Sales Forecast for Amazon Sales with Time Series Modeling

Balpreet Singh
Dept. Computer Science and Engineering
Dr. B R Ambedkar National Institute of
Technology, Jalandhar
Punjab, India
singh.balpreet@hotmail.com

Pawan Kumar
Dept. Computer Science and Engineering
Dr. B R Ambedkar National Institute of
Technology, Jalandhar
Punjab, India
pawankt10@outlook.com

Dr. Nonita Sharma
Dept. Computer Science and Engineering
Dr. B R Ambedkar National Institute of
Technology, Jalandhar
Punjab, India
nsnonita@gmail.com

Dr. K P Sharma
Dept. Computer Science and Engineering
Dr. B R Ambedkar National Institute of
Technology, Jalandhar
Punjab, India
sharmakp@nitj.ac.in

***Abstract*—Accurate sales prediction plays an important role in reducing costs and improving customer service levels, especially for B2C(Business to consumer) e-commerce. This paper attempts to forecast future sales at Amazon.com, Inc. based on historical sales data. Firstly, it proposes three possible forecasting approaches according to the historical data pattern, that is Holt-Winters exponential smoothing, neural network auto regression model and ARIMA(Autoregressive integrated moving average). Secondly, it specifies certain accuracy measures using which well determine the suitability of the forecast methods on the available sales data. Finally the three methods will be implemented to forecast Amazons quarterly sales in 2019. The results can help Amazon well manage its future operations.**

***Index Terms*—Forecasting, ARIMA, Holt-Winters Exponential Smoothing, Neural Network Auto Regression Model**

I. Introduction

When it comes to retail companies, it is important to record and maintain sales-related measures like number of transactions, page hits, and revenue generated with respect to time & then use this data to accurately predict values for the above mentioned parameters in order to efficiently plan for future scenarios [1]. Our work is based on the revenue data of Amazon.com, Inc., a US based online retail Company that focuses on e-commerce, cloud computing, digital streaming and artificial intelligence. As online sales are rising at an immense rate, consistent projection of sales allows the company to efficiently prepare for unexpected situations like web traffic spikes or sudden disturbances to product stock.

For an online retailer, the special online shopping festival, such as Black Friday and Cyber Monday in USA, and NOV 11 in China, makes up a substantial percentage of sales. This is especially true for Amazon.com. In order to provide better service during the holiday shopping season, Amazon hired a lot of temporary staff and extra permanent employees. During this season, it is a major challenge for Amazon to allocate resources such as employees, the third part logistics and items, to attain higher customer satisfaction. Therefore, it is useful to forecast future quarterly net sales [2]–[5], which can help Amazon prepare for future Black Fridays and Cyber Mondays.

The company earns around 5% of its total revenue generated for the whole year during these special online shopping festivals. As a result of this, it is considered a major parameter to predict overall sales. Predicting sales on such days is particularly exigent, as there is generally huge spike (i.e. anomalies) relative to normal working days. For example, total revenue generated on Black Friday is generally more than 10 times of the median sales of the year.

In this paper, we will try to use different forecasting methodologies to forecast Amazons future quarterly net sales, based on its historic quarterly data. In particular, we use ETS (Error Trend & seasonality) that applies the Holt- Winters exponential smoothing model multiple times with varying parameters, ARIMA (Autoregressive integrated moving average) [6], [7] and neural network autoregression model [7]–[10]. First, we will use three different methods to forecast the quarterly net sales in 2019 [7]. Then we will test and justify different approaches by comparing forecasting data with the actual net sales in 2019, and find the best approach suitable for forecasting future quarterly net sales at Amazon. In order to measure the accuracy of predictions made by a particular model, well use measures such as MAPE(Mean absolute percentage error) [11], RMSE(Root mean square error) [12] and many more.

In the following sections we briefly explain the proposed models, a general architecture explaining various stages in the forecasting process and then provide the comparative performance of these models.

II. Methodology

In this section, we provided details regarding the proposed models that are used to generate quarterly sales forecasts along with a general process flow on how well be applying these models and what are the various transformations that the data will undergo.

A. Process flow

Firstly, Amazon's quarterly sales data is gathered and imported into R studio's working environment. Then the data is transformed into time series format and any inconsistencies pertaining to it are dealt with like missing values, noise etc [3].

