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' business sections 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 [2]. There are, obviously, numerous other examples related to time series. The figure 1.1 illustrates graphical representation of time series example.

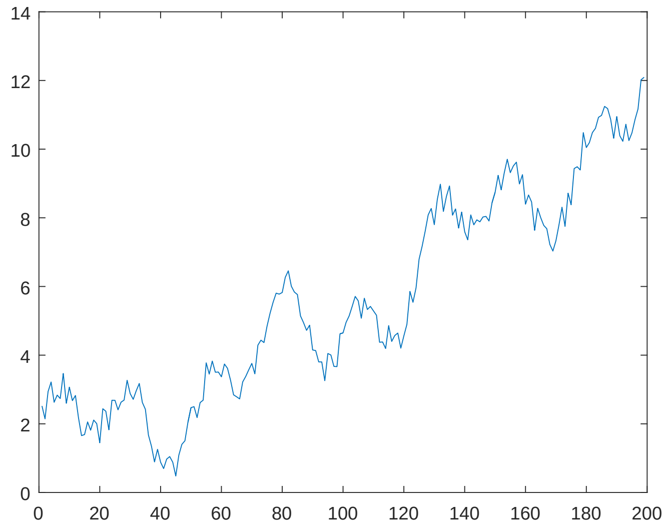

FIGURE 1.1: Time series example

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 [3].

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.

Chapter 2

Time Series Forecasting

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 not only the past values of the same time series, but also another external factors in addition, to be used for future values prediction. In this case, external factors are usually presented as another time series. Time series of external factors are not obliged to have the same time step's intervals, as the original time series data. It is expected that external factors have some influence on the original time series' progress. For example, intuitive external factors of energy consumption could be various meteorological indicators, air temperature of environment or air humidity [1].

2.1 Forecasting - Formal task definition

Let's assume that observations of some specific process are available at discrete units of time $t = 1, 2, \ldots,$ T. Then the sequence of values $Z_t = Z_1, Z_2, \ldots, Z_T$ is denoted as a time series.

At the moment of time unit T, it is necessary to make a forecast of L future values of the given process Z. In other words, it is needed to forecast time series values for each time unit $(T+1)$, $(T+2)$, $\ldots$ $(T+L)$. Time unit T is a moment, when the forecast is performed and it is called **origin**. The parameter L is denoted as a **lead time** and it represents the amount of future values to be forecasted.

In order to calculate the time series values for future time units, it is necessary to denote functional dependency that describes a relationship between past and future values of the given time series.

$$\begin{bmatrix} Z_{T+1} \\ Z_{T+2} \\ \vdots \\ Z_{T+L} \end{bmatrix} = f(Z_T, Z_{T-1}, Z_{T-2}, ...) + \begin{bmatrix} e_{T+1} \\ e_{T+2} \\ \vdots \\ e_{T+L} \end{bmatrix}$$

The functional dependency is usually called a forecast function or also a forecast model. The objective is to find forecast function such that the mean square of differences between the actual and forecasted values is as small as possible for each lead time l.

$$E = \frac{1}{L} \sum_{t=T+1}^{T+L} e_t^2 \rightarrow min$$

In addition to calculations of future values, sometimes it is also good to specify the limits of accuracy. The accuracy of the forecasts may be expressed by calculating probability limits on either side of each forecast.

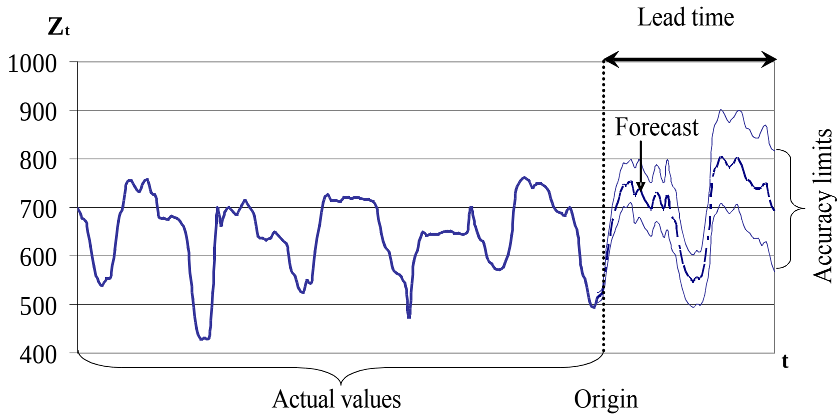

FIGURE 2.1: Time series forecasting example

2.2 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.

