

## THE DECOMPOSITION ISSUE OF A TIME SERIES IN THE FORECASTING PROCESS

Dariusz GRZESICA

University of Technology, Krakow, Poland  
dgrzesica@pk.edu.pl

**Abstract:** *Decomposition of time series is the estimate and extraction of deterministic part of the series – trend, cyclical and seasonal fluctuations in the hope that the rest of the data, that is, theoretically, a random variable will be stationary random process. During the process of predicting the time series elements affects significantly on the determination of the future values, which are characterized by a low forecast error. Therefore, the purpose of this article is to identify the elements of the time series decomposition and to determine the extent to which they affect the forecasting process. Problems that often appear when you run the forecast and methods of building models and forecasts based on time series will be presented. Observations will be described on the basis of nonparametric time series modeling.*

**Keywords:** time series decomposition, seasonal fluctuations, cyclicity.

### 1. Introduction

Nonparametric time series modeling can be used when the factors that affect the object under investigation can't be identified. The course of the phenomenon in a given period is recorded using time series. It allows to register values that occur at a certain point in time. To identify the character of changes of investigated phenomena, time series analysis should be conducted. It allows not only to retrieve elements that explains the formation of the values of the variables, but also to determine forecast for future periods. Two main goals of time series analysis should be considered as a definition of the nature described, through line observations of the phenomenon and prediction of future values.

In the time series can be identified, in particular:

- Developmental tendency (trend) – it expresses the tendency to a unidirectional change (increase or

reduction);

- Cyclical fluctuations – long-term rhythmic fluctuations around the trends;
- Seasonal fluctuations – fluctuations around trends in a period not exceeding one year.

Decomposition of the time series allows to foresee some phenomena, characterized by high dynamics of change. In the case of forecasting based on time series, this process is performed by creating the corresponding model. Such a model needs to be adapted well for the real variables, including elements of the decomposition, and then use it to build the forecast.

Many works use the elements of the decomposition of the time series to predict future values. In the work [9] an ensemble method composed of Empirical Mode Decomposition (EMD) algorithm and deep learning approach is presented. The load demand series were first decomposed and few algorithms were used like Intrinsic

Mode Functions (IMFs) and Deep Belief Network (DBN). In work [11] a hybrid time series adaptive network-based fuzzy inference system (ANFIS) model focused on empirical mode decomposition (EMD) to forecast stock prices were presented. Lots of works is using hybrid methods and models to analyze and forecast based on the decomposition of time series [8] [7].

## 2. Meaning of trend in modeling and forecasting time series

### 2.1. Trend models

In much of time series literature, trend is conceived as that part of a series which changes relatively slowly over time, and loosely defined as “long-term change in the mean level” [2]. Trend is a function of time, reflecting the tendency of the observed variable  $y_i$ . You can formulate it as:

$$\hat{y}_i = T_i = f(i)$$

A trend model is used to predict in the medium and long term time horizon, when the phenomenon is not subject to strong random effects. As a rule, trend are linear relative to coefficients, so that there are no problems with their estimation. By adopting a trend model, the problems associated with the choice of explanatory variables are eliminated. One of the methods of determining the functions is a trend graphic analysis. It allows to define the following dependencies in the time series:

- linear – the coordinate system  $(i, y)$ ,
- exponential – in the layout  $(i, \ln y_i)$ ,
- logarithmic – in the layout  $(\ln i, y_i)$ ,
- power – in the layout  $(\ln i, \ln y_i)$ ,
- polynomials.

Determination of the forecast based on the model of the trend is simple, as it is a simple extrapolation of the functions of the trend on the forecast horizon. The forecast with the horizon  $k$  is determined by the dependence:

$$\hat{y}_{i+k|i} = f(i+k)$$

Both the trend and its parameters can not undergo significant changes in the forecast

horizon as compared to the period for which the trend model was derived.

