

# Univariate Time Series Models for Forecasting Stationary and Non-stationary Data: A Brief Review

Bashirahamad Momin and Gaurav Chavan<sup>(✉)</sup>

Department of Computer Science and Engineering,  
Walchand College of Engineering, Sangli, Maharashtra, India  
bfmomin@yahoo.com, gauravgchavan91@gmail.com

**Abstract.** Due to advancements in domain of Information processing, huge amount of data gets collected which varies according to different time intervals. Structural models and Time-series models are used for analysing time series data. Time series models are very efficient as compared to structural models because modelling and predictions can be easily done. This paper gives a brief insight into Auto-regressive Models (AR), Moving Average Models (MA), Autoregressive Moving Average model (ARMA) and Autoregressive Integrated Moving Average Model (ARIMA). This paper also helps to understand the characteristics of the data which will be used for Time-series modelling.

**Keywords:** White noise · Stationarity · Auto Correlation Function (ACF) · Partial Autocorrelation Function (PACF) · Box Jenkins approach

## 1 Introduction

Data analytics has been a major area of Interest in the industry since the past decade. The amount of data will tend to grow 10-fold in upcoming few years having approximate storage of about 50 zetabytes [1]. Due to advancements in domain of IoT (Internet of Things), many devices send the data to central node or server for processing. These data is being dumped at regular time intervals. Time series data is a collection of observations that are sampled according to time. Some common examples of time series include the hourly readings of air temperature, monitoring of a person's heart rate, and daily closing price of a particular companies stock.

Time series data are analysed in order to identify patterns in data which can be trends or seasonal variations. Hence, understanding the time series mechanism helps to develop a mathematical model which can explain the data in such a way that control, monitoring and prediction can be done with ease. Structural models like regression models can also be used for modelling time series data. But, these models are not necessarily associated with time, for example stock price which varies according to time can be modelled as change in inflation or unemployment rate using structural models like linear regression. In time series model, there is no such concept of dependant or independent variable; there is only one variable which is a model variable which varies according to different periodic trends.

The significance of time series modelling is that same data can be modelled with different periodic trends which in case of structural model are tedious task. Also some factors can be unobservable, hence excluded from regression analysis. In time series analysis, each observation is somewhat dependent upon previous observation, and gets influenced by more than one previous observation. Error term too gets influenced from one observation to another. These influences are mapped using autocorrelation, which is used to either model trend itself or to model the underlying mechanisms. To model using time-series, four common models are being used which are Autoregressive Model (AR), Moving Average Model (MA), Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA). This paper highlights methodologies of each of these time series models and describes their behaviour.

In Sect. 2 a brief review related to time series modelling is discussed. Section 3 focuses on methodologies which are used for analysing time series data. Section 4 highlights observations pertaining to time series modelling. Finally, in Sect. 5, conclusions and future work are discussed.

## 2 Related Work

In contrast to time series models, great amount of literature is witnessed in the field of neural network approach [2–6]. However, these methods failed to identify the abnormal patterns which can occur in data. Models based on Support Vector Machines (SVM) and stochastic approaches are also utilized for modelling time series data [7, 8]. Genetic Algorithms along with Neural Network are also being used [9–11]. These models were unable to give effective analysis due to complex nature of data, noisy data and high dimensionality of data. Therefore effective Time series models were needed.

The time series models helps to interpret hidden patterns of the data and helps in analysis by fitting a model for forecasting [12, 18]. In this paper, not only time series models are described but also all the way these models are being analysed is also being showcased. But it will take plenty amount of time to any researcher to understand many important concepts relative to time series models. This paper provides in depth knowledge of each of these concepts.

## 3 Time Series Modelling and Forecasting

Regression without lags fails to account for the relationships through time and overestimates the relationship between dependant and independent variables. To overcome this time series analysis needs to be done. The upcoming sub-sections will help to understand the stationary and non-stationary data, white noise process, AR, MA, ARMA, ARIMA models.

### 3.1 Stationarity

One of the important properties of time-series process is stationarity of data. So a random process or a stochastic process is known to be stationary when its joint

distribution doesn't change over time. Time series data is stochastic or probabilistic in nature because there is no accurate formula when prediction needs to be done. But usually, time series data points are weakly stationary in nature i.e. those data-points which have constant mean  $\mu$ , constant variance  $\sigma^2$  and constant auto-covariance i.e. Auto-covariance  $(Y_t, Y_{t-1}) = \text{Auto-covariance } (Y_{t-2}, Y_{t-3})$  at regular periodic intervals.

