Applicability of ARIMA Models in Wholesale Vegetable Market: An Investigation

Manish Shukla and Sanjay Jharkharia
Quantitative Methods and Operations Management
Indian Institute of Management Kozhikode

Abstract

To investigate the applicability of ARIMA models in wholesale vegetable market models are built taking sales data of one perishable vegetable from Ahmedabad wholesales market in India. It is found that these models can be applied to forecast the demand with Mean Absolute Percentage Error (MAPE) in the range of 30%. This error is acceptable in fresh produce market where the demand and prices are highly unstable. The model is successfully validated using sales data of another vegetable from the same market. This model can facilitate the farmers and wholesalers in effective decision making.

Keywords: Forecasting, Supply Chain Management

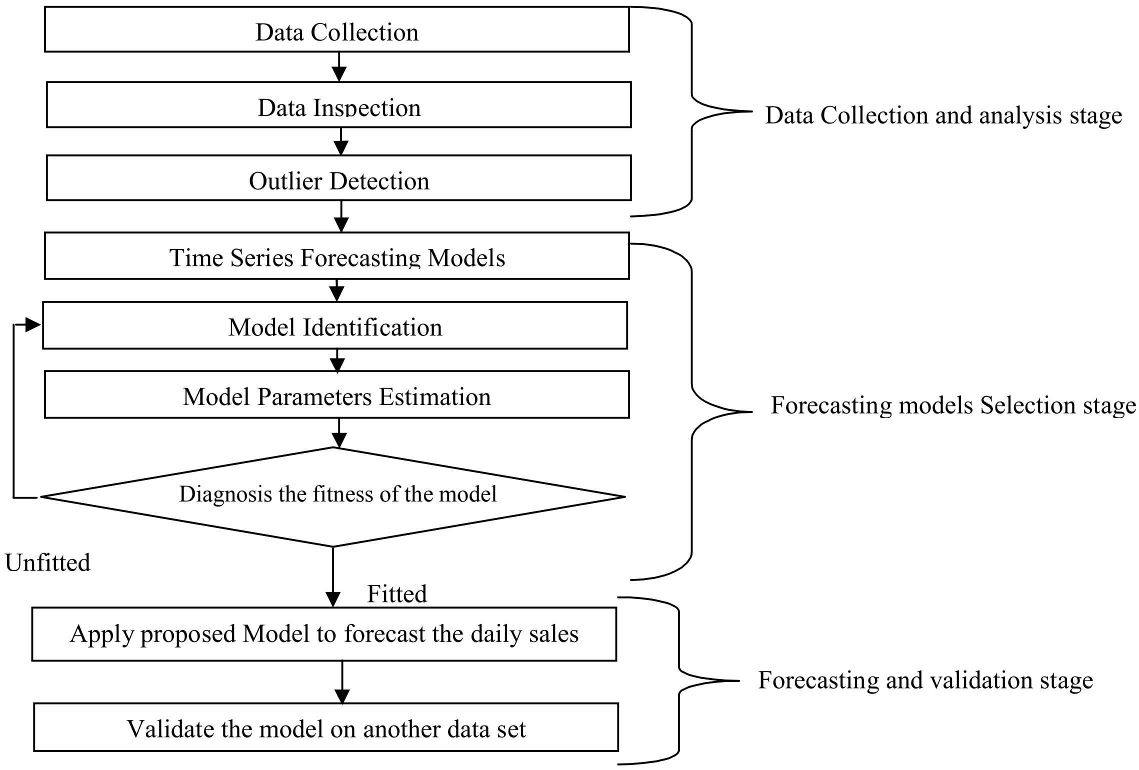
1. Introduction

The mechanism of vegetable trade is very different as compared to the trade of other lesser perishable agricultural commodities. In case of vegetables, the consolidator collects (purchases) the vegetables from the farmers and brings it to the wholesales market. The wholesales market follows a spot auction system, where commission agents participate in the auction on behalf of consolidators and retailers. These auctions happen only for few hours (e.g. 6am-11am). The consolidators are bounded to sell the vegetables by the end of the auctions, as there are almost no cold storages available and waiting for another day incurs extra cost. The perishable nature of the vegetables and lack of the infrastructure facilities can be attributed to the low bargaining power of the consolidator. This in turns results in low bargaining power and lesser profits to the farmers. One of the major reasons to this is the lack of demand visibility at the customer's end. Generally the farmers are unaware of the market trends and follow the traditional product mix (variety and volume of produce to be planted). This scenario results in a push rather than a pull system, and thus, a mismatch between demand and supply. This causes either waste of excess produce or unsatisfied customers.

