Análisis de Series Temporales y sus aplicaciones en R

Clase 1

Felipe Elorrieta López

Universidad de Santiago de Chile

Jun 28, 2023

Contenidos

¿Que es una serie temporal?

- En muchos fenómenos naturales o sociales podemos medir ciertas variables y asignar un valor numérico a cada observación. Cuando hacemos esto, podemos hablar de una serie de observaciones $\{y_t\}$ medidas secuencialmente en el tiempo.
- Por simplicidad, llamaremos serie de tiempo o serie cronologica a cualquier conjunto de observaciones numéricas $\{y_t, t \in T\}$ **ordenadas en el tiempo**, es decir la variable tiempo t es estrictamente creciente.

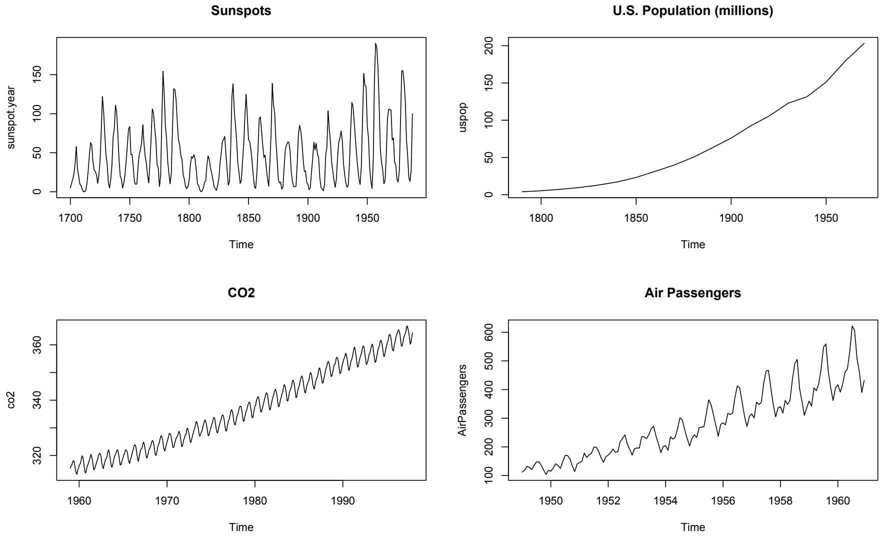
Componentes de series de tiempo

- Nivel
- Tendencia
- Estacionalidad
- Componente Aleatoria
- Lo cual puede ser expresado por la siguiente ecuación,

$$Y_t = N_t + T_t + S_t + \epsilon_t$$

- Este modelo lleva el nombre de modelo aditivo, ya que la serie de tiempo se puede descomponer en la suma de cada uno de estos componentes.

Componentes de series de tiempo



Procesos Estacionarios

- Un proceso estocástico $Y(t)$ se dice **estrictamente estacionario** (o estacionario en el sentido fuerte) si todas las distribuciones de dimensión finita permanecen iguales en el tiempo, es decir,

$$F(Y_{k+1}, Y_{k+2}, \ldots, Y_{k+n}) = F(Y_1, Y_2, \ldots, Y_n) \quad \forall n, k \in \mathbb{Z}$$

Procesos Estacionarios

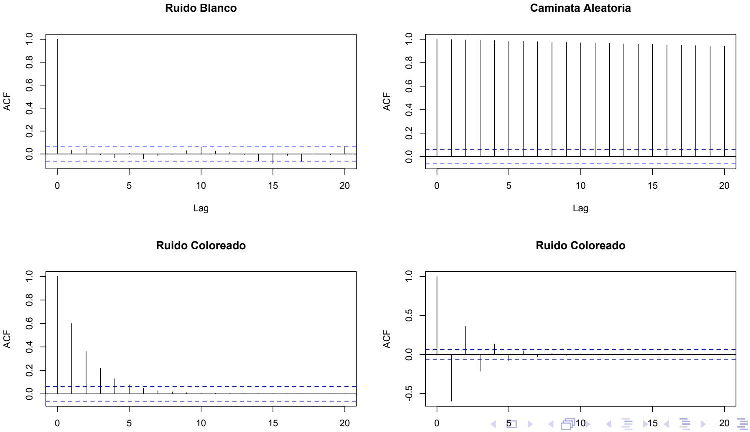
- Un proceso estocástico $Y(t)$ se dice **estrictamente estacionario** (o estacionario en el sentido fuerte) si todas las distribuciones de dimensión finita permanecen iguales en el tiempo, es decir,

$$F(Y_{k+1}, Y_{k+2}, \ldots, Y_{k+n}) = F(Y_1, Y_2, \ldots, Y_n) \quad \forall n, k \in \mathbb{Z}$$

- Un proceso estocástico $Y(t)$ se dice **debilemente estacionario** (o estacionario de segundo orden) si,
  1. $\mathbb{E}[Y_t] = \mu = cte < \infty \quad \forall t \in \mathcal{T}$
  2. $\mathbb{V}[Y_t] = \sigma^2 = cte < \infty \quad \forall t \in \mathcal{T}$
  3. Existe una función $\gamma(.)$ tal que $Cov(Y_t, Y_s) = \gamma(t - s) \quad \forall t, s \in \mathcal{T}$

Función de Autocorrelación

- La Función de Autocorrelación (ACF) nos permite estudiar la estructura de dependencia temporal de una serie de tiempo. En R se calcula como:
`acf(y,lag.max=20)`



Test de Box-Ljung

- Sea la Hipotesis

$$\begin{aligned} H_0 &: \rho(1) = \rho(2) = \ldots = \rho(L) = 0 \;\; \text{vs} \\ H_1 &: \exists \;\; i = 1, \ldots, L : \rho(i) \neq 0 \end{aligned}$$

- el estadístico de Box-Lung es el siguiente,

$$Q_y = n(n+2) \sum_{j=1}^{L} \frac{\hat{\rho}^2(j)}{n-j}$$

  donde $\hat{\rho}(j)$ es la autocorrelación estimada para el j-ésimo lag.
- bajo H_0, $Q \sim \chi^2_L$.