### 3.2 White Noise

A random process which has expectancy or mean, variance at any time as a constant and auto-covariance is 0. It implies that each observation is uncorrelated with other observations in the sequence (Fig. 1).

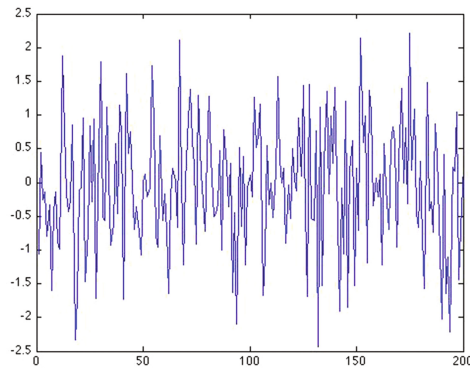


Fig. 1. White noise process [18].

The Ljung-Box statistics [13], which tests the “overall” randomness based on a number of lags, is a standard approach to determine whether the data points exhibit white noise property. Hence if white noise occurs in data, then there is no significance of doing time series modelling or estimation over such data.

### 3.3 Moving Average Model (MA)

Time series data at time interval  $t$  is given as  $y_t$ , where  $y_t$  which is current value at time  $t$  is considered as linear combination of different white noise processes  $u_t$ . The model can be represented as follows:

$$y_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \dots + \theta_q u_{t-q} \tag{1}$$

Where,

$\theta_1, \theta_2, \dots, \theta_q$  are parameters which are estimated using maximum likelihood function. Lag operator  $L$  used to represent same data for previous period.

For example,

$$u_{t-1} = Lu_t \text{ and } u_{t-2} = Lu_{t-1}$$

Therefore,

$$u_{t-2} = L^2u_t$$

The lag operator is used to make a representation of MA series much better as estimation interpretation becomes much easier. Hence equation is given by:

$$y_t = \sum_{i=1}^q L^i \theta_i u_t \tag{2}$$

### 3.4 Auto Regressive Model (AR)

The AR model depends on the past values and error terms.

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \mu_t \tag{3}$$

Where,

$\mu_t$  is a error term given as white noise and  $p$  is a lag term.

In terms of Lag operator AR model can be represented as:

$$y_t = \sum_{i=1}^p \phi_i L^i y_t + \mu_t \tag{4}$$

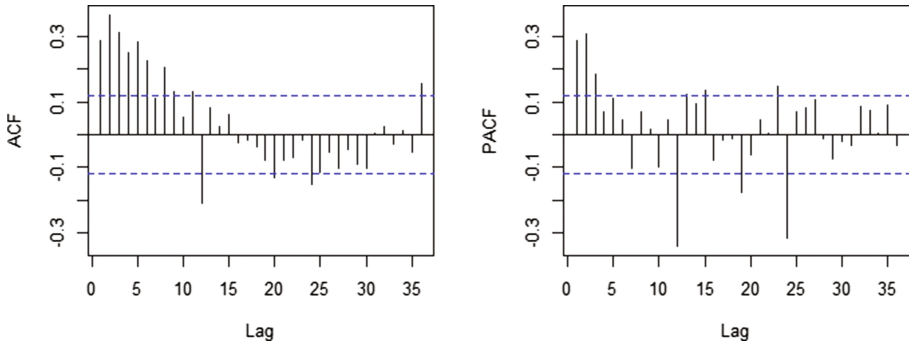
Stationarity condition matters a lot in AR estimation. Non stationary series has a non-declining effect in AR estimation which is undesirable in time series modelling.

### 3.5 Auto Regressive Moving Average Model (ARMA)

ARMA (p, q) is the process which combines AR series as well as MA series. Mathematically, it can be represented as:

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \dots + \theta_q u_{t-q} + u_t \tag{5}$$

The ARMA (p, q) process is detected by plotting correlogram of Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) against lags, which shows declining curve as lag increases in both ACF and PACF cases as shown in Fig. 2 below:



**Fig. 2.** The ACF and PACF correlogram for ARMA model.

As shown in Fig. 2. The coefficients of acf and pacf values go to negative axis when the coefficient or residual term comes out to be negative.