Vegetables are a seasonal produce and there is a lead time between the demand and supply. The harvesting decisions are based on experience or speculation rather than market demand. This is due to the lack of an accepted forecasting model. Due to these conditions farmers are forced to sell the vegetables to the consolidators at a very low price. This is more so in the developing countries such as, India where the major percentage (~98%) of the vegetables are sold in the spot markets. Hence, there exists a need to forecast the demand of vegetables in the wholesales market to avoid wastage of vegetables and to increase the profits of all the stake holders.

Some of the application of forecasting techniques includes, Collaborative Planning, Forecasting & Replenishment (CPFR)[1-2], Gray relation analysis[3], machine learning techniques [4], multi-agent based demand forecasting applying Genetic Algorithm (GA) [5]. But the literature addressing the demand forecasting of agricultural produce is very scares. Moreover, there are very few papers that have studied the demand forecasting of vegetables in the wholesales market. Among these, most of the studies are either focusing on price forecasting or forecasting the demand on an aggregate level. Researchers have generally considered all the vegetables as a single commodity [6] and have tried to forecast the demand of vegetables. But there is a need to forecast the demand of an individual vegetable such as onion, potato, tomato. It is also found that the papers are generally forecasting the demand on weekly or monthly level [7]. But, in real life situation the farmers may need the daily demand to take their harvesting decisions due to the short selling horizon and perishable nature of the vegetables. In case of wholesales market the demand and supply is beyond the control of the intermediary food traders [8]. Adding to it is the high price fluctuations and availability of substitute produce that further increases the uncertainty of forecasted demand of an individual produce on a daily basis. However, there exist no papers forecasting the demand of an individual fresh produce on a daily basis. In real life situation, researchers and practitioners prefer models that are easy to operate and have a user friendly interface to operate in terms of set-up and data requirements [9, 10]. There exist a large number of papers studying time series forecasting and comparing the results obtained from different forecasting techniques. It is found that AutoRegressive Integrated Moving Average (ARIMA) models are more preferred in literature for short term forecasting as compared to artificial intelligent models [7, 11]. The comparison shows that artificial intelligent models such as Artificial Neural Networks (ANN) performed almost equal but not better than ARIMA models [12, 13]. On the other hand, ARIMA models performed better than ANN models in short-term forecasting [14-16]. Exponential smoothing models are a special case of ARIMA models [17]. Box and Jenkins [18] first introduced the ARIMA model building methodology. These models are highly accepted and applied due to the statistical properties [7]. The steps of ARIMA model building methodology is presented in a flow chart in figure 1.

