Time Series Data Prediction and Analysis

by

Oleg Ostashchuk

in the
Faculty of ...

Contents

Chapter 1

Introduction

Time series is a common mathematical expression that can be frequently observed in various texts about statistics, signal processing or finances.

Every day newspapers contain business sections, that report daily stock prices, weekly interest rates, monthly rates of unemployment. Meteorology records hourly wind speeds, daily maximum and minimum temperatures and annual rainfall. Geophysics are continuously observing the shaking or trembling of the earth in order to predict possibly impending earthquakes. The social sciences survey annual death and birth rates, the number of accidents in home and various forms of criminal activities. [1] There are, obviously, numerous other examples related to time series. The figure 1.1 illustrates very common graphical representation of time series example.

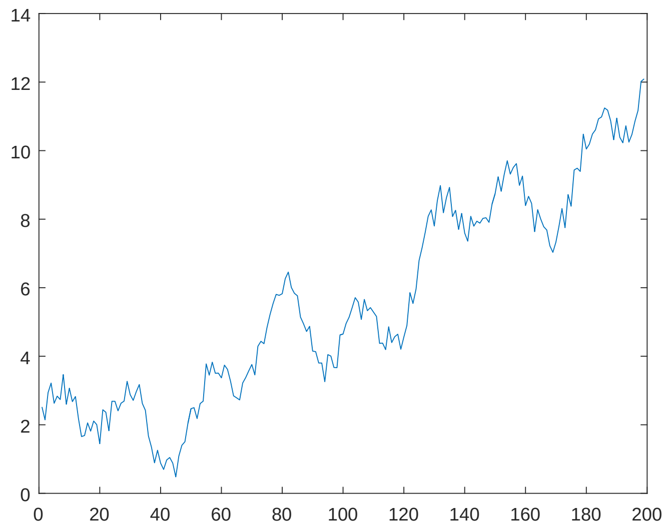

FIGURE 1.1: Time series example

The term time series itself, denotes a data storing format, which consists of a sequence of values recorded at specified time units. Values should represent the same meaning, in other words, there should be correlations among observed values, and at the same time there can be at most one value for each time unit. Therefore, not all sequences of values fulfill the time series requirements.

In theory, time series can be basically divided into two main groups, according to the definition of its time units' set.

- A discrete time series, are those, in which the set of time units, at which observations are made, is a discrete set. For example, it is a set, when observations are made repeatedly at some fixed time intervals. Different economic indicators, like stock prices, usually belongs to this group of time series data.
- Continuous time series are obtained, when observations are recorded continuously over some time interval. Typical examples are measurements from scientific sensors.

Despite the continuousness of some time series, in practice they are nevertheless stored in computer systems, in the manner of discrete values. Therefore, all test data for the future experiments will be perceived as a discrete time series. On the other hand, despite the discreteness of other time series, discrete time series are often illustrated in the graphs with the continuous curve.

1.1 Time Series Types Classification

There could be many various time series classifications based on specific characteristics. The most common of them are very intuitive and are well documented in publication [].

Based on the time steps of measurements, time series data are classified into:

- equidistant time series
- non-equidistant time series

Equidistant time series are formed by making observations of some process over the fixed intervals of time. A lot of physical or environmental processes are described by this kind of time series. Non-equidistant time series are those time series, which do not keep the constant distance between observations. For example, different markets' indicators demonstrate non-equidistant time series, as the price changes are not necessary performed within regular time intervals.

According to the rate of value dependency on their previously observations, the time series divided into:

- long memory time series
- short memory time series

Time series with long memory are those, for which the autocorrelation function decreases slowly. Here usually belong the processes, which don't have fast turnovers, like different physics measurements, air temperature. Short memory time series, those with autocorrelation function decreasing more rapidly, have typical examples from economic sphere, market prices indicators.

Another classification is:

- stationary time series
- non-stationary time series

Stationary time series are time series, for which statistical properties like mean value or variance, are constant over time. These time series stay in relative equilibrium relatively to the mean values. Other time series belong to non-stationary time series. In industry, trading or economy, time series usually belongs to the non-stationary category. In order to deal with the forecasting, non-stationary time series frequently have to be transformed to the stationary ones.

1.2 Time Series Analysis

Time series analysis unites a group of methods for analyzing time series data, the main goal of which is to understand the behavior or structure of data points, in order to extract potentially useful information.