### 2.2. Adaptive trend models

Adaptive trend models are created in order to accommodate changes in the structure or parameters of the phenomenon. They are a useful tool for short-term forecasting. Many adaptation models can be used. The most popular are exponential smoothing models. They assume that the observed sequence is the sum of the trend and the random component, and the trend might change and refraction at that time. The common record model has the form:

$$y_i = T_i + e_i$$

The form of the trend model taking into account the exponential smoothing method is as follows:

$$T_i = \alpha y_{i-1} + (1 - \alpha) T_{i-1}$$

Parameter  $\alpha$  is called the smoothing parameter and is defined from the interval  $(0; 1)$ . For the initial evaluation, trend usually takes the value of the first observation  $y_i$ . The value of the parameter  $\alpha$  is determined empirically, so that the forecast error is as small as possible. Forecast based on adaptive models is determined by the formula:

$$\hat{y}_{i+k|i} = T_i + (T_i - T_{i-1})k$$

### 2.3. Stationarity of the time series

In many practical cases of modeling and forecasting time series it is essential to determine its stationarity, i.e., a kind of statistical equilibrium around a constant mean level as well as a constant dispersion around that mean level. To examine the stationarity of the time series, different methods are used. The most commonly used method is Augmented Dickey-Fuller test (ADF) [4], Schwarz [9], Akaike [1] or Hannan & Quinn [5]. These characteristics need to be identified, removed, or modeled before further analysis can proceed. In order to eliminate non-stationarity, the method of differentiation is used, namely the calculation of further differences according to the formula:  $\Delta y_t = y_t - y_{t-1}$ . It may happen that the first differences will not create a time series to stationarity. Then

the method of differentiation is repeated to achieve time series stationarity (in practice not more than 3 times).

An example of bringing the non-stationary

time series into stationarity was presented on the example of the volume of transported goods (Fig. 1).

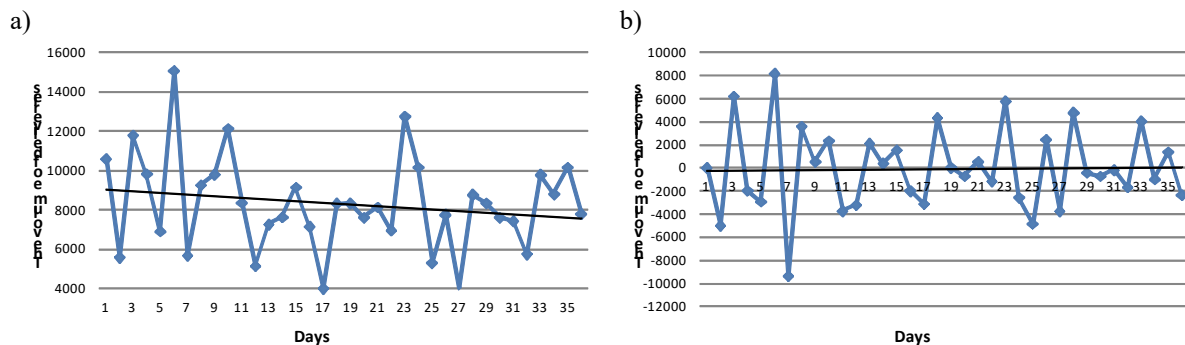


Figure 1: Transformation a time series to stationarity, a) non-stationary series  $p=0.4622$ , b) stationary series  $p=0.00071$

For the above data the extended stationary Dickey Fuller test was carried out to determine the degree of integration of the time series. The inference was made for the significance level  $\alpha = 0.05$ . In this case (Fig. 1b), the p-value at a level less than 0.05 indicates stationarity of the series. Therefore, the modeling and forecasting can be performed on non-stationary time series, which includes the changes coming from the trend.