Figure 1: Complete methodology of the forecasting approach



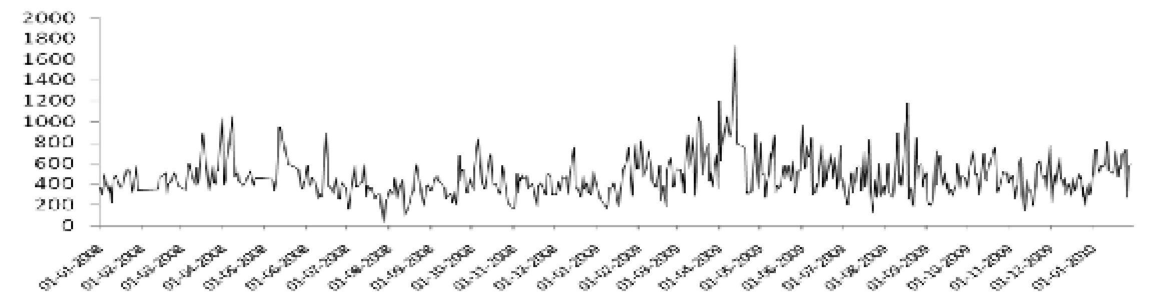
We applied ARIMA models due to its superior behaviour over other models in the same class [7] to forecast the demand of a fresh produce on a daily basis. The data is collected from Ahmedabad wholesales market over a period of twenty five months (1-Jan-07 to 31-Jan-2010) to build the models. This data is taken from the website of Agricultural Marketing Information Network-AGMARKNET portal (http://www.agmarknet.nic.in), designed, developed and maintained by Agricultural Informatics Division, Government of India. The model is identified using twenty four months of onion sales data. The left one month data is used to forecast the sales and to calculate the accuracy of the model. It is observed that the model is able to forecast the demand with an accuracy of Mean Absolute Percentage Error (MAPE) of 28.29%. In order to validate the model, data is collected for potato sales in the same market and the model is applied. Result shows that the proposed model is able to forecast the daily demand with approximately the same accuracy. Therefore, this model may be used to facilitate the farmers and wholesalers in effective decision making.

The rest of the paper is organized as follows: section two presents the data collection and ARIMA models building. Section three presents the description and implementation of the model. Section four presents forecasting and validation of the model and section five presents the result and discussion. The paper is concluded in section six with detail description of future research directions.

2. Data collection and analysis

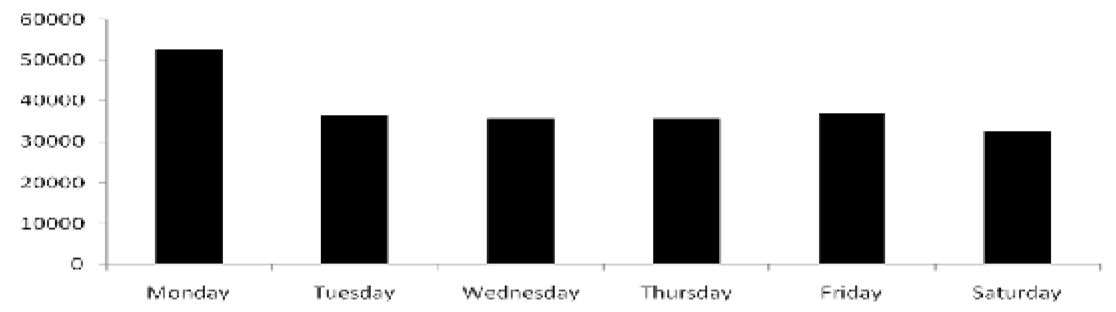
This section presents the data collection and analysis for ARIMA model building. The data is collected from the portal of Agricultural Marketing Information Network-AGMARKNET (http://www.agmarknet.nic.in). Total twenty-five months data is taken for onion sales in Ahmedabad market, India (1st January 2008 to 31st January 2010, see figure 2). Initial, twenty-four months' data is utilised for model identification. The underline assumption for taking twenty four months data is that the effect of annual cyclic factors such as changing expenditure patterns and harvesting seasons will diminish. Rest one month data is taken for comparing the forecast accuracy.

Figure 2: Daily sales for onion in Ahmedabad market (1/1/08–31/1/10)



In order to get a feel of the data, we plotted the data of mean sales across the weekdays. This reflected the importance of factors such as strikes, national holidays. From the figure 3, it is evident that the mean sale on Mondays is high as compared to other weekdays, which can be attributed to the Sundays being the scheduled holidays. Hence, on can infer that scheduled holidays effect the sales in the market where as unscheduled holidays such as strikes may not have the similar effect.

Figure 3 Mean sales across week days in Ahmedabad market



3. Forecasting Models

This section presents the ARIMA model identification and diagnosis stages.

3.1. Model Identification

This section presents the model identification stage. Among the several methods studied in the literature to judge the fitness of the models, we used Akaike information criterion (AIC) [19]. According to this the model with least AIC value will be selected. It was found that ARIMA (2, 0, 1) is the best suited model based on the AIC value.