Time series forecasting is one of the most important analysis method, performed over the time series data. General idea is based on the fact that information about the past events can be effectively exploited to build a predictions about the future events. In practice, it often happens, that processes required to be predicted are usually stored in time series format. From the point of view of time series data, this means that forecasting models use already measured values to predict future values before they are observed. Typical example is prediction of the product's future price, based on the last price changes.

There are many processes in different fields of business, industry and technology, which produce large amount of data in time series format. Usually, it could be different technical measurements from sensors, performance or load measurements. Another large group represents different economic indicators like electricity prices, number of sales, stock prices, interest ratings. Forecasting of all these mentioned processes is an important task.

Today, in the age of computers, the abilities of time series forecasting have increased. It is believed and observed that computers can make predictions much more precisely than humans. There have also been invented new forecasting methods, based on machine learning algorithms.

1.3 Aim and Structure of Thesis

The main aim of this thesis is to compare effectiveness of specific time series forecasting methods on the various practical test data from different fields. The specified aim is going to be achieved by accomplishing the following steps:

1. Make a research on time series forecasting methods and choose the most perspective methods.
2. Select practical sets of test data from different fields.
3. Implement forecasting models for each combination pair of test data x method.
4. Perform tests and compare prediction quality of methods for each set of test data.
5. Make a conclusion about methods' effectiveness for concrete set of test data and opportunities of improvements.

[Structure will be specified after thesis completion.]

Chapter 2

Time Series Forecasting

2.1 Forecasting Formal Definition

The word forecasting itself, usually means making predictions about future, by using scientific methods. According to the publication [1], the processes that are required to be predicted, are often described by time series data. Time series contains two mandatory sets of numbers, the set of time units, and the set of corresponding measurement's values.

While solving the time series forecasting problem, it is necessary to distinguish two interrelated terms, forecasting method and forecasting model.

- Forecasting method – represents a sequence of actions, that are necessary to perform, in order to obtain the time series forecasting model.
- Forecasting model – is a functional representation, that adequately describes time series. It is a basis of time series' future values prediction process.

There are two main ways, how time series forecasting tasks are performed. The first option is based on the computations, that use only the past values of the same time series, to predict the values in future. The second option allows to use not only the past values of the same time series, but also another external factors in addition, that can be useful for prediction. In this case, external factors are very often presented as another time series. Time series of external factors are not obliged to have the same time step intervals, as the original time series data, but on the other hand, it is expected that external factors should have some influence on the original time series' progress. For example, an intuitive external factors of energy consumption could be various meteorological indicators, like air temperature or air humidity.

2.1.1 Forecasting without external factors

Time series forecasting without external factors. If the observations of some stochastic process are available at discrete units of time $t = \{1, 2, \ldots, T\}$, then the sequence of values $Z(t) = \{Z(i) \mid i \in T\} = \{Z(1), Z(2), ..., Z(T)\}$ is denoted as a time series.

Let's assume that at the moment of time unit $- T$, it is necessary to make a forecast of $- l$ future values of the given process $Z(t)$. In other words, it is needed to determine the most probable future values for each of the time units $\{T+1, \ldots T+l\}$. Time unit $-$ **T** is a moment when the forecast is performed, it is usually named by term "**origin**". The parameter $- l$ is denoted as a "**leadtime**", it represents the number of future values that are going to be predicted.

In order to calculate the time series values at future time units, it is necessary to determine functional dependency that describes a relationship between past and future values of the given time series. The forecast is based on $- k$ past values, denoted as an input vector Z_T. As a result, the vector of $- l$ future predictions will be obtained, denoted as an output vector $\hat{Z}_T$. All predicted values $\hat{Z}(i)$ will be marked with sign ˆ in order to label them as predictions, not the real values.

$$Z_T = \begin{pmatrix} Z(T) \\ Z(T-1) \\ Z(T-2) \\ \vdots \\ Z(T-k) \end{pmatrix} \qquad \hat{Z}_T = \begin{pmatrix} \hat{Z}(T+1) \\ \hat{Z}(T+2) \\ \vdots \\ \hat{Z}(T+l) \end{pmatrix} \tag{2.1}$$

$$f(Z_T) = \hat{Z}_T \tag{2.2}$$

The functional dependency (2.2) is usually denoted as forecast function and it represents the forecast model. The intuitive aim is to find the forecast function such that the deviations between predicted values and actual values, that will be observed later in future, are as small as possible.

$$\varepsilon_T = \begin{pmatrix} Z(T+1) \\ Z(T+2) \\ \vdots \\ Z(T+l) \end{pmatrix} - \begin{pmatrix} \hat{Z}(T+1) \\ \hat{Z}(T+2) \\ \vdots \\ \hat{Z}(T+l) \end{pmatrix} \tag{2.3}$$

Analysis of deviations vector (2.3) represents a basis of so called "loss function" or "error function". This function measures the quality of forecast, based on the measured deviations. There are more options, how to calculate rate of quality from the deviations vector, usually root mean square error or mean absolute deviation are calculated. More details about error functions will be discussed in section 2.2. The formal objective of time series forecasting is then formulated as a minimization of loss function.

