'_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \dots \\ &= [e'_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \dots] + \psi_{t-T} \omega_I = Y_t + \psi_{t-T} \omega_I \end{aligned}$$

# Detección de outliers

---

En lo que sigue vamos a ver técnicas para detectar la presencia de ambos tipos de outliers a través de tests estadísticos.

Para ello es conveniente utilizar la representación  $AR(\infty)$  de las series de tiempo para definir los residuos en términos de la serie con outliers:

$$a_t = Y'_t - \pi_1 Y'_{t-1} - \pi_2 Y'_{t-2} - \dots$$

Una vez que se encuentra un outlier se puede incorporar al modelo, y continuar el proceso de detección de outliers hasta que no se encuentren nuevos eventos.

# Detección de IO

---

Si tenemos un único IO en  $t=T$ ,  $a_t = \omega_I \mathbf{I}\{T = t\} + e_t$  y un buen estimador para  $\omega_I$  es  $\tilde{\omega}_I = a_T$ , con varianza  $\sigma_e^2$ . Luego, se puede definir el estadístico  $\lambda_{1,T} = \frac{a_T}{\sigma_e}$  para testear si hay un IO en  $t=T$ , donde

**H<sub>0</sub>**: no hay outliers en  $t=T$       vs      **H<sub>1</sub>**: hay un outlier en  $T=t$ .

Bajo  $H_0$ ,  $\lambda_{1,T} \approx \mathcal{N}(0, 1)$ .

Si conocemos el instante  $T$ , para un test de nivel significación  $\alpha$ , rechazaremos el test cuando  $|\lambda_{1,T}| > z_{1-\alpha/2}$ .

Si no conocemos el valor de  $T$  debemos calcular el test para todas las observaciones. Luego rechazaremos  $H_0$  cuando  $\lambda_1 = \max_{1 \leq t \leq n} |\lambda_{1,t}| > z_{1-\alpha/(2n)}$

Corrección de Bonferroni

# Detección de AO

Para detectar los AO, se puede mostrar que los residuos se pueden escribir como

$$a_t = -\omega_A \pi_{t-T} + e_t, \quad \pi_0 = -1, \quad \pi_j = 0 \text{ si } j < 0$$

Luego, se puede escribir de forma recursiva

$$a_T = \omega_A + e_T, \quad a_{T+1} = -\pi_1 \omega_A + e_{T+1}, \quad \dots, \quad a_n = -\pi_{n-T} \omega_A + e_n$$

y resolver el problema por cuadrados mínimos, obteniendo el estimador

$$\hat{\omega}_A = -\rho^2 \sum_{t=1}^n \pi_{t-T} a_t, \quad \rho = 1 + \pi_1^2 + \dots + \pi_n^2$$

con varianza  $\rho^2 \sigma_e^2$ . De esta forma,  $\lambda_{2,T} = \frac{\hat{\omega}_A}{\rho^2 \sigma_e^2}$  es un estadístico para las test

**H<sub>0</sub>**: la serie no presenta outliers      vs.      **H<sub>1</sub>**: existe un AO en T.

Bajo H<sub>0</sub>  $\lambda_{2,T} \approx \mathcal{N}(0, 1)$ . Nuevamente, si desconocemos el valor de T aplicamos la corrección de Bonferroni.

Si  $|\lambda_{1,T}| > |\lambda_{2,T}|$  diremos que es de tipo IO, y en caso contrario de tipo AO.

# Heterocedasticidad

# Motivación

---

Hasta ahora nos concentramos en la estructura de  $\mathbb{E}[Y_t | Y_{t-1}, Y_{t-2}, \dots, Y_1]$

Sin embargo también se puede hacer un estudio de la varianza condicional.

$\mathbb{V}[Y_t | Y_{t-1}, Y_{t-2}, \dots, Y_1]$  mide la incerteza de la desviación de  $Y_t$  de su esperanza condicional. Este tipo de análisis se gana relevancia cuando se comienza a querer modelar series financieras.

Si  $Y_t$  sigue un modelo ARIMA, 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 los valores presente y pasados de la serie.

Se desarrollan entonces modelos para la varianza condicional.

