Coherence studies in time series

Vytautas Pankūnas[1], Julija Janeiko[2] and Danutė Krapavickaitė[3]

[1]Vilnius Gediminas Technical University, e-mail: vytautas.pankunas@stud.vgtu.lt
[2]Vilnius Gediminas Technical University, e-mail: julija.janeiko@stud.vgtu.lt
[3]Vilnius Gediminas Technical University, e-mail: danute.krapavickaite@vgtu.lt

Abstract

The aim of this paper is to present a way to measure strength of a relationship between the two time series by a coefficient of coherence. A definition of the coherence coefficient is given and an example of its application is provided.

1 Introduction

A dictionary says that coherence means integration of diverse elements, relationships, or values. One of the principles in official statistics is coherence of statistical information in the sense of possibility to combine it together. A measure of coherence in official statistics has to show a degree to which statistical results arising from different statistical processes can be combined together. In the case of the time series coherence shows the degree to which different time series reflect the same phenomenon in economy.

2 A concept of coherence

A concept of coherence may be met in the different fields of science, like physics, geophysics, classical time series and elsewhere.

We follow the definition of the coherence as it is presented in Stoffer & Shumway (2006) and Wei (2006). Let we have two finite stationary time series $x_1, x_2, ..., x_n$ and $y_1, y_2, ..., y_n$. Define their cross-covariance function

$$\gamma_{xy}(h) = \mathrm{E}\big((x_{t+h} - \mu_x)(y_t - \mu_y)\big) \quad \text{estimated by} \quad \frac{1}{n}\sum_{t=1}^{n}\big((x_{t+h} - \bar{x})(y_t - \bar{y})\big),$$

$\mu_x = Ex_t$, $\mu_y = Ey_t$, $\bar{x} = 1/n\sum_{i=1}^n x_i$, $\bar{y} = 1/n\sum_{i=1}^n y_i$, $t = 1, 2, ..., n$, $h = 0, 1, 2, ..., n$, $x_{t+n} = x_t$. The follwing representation is applied to the cross-covariance function:

$$\gamma_{xy}(h) = \sum_{k=-m_h}^{[n/2]} c_k \mathrm{e}^{i\omega_k t}, \quad m_h = \begin{cases} [n/2], & n \text{ is odd,} \\ [n/2] + 1, & n \text{ is even} \end{cases}$$

with the frequencies $\omega_k = 2\pi k/n$. This representation is called Fourier representation, and frequencies $\omega_k = 2\pi k/n$ are called Fourier frequencies. The Fourier coefficients c_k (cross-spectrum) are given by the formula

$$c_k = \frac{1}{n}\sum_{h=1}^{n} \gamma_{xy}(h)\mathrm{e}^{-i\omega_k t}, \quad k = -m_h, -m_h + 1, ..., 0, 1, ..., [n/2].$$

The cross-spectrum is generally a complex-valued function.

The energy associated with the cross-covariance sequence $\gamma_{xy}(h)$ is defined by $\sum_{h=1}^{n} \gamma_{xy}(h)$. The energy of $\gamma_{xy}(h)$ per unit time is called the power of the sequence:

$$Power = \frac{1}{n}\sum_{h=1}^{n} \gamma_{xy}(h) = \sum_{k=-[m_h]}^{[n/2]} c_k^2.$$

The quantities $f_0 = c_0^2$, $f_{[n/2]} = |c_{[n/2]}|^2$, $f_k = |c_{-k}|^2 + |c_k|^2 = 2|c_k|^2$, $k = 1, 2, ..., [n/2-1]$, which are obtained from the cross-covariance function $\gamma_{xy}(h)$ Fourier representation at the k-th frequency $\omega_k = 2\pi k/n$, are interpreted as the contribution of this frequency to the total power. The quantity f_k plotted as a function of ω_k is called a periodogram.

An important example of the application of the cross-spectrum is the problem of predicting an output series y_t from some input series x_t through a linear filter relation. A measure of strength of such a relation is the squared coherence function (or coherence coefficient), defined as

$$\mathrm{coh}_{xy}^2(\omega_k) = \frac{|f_{xy}(\omega_k)|^2}{f_{xx}(\omega_k) f_{yy}(\omega_k)} \tag{1}$$

where $f_{xx}(\omega_k)$ and $f_{yy}(\omega_k)$ are the individual spectra of the series x_t and y_t, respectively; $f_{xy}(\omega_k) = f_k$.