After data transformation, in case of ARIMA [6], [13], determine if the data is stationary and if not, make it stationary so that it is suitable for applying the ARIMA Model. For other models under consideration, no further pre-processing or the need of stationarity is required. Once all the models are applied, determine the accuracy of the predictions made by different models and compare as to which fits the data better.

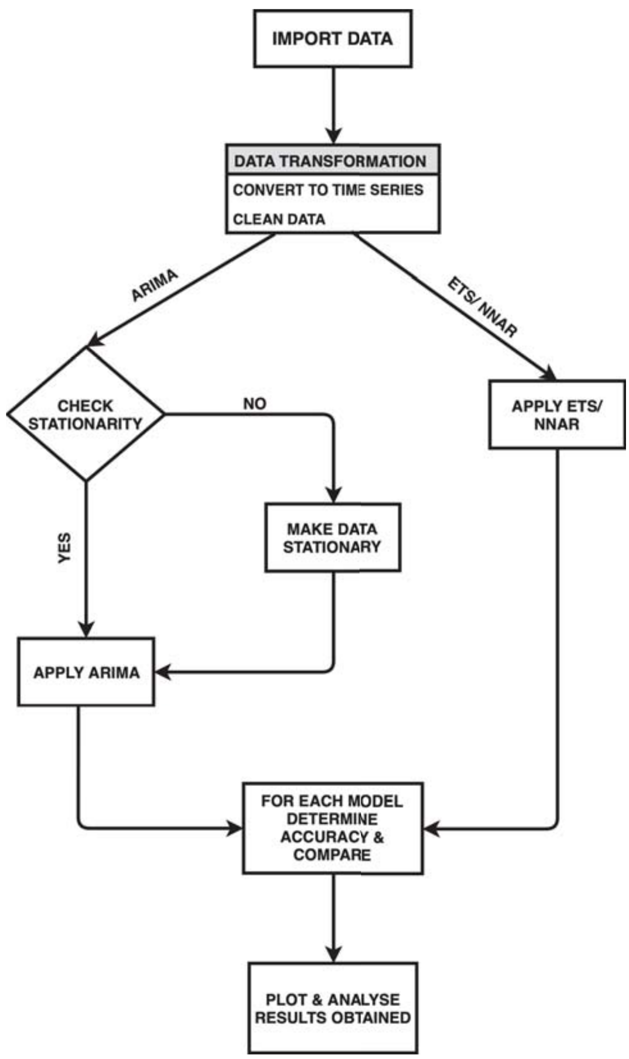

Fig. 1. Process flow diagram of forecasting.

B. Holt-Winters exponential smoothing

Exponential smoothing forecast methods base their predictions on weighted sum of past observations. Here, instead of assigning equal weightage to each past value, we assign weights in an exponentially decreasing order. It is used to perform forecasting on uni variate time series data [14].

Holt-Winters exponential smoothing is an extension of Exponential Smoothing that in addition to accommodating effects of trend, also incorporates seasonality present in the time series data. It includes three smoothing parameters:-

- Alpha - The value of alpha ranges between 0 & 1. It specifies the proportion of weightage assigned to immediate past value and (1 - alpha) is the weightage assigned to the rest of the historical data. Large values indicate that we depend more on recent past values whereas small values indicate that more attention is given to historical data.
- Beta - It is used to dampen the effects of trend.
- Gamma - It is used to dampen the effects of seasonality.

If trend and seasonality change linearly, we can model it in an additive manner or else if the trend and seasonality present in data change exponentially, they can be modelled in a multiplicative manner.

C. ARIMA

ARIMA stands for AutoRegressive Integrated Moving Average [7], [15]. It incorporates the principles of the simpler AutoRegressive method and Moving Average method. To this it adds the concept of integration.

The name of the model itself captures the main concept on which the model is based. These are:-

- AR: Autoregression. It predicts future values based on p number of past values often referred to as lags.
- I: Integrated. ARIMA works on stationary data i.e the mean, variance and covariance of the data needs to be time invariant. The concept of integration refers to differencing of data points i.e. subtracting a data point with its immediate predecessor in order to make the time series data stationary.
- MA: Moving Average. It is similar to AR with the notable difference that instead of past values/ lags, we depend on the associated error terms for these past values. We base our prediction on q number of error terms.

