Forecasting BIST 100 Index with Artificial Neural Networks and Regression Analysis

Yüksel Akay Ünvan[a], Cansu Ergenç[b]

a Ankara Yıldırım Beyazıt University, Turkey, akay.unvan@gmail.com, https://orcid.org/ 0000-0002-0983-1455
b Ankara Yıldırım Beyazıt University, Turkey, cansuergenc7@gmail.com, https://orcid.org/ 0000-0002-4722-0911

Forecasting BIST 100 Index with Artificial Neural Networks and Regression Analysis

Abstract

Making reliable forecasts is very important for financial analysis. For this reason, financial analysts make analyzes using different models. Financial analysts try to make the most accurate estimation in these analyzes. The Artificial Neural Network model is a widely used method in the field of finance. In this study, BIST 100 index was estimated using Artificial Neural Networks and Regression model. By using the closing prices of the BIST 100 index between 2010 and 2020, the closing values of the BIST 100 index for 2021 were estimated. Moreover, the Regression model and Artificial Neural Network model predictions were obtained. The mean square error of the neural networks and Regression model was also found. Finally, according to the result of the mean of error squares, the performance of the models was compared and seen that the Artificial Neural Network model was better.

1. Introduction

The concept of forecasting, which is a large part of the financial and economic world, is crucial to investors and governments. Financial and economic forecasts have evolved for centuries. Economic forecasting uses historical data published by countries or geographic regions in previous economic reports. (Lenel et al., 2020)

Economic forecasting is the process used to try to predict or anticipate future economic conditions by using various economic variables and indicators. Indeed, the methods for selecting a forecasting model, estimating its parameters, communicating the resulting forecasts, and evaluating their accuracy have improved in many ways over the past 20 years. On the one hand, economic forecasts are far from perfect. It is very important to understand the limitations of these estimates in order to be interpreted correctly (Elliott & Timmermann, 2008).

One of the key points in making economic forecasting so important is that it helps in making better policy decisions. For example, if the economy is in a recession, the government may implement expansionary fiscal policy (increased spending financed by borrowing) to protect economic demands and prevent the economy from recession (Wieland & Wolters, 2013).

Economic forecasts are critical in setting monetary policy/fiscal policy. In situations where the economy needs to be fixed and made better, inflation may rise and there may be situations where banks' interest rates need to rise. If the economy is likely to contract further, the Bank may need to undertake further quantitative easing.

Companies also use forecasts to develop business strategies. Financial decisions are determined by economic conditions and what the future will look like, even in uncertain situations. Data from the past is collected and analyzed so patterns can be found. Today, big data and artificial intelligence have changed the methods of business forecasting. Business forecasting is vital for companies as it allows them to plan production, financing, and other strategies.

There is a long history of research in finance and economic modeling. Time Series Analysis and Regression are the most commonly used traditional approaches in financial and economic modeling. However, in some cases, these models are not sufficient. Recent evidence shows that financial markets are nonlinear; however, these linear methods still provide good opportunities to describe nonlinear systems found in Time series analysis of the financial market (Maciel & Ballini, 2008). For this reason, financial analysts started to look for other models. In recent years, Artificial Neural Networks have been widely used as a powerful modeling technique in financial and economic forecasting.

Various variables such as exchange rates, gold prices, daily US dollar returns, bonds and oil prices have been used in many different studies to predict the returns of investment instruments through Artificial Neural Network models and to create models for these situations (Zhang & Berardi, 2001; KangaraniFarahani & Mehralian, 2013; Khodayari et al., 2020; Diaz and Nguyen, 2021; Mir et al., 2021; Le et al. 2021).

There are many studies that make financial forecasts and obtain models by combining Regression models with Artificial Neural Network models (Cao et al., 2011; Pradeepkumar & Ravi, 2017; Siami-Namini & Namin, 2018; Cao & Li, 2019). When other studies are examined, it has combined Time series models and Artificial Neural Network models to analyze financial data and make predictions by creating financial models (Zhang & Berardi, 2001; Lee et al., 2008; Patra et al., 2017; Altan & Karasu, 2019; Benrhmach, 2020) Some studies have compared the results of the different methods, such as Regression and Time series models, with the neural network by using the same data. (Carvalhal & Ribeiro, 2008; Kyung et al., 2008).

In the literature, there are studies that predict the index BIST 100 using the Artificial Neural Network model and compare it with other models (Kılıc et al., 2014; Ozbey et al., 2020; Molla et al., 2021; Karakul, 2020).