Another fact worth mentioning is that squared coherence coefficient is related with the conventional squared Pearson correlation coefficient in the form

$$\rho_{xy}^2 = \frac{\sigma_{yx}^2}{\sigma_x^2 \sigma_y^2}, \tag{2}$$

where σ_x^2 and σ_y^2 are variances of random variables x and y and $\sigma_{yx} = \sigma_{xy}$ is their covariance.

3 A coherence coefficient in practice

Data sets of quarterly aggregated statistics from different statistical surveys and administrative data sources in 2008-2017 are used for a case study. They are presented in Figure 1.

- Statistics Lithuania, Labor Force Survey data: variable for a number of employed (LFE); a number of unemployed (LFU) (in thousands).
- Statistics Lithuania, Labour remuneration survey data: number of employees (Emp); resource for remuneration (RRS) (in millions of Euros).
- Labour Exchange office data: number for registered unemployment (EU) (in thousands).
- Administrative data of the Social insurance institution Sodra: enterprise remuneration, from which taxes are payed (RSI) (in millions of Euros).

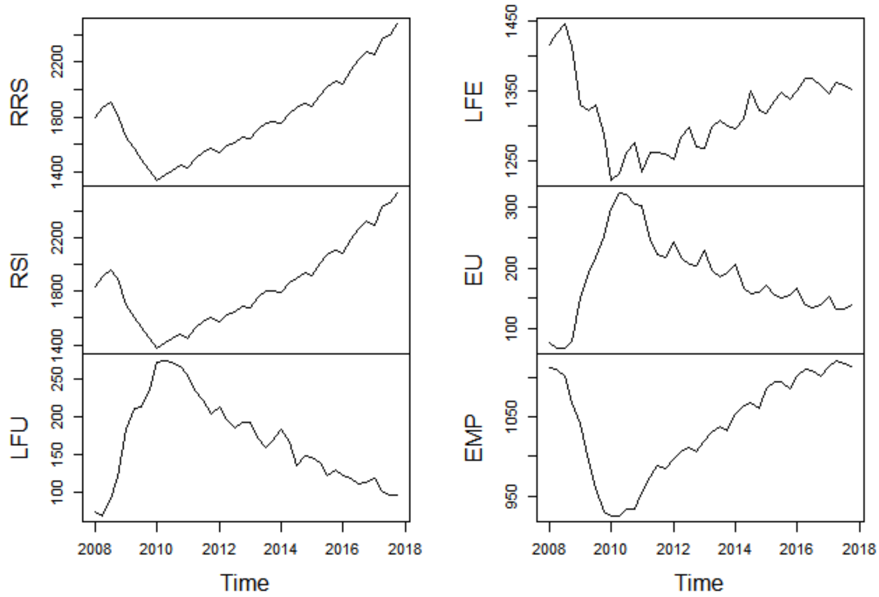

Figure 1: Graphical view of the quarterly data in 2008-2017

The common tendencies of change are observed despite the time series are generated by the different statistical processes. Some periodic fluctuations, possibly yearly, come to notice.

The coherence coefficients between the estimate for a number of unemployed LFU and enterprise remuneration RSI by Sodra are calculated for frequencies $\omega = 2\pi k/n$, $k = 1, 2, ..., [n/2]$. We consider this to be the best example to illustrate coherence as these two time series are generated by the different processes and they behave completely contrariwise: when LFU increases, RSI decreases and vice versa.

Before approaching to the coherence of RSI and LFU, let us draw the periodograms (Figure 2) for each *differentiated* series separately to find out which frequencies $\omega = \omega_k/2\pi$ have the highest contribution to the power of differentiated time series $\gamma_{xy}(h)$. The `R` package *astsa* is used to draw the periodograms and to calculate the coherence coefficients.