In addition to calculations of future values, sometimes it is required to determine accuracy limits. The accuracy of the forecasts may be expressed by calculating probability limits on either side of each forecast. These limits may be calculated for any convenient set of probabilities. They are such that the realized value of the time series, when it eventually occurs, will be included within these limits with the stated probability. [2]

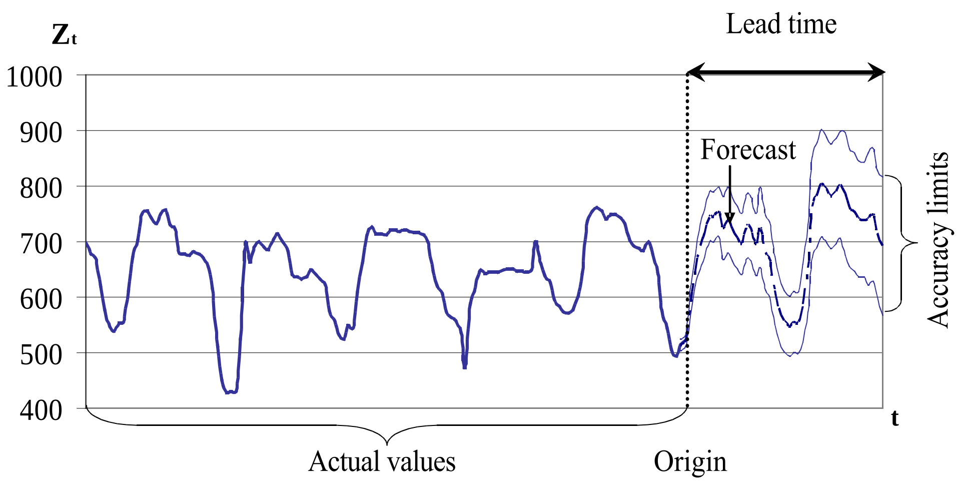

Figure 2.1: Time series forecasting without external factors

2.1.2 Forecasting with external factors

Time series process $Z(t)$ is specified at the discrete time units $t = \{1, 2, \ldots, T\}$. It is assumed, that this time series is affected by a set of external factors $\{X_1(t_1), X_2(t_2), \ldots X_m(t_m)\}$. Each external factor is represented as an independent time series process. For example, an external factor $X_1(t_1)$ is specified at the corresponding discrete time units $t_1 = \{1, 2, \ldots, T_1\}$.

The original time series $Z(t)$ and external factors $X_i(t_i)$ are not obliged to be specified at same time units. If the time units $t, t_1, t_2, \ldots, t_m$ are not equal, then it is necessary to recalculate the values of external factor to a single scale t.

Let's assume that at the moment of time unit T, it is necessary to make a forecast of $-\ l$ future values of the given process $Z(t)$. In order to calculate the predictions, it is necessary to determine functional dependency, that describes a relationship between past and future values, also considering the impact of external factors.

$$Z_T = \begin{pmatrix} Z(T) \\ Z(T-1) \\ Z(T-2) \\ \vdots \\ Z(T-k) \end{pmatrix} \qquad X_{i,\,T} = \begin{pmatrix} X_i(T+l) \\ \vdots \\ X_i(T+1) \\ X_i(T) \\ X_i(T-1) \\ \vdots \\ X_i(T-k) \end{pmatrix} \qquad \hat{Z}_T = \begin{pmatrix} \hat{Z}(T+l) \\ \vdots \\ \hat{Z}(T+2) \\ \hat{Z}(T+1) \end{pmatrix} \tag{2.4}$$

$$f(Z_T,\ X_{1,T},\ X_{2,T},\ ...,\ X_{m,T}) = \hat{Z}_T \tag{2.5}$$

The functional dependency (2.5) is a forecast function and it represents the forecast model with external factors. The rest tasks are performed in the same way as they were in the case of forecasting without external factors. The main objective is to find the forecast function such that the deviations between predicted values and actual values, that will be observed later in future, are as small as possible. This objective formulates minimization task of so called ”loss function” or ”error function”. More details about error functions will be discussed in section 2.2.