3.2. Model Parameters Estimation

The ARIMA (2, 0, 1) model is formulated using onion sales data. Gnu Regression, Econometrics and Time-series Library (gretl) software is used for model identification and forecasting. Total 64 function evaluations and 18 gradient evaluations are performed by gretl to find the model parameters.

3.3. Diagnosis the fitness of the Model

The model is diagnosed using the Ljung-Box Q statistic to check the overall adequacy of the model [20]. The test statistic, Q, is:

$$Q_n = nr(nr+2)\sum_{l=1}^{n}\frac{r_l^2(e)}{nr-l}; \qquad \cdots\cdots(7)$$

Where $r_l(e)$ = the residual autocorrelation at lag l

nr= the number of residuals

n= the number of time lags includes in the test

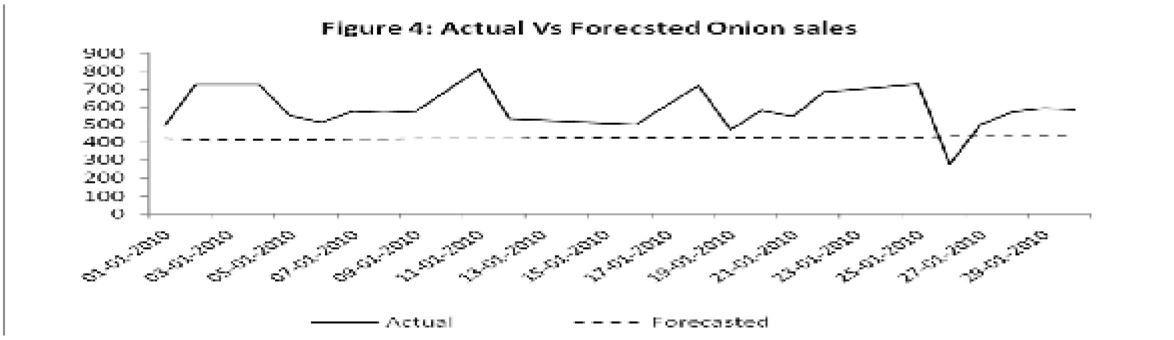
For model to be adequate, p-value associated with Q statistics should be large (p-value> α).

4. Forecasting and Model validation

This section presents the forecasting and model validation stages of the ARIMA model building methodology. Here, one month demand is forecasted for onion for January 2010. The model is later validated taking potato sales data.

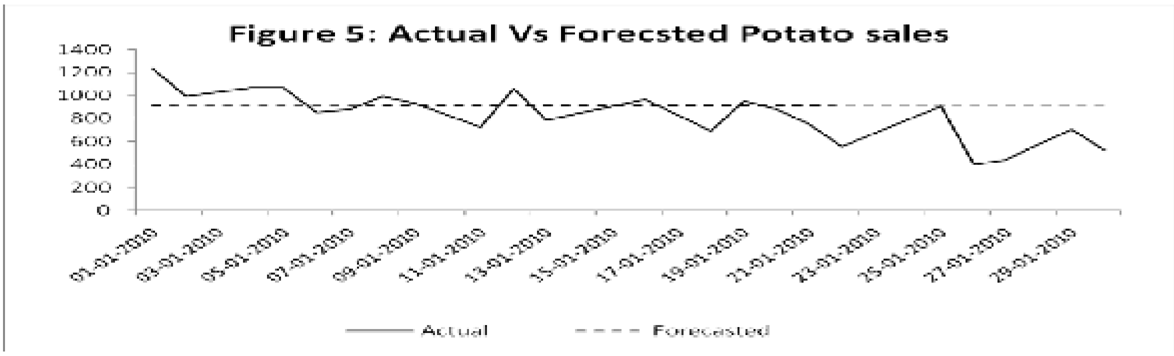
4.1. Forecasting

After the identification of the model and its adequacy check, it is used to forecast the demand of onion in the next periods. Hence, we used the identified ARIMA model to forecast the demand of onion for the month of January 2010. The forecasting results are presented in figure 4.



