Análisis de Series Temporales y sus aplicaciones en R

Clase 2

Felipe Elorrieta López

Universidad de Santiago de Chile

Jul 5, 2023

Contenidos

Análisis Multivariado de Series de Tiempo

- La globalización económica y la comunicación por Internet han acelerado la integración de los mercados financieros mundiales en los últimos años. Los movimientos de precios en un mercado pueden propagarse fácil e instantáneamente a otro mercado. Por esta razón, los mercados financieros son más dependientes entre sí que nunca, y hay que considerarlos conjuntamente para comprender mejor la estructura dinámica de las finanzas mundiales

ARIMAX (Transfer Function Model)

- El modelo de función de transferencia general fue discutido por Box y Tiao (1975). Cuando un modelo ARIMA incluye otra serie de tiempo como variable de entrada, el modelo se denomina a veces como un modelo Arimax.
- Varios nombres diferentes se utilizan para describir modelos ARIMA con una serie de entrada independiente.
  - Distributed lag models: $DL(p) : y_t = \nu(B)x_t + \epsilon_t$.
  - Autoregressive distributed lag models: $ADL(p) : \phi(B)y_t = \nu(B)x_t + \epsilon_t$.
  - ARMA model with exogenous explanatory variable ARMAX (ARIMAX): $\phi(B)y_t = \nu(B)x_t + \theta(L)\epsilon_t$.
  - Rational distributed lag model RDL: $y_t = \frac{\nu(B)}{\phi(B)}x_t + \theta(L)\epsilon_t$.
  - Transfer function: $y_t = \frac{\nu(B)}{\phi(B)}x_t + \epsilon_t$.

Modelo VAR

- El modelo de series de tiempo multivariado más comúnmente usado es el vector autoregressive (VAR) model. Una serie de tiempo multivariada r_t sigue un modelo VAR de orden p, VAR(p) si,

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + \epsilon_t$$

- donde ϕ_0 es un vector constante k-dimensional y ϕ_i es una matriz de $k \times k$. ϵ_t es una secuencia de vectores de ruido blanco con media cero y matriz de covarianza Σ.

Modelo VAR(1) bivariado

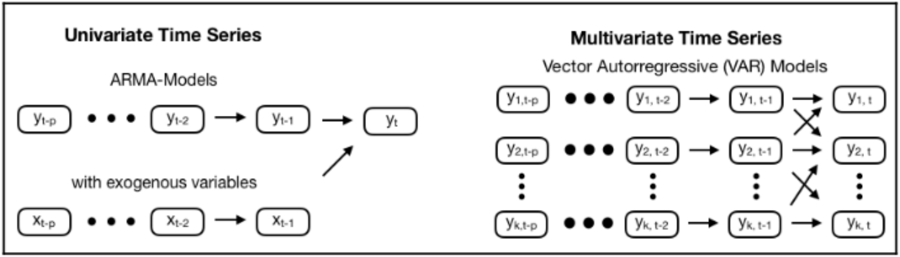
- Considere el modelo VAR(1) bivariado definido por:

$$r_t = \begin{pmatrix} r_{1t} \\ r_{2t} \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} r_{1,t-1} \\ r_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

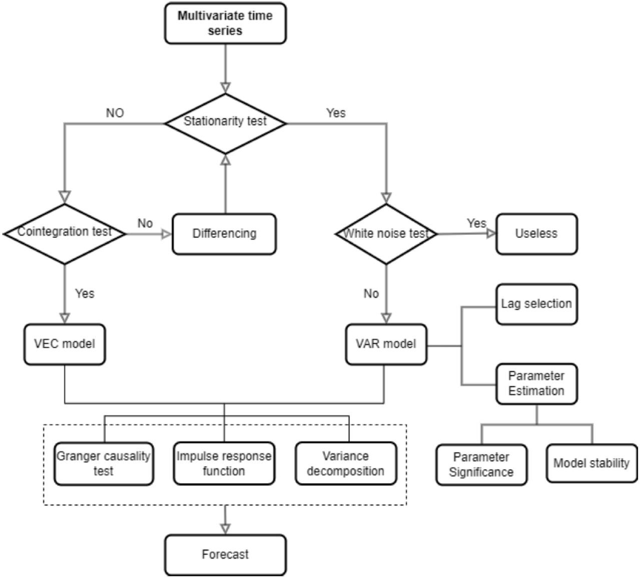
- o equivalentemente:

$$\begin{aligned} r_{1t} &= \phi_{11} r_{1,t-1} + \phi_{12} r_{2,t-1} + \epsilon_{1t} \\ r_{2t} &= \phi_{21} r_{1,t-1} + \phi_{22} r_{2,t-1} + \epsilon_{2t} \end{aligned}$$

Modelo VAR


Univariate Time Series
ARMA-Models
y_{t-p}
y_{t-2}
y_{t-1}
y_t
with exogenous variables
x_{t-p}
x_{t-2}
x_{t-1}
Multivariate Time Series
Vector Autorregressive (VAR) Models
$y_{1,t-p}$
$y_{1,t-2}$
$y_{1,t-1}$
$y_{1,t}$
$y_{2,t-p}$
$y_{2,t-2}$
$y_{2,t-1}$
$y_{2,t}$
$y_{k,t-p}$
$y_{k,t-2}$
$y_{k,t-1}$
$y_{k,t}$