In this study, the BIST 100 index was estimated by using Artificial Neural Networks and a Regression model. Using the closing prices of the BIST 100 index between 2010 and 2020, the closing values of the BIST 100 index for 2021 were estimated. Also, the predictions of the Regression model and the Artificial Neural Network model were obtained. The mean square errors of the neural networks and the Regression model were also obtained. In the analysis, estimates were first made with the Regression model, and then the Artificial Neural Network model was applied.

2. Material and Method

It is known that the Regression method was first proposed by the English statistician Francis Galton in the 19th century. He proposed this method for the purpose of a biological study (Galton, 1877; as cited in Stanton, 2001). Udny Yule and Karl Pearson applied and developed this method to broader general statistical fields (Yule, 1897; as cited in Hepple, 2001). In this period, the dependent and independent variables are assumed to have normal distributions. This assumption was extended with Fisher's publications in 1922 and 1925 to apply only to cases in which the conditional distribution of the dependent variable is normal (Fisher, 1922; as cited in Aldrich, 2005).

Regression analysis basically consists of statistical operations to predict the relationships between a dependent variable and one or more independent variables. In simple linear Regression analysis, one dependent variable and one independent variable are assumed to be linearly related. In multiple Regression

analysis, it is assumed that there is a linear relationship between more than two variables. (Schroeder et al., 2016).

The general Linear Regression model can be stated by the equation below (Montgomery, 2021).

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i \tag{1}$$

where,

y_i=dependent variable

x_i=explanatory variables

β_0= constant term

β_k=slope coefficients for each explanatory variable

ε_i=the model's error term

The multiple regression model is based on the following assumptions (Seber and Lee, 2012):

- There is a linear relationship between the dependent variables and the independent variables.
- The observations are independently and randomly selected from the population.
- $\varepsilon_i \sim N(0, \sigma^2)$

Sometimes data can have a non-linear relationship. One way to try to explain such a relationship is to use a polynomial Regression model (Aiken, 1991);

$$y_i = \beta_0 + \beta_1 X + \beta_2 X^2 + \ldots + \beta_h X^h + \varepsilon_i \tag{2}$$

Here h is called the degree of the polynomial. For lower degree, the relation has a specific name (i.e. h=2 means quadratic, h = 3 means cubic, h = 4 means quartic, etc.). Although this model enables a nonlinear relationship between Y and X, polynomial regression is still considered linear regression as the regression coefficients $\beta_1, \beta_2 \ldots, \beta_h$ are linear.

While applying the Regression model in time-dependent series, various transformations are applied. Time is considered as an independent variable in Time series regression analysis. Models such as Cubic regression model, Exponential regression model, Quadratic Regression model and Logistic Regression model can be used to transform time series. The model with the smallest mean squared error is selected and Cubic Regression was used in this study.

In 1997, Ramon Lawrence tested the performance of Artificial Neural Networks in price prediction in his article "Using Neural Networks to Forecast Stock Market Prices". As a result of the study, he said that price prediction with Artificial Neural Networks is not perfect, but outperforms regression and other mathematical models. In 2003, Chen et al. estimated stock returns using Taiwan Stock Exchange data for January 1982 and August 1992. As a result of the study, it is emphasized that the returns can be predicted with Artificial Neural Networks and investment decisions can be made with this analysis.

In Artificial Neural Networks, there are 3 layers. These are the input layer, the output layer and the hidden layers. The input layer is the first layer in the model. In statistics, independent variables can be used as the input layer. The output layer is the last layer in the model. In statistics, dependent variables can be used as the output layer. Hidden layers transfer the information from the input layer to the output layer.

Figure 1: Artifical neural networks layers (Maind et al., 2014)
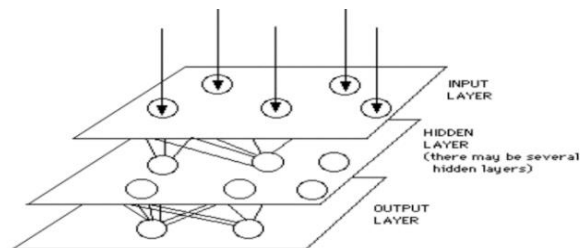
The following are the three most commonly used types of neural networks in artificial intelligence. (Patel & Gaurav, 2020):

Feedforward Neural Networks: Feedforward neural networks are the first type of Artificial Neural Networks created and can be considered the most used today. These neural networks are called feedforward neural networks because the flow of information through the network is unidirectional and does not loop.