4.2. Model Validation

In order to validate the proposed model, data is taken for potato sales in Ahmedabad market. This provides the confirmation that the proposed model is not just a fit to the current dataset and can be used to forecast the demand in future. The ARIMA model (2, 0, 1) is applied on the potato sales data for two years for parameter estimation and forecasting is done using the rest one month data. The data shows that the original potato sales series is non-stationary so first it is converted to stationary and then the proposed model is applied. Figure 5 presents the forecasting results obtained from the model.



5. Results and Discussion

This section discusses the outcome of the application of ARIMA models on onion sales for forecasting the demand for January 2010. The parameters of the ARIMA (2, 0, 1) model shows that sales in the last two period is highly influencing the sales in the current period**.** Hence, it can be interpreted that, the sales in the last periods and the week days affects the onion demand in the current period. It was surprising to note that the price, temperature and other variables does not significantly affect the onion sales. The effect of weekdays is mainly due to Sundays being a scheduled holiday. Other reason can be the lack of proper storage facilities and market information. The efficiency of the proposed models is judged based on the MAPE values. The outcome shows that the proposed model can forecast the onion demand with an accuracy of MAPE value 28.29. This value is highly significant in the vegetable market where there is almost no information of the future demand. The proposed model is validated on the potato sales data from the same market. The parameters of the model show that the week days significantly affect the potato demand. From this it is evident that the demand in the current period is dependent on the sales in the last period and week days. The accuracy of this model is calculated based on the MAPE. It is found that the MAPE value obtained (29.511) is comparable to the MAPE value for onion sales. The results and calculation of MAPE values for onion and potato sales is presented in table 1. Thus, the proposed model can be used to facilitate the farmers and wholesalers in effective decision making.

Table 1: The forecasted values obtained from ARIMA (2,0,1) for Onion/ Potato sales

	Onion demand forecast				Potato demand forecast		
Dates	Actual	Forecasted	APE		Actual	Forecasted	APE
01-01-2010	497.4	419.76	0.156092		1235.2	911.41	0.262136
02-01-2010	725.2	412.11	0.431729		992.6	906.92	0.086319
04-01-2010	724.9	411.91	0.43177		1064.5	906.62	0.148314
05-01-2010	549.8	413.28	0.248308		1065.1	907.71	0.14777
06-01-2010	514.13	414.92	0.192967		852.59	908.53	0.065612
07-01-2010	575.1	416.56	0.275674		878.6	909.17	0.034794
08-01-2010	565.2	418.15	0.260173		991	909.82	0.081917
09-01-2010	574.2	419.67	0.269122		930.6	910.49	0.02161
11-01-2010	810.76	421.12	0.480586		722.98	911.16	0.260284
12-01-2010	535.4	422.52	0.210833		1062.8	911.83	0.142049
13-01-2010	523.7	423.85	0.190663		786.9	912.49	0.159601
16-01-2010	501.7	425.13	0.152621		964.51	913.16	0.053239
18-01-2010	721.7	426.35	0.409242		687.9	913.83	0.328434
19-01-2010	471.2	427.51	0.092721		954.08	914.49	0.041495
20-01-2010	580.73	428.63	0.261912		879.63	915.16	0.040392
21-01-2010	544.95	429.7	0.211487		753.11	915.83	0.216064
22-01-2010	683.41	430.72	0.369749		555.51	916.49	0.649817
25-01-2010	727.5	431.7	0.406598		903.32	917.16	0.015321
26-01-2010	275.5	432.64	0.570381		402.2	917.83	1.282024
27-01-2010	500.5	433.53	0.133806		443.7	918.49	1.07007
28-01-2010	570.2	434.39	0.23818		574.08	919.16	0.601101
29-01-2010	590.6	435.21	0.263105		701.5	919.83	0.311233
30-01-2010	581.6	435.99	0.250361		520.62	920.5	0.768084
			∑APE=6.50X100				∑APE=6.78X100
			MAPE=28.296				MAPE=29.51

6. Conclusion and Future Scope

This paper studies the applicability of ARIMA Models in wholesale vegetable market to forecast the demand of a vegetable on daily basis. The literature addressing this problem is very scarce. Whereas, there exists a real requirement of daily demand forecast to support the farmers and wholesalers. In order to achieve this objective, we collected data for onion sales in Ahmedabad market and applied ARIMA models to forecast the demand. The proposed models are externally validated using data for potato sales in the same market. The forecasted values obtained shows that the model is highly efficient in forecasting the demand of vegetables on a day-to-day basis. The novelty of this paper is in studying the application of ARIMA models on vegetable wholesales data, and to forecast the future demand with such an accuracy. Future research may try to study the effect of one vegetable demand over other or can study a different forecasting technique and taking the results of this study as base for comparison.