# Ejemplo

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

El rendimiento del día  $t$  se define como

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

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

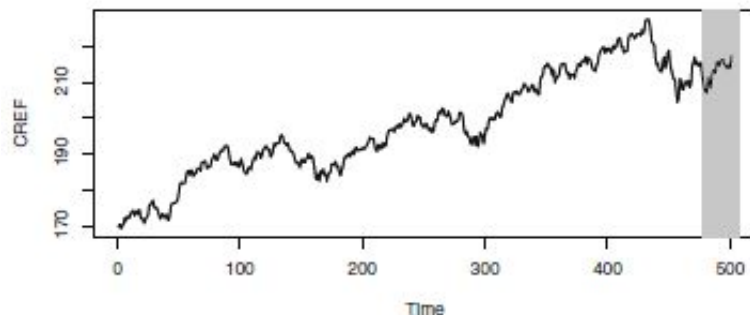
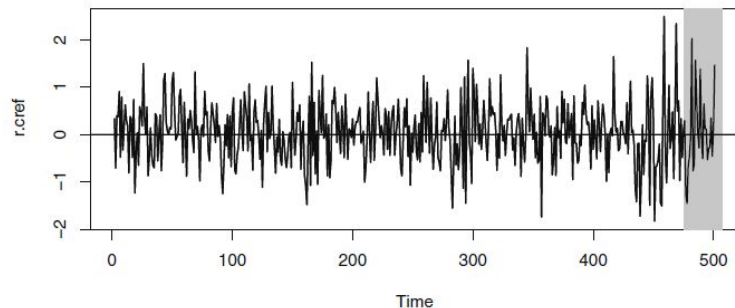


Exhibit 12.2 Daily CREF Stock Returns: August 26, 2004 to August 15, 2006





# Ejemplo

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

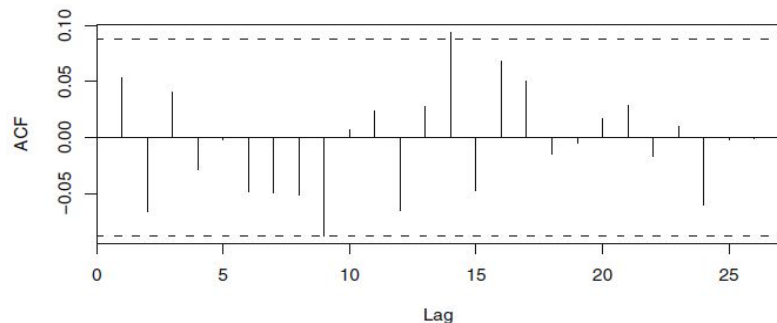
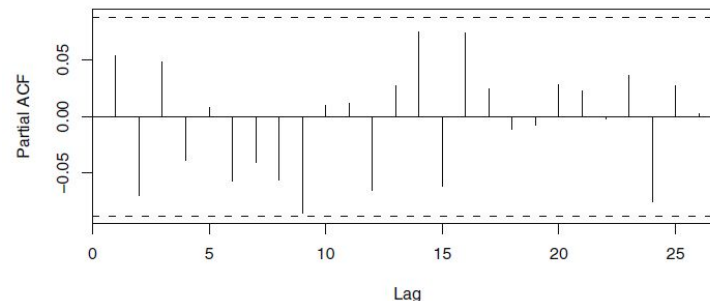


Exhibit 12.4 Sample PACF of Daily CREF Returns: 8/26/04 to 8/15/06



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

Vemos claramente que **independencia  $\neq$  descorrelación**.

Si la serie es independencia, la misma 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.
- **Tests estadísticos:** Podemos aplicar el test de Box-Ljung para determinar si el cuadrado de los datos es autocorrelacionada. En este caso, los grados de libertad para el test será el número de *taps* considerados. 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**

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

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

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

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

donde  $\{\varepsilon_t\}$  es una secuencia de v.a. i.i.d con distribución normal estándar, con  $\varepsilon_t$  indep. de  $r_{t-j}$ ,  $j \geq 1$ ;  $\alpha$  y  $\omega$  son constantes desconocidas.

# 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  $\eta_t$  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:

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

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:

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

o en términos del operador de backshift:

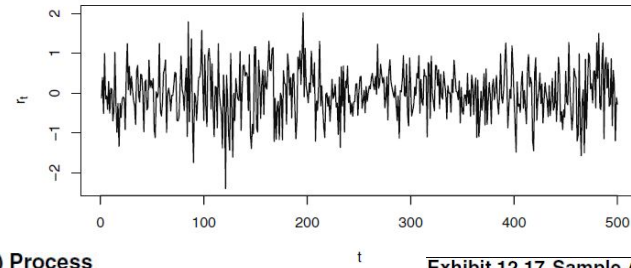
$$(1 - \beta_1 B - \dots - \beta_p B^p) \sigma_{t|t-1}^2 = \omega + (\alpha_1 B + \dots + \alpha_q B^q) r_t^2$$

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

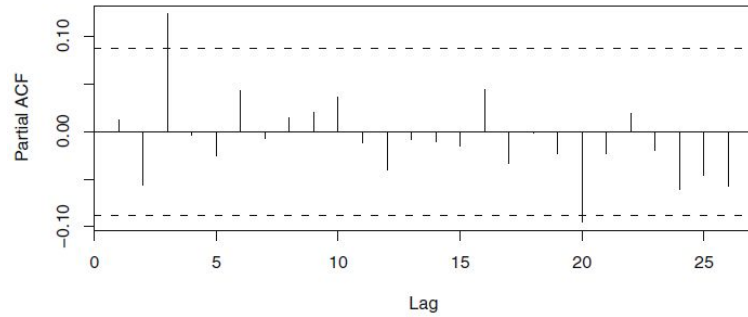
$$r_t^2 = \omega + (\beta_1 + \alpha_1) r_{t-1}^2 + (\beta_{\max(p,q)} + \alpha + \max(p,q) r^2) t - \max(p,q) + \eta_t - \beta_1 \eta_{t-1} - \dots - \beta_p \eta_{t-p}$$