Recurrent Neural Networks: Recurrent neural networks (RNN), as the name suggests, involve the repetition of operations in the form of loops. These are much more complicated than feedforward networks and can perform more complex tasks than simple image recognition.

Convolutional Neural Networks: Convolutional neural networks have been associated almost exclusively with computer vision applications since their inception. This is because their architecture is specifically suited for performing complex visual analysis.

3. Research Findings and Discussion

In the analysis, firstly the Regression model and then the Artificial Neural Network model estimation were made. BIST100 Index was used as the dependent variable

and gold price, USD rate, inflation rate, exchange rate and treasury bond were used as the independent variables. At the beginning of the study, the descriptive statistics were calculated. Then, it was tested whether the variables have normal distribution or not. Before performing the Regression analysis, it was tested whether the variables have normal distribution or not. Finally, Regression and Artificial Neural Network models were developed. The descriptive statistics used in the study are given in Table 1 and the normality test is given in Table 2.

Table 1: Descriptive statistics

	Mean	Std. Deviation	Minimum	Maximum	Median	Skewness	Kurtosis
BIST100	824,55	186,55	595,67	1133,56	773,15	0,355	-1,303

Table 2: Tests of Normality

Tests of Normality						
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
BIST100	,196	11	,200[*]	,909	11	,237

Testing the normality of variables is an important early step in almost every multivariate test (Tabachnick and Fidell, 2007). Skewness and kurtosis as shown in Table 1 are two important components used to measure the univariate normality of the data. According to Onwuegbuzie and Daniel (2002), the values of skewness $<\pm 2$ and kurtosis $<\pm 2$ are considered as acceptable. Since the measures of kurtosis and skewness for all value are within the range given by Onwuegbuzie and Daniel (2002), thus it is assumed that the distribution of data does not depart from normality. When the results given in Table 2 are examined, we can say that the data are normally distributed. "Curve Estimation" was applied to decide which method to use in the Regression analysis. The model with the smallest mean square error is selected. According to the results of the calculations, it was decided to use Cubic Regression. The results obtained by using Cubic Regression for the year 2021 are given in Table 3.

Table 3: The results for the year 2021 obtained using the Cubic Regression

DATE	BIST 100	ESTIMATION WITH CUBIC REGRESSION	ABSOLUTE PERCENTAGE ERROR (APE)
2020-01	1191,4	1143,1	4,054
2020-02	1059,9	1152	8,689
2020-03	896,4	1160,9	29,507
2020-04	1011,1	1170	15,716
2020-05	1055,2	1179,2	11,751
2020-06	1165,2	1188,5	2
2020-07	1126,9	1197,9	6,3
2020-08	1078,6	1207,4	11,941
2020-09	1145,2	1217	6,27
2020-10	1112,3	1226,7	10,285
2020-11	1283,5	1236,5	3,662
2020-12	1476,7	1246,5	15,589
2021-01	1473,4	1256,5	14,721
2021-02	1471,3	1266,7	13,906
2021-03	1391,7	1277	8,242
2021-04	1397,8	1287,4	7,898
2021-05	1420,4	1297,9	8,624

Figure 1: Forecast chart with BIST 100 and Cubic Regression model
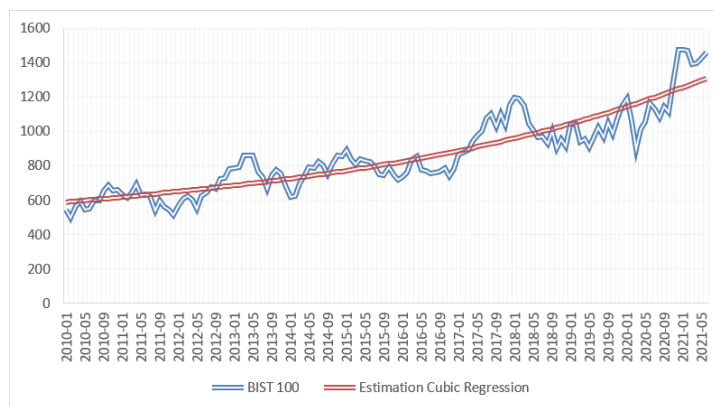