The accuracy limits may be calculated for any convenient set of probabilities. Accuracy limits are such that the realized value of the time series, when it eventually occurs, will be included within these limits with the stated probability. [2]

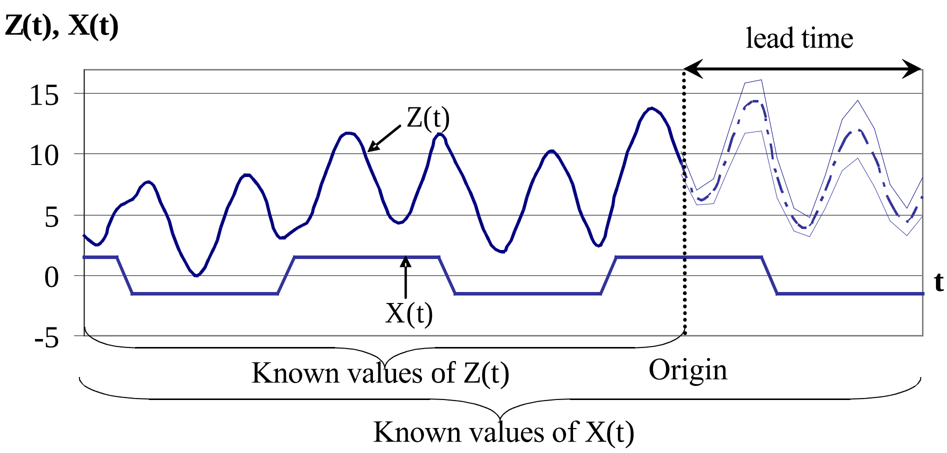

FIGURE 2.2: Time series forecasting with external factors

2.2 Forecasting Accuracy

The forecast error is defined as a deviation of predicted value and actual value:

$$\varepsilon(t) = Z(t) - \hat{Z}(t) \tag{2.6}$$

This error corresponds for one predicted value. If the forecast is performed for more than one future value, then the individual errors formulate a vector of errors. Error function, mentioned in the previous section, aggregates individual errors into one summary error, that represents the forecast's accuracy rate. Minimization of the error function is a general objective of the forecasting task.

Measures of aggregate error:

- Root Mean squared error (RMSE)

$$RMSE = \sqrt{\frac{1}{l}\sum_{t=1}^{l}(Z(t) - \hat{Z}(t))^2} \tag{2.7}$$

- Mean squared error (MSE)

$$MSE = \frac{1}{l}\sum_{t=1}^{l}(Z(t) - \hat{Z}(t))^2 \tag{2.8}$$

- Mean absolute error (MAE)

$$MAE = \frac{1}{l}\sum_{t=1}^{l} \mid Z(t) - \hat{Z}(t) \mid^2 \tag{2.9}$$

- Sum of squared errors (SSE)

$$SSE = \sum_{t=1}^{l}(Z(t) - \hat{Z}(t))^2 \tag{2.10}$$

The suitability of these quality measures are very similar and they differs only a little bit, for example strong errors are penalized by RMSE less than by other measures. All these measures are suitable for comparison of different forecasting models on the same test data, but not suitable for comparison of one forecasting model on various test data. Therefore, each test data have to be at first preprocessed, from the time series of absolute values into the time series of relative changes. [3]

2.3 Forecasting Models Overview

It is necessary to remark, that the names of time series forecasting model and the corresponding forecasting methods, often coincide. For example, one of the most popular forecasting model ARIMAX (Auto Regression Moving Average External). This model and its corresponding method is usually called ARIMAX. Today, it is expected that exist more than 100 different forecasting models, but the number of the common models' classes is far less.

2.3.1 Regression models

Regression analysis is a method for investigating functional relationships among variables. The relationship is expressed in the form of an equation or a model connecting the response or dependent variable and one or more explanatory or predictor variables.

The process, that is required to be predicted is denoted by the response variable Z. Set of predictor variables are denoted as $X_1, X_2, \ldots, X_p$, where p denotes the number of predictor variables. The relationship between Z and $X_1, X_2, \ldots, X_p$ can be approximated by the regression model:

$$Z = f(\ X_1, X_2, \ldots, X_p\ ) + \varepsilon \tag{2.11}$$

Where epsilon is assumed to be a random error representing the discrepancy in the approximation. [4]

The linear regression model is the simplest and the most widely used regression model. It assumes, that there is set of external factors $X_1(t), X_2(t), \ldots, X_p(t)$, which have an impact on the given process $Z(t)$ and the relationship between them is linear. Forecasting model based on the linear regression is determined by an equation (2.12).

$$Z(t) = \alpha_0 + \alpha_1 X_1(t) + \alpha_2 X_2(t) + \ldots + \alpha_p X_p(t) + \varepsilon_t \tag{2.12}$$

Where $\alpha_i, i = 0 \ldots p$ are regression coefficients (parameters), ε is the approximation error. In order to obtain a forecasted values $Z(t)$ at time units t, it is necessary to have values $X_i(t)$ at time moment t, sometimes in practice this can be impossible in some kind of problems.