Modelo ARMA

- Sea un proceso estocástico $\{Y_t\}, \quad t \in T$ definido por la ecuación

$$Y_t - \phi_1 Y_{t-1} - \ldots - \phi_p Y_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} \quad (1)$$

el cual es una combinación de un proceso AR(p) y un proceso MA(q) se conoce como "Modelo Autoregressivo de medias móviles"(ARMA(p, q)),
donde $\{\epsilon_t\} \sim RB$ y $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ son coeficientes fijos (a estimar).

Modelo ARMA

- Sea un proceso estocástico $\{Y_t\}, \quad t \in T$ definido por la ecuación

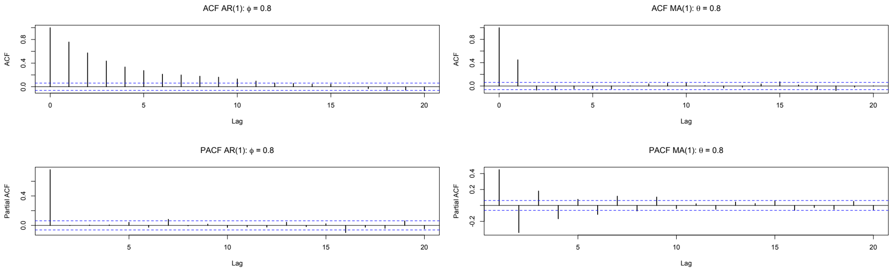
$$Y_t - \phi_1 Y_{t-1} - \ldots - \phi_p Y_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} \quad (1)$$

  el cual es una combinación de un proceso AR(p) y un proceso MA(q) se conoce como "Modelo Autoregressivo de medias móviles"(ARMA(p, q)),
  donde $\{\epsilon_t\} \sim RB$ y $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ son coeficientes fijos (a estimar).
- Note que si $p = 0$ el modelo ARMA(p, q) se reduce a un modelo MA(q) y si $q = 0$ el modelo ARMA(p, q) se reduce a un modelo AR(p)

Modelo ARMA

- Se puede identificar el proceso ARMA apropiado para nuestros datos usando la ACF y la PACF bajo la siguiente estructura,

Modelo	ACF	PACF
AR(p)	Decaim. Exp	$0 \quad \forall k > p$
MA(q)	$0 \quad \forall k > q$	Decaim. Exp
ARMA(p,q)	Decaim. Exp	Decaim. Exp



Raiz Unitaria

- En un proceso $y_t \sim AR(1)$ definido por $y_t = \phi y_{t-1} + \epsilon_t$, si $|\phi| < 1$ entonces $\{y_t\}$ es estacionario.
- Que sucede si $\phi = 1$?

Raiz Unitaria

- En un proceso $y_t \sim AR(1)$ definido por $y_t = \phi y_{t-1} + \epsilon_t$, si $|\phi| < 1$ entonces $\{y_t\}$ es estacionario.
- Que sucede si $\phi = 1$?
- Tenemos que $y_t = y_{t-1} + \epsilon_t$. En este caso decimos que y_t es un proceso de caminata aleatoria.
- Un proceso de caminata aleatoria es un proceso no estacionario.

Raiz Unitaria

- En un proceso $y_t \sim AR(1)$ definido por $y_t = \phi y_{t-1} + \epsilon_t$, si $|\phi| < 1$ entonces $\{y_t\}$ es estacionario.
- Que sucede si $\phi = 1$?
- Tenemos que $y_t = y_{t-1} + \epsilon_t$. En este caso decimos que y_t es un proceso de caminata aleatoria.
- Un proceso de caminata aleatoria es un proceso no estacionario.
- Los test de Raíz unitaria permiten verificar la siguiente hipótesis:

$$H_0 : \phi = 1 \quad \text{vs} \quad H_1 : \phi < 1$$

- o equivalentemente,

$$H_0 : \text{Caminata Aleatoria} \quad \text{vs} \quad \text{Proceso Estacionario}$$

el más conocido de estos test, es el test de Dickey Fuller.

Modelos ARIMA

- Sea $x_t = (1-B)^d y_t$ y suponga que $x_t \sim ARMA(p,q)$, es decir

$$\Phi_p(B)x_t = \Theta_q(B)\epsilon_t$$

  donde $\epsilon_t \sim RB(0,\sigma^2)$, entonces podemos reescribir el proceso como,

$$\Phi_p(B)(1-B)^d y_t = \Theta_q(B)\epsilon_t$$

- Este proceso $\{y_t\}$ se conoce como "Autoregresivo integrado de medias móviles" y se denota como $\{y_t\} \sim$ ARIMA(p,d,q)

Modelo SARIMA

- Decimos que $\{y_t\}$ sigue un proceso SARIMA$(p, d, q) \times (P, D, Q)_S$ si puede escribirse como,

$$\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D Y_t = \theta_q(B)\Theta_Q(B^S)\epsilon_t \quad (2)$$

- donde

$$\begin{aligned}
\phi_p(B) &= 1 - \phi_1 B - \ldots - \phi_p B^p \\
\Phi_P(B^S) &= 1 - \Phi_1 B^S - \ldots - \Phi_P B^{PS} \\
\theta_q(B) &= 1 + \theta_1 B + \ldots + \theta_q B^q \\
\Theta_Q(B^S) &= 1 + \Theta_1 B + \ldots + \Theta_Q B^{QS}
\end{aligned}$$

