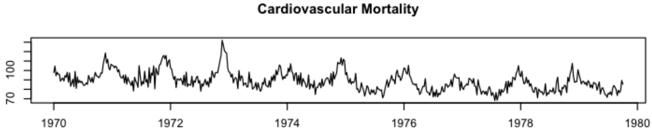
Time Series and Forecasting
Cross-correlation

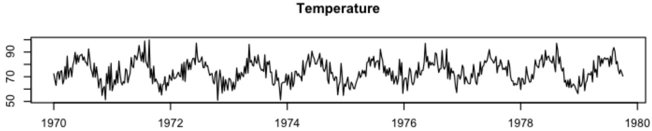
Maria Eduarda Silva
mesilva@fep.up.pt

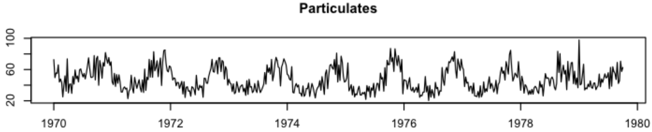
School of Economics, University of Porto

LMU, July 1 2022

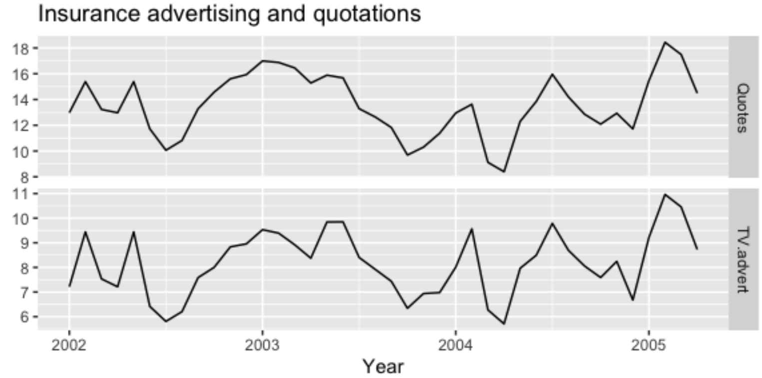
Daily Pollution, Temperature and Cardiovascular Mortality in Los Angeles County



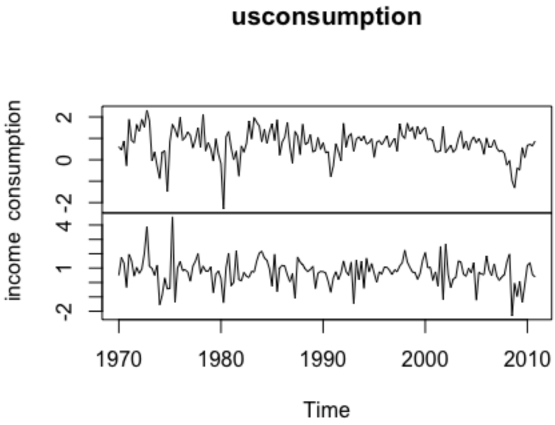




TV advertising and insurance quotations



Percentage changes in quarterly personal consumption expenditure and personal disposable income for the US, 1970 to 2010.



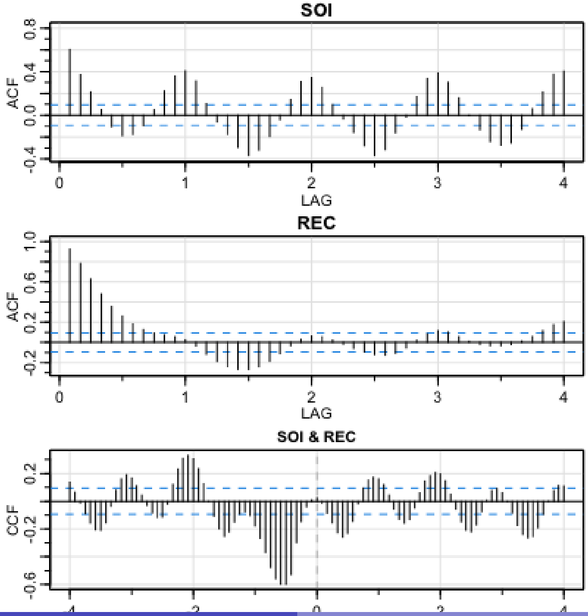
Sample Cross correlation

- $\hat{\rho}_{XY}(h) = Corr(X_{t+h}, Y_t) \quad h = 0, \pm 1, \pm 2, \ldots$
- $\hat{\rho}_{XY}(h) = \hat{\rho}_{YX}(-h)$
- If X_t and Y_t are independent , then, for large sample sizes