The nonlinear regression models are based on assumptions, that there is given a mathematical function, that describes relationship between given process $Z(t)$ and the external factor $X(t)$.

$$Z(t) = f(X(t), \alpha) + \varepsilon_t \tag{2.13}$$

While constructing the forecast model, it is necessary to determine the function parameters α. For example, $Z(t)$ dependency on $sin(X(t))$

$$Z(t) = \alpha_1 sin(X(t)) + \alpha_0 + \varepsilon_t \tag{2.14}$$

In order to construct this model it is sufficient only to determine the parameters $\alpha = (\alpha_0, \alpha_1)$. However in practice it is not very common, that type of functional dependency between process $Z(t)$ and external factor $X(t)$ is already known in advance. Therefore, nonlinear regression models are used less frequently, than the linear ones.

2.3.2 Autoregressive models

Autoregressive models are based on the idea, that values of process Z(t) are linearly dependent on some number of past values of the same process Z(t). Autoregressive and moving average models belongs to the most popular time series analysis. Autoregressive model is known due to its high effectiveness for some kind of time series. In this model, the current value of process is expressed as a sum of finite linear combination of previous values and the impulses, called white noise.

$$Z(t) = c + \varphi_1.Z(t-1) + \varphi_2.Z(t-2) + ... + \varphi_p.Z(t-p) + \varepsilon_t \tag{2.15}$$

The formula describes the autoregressive model of order p. In written texts this model is frequently marked as $AR(p)$. Individual signs meanings: φ_i - parameters of the model, c - a constant, ε - white noise (error of the model). To determine the values of c and φ_i parameters, mean least squares or maximum likelihood methods are used.

The second model, moving average model. It plays very important role in time series description and it is very often used in relation with the autoregressive models. Moving average model of order q is described by formula:

$$Z(t) = \frac{1}{q}[Z(t-1) + Z(t-2) + ... + Z(t-q)] + \varepsilon_t \tag{2.16}$$

In literature, moving average model of order q is often denoted as $MA(q)$. Signs' meanings, q - order of moving average, ε_t - prediction error. Moving average model is in fact a finite impulse response filter applied to white noise.

In order to achieve better prediction quality, two previous models are often merged into one model, autoregressive and moving average model. Common model is denoted as $ARMA(p, q)$ and it unites a moving average filter of order q and autoregression of filtered values of order p.

If the time series data show evidence of non-stationarity, then the initial differencing step can be applied to reduce the non-stationarity. This model is usually denoted as $ARIMA(p, d, q)$. The parameter d - the degree of differencing, corresponds to the *integrated* part of the model.

Another option is an $ARIMAX(p, d, q)$ model, that is an extension of $ARIMA(p, d, q)$ model. It is described by formula:

$$Z(t) = AR(p) + \alpha_1 X_1(t) + ... + \alpha_S X_S(t) \tag{2.17}$$

This model is extended by impact of external factors. In this model, the process $Z(t)$ is a result of model $MA(q)$, that are filtered values of the original process. Subsequently autoregressive forecasting, with additional regression parameters corresponding to external factors, is applied.

2.3.3 Exponential smoothing models

Exponential smoothing models are widely used for modeling finance and economical processes. The basis of exponential smoothing, is an idea of repetitive revision of forecasting function, with each income of newly observed value. Exponential smoothing model assigns exponentially decreasing weights to past values, according to the age. Therefore, newly observed values have higher impact on forecasted value, than the elder ones. Model function:

$$Z(t) = S(t) + \varepsilon_t \tag{2.18}$$

$$S(t) = \alpha \cdot Z(t-1) + (1-\alpha) \cdot S(t-1) \tag{2.19}$$

α - smoothing coefficient, $0 < \alpha < 1$. Initial conditions are set as $S(1) = Z(0)$. In this model, each subsequently smoothed value $S(t)$ is a weighted combination of previous time series value $Z(t-1)$ and previously smoothed value $S(t-1)$.