Modelo SARIMA

usach

- Decimos que $\{y_t\}$ sigue un proceso SARIMA$(p, d, q) \times (P, D, Q)_S$ si puede escribirse como,

Polinomio Autoregresivo Orden de Integracion Polinomio de Medias Moviles

$$\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D Y_t = \theta_q(B)\Theta_Q(B^S)\epsilon_t \quad (2)$$

- donde

Polinomio Autoregresivo Estacional

Orden de Integracion Estacional

Polinomio de Medias Moviles Estacional

$$\begin{aligned}
\phi_p(B) &= 1 - \phi_1 B - \ldots - \phi_p B^p \\
\Phi_P(B^S) &= 1 - \Phi_1 B^S - \ldots - \Phi_P B^{PS} \\
\theta_q(B) &= 1 + \theta_1 B + \ldots + \theta_q B^q \\
\Theta_Q(B^S) &= 1 + \Theta_1 B + \ldots + \Theta_Q B^{QS}
\end{aligned}$$

Metodología de Box y Jenkins

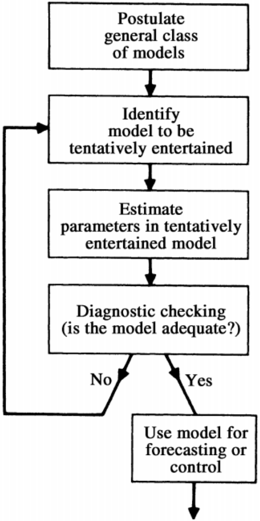
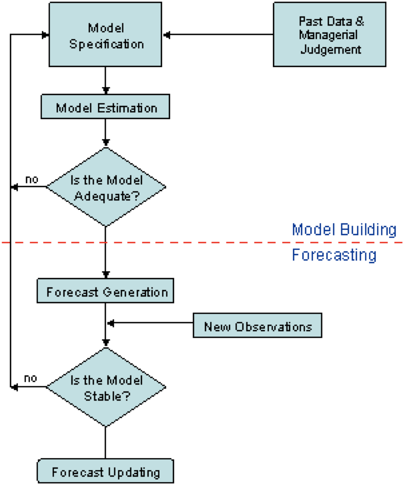


FIGURE 1.7 Stages in the iterative approach to model building.

Metodología de Box y Jenkins



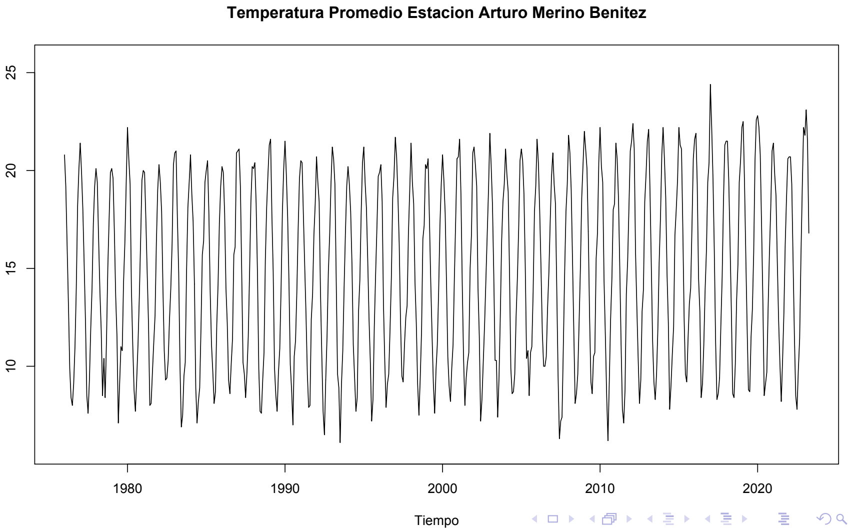
Forecasting System:
The Model-Building and The Forecasting Phases

Librerias de Series Temporales en R

- **TSA**: Cryer and Chan (2010, 2nd ed) Time series analysis with applications in R
- **astsa**: Shumway and Stoffer (2017, 4th ed) Time Series Analysis and Its Applications: With R Examples
- **forecast**: by Rob Hyndman et al.
- **fUnitRoots**: by Diethelm Wuertz et al.
- **rugarch**: by Alexios Galanos
- **lmtest**: by Torsten Hothorn et al.

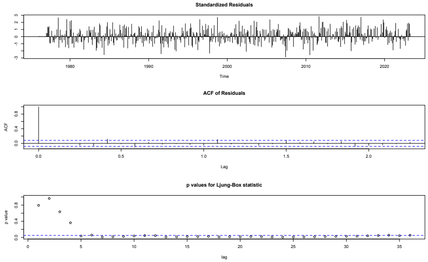
Aplicación

- Serie de Temperatura Promedio Mensual medida en la estación Arturo Merino Benitez entre Enero de 1976 y Abril de 2023.



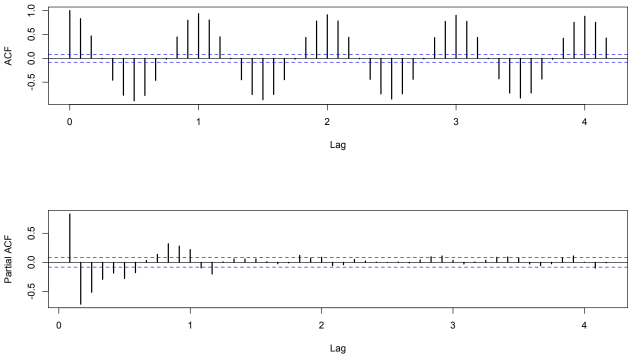
Función Auto-ARIMA

```
model=auto.arima(Temp)
Series: Temp
ARIMA(2,0,2)(0,1,1)[12]
Coefficients:
         ar1      ar2      ma1     ma2     sma1
      1.4955  -0.5991  -1.2706  0.4657  -0.8675
s.e.  0.2599   0.2224   0.2723  0.1886   0.0272
sigma^2 = 0.6877:  log likelihood = -690.89
AIC=1393.78   AICc=1393.93   BIC=1419.7
```