$$\Rightarrow \{r_t^2\} \text{ sigue un modelo ARMA}(\max(p,q), q)$$

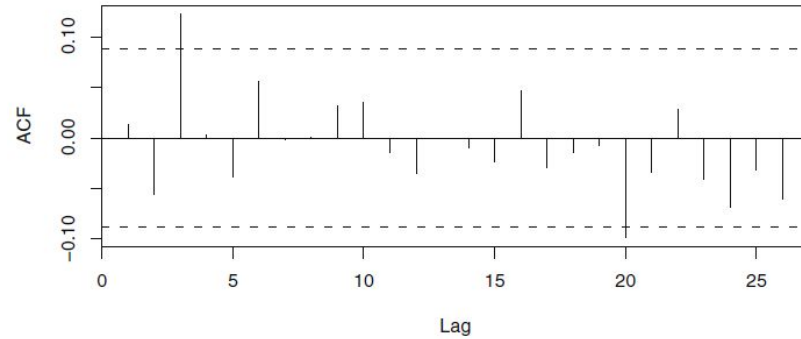
**Exhibit 12.12 Simulated GARCH(1,1) Process**



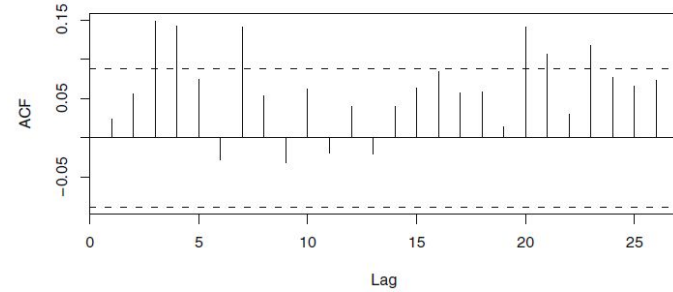
**Exhibit 12.14 Sample PACF of Simulated GARCH(1,1) Process**



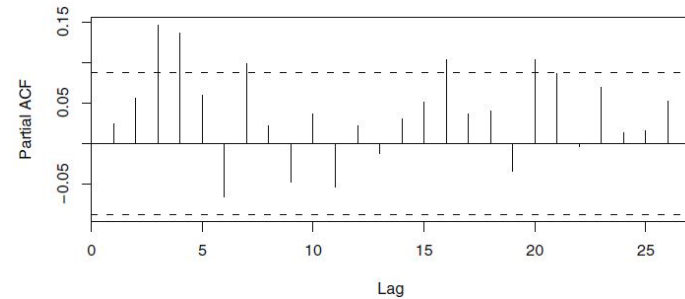
**Exhibit 12.13 Sample ACF of Simulated GARCH(1,1) Process**



**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

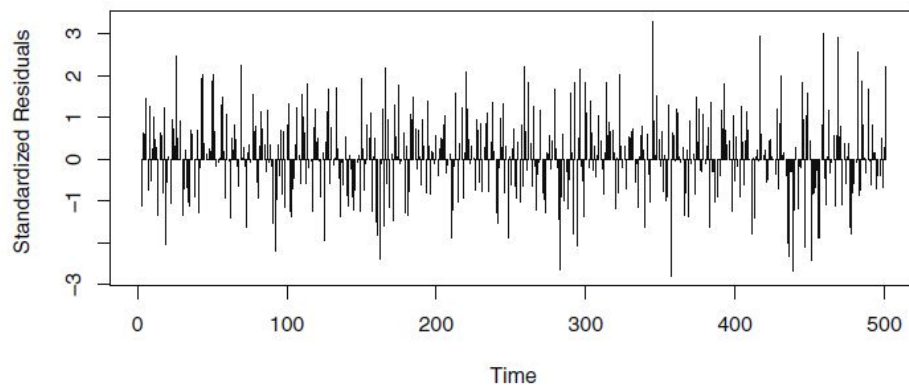
$$\hat{\varepsilon}_t = r_t / \hat{\sigma}_{t|t-1}$$

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

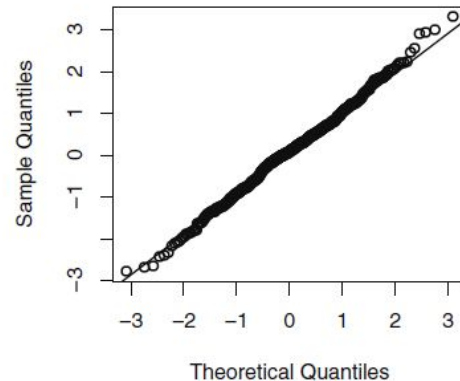
Para verificar el modelo podemos estudiar estos residuos estandarizados vamos a analizar

- 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**