$$\hat{\rho}_{xy}(h) \approx N(0, \frac{1}{n})$$

if at least one of the series is independent white noise

SOI and REC correlation analysis



Cross correlation and pre-whitening

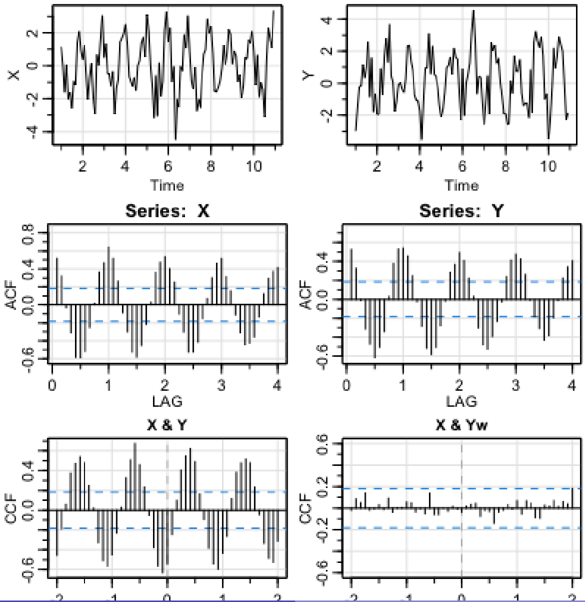
- To assess the significance of the cross correlations we need that (at least) one the series is white noise.
- Therefore we should pre-whiten a series prior to a cross-correlation analysis.
- Otherwise we may get spurious correlations.

Example of spurious correlations: 2 time series generated independently but showing high cross correlation.

$$x_t = 2\cos(2\pi t \frac{1}{12}) + e_{t1}, \quad e_{t1} \sim N(0,1)$$

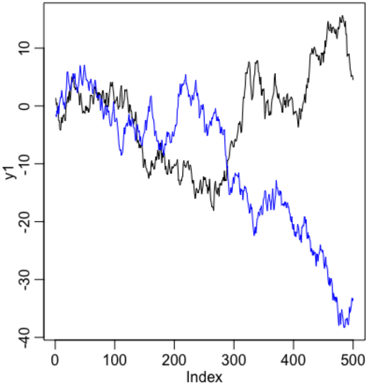
$$y_t = 2\cos(2\pi (t+5) \frac{1}{12}) + e_{t2}, \quad e_{t2} \sim N(0,1)$$

Cross correlation and pre-whitening

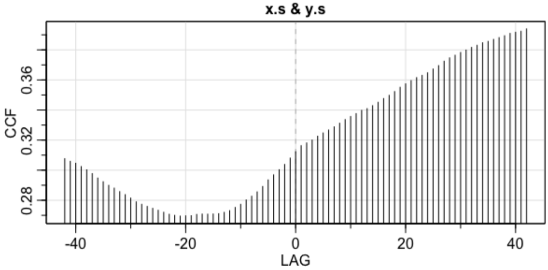


More on Spurious correlation

- A word of caution: spurious relationships arise also between trending time series
- Two independent random walks



High correlation



Significant linear relationship

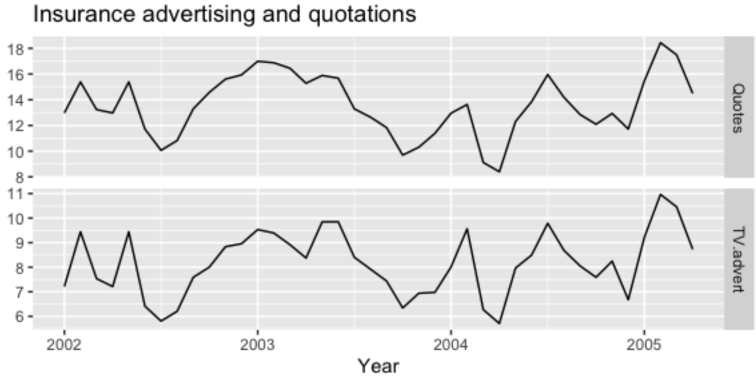
```
> summary(sr.reg1)

Call:
lm(formula = y1 ~ y2)

Residuals:
     Min       1Q   Median       3Q      Max 
-12.6878  -5.3229  -0.0348   4.9765  13.1047 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -6.01187    0.35794  -16.80   <2e-16 ***
y2          -0.44224    0.02312  -19.13   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.107 on 498 degrees of freedom
Multiple R-squared:  0.4235,	Adjusted R-squared:  0.4223 
F-statistic: 365.8 on 1 and 498 DF,  p-value: < 2.2e-16
```

Prediction via Cross-correlation

- Determining leading/lagging relations between stationary time series $x_t,\ y_t$ is important for predicting y_t from x_t purposes

- If for some unknown l the model $y_t = Ax_{t-l} + e_t$ holds, then
  - $l > 0$ then x_t leads y_t
  - $l < 0$ x_t lags y_t

TV advertising and insurance quotations: which is the leading series?