Aplicación

```
y=log(Temp)
par(mfrow=c(2,1))
acf(y,lag.max=50,lwd=2,main="")
pacf(y,lag.max=50,lwd=2,main="")
```



Test de Raíz Unitaria

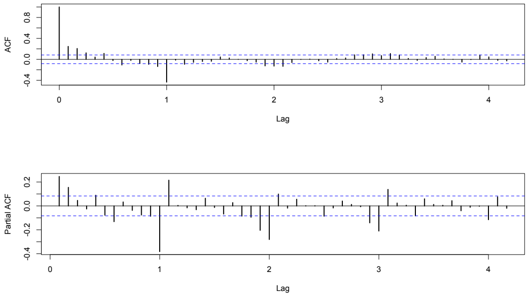
```
require(fUnitRoots)
adfTest(y,lags = 1,type = c("nc"))

Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 1
  STATISTIC:
    Dickey-Fuller: -1.7525
  P VALUE:
    0.07992
```

```
adfTest(y,lags = 12,type = c("nc"))

Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 12
  STATISTIC:
    Dickey-Fuller: 0.4029
  P VALUE:
    0.7447
```

Test de Raíz Unitaria

Serie	X_t	∇X_t	$\nabla_{12} X_t$	$\nabla_{12}\nabla X_t$	$\nabla^2 X_t$	$\nabla_{12}^2 \nabla X_t$
adfTest(., lags = 1)	0.08	0.01	0.01	0.01	0.01	0.01
adfTest(., lags = 12)	0.75	0.01	0.01	0.01	0.01	0.01
Stdev	0.116	0.039	**0.009**	0.014	0.033	0.027

- La solucion de minima varianza es $d = 0$ y $D = 1$.



Modelo Propuesto

- $\text{SARIMA}(1, 0, 1) \times (0, 1, 1)_{12}$

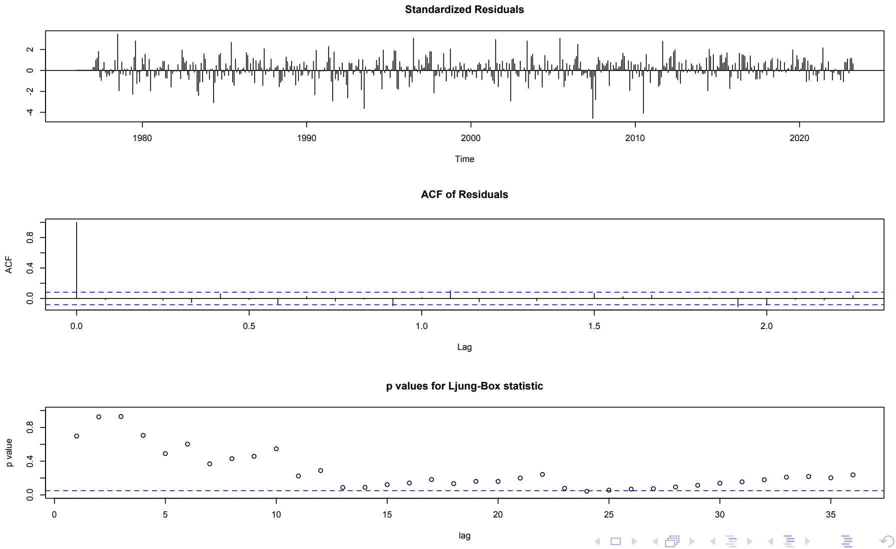
```
model4=arima(y,order=c(1,0,1),seasonal=list(order=c(0,1,1),period=1
summary(model4)
Call:
arima(x = y, order = c(1, 0, 1), seasonal = list(order = c(0, 1, 1)
, period = 12))
Coefficients:
         ar1      ma1     sma1
      0.7471  -0.5213  -0.8958
s.e.  0.0836   0.1033   0.0264
```

- Significancia de Parametros

```
require(lmtest)
coeftest(model4)
z test of coefficients:

      Estimate Std. Error  z value  Pr(>|z|)
ar1   0.747089   0.083583   8.9383 < 2.2e-16 ***
ma1  -0.521283   0.103272  -5.0477 4.472e-07 ***
sma1 -0.895826   0.026441 -33.8801 < 2.2e-16 ***
```

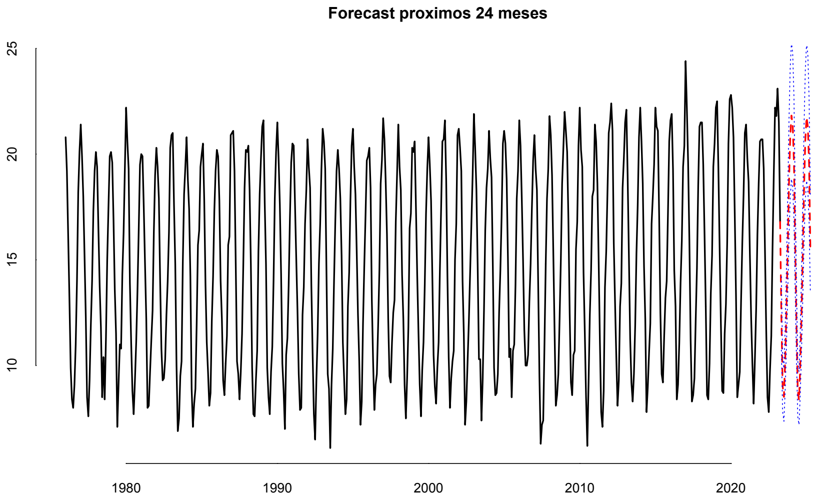
Diagnostico del Modelo

```
tsdiag(model4,gof.lag=36)
```