2.2.1 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.

$$Z(t) = c + \varphi_1.Z(t-1) + \varphi_2.Z(t-2) + ... + \varphi_p.Z(t-p) + \varepsilon_t \tag{2.1}$$

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:

$$Z(t) = \frac{1}{q}[Z(t-1) + Z(t-2) + ... + Z(t-q)] + \varepsilon_t \tag{2.2}$$

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:

$$Z(t) = AR(p) + \alpha_1 X_1(t) + ... + \alpha_S X_S(t) \tag{2.3}$$

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.

2.2.2 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:

$$Z(t) = S(t) + \varepsilon_t \tag{2.4}$$

$$S(t) = \alpha \cdot Z(t-1) + (1-\alpha) \cdot S(t-1) \tag{2.5}$$

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 $S(t-1)$.

2.2.3 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.

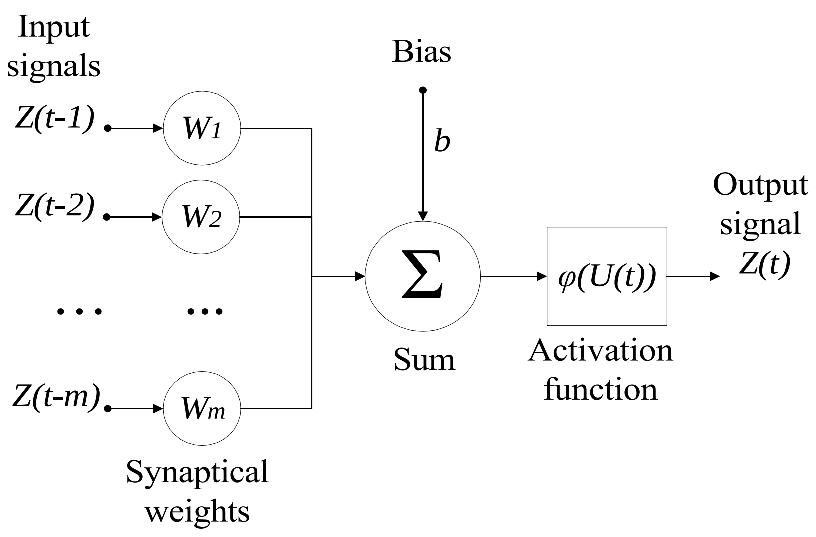

FIGURE 2.2: Neuron model

The neuron model can be described by the following equations:

$$U(t) = b + \sum_{i=1}^{m} w_i \cdot Z(t-i) \tag{2.6}$$

$$Z(t) = \varphi(U(t)) \tag{2.7}$$

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.

2.2.4 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.

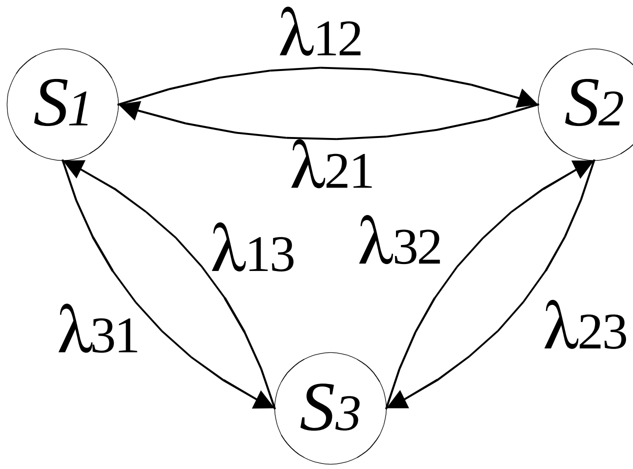

FIGURE 2.3: Markov chain model

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.

Bibliography