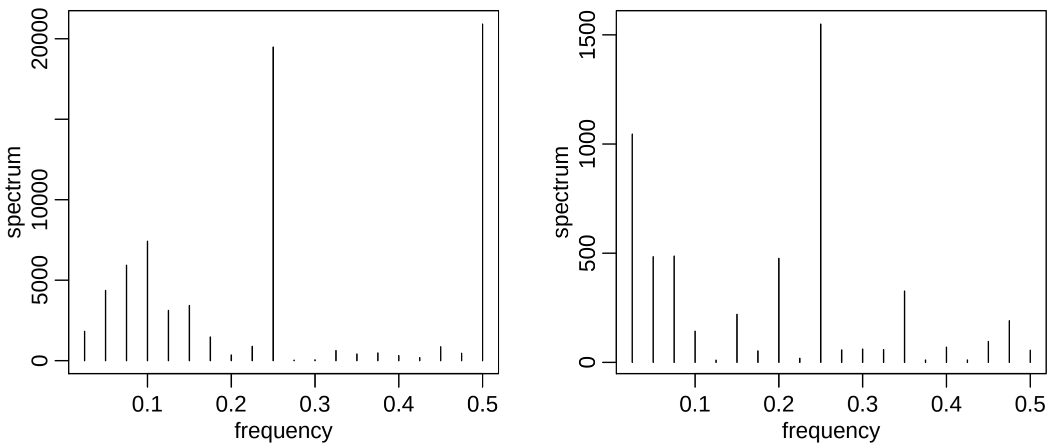

Figure 2: Periodograms of LFU (left) and RSI (right)

In the context of the study data the frequency $\omega = 0.25 = \omega_{10}/2\pi$ with the corresponding period $T = 1/\omega = 4$ – four quarters is the most important. We pay attention to the frequencies $\omega = 0.025 = \omega_1/2\pi$ (period 40 quarters) and $\omega = 0.125 = \omega_5/2\pi$ (period 8 quarters) as well.

Now we can calculate and portray the coherence between LFU and RSI in Figure 3.

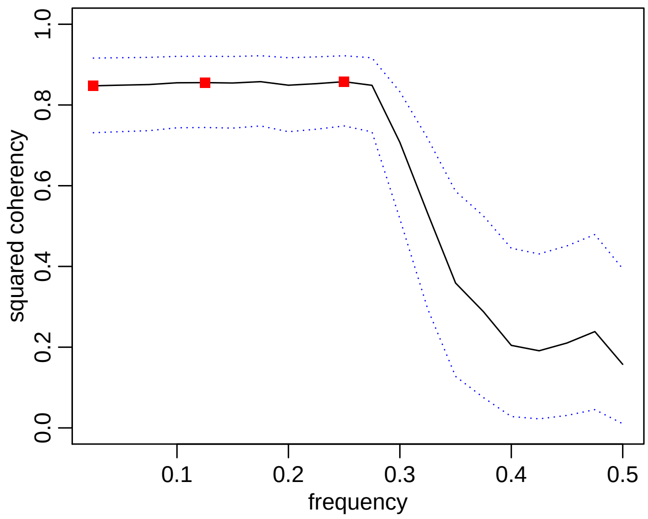

Figure 3: Coherence between LFU and RSI

The numeric values of the coherence coefficients and squared correlation coefficient between LFU and RSI are presented in Table 1.

Table 1: Coherence coefficients and squared correlation coefficient

ω	$\mathrm{coh}^2_{xy}(\omega)$
0.025	0.8475
0.125	0.8554
0.25	0.8577
$\rho(x, y)$	-0.8624

4 Conclusion

The simulation study shows that both kinds of indicators: squared correlation coefficient (2) and coherence coefficients (1) suit well for assessment of the linear dependency or similarity between the time series. At the moment it is hard to say anything about the preferences of some of them.

References

Shumway R. H. and Stoffer D. S. (2006), *Time Series Analysis and Its Applications*, Springer Texts in Statistics, DOI = 10.1007/978-3-319-52452-8.

Stoffer, D. (2017), *astsa: Applied Statistical Time Series Analysis*, R package version 1.8, `https://CRAN.R-project.org/package=astsa`.

Wei W. W. S. (2006), *Time Series Analysis: Univariate and Multivariate Methods*, Pearson Education, `https://books.google.lt/books?id=aY0QAQAAIAAJ`.