Modelo VAR



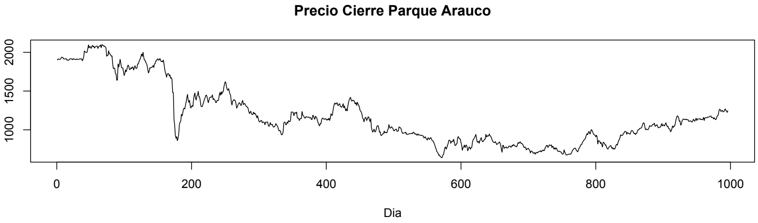
Multivariate time series
Stationarity test
NO
Yes
Cointegration test
No
Differencing
White noise test
Yes
Useless
Yes
No
Lag selection
VEC model
VAR model
Parameter Estimation
Granger causality test
Impulse response function
Variance decomposition
Parameter Significance
Model stability
Forecast

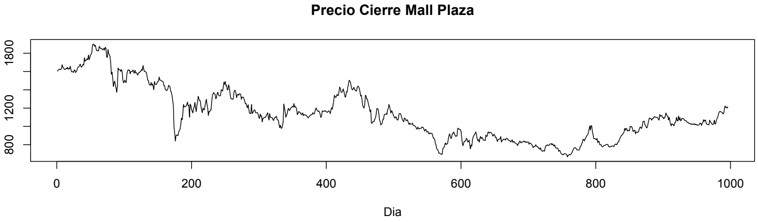
Librerías de Series Temporales en R

- **vars**: VAR Modelling by Bernhard Pfaff and Matthieu Stigler.
- **fUnitRoots**: by Diethelm Wuertz et al.
- **lmtest**: by Torsten Hothorn et al.
- **strucchange**: Testing, Monitoring, and Dating Structural Changes by Achim Zeileis et al.
- **RobPer**: Robust Periodogram and Periodicity Detection Methods by Jonathan Rathjens et al.
- **iAR**: Irregularly Observed Autoregressive Models by Felipe Elorrieta et al.

Aplicación en Series de Tiempo Financieras

- Serie del Precio de cierre diario de las acciones de Parque Arauco y Mall Plaza en la Bolsa de Santiago observado entre el 3 jul 2019 y el 29 jun 2023.





Cointegración

- Dos procesos integrados x_t y y_t se dicen cointegrados si existe un parámetro α tal que:

$$u_t = y_t - \alpha x_t \tag{1}$$

  es un proceso estacionario.
- Por lo tanto, una forma de chequear si dos series temporales están cointegradas es usando un test de raíz unitaria.
- Si rechazamos la raíz unitaria en u_t, entonces x_t y y_t están cointegradas. Si no rechazamos, x_t y y_t no están cointegradas
- El parámetro α se conoce como el parámetro de cointegración.
- El parámetro de cointegración se puede estimar usando Mínimos Cuadrados Ordinarios (MCO).

Cointegración

- Antes de implementar la prueba de cointegración de Engle y Granger debemos chequear que ambas series son integradas:

```
adfTest(data, lags = 1, type = c("nc"))
  P VALUE:
    0.2246
adfTest(data2, lags = 1, type = c("nc"))
  P VALUE:
    0.326
```

Cointegración

- Antes de implementar la prueba de cointegración de Engle y Granger debemos chequear que ambas series son integradas:

```
adfTest(data, lags = 1, type = c("nc"))
  P VALUE:
    0.2246
adfTest(data2, lags = 1, type = c("nc"))
  P VALUE:
    0.326
```

- Como ambas series son integradas, entonces revisamos si los residuos de una regresión lineal entre ambas series es un proceso estacionario:

```
model=lm(data~data2)
adfTest(residuals(model), lags = 1, type = c("nc"))
  P VALUE:
    0.01
```

Test de Causalidad de Granger

- Para estudiar la causalidad de Granger en R se debe usar el siguiente código:

  ```
  grangertest(formula, order, data)
  ```

- Si dos series temporales, x_t y y_t, están cointegradas, debe existir causalidad de Granger de x_t a y_t, de y_t a x_t, o en ambos sentidos.
- La presencia de causalidad de Granger en una o ambas direcciones entre x_t y y_t no implica necesariamente que las series están cointegradas.