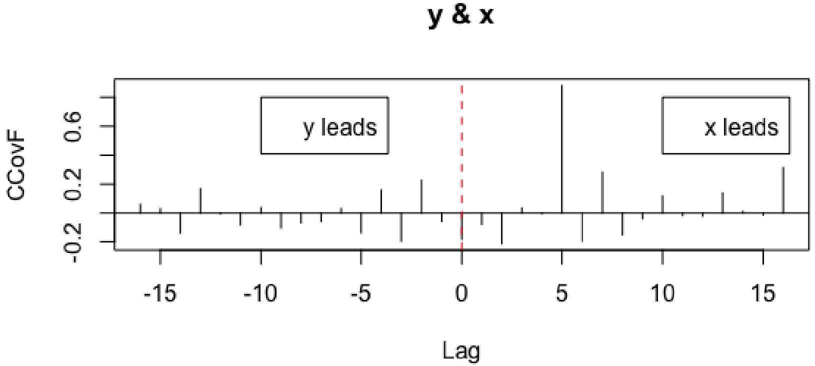


Prediction via Cross-correlation

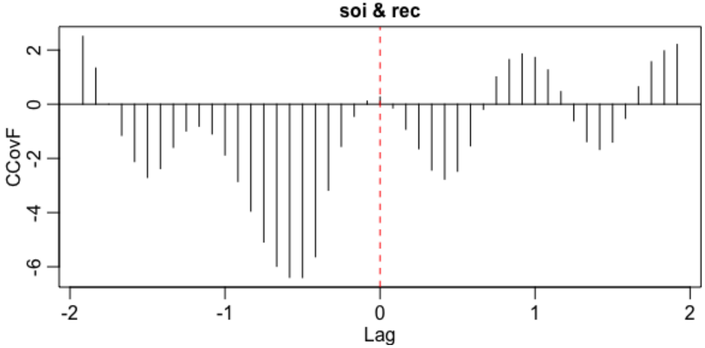
- $\gamma_{yx}(h) = A\gamma_x(h - l)$
- $\gamma_{yx}(h)$ will look like $\gamma_x(h - l)$
- Largest value of $\gamma_{yx}(h)$ is $\gamma_x(0)$ when $h = l$
- That is $\gamma_{yx}(h)$ has a peak at lag $h = l$
- Peak on positive lags h if x_t leads y_t
- Peak on negative lags if x_t lags y_t

Prediction via Cross-correlation

$y_t = x_{t-5} + e_t$

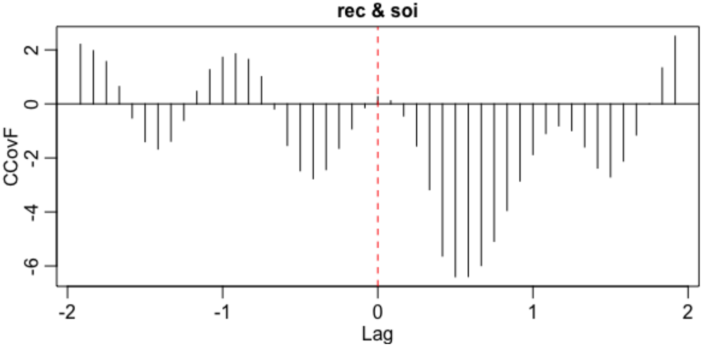


Prediction via Cross-correlation



This plot indicates a lag $l < 0$ around -6 (months) and so *soi* leads

Prediction via Cross-correlation



This plot indicates a lag $l > 0$ around 6 (months) and so *soi* leads

Measures of Dependence - bivariate context

If we have two time series X_t and Y_t we define

- Cross covariance function

$$\gamma_{XY}(s,t) = cov(X_s, Y_t) = E\left[(X_s - \mu_{X_s})(Y_t - \mu_{Y_t})\right] \tag{1}$$

- Cross Correlation function (CCF)

$$\rho_{XY}(s,t) = \frac{\gamma_{XY}(s,t)}{\sqrt{\gamma_X(t,t)\gamma_Y(s,s)}} \tag{2}$$

Jointly stationary time series

$\{X_t\}$ and $\{Y_t\}$ time series are jointly stationary if

- $\{X_t\}$ is stationary
- $\{Y_t\}$ is stationary
- The cross-correlation between $\{X_t\}$ and $\{Y_t\}$ depends only on the lag

Jointly stationary time series

Formaly

- $E(X_t) = \mu_X$
- $E(Y_t) = \mu_Y$
- $\gamma_X(h)$ ACF of X_t
- $\gamma_Y(h)$ ACF of Y_t
- **Cross covariance:**

$$Cov(X_s, Y_t) = E[(X_s - \mu_X)(Y_t - \mu_Y)] = \gamma_{XY}(s-t)$$

$$\gamma_{XY}(h) = Cov(X_{t+h}, Y_t) = E[(X_{t+h} - \mu_X)(Y_t - \mu_Y)]$$

- **Cross correlation, CCF:**
  $\rho_{XY}(h) = \frac{\gamma_{XY}(h)}{\sqrt{\gamma_X(0)\gamma_Y(0)}}$

Properties of CCF

Note that

$$\gamma_{XY}(-h) = E[(X_{t-h} - \mu_X)(Y_t - \mu_Y)] = E[(Y_t - \mu_Y)(X_{t-h} - \mu_X)] = \gamma_{YX}(h)$$

- $\gamma_{XY}(h) = \gamma_{YX}(-h)$.
- $\rho_{XY}(h) = \rho_{YX}(-h)$.