The standard notation used to specify an ARIMA model is of the form: ARIMA(p, d, q). The parameters of the ARIMA model are defined as follows:-

- p: The number of past values/ lag included to make the prediction.
- d: The number of times that the time series data needs to be differenced in order to make it stationary.
- q: The number of past error terms included in our model, also referred to as the size of the moving average bracket.

D. Neural network model

A neural network is an attempt to simulate/ model how a human mind works [8], [9]. It can be visualised as a graph where each node is referred to as a neuron and edges connect a pair of neurons. An incoming edge to a neuron represents input along with certain weight assigned to that input and an outgoing edge represents the output generated as a result of applying a non linear function on the weighted sum of all its inputs. Given a suitable number of nonlinear processing units, a neural network can approximate any complex target function in time by learning from its experience and that too with satisfactory accuracy.

A Feedforward Neural Network is a widely adopted model used for time series forecasting applications. Fig. 2 shows a typical three-layered feedforward neural network in which nodes at the input layer acquire the past observations, perform certain processing and forward the results to nodes at the subsequent layer while the node at the output layer furnishes the forecast for the future values. Nodes at the hidden layer are used for processing the data and applying certain non linear operations on them.

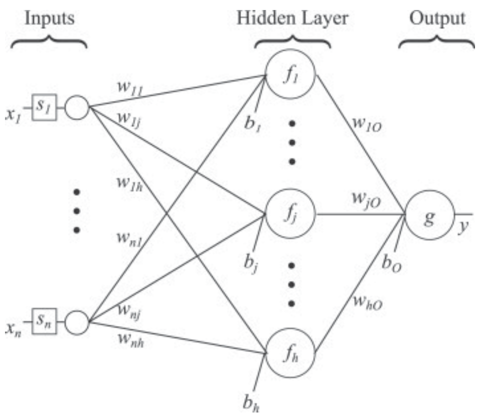

Fig. 2. Architecture of neural network model with single hidden layer.

E. Accuracy measures

Whenever we make certain estimates or projections, there are bound to be some error associated with it. Similar is the case with any forecasting technique. In order to find out how good a prediction is, we need to have certain measures that indicate the accuracy of the results [12]. These measures check the difference between the actual values and predicted values. In this paper we evaluate and compare the accuracy of a forecasting technique based on MAPE [11] (Mean absolute percentage error). It measures accuracy as a percentage, and can be calculated as the average absolute percent error for each time period minus actual values divided by actual values. Where A_t is the actual value and F_t is the forecast value, this is given by (1):-

$$M = \frac{1}{n}\sum_{t=1}^{n}|\frac{A_t - F_t}{A_t}| \tag{1}$$

III. Results and Discussions

In this paper, we will try to forecast the quarterly net sales in 2018 and then compare the actual and forecast data. Usually, there is seasonality present in a retailers sales, resulting from the holiday shopping season. Thus, it is important to collect data continuously. We obtained Amazons net sales of the first quarter to the fourth quarter from 2005 to 2018, which amounted to a total of 56 observations (actual net sales data).

There is seasonality present in the sales of Amazon. The revenue generated during the 4th quarter were the highest as compared to other quarters for each year throughout the data. This is typical for a retailer since the fourth quarter encompasses the holiday shopping season that typically runs from late November through the end of December. In addition, sales have been increasing over time. The deseasonlized sales chart shows that sales have been rising at an increasing rate.

Based on the analysis of the data pattern, we can use Holt-Winters exponential smoothing, neural network autoregression model and ARIMA (Autoregressive Integrated Moving Average model) to forecast Amazons quarterly net sales. We used R Studio to perform all of our analysis.

In this section we give a brief description of the dataset in terms of data collection process, attributes present in the data as well as the frequency at which the data is recorded on a yearly basis. We also present the forecasting results obtained by applying the above mentioned models along with a comparative analysis based on certain accuracy measures [8].

A. Dataset Description

The sales data was obtained from the 10-k and Annual report released by Amazon.com at the start of 2019. The report details it's financial performance. Every company in the United States that trades publicly need to file this report to the SEC (US Securities and Exchange Commission). We obtained quarterly sales data from the first quarter of 2005 to the fourth quarter of 2018 in Millions of US \$. Fig. 3 shows a snapshot of the dataset.