2.3.4 Neural network models

Artificial neural network forecasting models and artificial neural networks itself, represent the class of machine learning algorithms, that are inspired by the concept of human's brain and its building structure. As it is known from the biology, human's brain and other parts of the nervous system are built from nerve cells, so called neurons. Neurons are used to process and transmit the information of the electrical or chemical signals.

In the sphere of information technology, artificial neural networks (ANN) are being used for solving various complex problems. ANNs are built from artificial neurons. The model artificial neuron is illustrated on the figure 1.2.

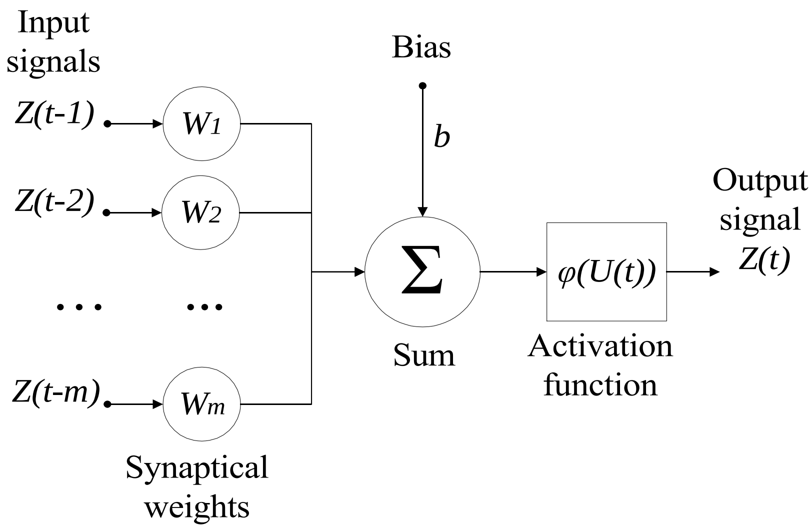

Figure 2.3: Neuron model

The neuron model can be described by the following equations:

$$U(t) = b + \sum_{i=1}^{m} w_i \cdot Z(t-i) \tag{2.20}$$

$$Z(t) = \varphi(U(t)) \tag{2.21}$$

Where $Z(t-1), \ldots Z(t-m)$ - input signals, $w_1, \ldots w_m$ – synaptical weights, b - bias (threshold), $\varphi()$ – activation function. In theory, activation function could be any differentiating function, but in practice it is mostly used one of these functions:

- binary step function
- sigmoid function
- hyperbolic tangent function

ANN consists of interconnected neurons. The way, how the neurons are connected between each other, determines the structure of the ANN. According to the structure, ANNs can be divided into:

- one layer artificial neural network
- multi-layer artificial neural network
- recurrent neural network

In theory, ANNs can be used for approximation of any non-linear function. Therefore, ANNs can be used for modeling of non-linear dependency function for forecasting time series future values based on the its past values. ANN forecasting models also can be easily applied for time series forecasting with impact of additional external factors.

2.3.5 Markov chain models

Forecasting models based on the Markov chains assume, that future state of the process is dependent only on its current state and is not dependent on its elder states. Markov chain models are applicable on the short-memory time series. Example of Markov chain for process with 3 states is illustrated on figure 1.3.

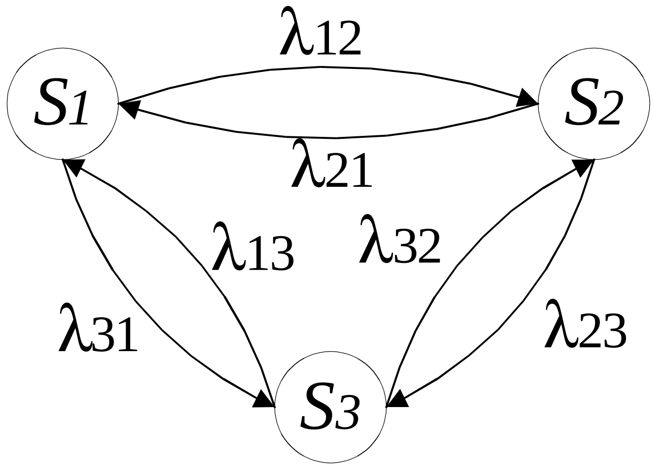

FIGURE 2.4: Markov chain model

In this model, $S1, S2, S3$ - are states of process $Z(t)$, α_{xy} – probability of transition from state x to state y. By building the Markov chain model, the set of states and corresponding transitions' probabilities are defined. If the current process state is defined, the future state is selected as the state with maximal transition probability. If the transition probabilities are properly stored in matrix, subsequent future values can be determined by probability matrix's multiplication and maximum probability selection.