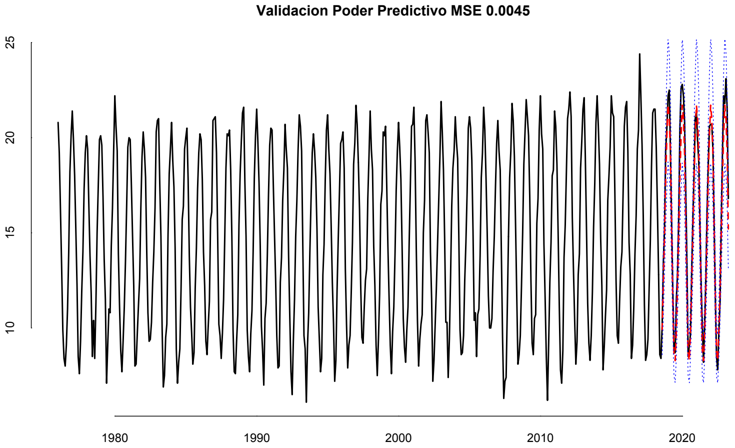
Forecast

```
Forecast<-predict(model4,n.ahead=24)
```



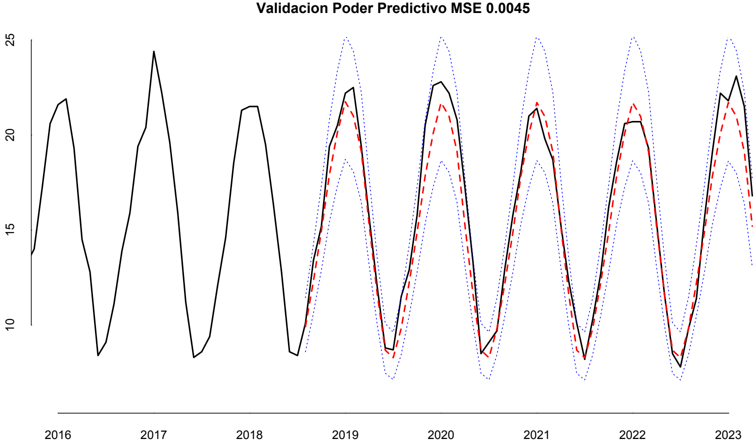
Validacion

- Para evaluar la capcidad predictiva del modelo propuesto, usaremos el primer 90 % de la serie para entrenar el modelo propuesto y el 10 % para validarlo.



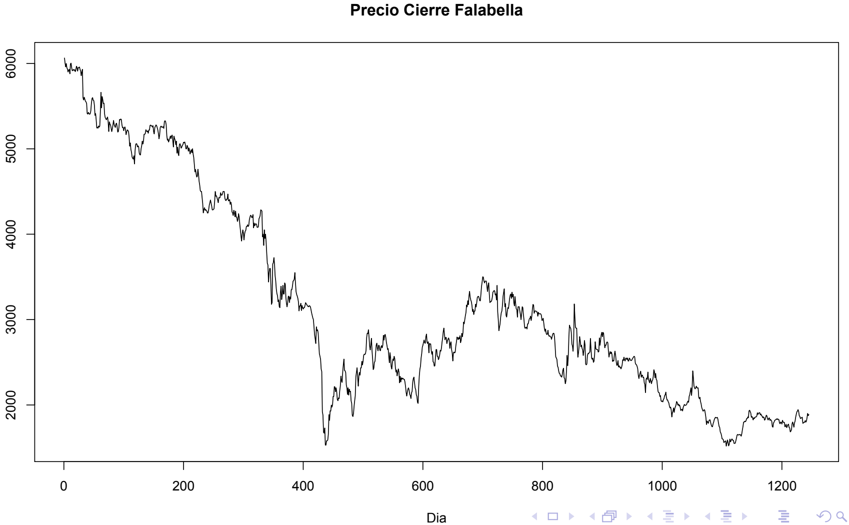
Validacion

- Para evaluar la capcidad predictiva del modelo propuesto, usaremos el primer 90 % de la serie para entrenar el modelo propuesto y el 10 % para validarlo.



Aplicación en Series de Tiempo Financieras

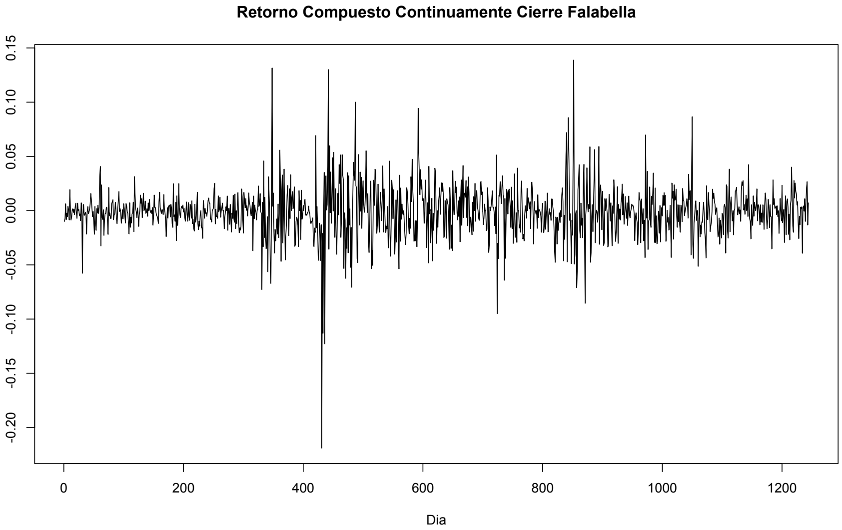
- Serie del Precio de cierre diario de Falabella en la Bolsa de Santiago observado entre el 14 jun 2018 y el 14 jun 2023.