When Table 3 and Figure 2 are examined, it is seen that there is a difference between BIST 100 index values and forecast values. It was found that the estimated values for the year 2021 data were lower than the index values of BIST 100. When the years 2020-2021 were examined, there was an increase in the results of both models until February. However, despite the decrease in BIST 100 index values after February, an increase in forecast values was observed. In May 2021, the BIST 100 index values also began to increase. After testing the results of the Cubic Regression model, the Artificial Neural Network model was constructed. The Artificial Neural Network model was obtained using MATLAB software. 20% of the data was used as test data and 80% as training data. These data were selected randomly. The mean

absolute percentage error (MAPE) is calculated for the Artificial Neural Network and the Cubic Regression. MAPE is the mean or average of the absolute percentage errors of forecasts (Swamidass, 2000). The MAPE of the Artificial Neural Network model is found to be 3,737. The MAPE of the Cubic Regression model was found to be 9,953. Based on these results, it was found that the Artificial Neural Network model has better performance. In addition, when the graphics are examined, it is seen that the Artificial Neural Network model fits better with the BIST100 data set. The results for 2021 obtained by using the Artificial Neural Network model are given in Table 4.

Table 4: The results for the year 2021 obtained using the Artificial Neural Network

DATE	BIST 100	ESTIMATION WITH ARTIFICIAL NEURAL NETWORK	ABSOLUTE PERCENTAGE ERROR (APE)
2020-01	1191,4	1112,2	6,648
2020-02	1059,9	1076,6	1,576
2020-03	896,4	1026,6	14,525
2020-04	1011,1	988,6	2,225
2020-05	1055,2	1055,9	0,066
2020-06	1165,2	1125,4	3,416
2020-07	1126,9	1149,5	2,006
2020-08	1078,6	1151,8	6,787
2020-09	1145,2	1149,2	0,349
2020-10	1112,3	1201,5	8,019
2020-11	1283,5	1278,3	0,405
2020-12	1476,7	1345,8	8,864
2021-01	1473,4	1416,3	3,875
2021-02	1471,3	1435,4	2,440
2021-03	1391,7	1433,9	3,032
2021-04	1397,8	1431,5	2,411
2021-05	1420,4	1429,3	0,627

Figure 2: Forecast chart with BIST 100 and Artificial Neural Network model
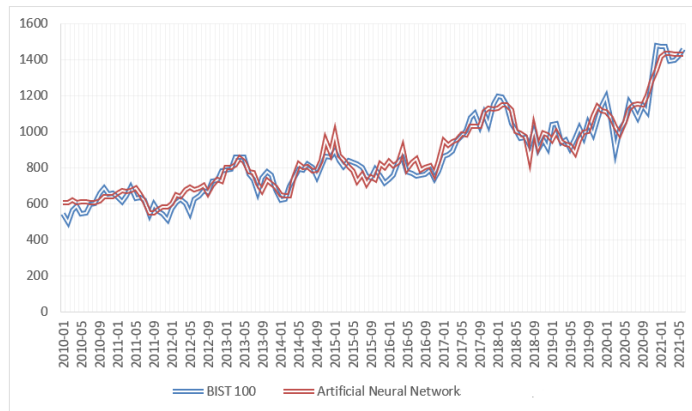

When Table 4 and Figure 3 were examined for the dataset obtained with the Artificial Neural Network model, it could be seen that the index BIST 100 and the prediction results were very close. Similar to the Regression model, an increase was observed in the Artificial Neural Network model until the end of the period. However, in the real data of BIST 100 index, a decrease was observed. When the Artificial Neural Network model and the Regression model were compared, it was seen that the predictions obtained by the Artificial Neural Network were closer to the real values. The prediction values obtained by the Artificial Neural Network model and the Regression model are given are given in Table 5.

Table 5: The results for the year 2021 obtained using the Artificial Neural Network

DATE	BIST 100	ESTIMATION WITH CUBIC REGRESSION	ESTIMATION WITH ARTIFICIAL NEURAL NETWORK
2020-01	1191,4	1143,1	1112,2
2020-02	1059,9	1152,0	1076,6
2020-03	896,4	1160,9	1026,6
2020-04	1011,1	1170,0	988,6
2020-05	1055,2	1179,2	1055,9
2020-06	1165,2	1188,5	1125,4
2020-07	1126,9	1197,9	1149,5
2020-08	1078,6	1207,4	1151,8
2020-09	1145,2	1217,0	1149,2
2020-10	1112,3	1226,7	1201,5
2020-11	1283,5	1236,5	1278,3
2020-12	1476,7	1246,5	1345,8
2021-01	1473,4	1256,5	1416,3
2021-02	1471,3	1266,7	1435,4
2021-03	1391,7	1277,0	1433,9
2021-04	1397,8	1287,4	1431,5
2021-05	1420,4	1297,9	1429,3