References

1. Rodriguez, R. R., Escoto, R. P., Bru, J. M., and Bas, A. O., 2008, “Collaborative forecasting management: fostering creativity within the meta value chain context,” Supply Chain Management: An International Journal, 13(5), 366-374.
2. Du, X. F., Leung, S. C., Zhang, J. L., and Lai, K., 2009, Procurement of agricultural products using the CPFR approach,” Supply Chain Management: An International Journal, 14(4), 253-258.
3. Chen, F., and Ou, T., 2009, “Gray relation analysis and multilayer functional link network sales forecasting model for perishable food in convenience store,” Expert Systems with Applications, 36(3), 7054-7063.
4. Carbonneau, R., Laframboise, K., and Vahidov, R., 2008, “Application of machine learning techniques for supply chain demand forecasting,” European Journal of Operational Research, 184(3), 1140-1154.
5. Liang, W., and Huang, C., 2006, “Agent-based demand forecast in multi-echelon supply chain,” Decision Support Systems, 42(1), 390-407.
6. Mutuc, M. E., Pan, S., and Rejesus, R. M., 2007, “Household Vegetable Demand in the Philippines: Is There an Urban-Rural Divide?” Agribusiness, 23(4), 511-527.
7. Zou, H., Xia, G., Yang, F., and Wang, H., 2007, “An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting,” Neurocomputing, 70(16-18), 2913-2923.
8. Adebanjo, D., 2009, “Understanding demand management challenges in intermediary food trading: a case study,” Supply Chain Management: An International Journal, 14(3), 224-233.
9. Bucklin, R. E., and Gupta, S., 1999, “Commercial use of UPC scanner data: Industry and academic perspectives,” Marketing Science, 18(3), 247-273.
10. Ali, Ö. G., Sayın, S., van Woensel, T., and Fransoo, J., 2009, “SKU demand forecasting in the presence of promotions,” Expert Systems with Applications, 36(10), 12340-12348.
11. Co, H. C., and Boosarawongse, R., 2007, “Forecasting Thailand’s rice export: Statistical techniques vs. artificial neural networks,” Computers and Industrial Engineering, 53(4), 610-627.
12. Church, K. B., and Curram, S. P., 1996, “Forecasting consumers’ expenditure: a comparison between econometric and neural network models,” International Journal of Forecasting, 12(2), 255-267.
13. Ntungo, C., and Boyd, M., 1998, “Commodity futures trading performance using neural network models vs. ARIMA models,” The Journal of Futures Markets, 18(8), 965-983.
14. Maier, H. R., and Dandy, G. C., 1996, “Neural network models for forecasting univariate time series,” Neural Networks World, 6(5), 747-772.
15. Kirby, H. R., Watson, S. M., and Dougherty, M. S., 1997, “Should we use neural networks or statistical models for short-term motorway traffic forecasting,” International Journal of Forecasting, 13(1), 43-50.
16. Darbellay, G. A., and Slama, M., 2000, “Forecasting the short-term demand for electricity - Do neural networks stand a better chance,” International Journal of Forecasting, 16(1), 71-83.
17. Mckenzie, E. D., 1984, “General exponential smoothing and the equivalent ARMA process,” Journal of Forecasting, 3(3), 333-344.
18. Box, G. E., and Jenkins, G. M. 1970, “Time series analysis: Forecasting and control,” Holden Day, San Francisco.
19. Akaike, H., 1974, “A new look at the statistical model identification,” IEEE Transactions on Automatic Control, 19(6), 716-723.
20. Bowerman, B. L., Connell, R. T., and Koehler, A. B. 2005, Forecasting, Time Series, and Regression: An Applied Approach 4th, Thomson, Belmont, CA.