Retornos de Activos

- Sea P_t: El precio de un activo.
- Se define el retorno simple a un periodo: desde la fecha $t-1$ a t como,
$$R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}}$$
- También es posible obtener el retorno compuesto continuamente a través de: $R_t = \log P_t - \log P_{t-1}$. Este es el retorno comunmente usado.

Retornos de Activos



Hechos Estilizados

- Las series financieras suelen presentar las siguientes características, generalmente conocidos como los hechos estilizados (stylized-facts).
  1. **Ausencia de autocorrelación significativa:** Las series financieras (retornos), exhiben bajo nivel de autocorrelación, por lo cual tienen bajo nivel de predicción.
  2. **Distribuciones con colas pesadas:** Por lo general las series financieras tienen mucho más kurtosis que la de un ruido blanco gaussiano.
  3. **Agrupamiento de la volatilidad:** La volatilidad es persistente y puede ser alta para ciertos periodos de tiempo y baja para otros. Esta característica se puede ver reflejada en las autocorrelaciones de la serie al cuadrado significativamente distintas de cero.

Proceso GARCH

- El proceso GARCH(p,q) está dado por.

$$\begin{aligned} r_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \end{aligned}$$

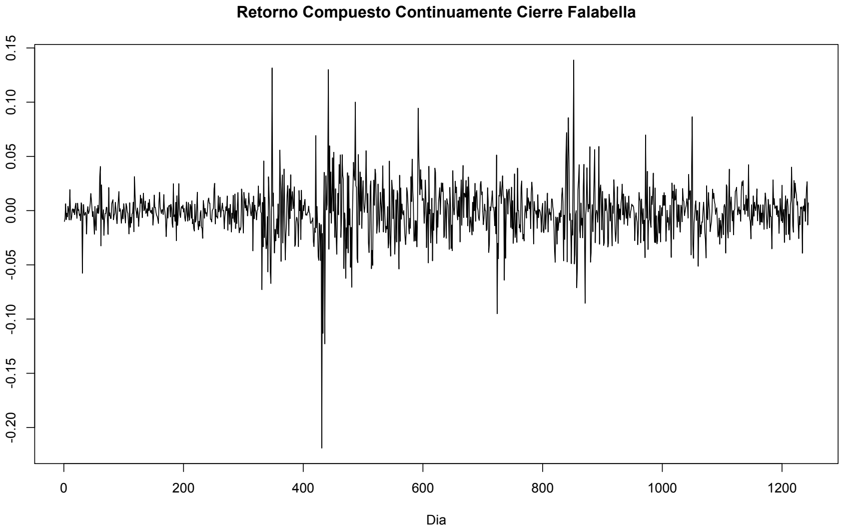
.

- donde ϵ_t *iid* $N(0,1)$, $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, \ldots, p$, $\beta_j \geq 0$, $j = 1, \ldots, q$ y $\sum_{i=1}^{\text{máx}\{p,q\}} (\alpha_i + \beta_j) < 1$.
- Si $q = 0$, el modelo anterior se reduce al siguiente modelo ARCH(p).

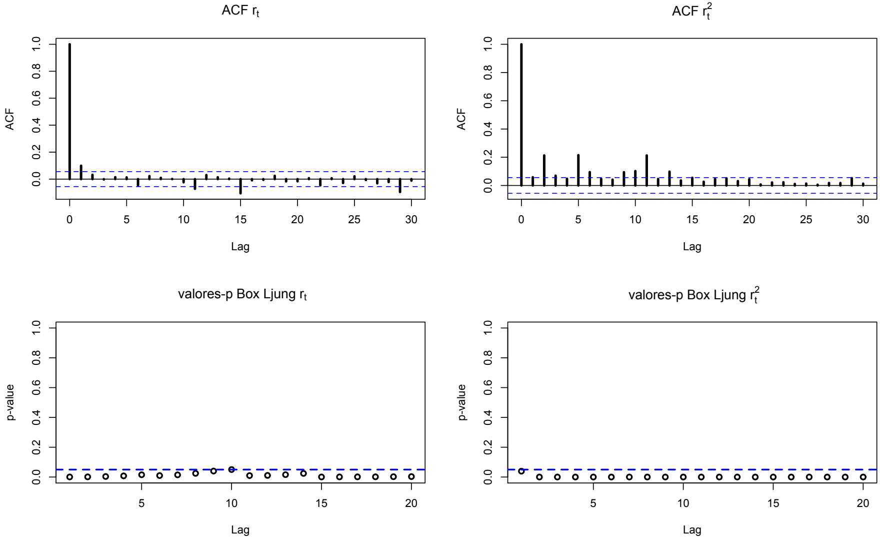
$$\begin{aligned} r_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 r_{t-1}^2 + \alpha_2 r_{t-2}^2 + \ldots + \alpha_p r_{t-p}^2 \end{aligned}$$

.