B. Holt-Winters Exponential smoothing (Using ETS)

Holt-Winters Exponential Smoothing takes into account both trend and seasonal patterns of the data as the smoothing process is applied [7]. In this model sales is chosen as the dependant/ target variable, time is set as the independent variable and due to the presence of quarterly data, the seasonality present is four.

ETS(M, A, A) showed the most promising results. Here, M indicates multiplicative level, A indicates additive trend & A indicates additive seasonality component in the order specified above. The Smoothing parameters of the above model are:-

- α = 0.4485
- β = 0.2138
- γ = 0.1852

The forecast results are shown in TABLE I below along with supporting graph in Fig. 3:-

C. ARIMA

In ARIMA model, a number of ARIMA models will be compared, and the model that produces random residuals with the lowest RMSE for the 2005 to 2017 sales will be used to

AmazonRevenue2005-18

Quarter	Amazon Quarterly Revenue (Millions of US $)	Quarter	Amazon Quarterly Revenue (Millions of US $)
2005	1,902	2012	13,185
2005	1,753	2012	12,834
2005	1,858	2012	13,806
2005	2,977	2012	21,268
2006	2,279	2013	16,070
2006	2,139	2013	15,704
2006	2,307	2013	17,092
2006	3,986	2013	25,586
2007	3,015	2014	19,741
2007	2,886	2014	19,340
2007	3,262	2014	20,579
2007	5,672	2014	29,328
2008	4,135	2015	22,717
2008	4,063	2015	23,185
2008	4,264	2015	25,358
2008	6,704	2015	35,746
2009	4,889	2016	29,128
2009	4,651	2016	30,404
2009	5,449	2016	32,714
2009	9,520	2016	43,741
2010	7,131	2017	35,714
2010	6,566	2017	37,955
2010	7,560	2017	43,744
2010	12,947	2017	60,453
2011	9,857	2018	51,042
2011	9,913	2018	52,886
2011	10,876	2018	56,576
2011	17,431	2018	72,383

Fig. 3. Quarterly Amazon sales data (2005-18).

TABLE I
ETS 2018

Period	Forecast (M USD)	Lower 95% Limit	Upper 95% Limit
Q1/2018	43816.90	38568.81	49064.99
Q2/2018	45144.66	38407.84	51881.48
Q3/2018	48297.51	39348.48	57246.54
Q4/2018	58302.87	45124.29	71481.46

forecast the 2018 sales. From the data pattern, we can find that the data is non-stationary because of the upward trend. Thus, we will transform the data to make it stationary. Then the ARIMA model, with chosen p, q and d, can be used to forecast the Amazon quarterly sales in 2018.

ARIMA(1, 2, 2) was selected in order to make the necessary predictions. Here, the parameter p = 1 indicates that we are only dependant on the past one value. The parameter d = 2 represents the differencing order which indicates that the original series was non stationary & the parameter q = 2

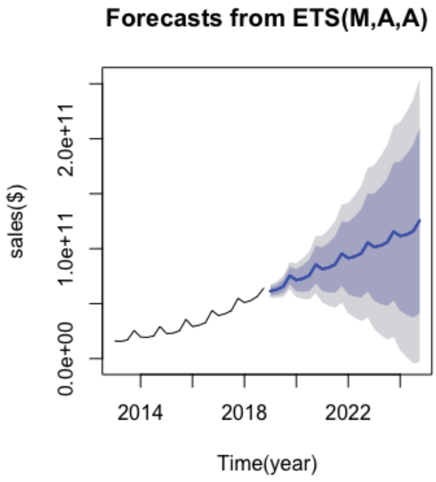

Fig. 4. Graph depicting the forecasting results obtained by applying ETS.

indicates that we are only dependant on the past two error terms. In case of seasonal ARIMA, ARIMA(0, 1, 0)(0, 1, 0) [16] was selected.