Figure 3: Forecast chart with BIST 100 and Artificial Neural Network model and Cubic Regression model
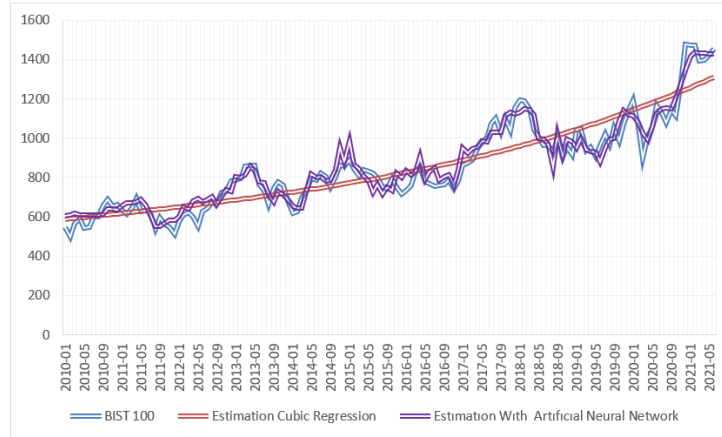
4. Conclusions

In this study, Cubic Regression and Artificial Neural Network models were developed by using BIST 100 index data between 2010-2020. First, the mean square error values of the models were applied to decide which Regression model to use. According to these results, it was decided to select the Cubic Regression model. Then, the results of the Cubic Regression model and the actual values of the index BIST 100 were compared. There was a difference between the estimated values and the actual values. Then the Artificial Neural Network model was created.

As a result of the study, it can be seen that the Artificial Neural Network model gives better results and performs better than the Cubic Regression model. On examining the tables and graphs, a decrease in BIST 100 values is observed in March 2020. The reason for this decline is believed to be the Covid 19 epidemic that started in Wuhan, China in late 2019 and spread around the world. This is a variable that is not added to the model. Therefore, both Artificial Neural Networks and Regression models were not affected by this situation and did not decrease.

A similar decrease was observed in February 2021. The reason for this is probably the quarantine process that occurred due to the Covid19 outbreak. Since unexpected situations are not taken into account in the model, there are discrepancies between the estimates and the actual values. However, it is shown that the predictions obtained with the Artificial Neural Network model give better results than the Cubic Regression model. If it is assumed that there will be no factors that are not added to the model in the future; it can be said that the BIST 100 index can be estimated more accurately with the Artificial Neural Network model.

References

Aiken, L. S., & West, S. G. Reno., RR (1991). Multiple regression: Testing and interpreting interactions.

Aldrich, J. (2005). Fisher and regression. Statistical Science, 401-417.

Altan, A., & Karasu, S. (2019). The effect of kernel values in support vector machine to forecasting performance of financial time series. The Journal of Cognitive Systems, 4(1), 17-21.

Benrhmach, G., Namir, K., Namir, A., & Bouyaghroumni, J. (2020). Nonlinear autoregressive neural network and extended Kalman filters for prediction of financial time series. Journal of Applied Mathematics, 2020.

Cao, J., Li, Z., & Li, J. (2019). Financial time series forecasting model based on CEEMDAN and LSTM. Physica A: Statistical Mechanics and its Applications, 519, 127-139.

Cao, Q., Parry, M. E., & Leggio, K. B. (2011). The three-factor model and artificial neural networks: predicting stock price movement in China. Annals of Operations Research, 185(1), 25-44.

Carvalhal, A., & Ribeiro, T. (2008). Do artificial neural networks provide better forecasts? Evidence from Latin American stock indexes. Latin American Business Review, 8(3), 92-110.

Diaz, J. F., & Nguyen, T. T. (2021). Application of grey relational analysis and artificial neural networks on corporate social responsibility (CSR) indices. Journal of Sustainable Finance & Investment, 1-19.

Elliott, G., & Timmermann, A. (2008). Economic forecasting. Journal of Economic Literature, 46(1), 3-56.

Hepple, L. W. (2001). Multiple regression and spatial policy analysis: George Udny Yule and the origins of statistical social science. Environment and Planning D: Society and Space, 19(4), 385-407.

KangaraniFarahani, M., & Mehralian, S. (2013, August). Comparison between artificial neural network and neuro-fuzzy for gold price prediction. In 2013 13th Iranian Conference on Fuzzy Systems (IFSC) (pp. 1-5). IEEE.