**Box Jenkins Approach for ARMA Model**

Box and Jenkins [14] suggested that differencing non-stationary series one or more times can achieve stationarity. Doing so leads to an ARIMA model, with the “I” standing for “Integrated”. The approach consists of three main stages which are:

1. Identification.
2. Estimation and
3. Diagnostic checking.

In first stage, the orders of the model are determined. We plot correlogram of ACF and PACF to identify the lag values. In second stage, parameters  $\phi$  and  $\theta$  are estimated. The parameters are estimated using Ordinary Least squares [15] and maximum likelihood function [16]. The third stage checks whether the model fit is good. On the other hand, the residual diagnostics, checks whether the residuals have a certain correlation.

**Information Criteria for ARMA (p, q) Model Selection**

The information criteria for ARMA (p, q) model helps to determine the order of the time series based on Residual Sum of Square (RSS) [17]. The model which gives minimum RSS for value of p and q is best suitable model. When more number of lags are added the one which gives less RSS is best model. But when numbers of lags are added the RSS goes down. Hence, penalty term is considered while finding suitable model which is given as:

$$\min(f(RSS) + \text{penalty term}) \tag{6}$$

Thus adding extra terms will also add penalty while finding the p and q. The standard methods which are used for computing information criteria are:

- Akaike’s Information Criteria (AIC).
- Schwarz’s Bayesian Information Criteria (SBIC).
- Hannon Quinn Criteria (HQIC).

Where,

$$AIC = \ln(\hat{\sigma}^2) + \frac{2k}{T},$$

$$SBIC = \ln(\hat{\sigma}^2) + \frac{k}{T} \ln(T),$$

$$HQIC = \ln(\hat{\sigma}^2) + \frac{2k}{T} \ln(\ln(T))$$

Where,

$$\hat{\sigma}^2 = \frac{RSS}{T}, k = p + q + 1, \text{ and } T = \text{Sample size.}$$

The model which gives the less value of AIC is best model.

### 3.6 ARIMA Model

The AR, MA, ARMA models cannot handle non-stationarity that is the series that has trend. In stationary series, the model reverts around its mean value. To make non-stationary series into stationary, we difference it i.e. taking the difference of the series with its own lag. Differencing will give the series leads to a new set of values with constant mean and constant variance. Example,

As shown in table above  $\nabla \mathbf{y} = \mathbf{y}_t - \mathbf{y}_{t-1}$  is first order difference, and  $\nabla^2 \mathbf{y}$  is a second order difference, which gives constant mean and 0 or constant variance. Thus differencing ARMA model makes it integrated, which gives its name ARIMA. It can be represented as ARIMA (p, q, r) (Table 1),

**Table 1.** First order differencing and second order differencing.

$\mathbf{y}_t$	$\nabla \mathbf{y}$	$\nabla^2 \mathbf{y}$
25		
16	9	2
9	7	2
4	5	

Where,

p = number of AR terms, q = order of differencing and r = number of MA terms.

Also, ARIMA can be represented AR, MA and ARMA model ARIMA(1,0,0) gives AR(1) model, ARIMA(0,0,1) gives MA(1) model and ARIMA(1,0,1) gives ARMA(1,1) model where only differencing is 0.

The main query which arises is to determine the order of differencing. It can be given by observing the following characteristics:

1. Positive auto-correlation at higher lags will need higher order differencing.
2. Zero or negative auto-correlation for first order itself then no differencing is needed.

3. The optimal order of differencing is the one whose prediction gives less value for RMSE (root mean square error) or MAE (Mean Absolute error).

### 4 Observations

In time series modelling, depending upon characteristics of periodic patterns, predictions are done. Different trends will lead to different periodic patterns which can be classified as hourly, daily, weekly, monthly data. The goal is to find the appropriate statistical relationship of the given series with its own past values. We build the model in training data, and test it on new dataset (holdout) to check whether predicted values match the actual values. Thus, accuracy is computed by computing forecast errors which can be represented as:

$$\text{Forecast error} = \text{Actual value} - \text{Predicted value.}$$

Accuracy can be computed using mean square error (MSE). The model which gives less MSE for Holdout sample is good model for prediction.