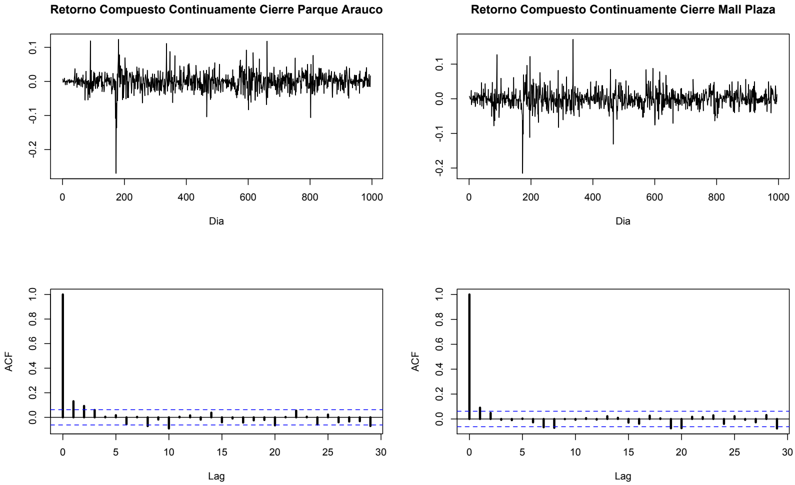
Test de Causalidad de Granger

- En nuestro ejemplo el resultado fue el siguiente:

```
grangertest(data ~ data2,order=1,multidata)
Granger causality test

Model 1: data ~ Lags(data, 1:1) + Lags(data2, 1:1)
Model 2: data ~ Lags(data, 1:1)
  Res.Df Df      F Pr(>F)
1    992
2    993 -1 1.5171 0.2183
grangertest(data2 ~ data,order=1,multidata)
Granger causality test

Model 1: data2 ~ Lags(data2, 1:1) + Lags(data, 1:1)
Model 2: data2 ~ Lags(data2, 1:1)
  Res.Df Df      F   Pr(>F)
1    992
2    993 -1 10.454 0.001265 **
```

Retornos

- Como la series de precios son integradas, intentaremos modelar su relación a partir de los retornos compuestos continuamente:



Test de Causalidad de Granger

- En el caso de los retornos compuestos continuamente notamos que la causalidad existe en ambas direcciones:

```
adfTest(rtn, lags = 1, type = c("nc"))
  P VALUE:
    0.01
adfTest(rtn2, lags = 1, type = c("nc"))
  P VALUE:
    0.01
grangertest(rtn ~ rtn2,order=1)
  Res.Df Df     F    Pr(>F)
1    991
2    992 -1 14.46 0.0001519 ***
  Res.Df Df      F    Pr(>F)
1    991
2    992 -1 30.999 3.324e-08 ***
```

Función de Correlación Cruzada

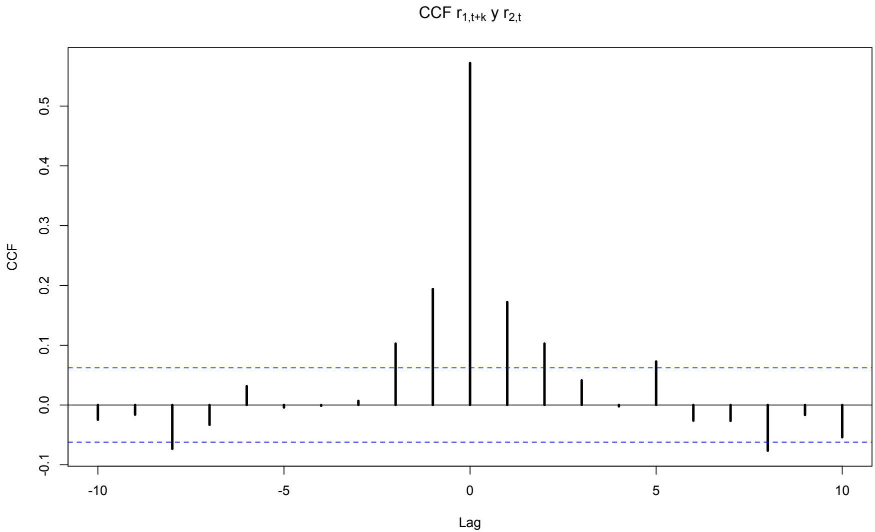
- Considere una serie de tiempo k-dimensional $r_t = (r_{1t}, \ldots, r_{kt})'$. Suponga además que la serie de tiempo r_t es débilmente estacionaria.
- La matriz de correlación contemporánea (o lag-zero) de r_t es la matriz cuyo (i,j)-ésimo elemento se define como,

$$\rho_{ij}(0) = \frac{\Gamma_{ij}(0)}{\sqrt{\Gamma_{ii}(0)}\sqrt{\Gamma_{jj}(0)}} = \frac{\mathbb{C}(r_{it}, r_{jt})}{\sqrt{\mathbb{V}[r_{it}]}\sqrt{\mathbb{V}[r_{jt}]}}$$

- Mientras que la matriz de de correlación cruzada para el rezago l se define como,

$$\begin{aligned} \rho(l) &= [\rho_{ij}(l)] \\ \rho_{ij}(l) &= \frac{\Gamma_{ij}(l)}{\sqrt{\Gamma_{ii}(0)}\sqrt{\Gamma_{jj}(0)}} = \frac{\mathbb{C}(r_{i,t}, r_{j,t-l})}{\sqrt{\mathbb{V}[r_{it}]}\sqrt{\mathbb{V}[r_{jt}]}} \end{aligned}$$

Función de Correlación Cruzada



Modelo VAR

- Para implementar el modelo VAR debemos seleccionar el orden del modelo.
- Una forma de hacerlo es a partir de la función VARselect de acuerdo al siguiente código:

```
require(vars)
Y=cbind(rtn,rtn2)
VARselect(Y,type="none")
$selection
AIC(n)  HQ(n)  SC(n) FPE(n)
     7      2      1      7
```

- Para implementar el modelo VAR, usaremos los dos órdenes más bajos obtenidos en el resultado anterior.