### 3. Seasonal fluctuations in time series

Seasonal fluctuations are one of the components of the time series, which are characterized by a short period of duration (up to 1 year) and reflect the influence of behaviors associated with the occurrence of certain periodic changes, for example, the rhythm of work throughout the week or day, time of year, holidays, etc. Variation over a period of one year, though results can readily be adopted to Rother periods (e.g. variation on different days of the week) [3]. To use the seasonal fluctuations, it is necessary to perform several assumptions. The time units used for the dimensions should be equal. Historical data should cover a period in which the parameters of the model describing the process were constantly. The number of

observations should be as large as possible. There are many problems in identifying seasonality in the time series. The greatest concern of standardization of the length of the study period or the inclusion of public holidays. Serious problems in cases where out-of-date observations occur or when they are missing can be observed. Then on the built model a strong impact may have the rare factor. Outliers can be determined by calculating the first quartile ( $Q_1$ ) and third ( $Q_3$ ) of stationary time series and setting the magnitude of the likely deviation  $R_Q = Q_3 - Q_1$ . All outgoing observations are outside the range  $(Q_1 - 3R_Q; Q_3 + 3R_Q)$ .

The seasonal component of time series can be performed using various methods. For known period of changes it is a method of indicators, harmonic analysis [5], exponential smoothing models (Winters model) [12]. For an unknown period of changes, models with delayed predicted values like ARMA, autocorrelation analysis are used.

### 4. Cyclical fluctuations in a time series

Cyclic fluctuations are recurrent. The changes are not mechanical – in spite of its frequency, occur at different times. It is difficult to predict the time of its occurrence, but also the factor that causes

this change. Research into the nature of variability of the time series has led to the formation of many theoretical models, on the one hand, to identify empirical regularities that can be observed in subsequent cycles.

Cyclicity can occur in cycles of different frequencies (high - frequent fluctuations, low - fluctuations rarely occur). It is easier to observe the cyclicity when the oscillation frequency is high and the length is small. It is more difficult to observe fluctuations when their frequency is low and the fluctuation cycle is long, often multiyear. Hence came the conclusion that cyclical fluctuations are those that have a term longer than one year. However, these fluctuations can be observed also when the number of observations in the time series is very large.

Spectral analysis in time series uses spectral analysis, which consists of transforming the time series into a series in the frequency domain. It is an analysis of the time series by decomposition into sines, cosines, amplitudes, wavelengths, and phase angles. This is accomplished using a discrete Fourier transform in accordance with the formula:

$$f(t) = \sum_{k=-\infty}^{+\infty} a_k e^{jk\omega_0 t} \quad (4)$$

for  $-T/2 \leq t \leq T/2$

where  $\omega_0 = 2\pi / T$ , and the coefficients  $a_k$  are determined by the following formula:

$$a_k = \frac{1}{T} \int_{-T/2}^{T/2} f(t) e^{-jk\omega_0 t} dt \quad (5)$$

The Fourier transform apply to periodicity in time series. Periodic function  $f(t)$  continuous independent variable  $t$  with period  $T$ .

Easier method of determining the periodicity of a time series is the use of Fast Fourier Transform (FFT). FFT is based on Divide and Conquer Algorithm by dividing the transform value  $N = N_1 N_2$  by the magnitude of  $N_1$  and  $N_2$  using  $O(N)$  operations of multiplication. To built the model FFT is determined by the formula:

$$y_t = \alpha_0 + \sum_{i=1}^{n/2} \left( \alpha_i \sin\left(\frac{2\pi}{n} \cdot i \cdot t\right) + \beta_i \cos\left(\frac{2\pi}{n} \cdot i \cdot t\right) \right) \quad (4)$$

where:

$i$  - number of harmonics,

$\alpha_0, \alpha_i, \beta_i$  - model parameters,

The parameter  $\alpha_0$  may be equal to  $f(t)$ , if the time series is non-stationary. The  $f(t)$  represent the function of a trend.