The brief behaviours of white noise AR, MA, ARMA and ARIMA for time-series data are shown in Table 2 below:

**Table 2.** Behaviour of time series models

Model	Characteristic	PACF correlogram	ACF correlogram	Data Characteristic
White noise	Cannot be used for Time series modelling	No spikes	No spikes	Random in nature
AR(p)	1. $y_t$ depends on its own past values 2. P is computed using PACF function	Spikes till $p^{\text{th}}$ lag then cuts off to zero	Spikes then decays to zero	Data should be stationary in nature
MA(q)	1. $y_t$ depends on error term which follows a white noise process 2. q is computed using ACF function	Spikes then decays to zero	Spikes till $p^{\text{th}}$ lag then cuts off to zero	Data should be stationary in nature
ARMA (p, q)	1. ARMA = AR+MA 2. Value of p and q are determined using AIC criteria	Spikes then decays to zero	Spikes then decays to zero	Data should be stationary in nature
ARIMA (p, d, q)	1. Data is made stationary by differencing it 2. Box-Jenkins approach is used to determine model	Spikes then decays to zero	Spikes then decays to zero	Data should be non-stationary in nature

### 5 Conclusion

Thus, while doing time series analysis, the data should not be a white noise process or random in nature. Thus only stationary or non-stationary will lead to identification of proper periodic pattern which can be used to determine appropriate forecasting model. The accuracy of forecasting is determined using Root mean square error or Mean Absolute error.

As a part of future work, major focus will be put upon understanding multivariate time series analysis. Also, forecasting over white-noise process can be done by using unsupervised algorithms like Neural Networks. But, it is very complex in nature. Hence, further part of research will be emphasized upon using semi-supervised algorithm like Support Vector Regression.

**Acknowledgements.** I would like to thank my Guide Dr. B.F. Momin for their continuous help and encouragement. Also thanks to TEQIP-II for financial support in presenting this paper. Also I would like to thank Department of CSE, Walchand College of Engineering, for providing useful resources.

## References

1. Barnaghi, P., et al.: Semantics for the Internet of Things: early progress and back to the future. *Int. J. Semant. Web Inf. Syst. (IJSWIS)* **8**(1), 1–21 (2012)
2. Zhang, C., et al.: A multimodal data mining framework for revealing common sources of spam images. *J. Multimed.* **4**(5), 313–320 (2009)
3. Zhang, K.: Research on key technologies of college computer room billing management system. In: *Key Engineering Materials*, vol. 474. Trans Tech Publications (2011)
4. Zhang, M.: Application of data mining technology in digital library. *JCP* **6**(4), 761–768 (2011)
5. Shen, C.-W., et al.: Data mining the data processing technologies for inventory management. *J. Comput.* **6**(4), 784–791 (2011)
6. Danping, Z., Deng, J.: The data mining of the human resources data warehouse in university based on association rule. *J. Comput.* **6**(1), 139–146 (2011)
7. Jiang, J., Long, B.G., Wei, M., Feng, K.F.: Block-based parallel intra prediction scheme for HEVC. *J. Multimed.* **7**(4), 289–294 (2012)
8. Yang, S.-Y., et al.: Incremental mining of closed sequential patterns in multiple data streams. *JNW* **6**(5), 728–735 (2011)
9. Fu, Z., Juanjuan, B., Qiang, W.: A novel dynamic bandwidth allocation algorithm with correction-based the multiple traffic prediction in EPON. *JNW* **7**(10), 1554–1560 (2012)
10. Zhi, Q., Lin, Z.-W., Ma, Y.: Research of Hadoop-based data flow management system. *J. China Univ. Posts Telecommun.* **18**, 164–168 (2011)
11. Cui, J., Taoshen, L., Hongxing, L.: Design and development of the mass data storage platform based on Hadoop. *J. Comput. Res. Dev.* **49**(12), 12–18 (2012)
12. Krishna, G.V.: An integrated approach for weather forecasting based on data mining and forecasting analysis. *Int. J. Comput. Appl.* **120**(11), 0975–8887 (2015)
13. Ljung, G.M., Box, G.E.P.: On a measure of lack of fit in time series models. *Biometrika* **65**(2), 297–303 (1978)
14. Box, G.E.P., Jenkins, G.M.: Some comments on a paper by Chatfield and Prothero and on a review by Kendall. *J. R. Stat. Soc. Ser. A (Gen.)* **136**(3), 337–352 (1973)
15. Noreen, E.: An empirical comparison of probit and OLS regression hypothesis tests. *J. Account. Res.* 119–133 (1988)
16. Pan, J.-X., Fang, K.-T. Maximum likelihood estimation. In: *Growth Curve Models and Statistical Diagnostics*, pp. 77–158. Springer, New York (2002)
17. Johansen, S.: The Welch-James approximation to the distribution of the residual sum of squares in a weighted linear regression. *Biometrika* **67**(1), 85–92 (1980)
18. Hamilton, J.D.: *Time Series Analysis*, vol. 2. Princeton University Press, Princeton (1994)