Modelo VAR

- **Modelo 1: VAR(1)**

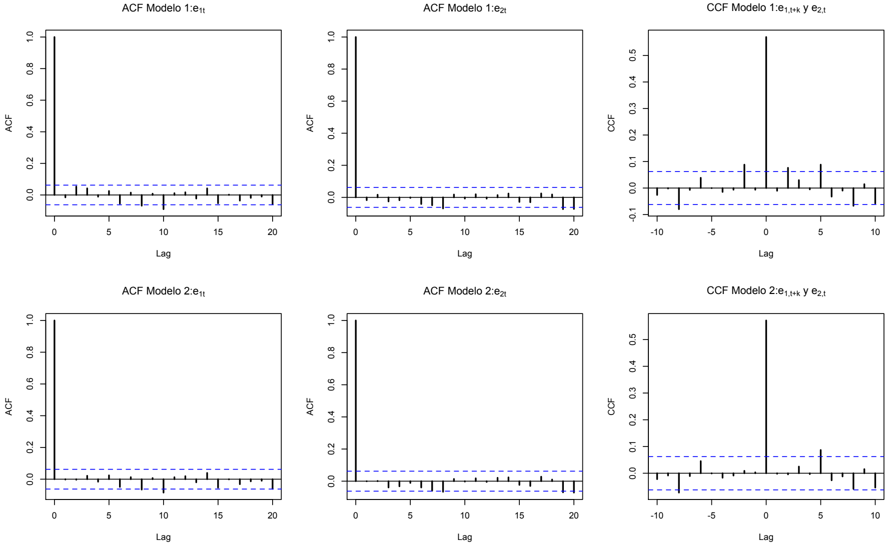
```
varmodel=VAR(Y,p=1,type="none")
summary(varmodel)
Estimation results for equation rtn:
====================================
rtn = rtn.l1 + rtn2.l1

        Estimate Std. Error t value Pr(>|t|)
rtn.l1   0.04812    0.03811   1.263 0.206933
rtn2.l1  0.14463    0.03801   3.805 0.000151 ***
Estimation results for equation rtn2:
=====================================
rtn2 = rtn.l1 + rtn2.l1

        Estimate Std. Error t value Pr(>|t|)
rtn.l1   0.21215    0.03806   5.574 3.21e-08 ***
rtn2.l1 -0.03030    0.03796  -0.798    0.425
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Modelo VAR

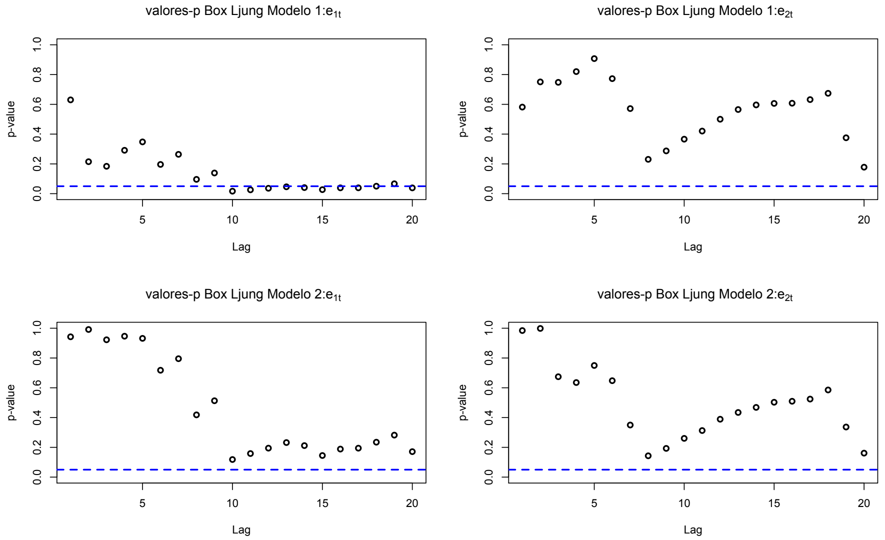
- **Modelo 2: VAR(2)**

```
varmodel2=VAR(Y,p=2,type="none")
summary(varmodel2)
Estimation results for equation rtn:
====================================
rtn = rtn.l1 + rtn2.l1 + rtn.l2 + rtn2.l2

        Estimate Std. Error t value Pr(>|t|)
rtn.l1   0.03439    0.03855   0.892 0.372588
rtn2.l1  0.14228    0.03859   3.687 0.000239 ***
rtn.l2   0.01829    0.03872   0.472 0.636768
rtn2.l2  0.07344    0.03845   1.910 0.056417 .
Estimation results for equation rtn2:
=====================================
rtn2 = rtn.l1 + rtn2.l1 + rtn.l2 + rtn2.l2