The model parameters are determined based on the method of least squares in accordance with the formulas:

$$\alpha_0 = \frac{1}{n} \sum_{t=1}^n y_t \quad (6)$$

$$\alpha_i = \frac{2}{n} \sum_{t=1}^n y_t \cdot \sin\left(\frac{2\pi}{n} \cdot i \cdot t\right) \quad (7)$$

for  $i = 1, \dots, \frac{n}{2} - 1$

$$\beta_i = \frac{2}{n} \sum_{t=1}^n y_t \cdot \cos\left(\frac{2\pi}{n} \cdot i \cdot t\right) \quad (8)$$

for  $i = 1, \dots, \frac{n}{2} - 1$

where:

$a_0, a_i, b_i$  - parameter estimates of  $\alpha_0, \alpha_i, \beta_i$ .

Using trigonometric functions the harmonics can be calculated. First harmonic has a period equal to  $n$ , the second  $n/2$ , one  $n/3$ , etc. The maximum number of harmonics is half the number of observations in the time series.

The forecast is determined by inserting in the  $i$ -th value, the next value of the time series.

The use of spectral analysis depends on the presence in the time series of cyclicity. This is accomplished by the periodogram, that is the frequency of sinusoidal waves that are inverse to duration of observation. When the periodogram shows no frequency changes, the use of spectral analysis is pointless.

## 5. Conclusions

To build a model mapping the actual mileage, and then determine on the basis of his forecast, it is necessary to perform a series of actions preparing the time series for the corresponding transformations. The first thing to do is to determine the trend in the time series. Not take into account trends may lead to incorrect implementation of the process of determining the forecast, which will be characterized by a small level of relevance. With this purpose it is necessary to investigate the stationarity of the time series, that is, to carry out a unit-element test, on the basis of which you can assign a level of significance of the studied time

series. In the case of cyclical fluctuations, for simplicity of calculations it is better to apply a Fast Fourier Transform. Based on the spectrum of time series, it can be stated whether the observations contained in a time series change cyclically over a period of time. Decomposition of the time series allows to pay attention to those elements that are not visually possible to determine, and which occur in series and which can be used in the forecasting process. The preparation process of the time series is as important as the method chosen to generate forecasts.

## References

- [1] Akaike H. A New Look at the Statistical Model Identification. *I.E.E.E. Transactions on Automatic Control*. No. 19, 1974.
- [2] Barbosa R. M., Silva M. E., Fernandes M. J. *Time Series Analysis of Sea-Level Records: Characterising Long-Term Variability*, edited by Donner R. V., Barbosa R. M. *Nonlinear Time Series Analysis in the Geosciences: Applications in Climatology, Geodynamics and Solar-Terrestrial Physics*. Springer-Verlag Berlin Heidelberg, 2008, pp. 158.
- [3] Chatfield C. *Time-series Forecasting*. CRC Press. 2001, pp. 21.
- [4] Dickey D.A., Fuller W.A. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, No. 74, 1979.
- [5] Hannan E.J., Quinn B.G. The Determination of the Order of an Autoregression. *Journal of the Royal Statistical Society, B*. 1979, No. 41.
- [6] Katznelson Y. *An introduction to Harmonic Analysis*, 3<sup>rd</sup> Edition. Cambridge University Press, 2004.
- [7] Khandelwal I., Adhikari R., Verma G. International Conference On Computer, Communication And Convergence (ICCC 2015). *Procedia Computer Science*. Vol. 48, 2015, pp. 173-179.
- [8] de Oliveira J., Ludermir T. *A hybrid evolutionary decomposition system for time series forecasting*. *Neurocomputing*. Vol. 180, 2015, pp. 27-34.
- [9] Qiu X., Ren Y., Suganthan P. N., Amaratunga G. A. J. Empirical Mode Decomposition based ensemble deep learning for load demand time series forecasting. *Applied Soft Computing*, Vol. 54, 2017, pp. 246-255.
- [10] Schwarz G. Estimating the Dimension of a Model. *Annals of Statistics*, 1978, No. 6.
- [11] Wei L. Y. A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting. *Applied Soft Computing*. Vol. 42, 2016, pp. 368-376.
- [12] Yaffee R., McGee M. *Introduction to Time Series Analysis and Forecasting with Applications of SAS and SPSS*, Academic Press, 2000, pp. 476.