Kayahan Karakul, A. (2020). Yapay Sinir Ağları İle Borsa Endeksi Tahmini-Forecastıng Stock Market Index Wıth Artıfıcıal Neural Networks.

Khodayari, M. A., Yaghobnezhad, A., & Khalili Eraghi, K. E. (2020). A Neural-Network Approach to the Modeling of the Impact of Market Volatility on Investment. Advances in Mathematical Finance and Applications, 5(4), 569-581.

Lawrence, R. (1997). Using neural networks to forecast stock market prices. University of Manitoba, 333, 2006-2013.

Ünvan, A. Y. & Ergenç, C. (2022). Forecasting BIST 100 Index with Artificial Neural Networks and Regression Analysis. BILTURK, The Journal of Economics and Related Studies, 4(1), 20-32. doi: 10.47103/bilturk.1039669.

Le, T. L., Abakah, E. J. A., & Tiwari, A. K. (2021). Time and frequency domain connectedness and spill-over among fintech, green bonds and cryptocurrencies in the age of the fourth industrial revolution. Technological Forecasting and Social Change, 162, 120382.

Lee, K. J., Chi, A. Y., Yoo, S., & Jongdae Jin, J. (2008). Forecasting Korean Stock Price Index (Kospı) Using Back Propagation Neural Network Model, Bayesian Chıao's Model, And Sarima Model. Academy Of Information & Management Sciences Journal, 11(2).

Lenel, L., Köster, R., & Fritsche, U. (2020). Introduction (Futures Past. Economic Forecasting in the 20th and 21st Century). Futures Past. Economic Forecasting in the 20th and 21st Century.

Maciel, L. S., & Ballini, R. (2008). Design a neural network for time series financial forecasting: Accuracy and robustness analysis. Anales do 9º Encontro Brasileiro de Finanças, Sao Pablo, Brazil.

Maind, S. B., & Wankar, P. (2014). Research paper on basic of artificial neural network. International Journal on Recent and Innovation Trends in Computing and Communication, 2(1), 96-100.

Mir, H., Zaraatgari, R., & Sotoudeh, R. (2021). Comparison of Risk Factors for Investing in Tehran Stock Exchange Using Smart Neural Network (Forecasting Tehran Stock Exchange with Neural Networks). Agricultural Marketing and Commercialization Journal, 5(1), 43-57.

Molla, B., Cagil, G., & Uyaroglu, Y. (2021). Chaotic analysis of BIST 100 return time series and short-term predictability with ANFIS.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introduction to linear regression analysis. John Wiley & Sons.

Onwuegbuzie, A. J., & Daniel, L. G. (1999). Uses and misuses of the correlation coefficient.

Ozbey, F., & Paksoy, S. (2020). GARCH Ailesi Modelleri ve ANN Entegrasyonu ile BİST 100 Endeks Getirisinin Volatilite Tahmini 1. Business and Economics Research Journal, 11(2), 385-396.

Patra, A., Das, S., Mishra, S. N., & Senapati, M. R. (2017). An adaptive local linear optimized radial basis functional neural network model for financial time series prediction. Neural Computing and Applications, 28(1), 101-110.

Pradeepkumar, D., & Ravi, V. (2017). Forecasting financial time series volatility using particle swarm optimization trained quantile regression neural network. Applied Soft Computing, 58, 35-52.

Schroeder, L. D., Sjoquist, D. L., & Stephan, P. E. (2016). Understanding regression analysis: An introductory guide (Vol. 57). Sage Publications.

Seber, G. A., & Lee, A. J. (2012). Linear regression analysis (Vol. 329). John Wiley & Sons.

Siami-Namini, S., & Namin, A. S. (2018). Forecasting economics and financial time series: ARIMA vs. LSTM. arXiv preprint arXiv:1803.06386.

Stanton, J. M. (2001). Galton, Pearson, and the peas: A brief history of linear regression for statistics instructors. Journal of Statistics Education, 9(3).

Swamidass, P. M. (2000). MAPE (mean absolute percentage error) Mean Absolute Percentage Error (MAPE). Encyclopedia of production and manufacturing management, 462-462.

Tabachnick, B. G., & Fidel, L. S. (2007). Using Multivariate Statistics. USA: Person Education Inc.

Zhang, G. P., & Berardi, V. L. (2001). Time series forecasting with neural network ensembles: an application for exchange rate prediction. Journal of the operational research society, 52(6), 652-664.