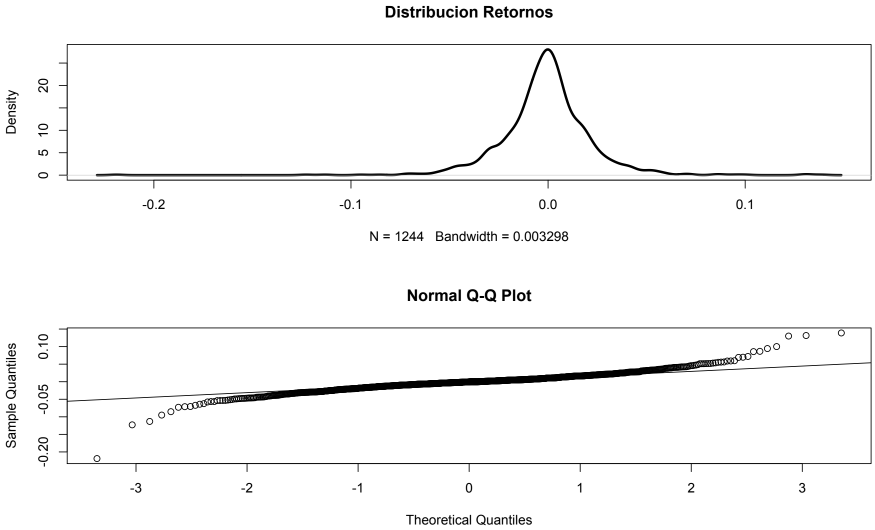
Hechos Estilizados



Hechos Estilizados



Hechos Estilizados



Ajuste Modelo GARCH

```
library(rugarch)
spec<-ugarchspec(variance.model=list(model="sGARCH",garchOrder = c(1, 1)),
mean.model=list(armaOrder=c(0,0),include.mean=FALSE),distribution.model="norm")
fit=ugarchfit(data=rtn,spec=spec)
show(fit)

Conditional Variance Dynamics
-----------------------------------
GARCH Model : sGARCH(1,1)
Mean Model : ARFIMA(0,0,0)
Distribution : norm
Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
omega   0.000014    0.000003   4.7746    2e-06
alpha1  0.171198    0.023661   7.2354    0e+00
beta1   0.816755    0.018999  42.9885    0e+00

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic  p-value
Lag[1]                      8.649 0.003272
Lag[2*(p+q)+(p+q)-1][2]     8.698 0.004225
Lag[4*(p+q)+(p+q)-1][5]     8.869 0.017864
```

Modelo ARMA-GARCH

- Las series financieras que presentan baja correlación se le ajusta un modelo ARMA cuyas innovaciones son un proceso GARCH. Por lo tanto el modelo $ARMA(p_1, q_1) - GARCH(p_2, q_2)$ a utilizar esta dado por,

$$
\begin{aligned}
r_t &= \phi_1 r_{t-1} + \phi_1 r_{t-2} + \ldots + \phi_{p_1} r_{t-p_1} + \theta_1 \epsilon_{t-1} + \ldots + \theta_{q_1} \epsilon_{t-q_1} \\
\epsilon_t &= \nu_t \epsilon_t \\
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p_2} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q_2} \beta_j \sigma_{t-j}^2
\end{aligned}
$$

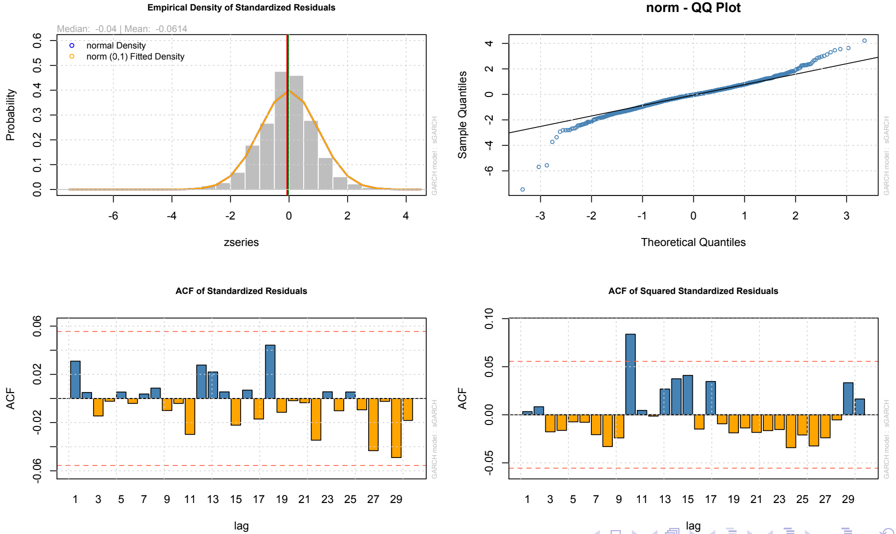
.

Ajuste Modelo ARMA-GARCH

```
spec<-ugarchspec(variance.model=list(model="sGARCH",garchOrder = c(1, 1)),
mean.model=list(armaOrder=c(1,0),include.mean=FALSE),distribution.model="norm")
fit=ugarchfit(data=rtn,spec=spec)
show(fit)
Conditional Variance Dynamics
-----------------------------------
GARCH Model : sGARCH(1,1)
Mean Model : ARFIMA(1,0,0)
Distribution : norm
Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
ar1     0.064653    0.031264   2.0680 0.038641
omega   0.000014    0.000003   4.7405 0.000002
alpha1  0.170418    0.023587   7.2252 0.000000
beta1   0.817501    0.019169  42.6472 0.000000
Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[4*(p+q)+(p+q)-1][5]     1.385  0.8737
Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[2*(p+q)+(p+q)-1][5]   0.46432  0.9630
```