The forecast results for ARIMA & SARIMA are shown in TABLE II & TABLE III respectively along with supporting graphs in Fig. 4 & Fig. 5 below:-

TABLE II
ARIMA 2018

Period	Forecast (M USD)	Lower 95% Limit	Upper 95% Limit
Q1/2018	44630.88	42776.15	46485.60
Q2/2018	48308.01	46168.07	50447.96
Q3/2018	53051.29	50628.67	55473.90
Q4/2018	53895.73	51096.32	56695.15

TABLE III
SEASONAL ARIMA 2018

Period	Forecast (M USD)	Lower 95% Limit	Upper 95% Limit
Q1/2018	43270.99	41389.48	45152.51
Q2/2018	45854.73	43717.48	47991.98
Q3/2018	52879.35	50513.85	55244.85
Q4/2018	58019.32	55445.74	60592.90

D. Neural network autoregression model

We constructed 20 feed forward networks. Each one of them was a 2-2-1 network (i.e 2 independent variables trying to map to the dependent variable and 1 hidden layer with 2 nodes) with 9 weight options. The final model was generated by taking the average of these 20 networks.

The lagged values of the uni-variate time series are supplied as input to the neural network whereas the output generated is the predicted revenue data point. The neural network that

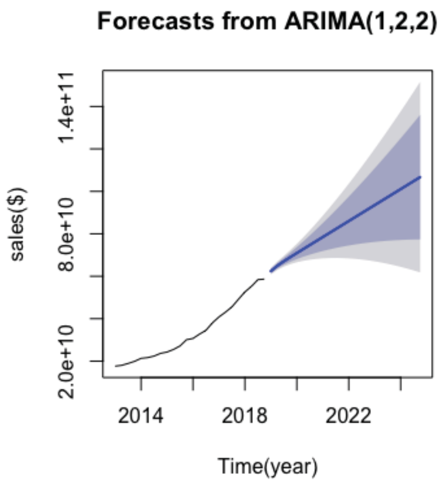

Fig. 5. Graph depicting the forecasting results obtained by applying ARIMA.

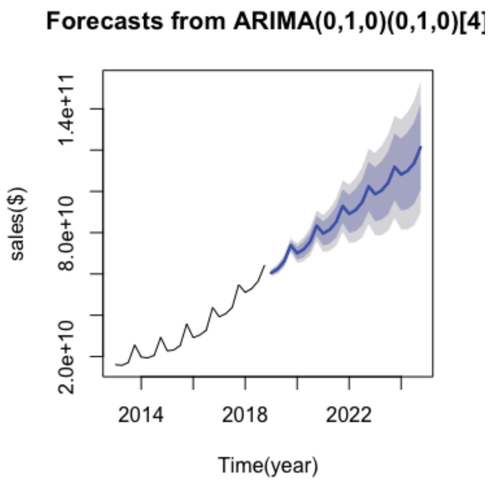

Fig. 6. Graph depicting the forecasting results obtained by applying SARIMA.

best fits our data was NNAR(1, 1, 2) [16] model. It indicated that we have one input layer, one hidden layer with 2 nodes and one output layer. It shows that we are only dependant on the previous value to act as input to the model i.e lag = 1.

The forecast results are shown in TABLE IV below along with supporting graph in Fig. 6:-

TABLE IV
NNAR 2018

Period	Forecast (M USD)
Q1/2018	44912.42
Q2/2018	46999.20
Q3/2018	51225.26
Q4/2018	54534.07

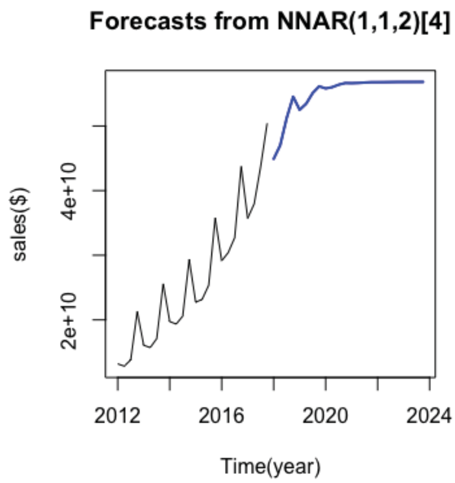

Fig. 7. Graph depicting the forecasting results obtained by applying NNAR.