2.4 Forecasting models comparison

In the previous section, there were described the most common time series forecasting models and methods. In this section, mentioned forecasting will be compared by specifying their advantages and disadvantages.

Regression models and methods. The advantages of the given models are simplicity, flexibility and uniformity of calculations and model's construction. Linear regression models are even accessible that the nonlinear ones. Another advantage is a transparency of all intermediate calculations. The main disadvantages of nonlinear regression models is a complex task of functional dependency determination. The disadvantages of the linear regression models are low adaptability and inefficiency with nonlinear processes.

Autoregressive models and methods. The advantages of the given class of models are transparency and uniformity of calculations and model's construction. Today, this class of models is the most frequently used, therefore, there is a lot of information how to apply this models for the specific problems. The disadvantages of this models are large number of parameters required to be determined, low adaptability, linearity, inefficiency with nonlinear processes.

Exponential smoothing models and methods. The advantages of the given models are simplicity and uniformity of calculations and model's construction. This models are often used for long term forecasting. The disadvantage of this model is inflexibility.

Artificial neural networks models and methods. The main advantage of these models is a nonlinearity. The neural networks are capable to deal with the nonlinear dependencies between future and past values of the processes. Another important advantages are adaptability and scalability (ability of parallel computations). The disadvantages of neural networks' models are absence of transparency, complexity of architecture, high performance requirements during the process of learning.

Markov chains models and methods. The advantages of Markov chains models are uniformity of design and analysis. The disadvantage of these models is an impossibility of long term forecasting.

Chapter 3

Time Series Data

3.1 The Additive and Multiplicative Time series models

The additive model of time series is based on the assumption, that a given time series $Z(t)$ is an aggregate combination of four components. This concept is a basis of time series decomposition, and plays an important role in time series data preprocessing, in order to reach a better forecasting quality. []

$$Z(t) = T(t) + L(t) + S(t) + R(t) \tag{3.1}$$

- $T(t)$ - Represents a (monotone) function of t, called a trend, that impacts the growth or decline of the values.
- $L(t)$ - Represents some nonrandom long term cyclic influence.
- $S(t)$ - Describes some nonrandom short term cyclic influence.
- $T(t)$ - Component of random deviations, that has a non-deterministic influence on the time series values.

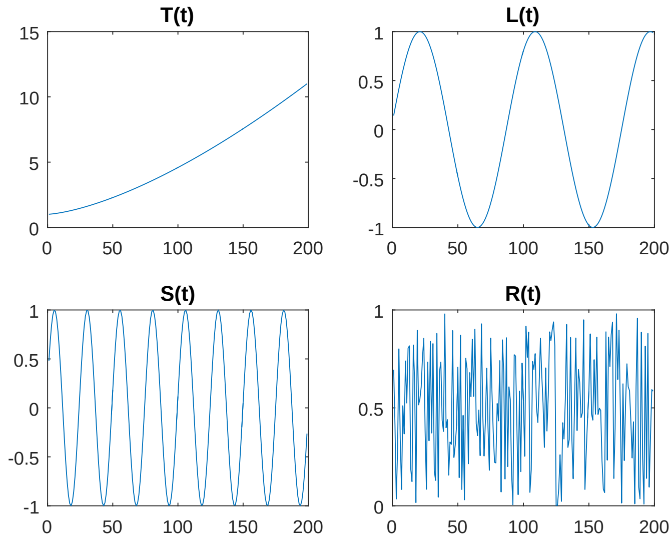

FIGURE 3.1: Additive model example - T(t), L(t), S(t), R(t) components

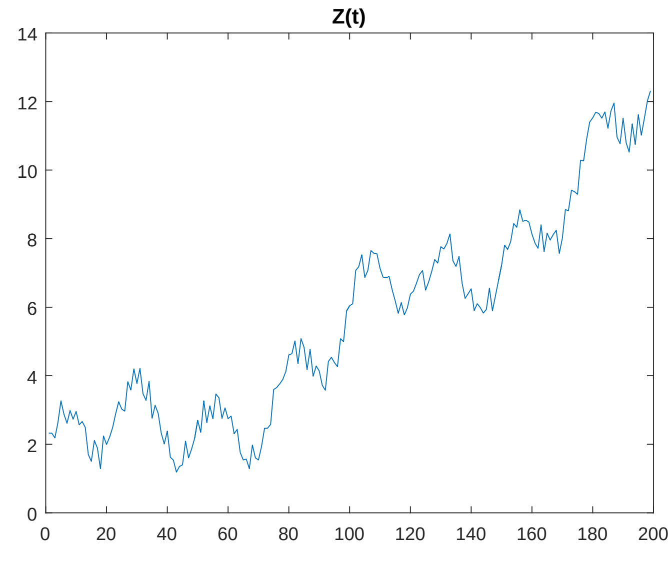

FIGURE 3.2: Additive model example - Z(t) time series

The multiplicative model is an analogy to the additive model, time series is a combination of the same four components, but instead of taking a sum, the multiplication of the components is considered.

