Lecture 3: Statistics of stationary time series

Josep Vives

March 3, 2024

In Lecture 1 we saw how to extract the trend and the cyclic components from a time series and to decompose it as the sum of the trend, the cyclic components and a residual time series. We can consider that our residual series is a second order stationary time series; this can be checked visually on the graph if we do not observe trends nor cycles.

So, the second step is to adjust a second order stationary model to our residual time series. We will see, in the following chapters, different second order stationary models. But at the end, our analysis of a time series finishes if we prove that the residuals are an IID noise with a certain law.

We have seen that a second order stationary process is determined by the pair (μ, γ) or the triplet (μ, σ^2, ρ) where μ is the constant mean, γ the autocovariance function, that is symmetric and depends only on the lag, σ^2 is the variance, that coincide with $\gamma(0)$, and ρ is the autocorrelation function, that is also symmetric, $\rho(0) = 1$ and depends only on the lag.

So, given a time series $x_1, \dots, x_n$, assumed to be stationary, we need methods to estimate these objects, so we need, adequate estimators of μ, σ^2, γ and ρ.

1 Basic estimators of the characteristics of a second order stationary time series

Consider $x_1, \cdots, x_n$, observed data from a stationary time series. To fit a stationary model we need to estimate the mean μ and the covariance function γ, or alternatively, to estimate the mean μ, the variance σ^2 and the autocorrelation function ρ. How to do it?

The usual estimators are

1. The empirical mean:

$$\bar{x}_n := \frac{1}{n} \sum_{i=1}^{n} x_i$$

2. The empirical variance:

$$\widehat{\sigma}^2 = \widehat{\gamma}(0) := \frac{1}{n} \sum_{j=1}^{n} (x_j - \bar{x}_n)^2$$

3. The empirical auto-covariance:

$$\widehat{\gamma}(l) := \frac{1}{n} \sum_{j=1}^{n-l} (x_j - \bar{x}_n)(x_{j+l} - \bar{x}_n), \quad l < n$$

4. The empirical auto-correlation:

$$\widehat{\rho}(l) := \frac{\widehat{\gamma}(l)}{\widehat{\gamma}(0)}$$

The graph of the empirical auto-correlation is the basic tool to detect what is the best model for our data. Given a model, we have the theoretical shape of its auto-correlation function and so, to identify a model that fits our data consist in to search a model with a theoretical shape of its auto-correlation function similar to the shape of the observed correlogram of our data. We will see how to do this with concrete models in the next lectures.

2 Testing the IID noise hypothesis

The effort to describe a time series finishes when after extracting different components of the series we obtain an IID noise and describe its law. Recall that an IID noise is nothing but a collection of independent and identically distributed random varibles, or in other words, a sample of a certain probability law.

So, to have at hand a method to determine if a series of residuals is or not an IID noise is of major importance. The following result solves the problem; the proof can be found in [1].

Theorem 2.1 *Let $y_1, y_2, \dots, y_n$ be n observations of an IID noise. Let be*

$$\{\widehat{\rho}(j),\ 1 \leq j \leq h\}$$

the graph of its auto-correlation function until the lag h. Then, for n big enough and h relatively small, values $\widehat{\rho}(j)$ can be considered as a random sample of a $N(0, \frac{1}{n})$.

So, if the IID hypothesis is true, 95% of the h observed auto-correlations have to be in the interval

$$\left[-1.96\frac{1}{\sqrt{n}}, 1.96\frac{1}{\sqrt{n}}\right].$$

This result is useful to test if a certain time series is or not an IID noise. For example, given the first 50 values of $\widehat{\rho}$, if 3 or more of them appears out of the interval it would be difficult to accept that the series is an IID noise.

Remark 2.2 *If we have a time series with length n it is commonly accepted that values of $\widehat{\rho}$ up to lag of order $\frac{n}{3}$ are significative. But some researchers consider only acceptable auto-correlations until lag $\log n$.*

A more precise method based on the previous idea in order to test if a time series is an IID noise is Ljung-Box test (1978). It is based on the statistic

$$Q_{LB} := n(n+2)\sum_{j=1}^{h} \frac{\widehat{\rho}^2(j)}{n-j}$$

where h is a parameter that have to be determined and $Q_{LB}(h) \sim \chi^2_h$. Recall that χ^2_h denotes the chi-square distribution with h degrees of freedom.

If Q_{LB} is big, we reject IID noise hypothesis. If we have obtained $Q_{LB} = q$,

$$p := P\{Q_{LB} \geq q\} = P\{\chi_h^2 \geq q\}$$

is the $p-$value. So, usually, if $p < 0.05$, we reject the IID character of the series.

Strictly speaking the p-value is the probability that what it has been observed or something worst with respect the null hypothesis, happens. If the p-value is small it means that something very unlikely if the null hypothesis is true, has been observed. So, it is more reasonable to reject the null hypothesis.

We can say also that the p-value is the risk to reject the null hypothesis when it is true. If it is small, the risk is small and we reject the null hypothesis. Another way to see it is to interpret the p-value as the reliability of the null hypothesis. If p is small, this means the null hypothesis is not reliable and we reject it.

3 Testing Gaussianity

If we know that our residuals are an IID noise it can be interesting to test if it is in fact a Gaussian white noise. The traditional way to do this is using the well-known Kolmogorov-Smirnov goodness of fit test.

An alternative is the so called $Q-Q$-plot:

Let $Y_1, \ldots, Y_k$ be a sample of a $N(\mu, \sigma^2)$ law and let $X_1, \ldots, X_n$ be a sample of a $N(0,1)$ law. Let $X_{(1)}, \ldots, X_{(k)}$ and $Y_{(1)}, \ldots, Y_{(k)}$ the corresponding order statistics. Denote by $x_{(i)}$ and $y_{(i)}$ the corresponding empirical observations.

We have of course

$$\mathbb{E}(Y_{(j)}) = \mu + \sigma\mathbb{E}(X_{(j)}).$$

If $m_i := \mathbb{E}(X_{(i)})$, we can draw the pairs

$$(m_i, y_{(i)})$$

and the graphics of these pairs should be approximately linear. And in particular, its correlation should be near to 1. This is the basis of the so-called Shapiro-Wilk test (1965).

The value m_i is usually approximated by

$$m_i \sim \Phi^{-1}(\frac{i-0.5}{n})$$

where Φ is the cumulative probability of the standard normal distribution.

Therefore, we have

$$R^2 := \frac{\left(\sum(Y_{(i)} - \bar{Y}_n)\Phi^{-1}(\frac{i-0.5}{n})\right)^2}{\sum(Y_{(i)} - \bar{Y}_n)^2 \cdot \sum(\Phi^{-1}(\frac{i-0.5}{n}))^2} \in [0,1],$$

and the $p-$value is the probability

$$P\{R^2 < r^2\}.$$

where r is the observed correlation.

The distribution of R^2 is the Shapiro-Wik distribution. Its values can be found at

http://www.real-statistics.com/statistics-tables/shapiro-wilk-table/.

For $n \geq 50$, for example, we have

$$P\{R^2 < 0.987\} = 0.05.$$

So, if the observed r^2 is less than 0.987, we reject the normal hypothesis.

References

[1] P. J. Brockwell and R. A. Davis (1991): *Time Series: Theory and Methods.* Springer.