E. Comparing results of the models applied

TABLE V contains value of Mean absolute percentage of error calculated in order to see which model fits our data better:-

TABLE V
MAPE

Model	MAPE
ETS	3.501521
NNAR	4.663975
ARIMA	3.469872
SARIMA	2.884046

Following is the sales forecast(Million USD) for the year 2019 in TABLE VI:-

TABLE VI
SEASONAL ARIMA 2018

Period	NNAR (M USD)	ETS (M USD)	ARIMA (M USD)	SARIMA (M USD)
Q1/2019	62670.08	61366.97	62405.49	60559.35
Q2/2019	63640.42	62808.49	64939.05	62403.35
Q3/2019	65728.01	65709.06	67048.13	66093.35
Q4/2019	69150.43	75518.32	69009.43	73847.44

Amazon on 25^{th} April 2019 reported earnings of \$59.7 billion for the 1^{st} quarter of 2019. In our study, seasonal ARIMA predicted it to be \$60.6 billion being the closest to the actual projections as compared to other models under consideration. This shows that the seasonal ARIMA model best fits the sales data and provides much better estimates/ forecast.

IV. Conclusion

In this paper, we analyze three methods to forecast sales for Amazon based on the historical data. The results show that seasonal ARIMA gives the most accurate results as compared tot he other applied methods. Based on the forecasting results, Amazon can have a big picture of the demand and then to take relevant measures to arrange resources, such as hiring more employees, storing more items or expanding shipping capacity, and thus to offer good service to improve customer satisfaction.

Though, the error percentage in forecasting (MAPE) [11] for the applied methods is not that significant and can be applied to the forecast of Amazon sales, there are still some obstacles to using these methods as follows. One major obstacle impeding the implementation of the forecast is the necessary data to precisely carry out the forecast. Amazons quarterly sales are influenced by many diverse factors, such as population, disposable household income, interest rate, macroeconomic trend and so on.

References

[1] R. A. Khan and S. Quadri, "Business intelligence: an integrated approach," *Business Intelligence Journal*, vol. 5, no. 1, pp. 64–70, 2012.

[2] J.-h. YU and X.-j. LE, "Sales forecast for amazon sales based on different statistics methodologies," *DEStech Transactions on Economics, Business and Management*, no. iceme-ebm, 2016.

[3] P. Sobreiro, D. Martinho, and A. Pratas, "Sales forecast in an it company using time series," in *2018 13th Iberian Conference on Information Systems and Technologies (CISTI)*, pp. 1–5, IEEE, 2018.

[4] A. Ribeiro, I. Seruca, and N. Durão, "Sales prediction for a pharmaceutical distribution company: A data mining based approach," in *2016 11th Iberian Conference on Information Systems and Technologies (CISTI)*, pp. 1–7, IEEE, 2016.

[5] G. Nunnari and V. Nunnari, "Forecasting monthly sales retail time series: A case study," in *2017 IEEE 19th Conference on Business Informatics (CBI)*, vol. 1, pp. 1–6, IEEE, 2017.

[6] P. A. Cholette, "Prior information and arima forecasting," *Journal of Forecasting*, vol. 1, no. 4, pp. 375–383, 1982.

[7] S. G. Makridakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting Methods for Management*. Wiley, 2008.

[8] I. Alon, M. Qi, and R. J. Sadowski, "Forecasting aggregate retail sales:: a comparison of artificial neural networks and traditional methods," *Journal of retailing and consumer services*, vol. 8, no. 3, pp. 147–156, 2001.

[9] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks:: The state of the art," *International journal of forecasting*, vol. 14, no. 1, pp. 35–62, 1998.

[10] E. Bouding, *Time Series Analysis: Forecast and Control*. Prentice-HallInc, 1994.

[11] Stephanie, "Mean absolute percentage error (mape)." https://www.statisticshowto.datasciencecentral.com/mean-absolute-percentage-error-mape/, 2017. Accessed: 2019-05-23.

[12] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International journal of forecasting*, vol. 22, no. 4, pp. 679–688, 2006.

[13] L. Zhou and X. Yong, "Electricity demand forecasting based on arima model and linear neural network," *J. Ludong Univ*, vol. 3, pp. 89–94, 2015.

[14] I. Naim and T. Mahara, "Comparative analysis of univariate forecasting techniques for industrial natural gas consumption," *International Journal of Image, Graphics and Signal Processing*, vol. 10, no. 5, p. 33, 2018.

[15] R. J. Hyndman, "Automatic time series forecasting," in *Book of Abstracts*, p. 75, 2007.

[16] L. Bianchi, J. Jarrett, and R. C. Hanumara, "Improving forecasting for telemarketing centers by arima modeling with intervention," *International Journal of Forecasting*, vol. 14, no. 4, pp. 497–504, 1998.