$$Z(t) = T(t) \times L(t) \times S(t) \times R(t) \tag{3.2}$$

3.2 Data preprocessing

Before the raw time series data can be applied to the forecasting methods, very often it have to undergo several transformations. The forecast quality greatly depends on the way, how the input data are preprocessed. Some forecasting methods, for example neural networks methods, have very strict requirements on the input data format, and the absence of the proper data preprocessing, leads to the inefficiency of the given forecasting model.

In this thesis will be compared forecast qualities of different forecasting models applied on the same data, as well as the forecast qualities of specific forecasting models applied on different time series data.

The first part, comparing the results of different forecasting models on the same data can be easily performed by using some of aggregate error calculations, mentioned in section 2.2. The second part is a little bit harder, comparing results of one model, applied to various data, requires a transformation of initial time series into time series of relative changes, this ensures the uniformity of comparison.

3.2.1 Detrending

Detrending is transformation, that removes the trend component of a series. The most common methods are calculations of the differences or relative differences (changes) between subsequent values of the time series. For the given time series $Z(t)$ the corresponding transformation will be expressed by the equations:

- Differences rate TS:
$$D(t) = Z(t) – Z(t-1) \tag{3.3}$$
- Relative difference rate TS:
$$R(t) = \frac{Z(t) – Z(t-1)}{Z(t-1)} \tag{3.4}$$

Another option is application of logarithmic return rate. This is very similar method, the only difference is taking the logarithmic values instead of absolute ones. Logarithmic return rate provides better scaling properties, which often results into better forecasting quality.

$$LR(t) = log(Z(t)) – log(Z(t-1)) = log(\frac{Z(t)}{Z(t-1)}) \tag{3.5}$$

3.2.2 Normalization

In practice, very often, there can be observed chaotic and non-stationary time series. Forecasting of this kind of time series usually contains an initial step, that is based on turning the non-stationary time series into the stationary ones. Forecasting models, like artificial neural network model usually have a bounded input requirement. Therefore, normalization of the input data is often an obligatory task, in the case of neural network models. There are various types of normalization, but one of the most common is represented by equation (3.6).

$$Z^{'}(t) = \frac{Z(t) - \mu}{\sigma} \tag{3.6}$$

Where μ is the mean value and σ is the standard deviation of the given time series.

$$\mu = \frac{1}{n}\sum_{t=1}^{n} Z(t) \qquad \sigma = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Z(t) - \mu)^2} \tag{3.7}$$

The general aim of normalization (3.6) is adjustment of the values by shifting and scaling in order to obtain a so called normal distribution of the values. This means obtaining a time series with mean property equal to 0 and standard deviation property equal to 1.

3.2.3 Scaling

Scaling is a transformation, that adjust scales of the values within some specific boundaries. The most common used scaling are transformations of values within $\langle -1, 1\rangle$ range or $\langle 0, 1\rangle$ range.

- Scaling range $\langle -1, 1\rangle$

$$Z^{'}(t) = \frac{2 \cdot Z(t) - (max + min)}{max - min} \tag{3.8}$$

- Scaling range $\langle 0, 1\rangle$

$$Z^{'}(t) = \frac{Z(t) - min}{max - min} \tag{3.9}$$

Where $min; max$ corresponds to $minimum; maximum$ values of the time series $Z(t)$.

Bibliography

[1] Michael Falk. *A First Course on Time Series Analysis — Examples with SAS*. Chair of Statistics, University of Wurzburg, aug 2012.

[2] Gregory C. Reinsel Greta M. Ljung George E. P. Box, Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control*. Wiley, fifth edition, aug 2015.

[3] Wikipedia. Forecasting, apr 2016. URL https://en.wikipedia.org/wiki/Forecasting#Forecasting_accuracy.

[4] Ali S. Hadi Samprit Chatterjee. *Regression Analysis by Example*. Wiley, fifth edition, 2012.