        Estimate Std. Error t value Pr(>|t|)
rtn.l1   0.21748    0.03847   5.653 2.07e-08 ***
rtn2.l1 -0.05077    0.03852  -1.318  0.18782
rtn.l2   0.11228    0.03865   2.905  0.00375 **
rtn2.l2 -0.04840    0.03838  -1.261  0.20750
```

Modelo VAR

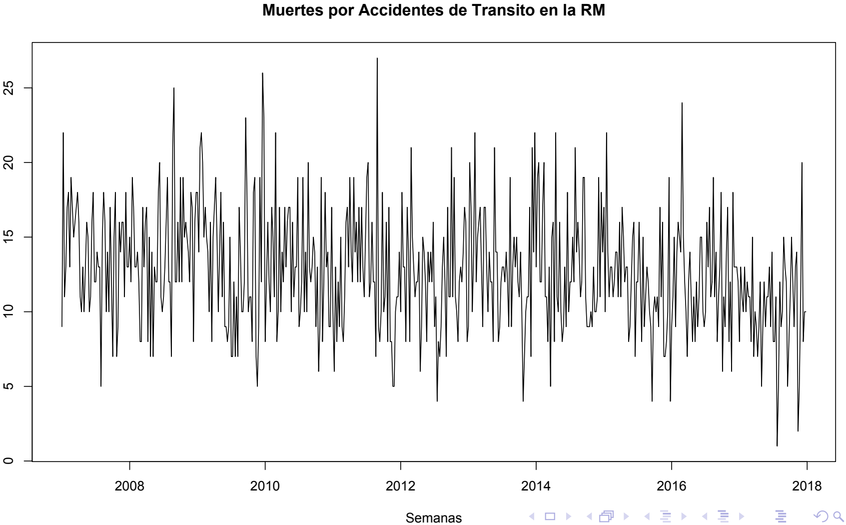

ACF Modelo 1:e_{1t}
ACF Modelo 1:e_{2t}
CCF Modelo 1:$e_{1,t+k}$ y $e_{2,t}$
ACF Modelo 2:e_{1t}
ACF Modelo 2:e_{2t}
CCF Modelo 2:$e_{1,t+k}$ y $e_{2,t}$
ACF
CCF
Lag

Modelo VAR


valores-p Box Ljung Modelo 1:e_{1t}
valores-p Box Ljung Modelo 1:e_{2t}
valores-p Box Ljung Modelo 2:e_{1t}
valores-p Box Ljung Modelo 2:e_{2t}
p-value
Lag
0.0
0.2
0.4
0.6
0.8
1.0
5
10
15
20

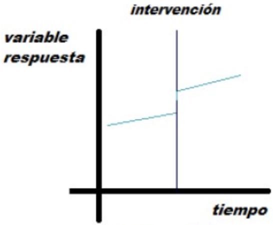
Aplicación en Datos de Salud

- Serie semanal de las Muertes por Accidentes de Tránsito en la Región Metropolitana entre el año 2007 y 2017:

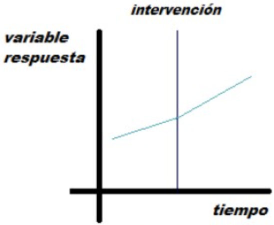


Series de Tiempo Interrumpidas

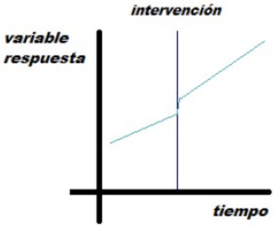
- El análisis de series de tiempo interrumpida (ITS) se caracteriza porque la serie se interrumpe, debido a la aplicación de un tratamiento. Los datos reflejan esta interrupción mostrando un cambio de nivel, tendencia o ambos. Gráficamente podemos ver estos patrones de cambio de la siguiente manera:



Cambio de Nivel



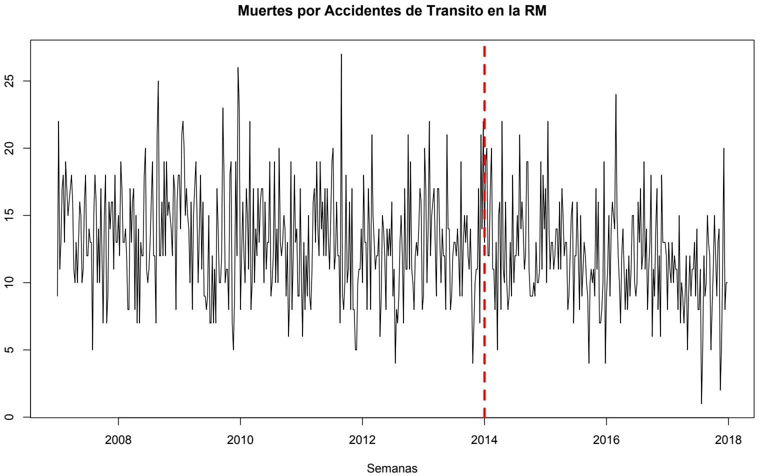
Cambio de pendiente



Cambio de nivel y de pendiente

Ley Emilia

- La Ley Emilia entró en vigencia el 16 de Septiembre del año 2014. Para analizar su impacto se puede usar un modelo de ITS sobre la serie de muertes por accidentes de tránsito:

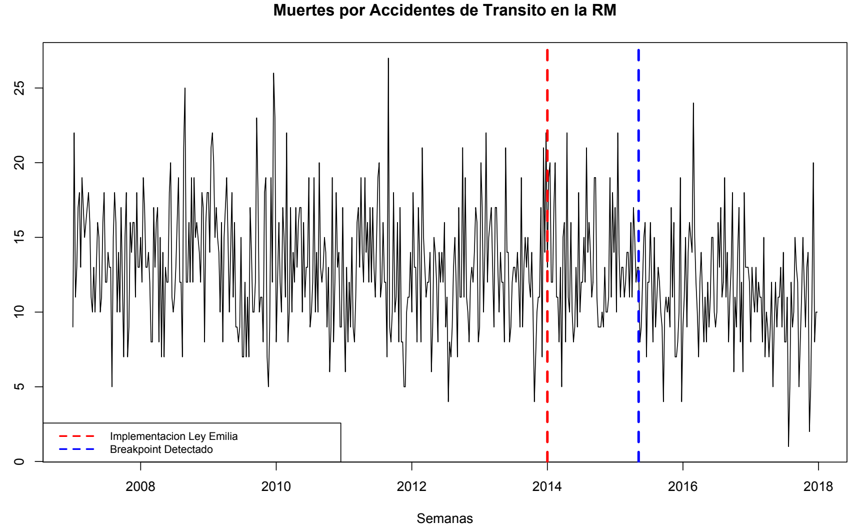


Detección de puntos de Quiebre

- Para detectar puntos de quiebre de una serie temporal se puede usar el siguiente código:

```
require(strucchange)
fs.serie <- Fstats(serie ~ 1)
bp=breakpoints(fs.serie)
bp

 Optimal 2-segment partition:

Call:
breakpoints.Fstats(obj = fs.serie)

Breakpoints at observation number:
435

Corresponding to breakdates:
2015(19)
```

Detección de puntos de Quiebre



Regresión con Cambio Estructural

- Supongamos que en el instante T la serie cambia su estructura de tendencia lineal con parámetros (a_0, b_0), a otra con parámetros (a_1, b_1) distintos de los anteriores.
- En este instante se produce un salto cuyo valor suponemos es: $(a_1 - a_0) + T(b_1 - b_0)$
- Este modelo se puede mostrar de forma matricial,

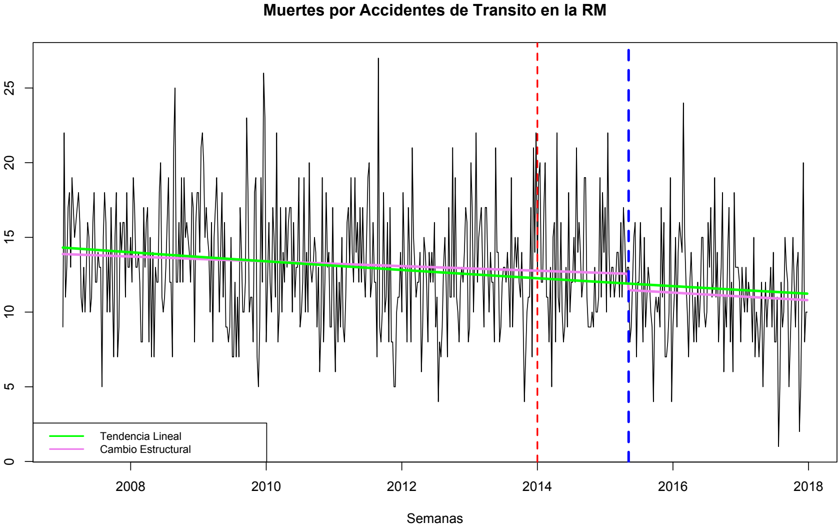
$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \\ y_{T+1} \\ y_{T+2} \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & T & 0 & 0 \\ 0 & 0 & 1 & T+1 \\ 0 & 0 & 1 & T+2 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & n \end{pmatrix} \begin{pmatrix} a_0 \\ b_0 \\ a_1 \\ b_1 \end{pmatrix} + \epsilon$$

Regresión con Cambio Estructural

- Para modelar este cambio en tendencia, usamos un GLM de familia Poisson como sigue:

```
t=1:length(serie)
dummy=c(rep(0,bp$breakpoints-1),rep(1,length(serie)-
bp$breakpoints+1))
tdummy=t*dummy
tdummy1=t*(1-dummy)
model2=glm(serie~tdummy+tdummy1,family="poisson")
summary(model2)
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.631e+00  2.604e-02 101.018  < 2e-16 ***
tdummy      -4.393e-04  7.217e-05  -6.087 1.15e-09 ***
tdummy1     -2.300e-04  1.052e-04  -2.187   0.0287 *
```

Regresión con Cambio Estructural

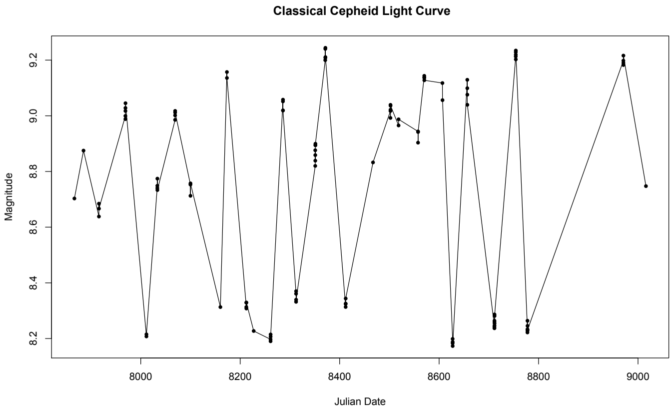


Aplicación en Datos Astronómicos

- Para cada objeto astronómico se tiene una serie de tiempo (llamada curva de luz) compuesta por los siguientes atributos,
  - **Tiempo:** En días Julianos
  - **Magnitud:** Medida del brillo de un objeto astronómico.
  - **Error:** Error de medición de la magnitud.

Curvas de Luz

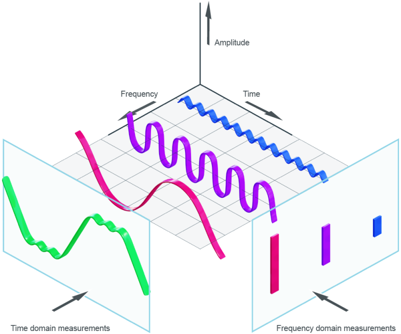
- Curva de Luz de una estrella Cefeida clásica observada por el estudio OGLE durante aproximadamente 1150 días:



- Note que en la práctica las curvas de luz se miden irregularmente en el tiempo.

Análisis en el Dominio de la Frecuencia

- Los métodos formulados en el dominio de la frecuencia se basan en la transformada discreta de Fourier. La principal diferencia con el análisis en el dominio del tiempo es que estos métodos analizan la respuesta del proceso en base a un conjunto determinado de frecuencias.



Estimación Periodo

- El periodograma de Lomb-Scargle Generalizado (GLS) es una extensión del periodograma de Fourier para series de tiempo irregularmente observadas que toma en cuenta el error de medición.
- Sean y_i las N mediciones de una serie de tiempo, en los tiempos t_i con un error de σ_i.

$$\chi^2_m(f) = min_\beta \chi^2(f) = \sum_i \frac{(y_i - c - a\cos(2\pi f t_i) - b\sin(2\pi f t_i))^2}{\sigma_i^2}$$

$$\chi^2_0 = \frac{\sum_i [y_i - \mu]^2}{\sigma_i^2}$$

- donde $\mu = \frac{\sum_i y_i/\sigma_i^2}{\sum_i 1/\sigma_i^2}$. El periodograma de Lomb-Scargle Generalizado (GLS) se obtiene como,

$$P_f(f) = \frac{(N-1)}{2}\frac{\chi_0^2 - \chi_m^2(f)}{\chi_0^2}$$

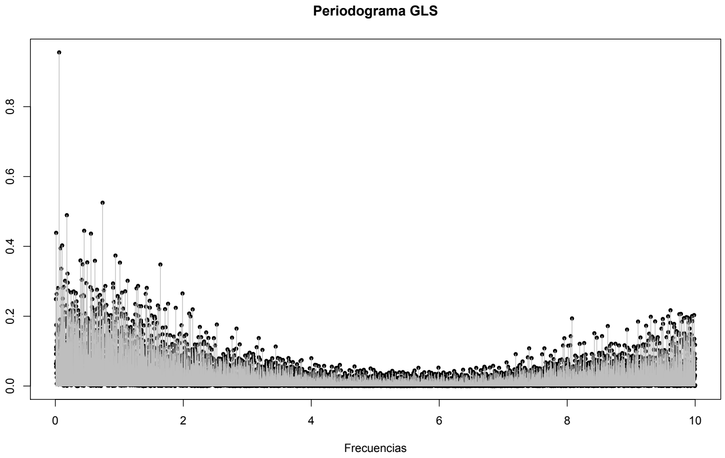
Periodograma GLS

- Para estimar el periodo usando el método de GLS se debe usar el siguiente código:

```
require(RobPer)
t=clcep[,1]
T_tot<-range(t)[2]-range(t)[1]
min_f = 0.01
max_f = 10
fr1=seq(from=min_f,to=max_f,by=0.001)
GenLS<-RobPer(clcep,weighting=TRUE,regression="L2",
model="sine",periods=1/fr1)
```

Periodograma GLS

- El valor máximo del periodograma $P = 0{,}955$ se obtiene para la frecuencia $f = 0{,}060$



Fase y Modelamiento Armónico

- Para tener una idea del comportamiento periódico de la curva de luz, se puede obtener la curva de luz doblada en su fase. Donde la fase, se obtiene como,

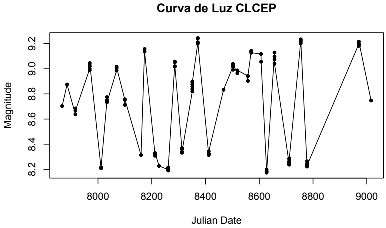
$$\phi = \left(\frac{t - t_0}{p}\right) - \mathbb{E}(t) \tag{2}$$

donde t_0 es el tiempo de referencia, t es el tiempo en que la observación fue tomada, p es el periodo de la curva de luz y $\mathbb{E}(t)$ es la parte entera de $\frac{t-t_0}{p}$.

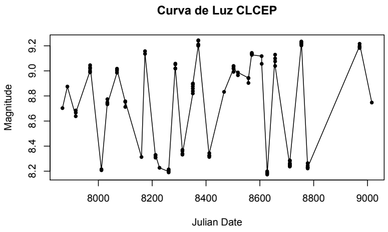
- Generalmente, las curvas de luz de estrellas periódicas se ajustan usando un modelo armónico. El modelo p-armónico está definido por:

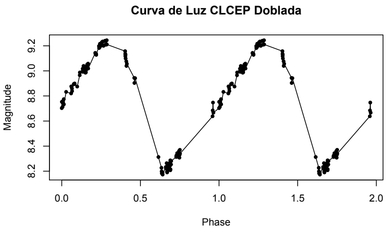
$$y_t = \beta_0 + \sum_{j=1}^{p}(\alpha_{1j} sin(2\pi f_1 jt) + \beta_{1j} cos(2\pi f_1 jt)) + \epsilon_t \tag{3}$$

Modelamiento Armónico

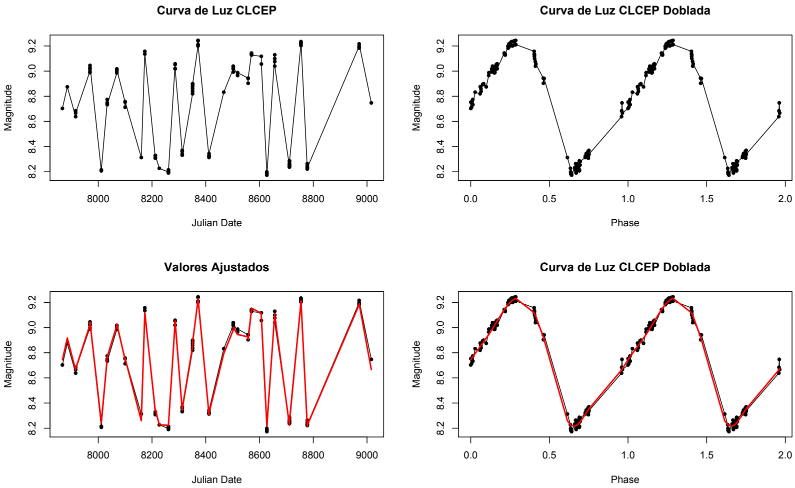


Modelamiento Armónico





Modelamiento Armónico



Irregular Autoregressive Model (IAR)

- En datos irregularmente observados no existe un test que nos permita evaluar si los residuos de nuestro modelo son blancos.
- Para abordar este problema usaremos una extensión del modelo AR para datos irregularmente observados.
- El modelo IAR es definido por,

$$y_{t_j} = \phi^{t_j - t_{j-1}} \, y_{t_{j-1}} + \sigma \sqrt{1 - \phi^{2(t_j - t_{j-1})}} \, \varepsilon_{t_j}, \tag{4}$$

  donde ϕ es el parámetro de autocorrelación y ε_{t_j} es una secuencia de ruido blanco con media cero y varianza uno.
- El proceso IAR es estrictamente estacionario bajo ningún supuesto distribucional.

Modelo iAR

- Para implementar el modelo iAR en los residuos del modelo armónico se debe usar el siguiente código:

```
require(iAR)
iar=IARloglik(y=fit$res,st=clcep[,1],standardized=F)
$phi
[1] 7.036522e-05
```

Modelo iAR

- Para implementar el modelo iAR en los residuos del modelo armónico se debe usar el siguiente código:

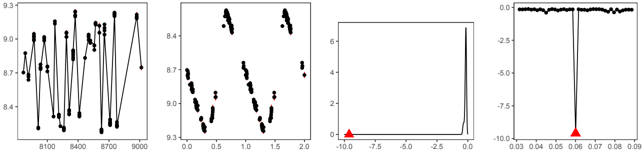
```
require(iAR)
iar=IARloglik(y=fit$res,st=clcep[,1],standardized=F)
$phi
[1] 7.036522e-05
```

- Si bien, el valor estimado del parámetro de autocorrelación es pequeño, podemos testear su significancia con el siguiente código:

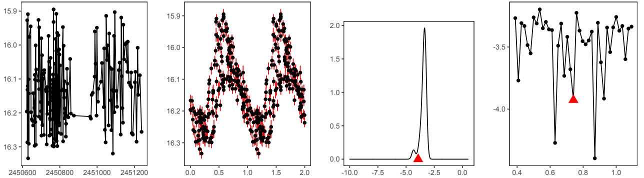
```
test<-IARTest(y=clcep[,2],st=clcep[,1],f=fr,phi=iar$phi,
model="iAR",plot=TRUE,xlim=c(-10,0.5))
test$pvalue
[1] 0
```

Interpretación Test iAR

- **Hipótesis:** Si ajustamos una curva de luz de una estrella periódica usando un modelo armónico bien especificado, entonces los residuos no deben tener dependencia en el tiempo ($\phi = 0$).

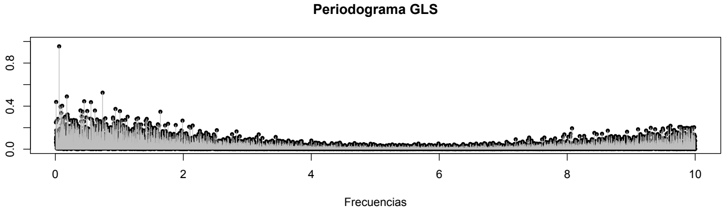


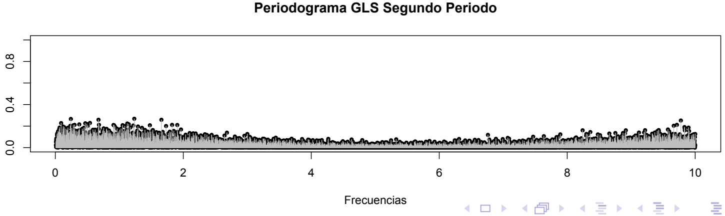
- Sin embargo, en algunos casos el parámetro de autocorrelación muestra que los residuos aún tienen dependencia temporal



Periodograma GLS

- En nuestro ejemplo el modelo está bien especificado con residuos no correlacionados, lo cual también puede ser apreciado a partir del periodograma GLS.





felipe.elorrieta@usach.cl

@felipeelorrieta