Análisis de Residuos

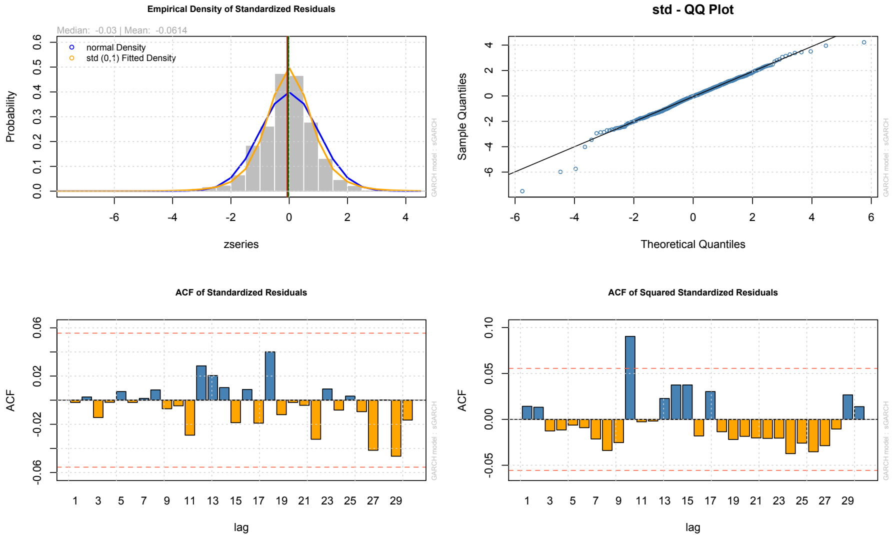
```
plot(fit)
```



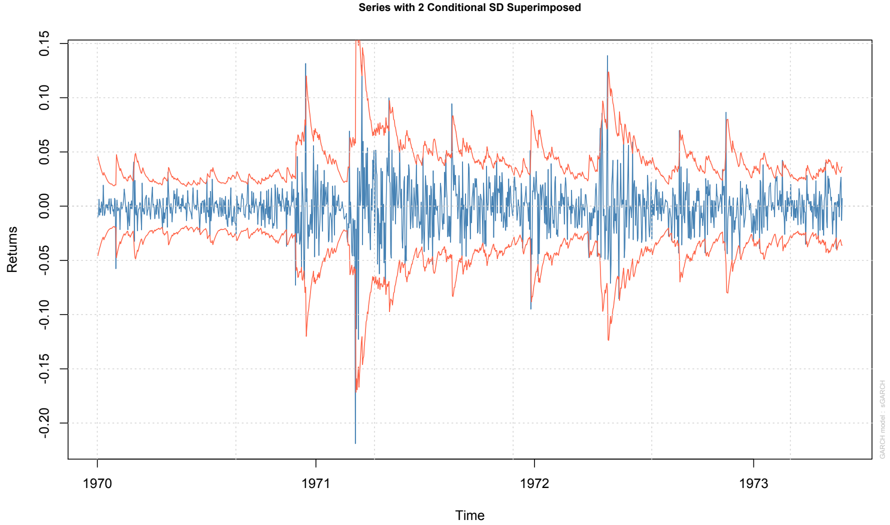
Ajuste Modelo ARMA-GARCH con errores no-normales

```
spec<-ugarchspec(variance.model=list(model="sGARCH",garchOrder = c(1, 1)),
mean.model=list(armaOrder=c(1,0),include.mean=FALSE),distribution.model="std")
fit=ugarchfit(data=rtn,spec=spec)
show(fit)
Conditional Variance Dynamics
-----------------------------------
GARCH Model : sGARCH(1,1)
Mean Model : ARFIMA(1,0,0)
Distribution : std

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
ar1     0.101349    0.029373   3.4504 0.000560
omega   0.000008    0.000004   1.9136 0.055676
alpha1  0.139902    0.028678   4.8784 0.000001
beta1   0.859096    0.026840  32.0085 0.000000
shape   4.742920    0.688817   6.8856 0.000000
```

Análisis de Residuos

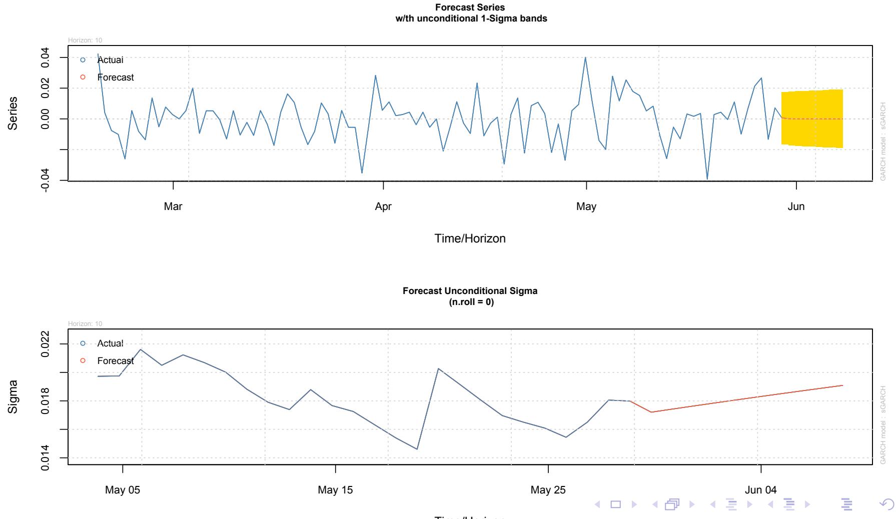


Volatilidad Estimada



Forecast

```
forc = ugarchforecast(fit, n.ahead=10)
plot(forc)
```




SOCHE
SOCIEDAD CHILENA DE ESTADÍSTICA
MINICURSO ONLINE
ANÁLISIS DE SERIES TEMPORALES Y SUS APLICACIONES EN R
FELIPE ELORRIETA
PROFESOR ASOCIADO USACH
Miércoles
5 de julio
11:30
TEMARIO
• Aplicaciones en Finanzas 2:
◦ Análisis Multivariado de Series Temporales.
◦ Modelos VAR-VARMA
• Aplicaciones en Salud
◦ Análisis de Series de Tiempo Interrumpidas
• Aplicaciones en Astronomía:
◦ Series de Tiempo Irregularmente Observadas.
◦ Periodograma
◦ Modelos iAR
CERTIFICADO GRATUITO PARA SOCIOS ACTIVOS SOCHE
PERSONAS EXTERNAS A LA SOCHE TIENE UN COSTO DE $5.000.
INSCRIPCIONES EN NUESTRA WEB WWW.SOCHE.CL

felipe.elorrieta@usach.cl
@felipeelorrieta