Forecasting Australian Red Wine Sales with SARIMA and ANNs

Wei Ye [a], Arsen V. Melkumian [b]

College of Foreign Language, University of Shanghai for Science and Technology, Shanghai 200093, China

[a]843609066@qq.com, [b]arsenmelkumyan@usst.edu.cn

Abstract. Three models, including Naive Forecast, Seasonal Auto Regressive Integrated Moving Average (SARIMA), and artificial neural networks (ANNs), have been used to forecast the demand for the Red wine in Australia. The evaluation of the precision of each forecasting model is based on the Mean Absolute Percentage Error (MAPE). By comparison, it is found that ARIMA (1,1,5)(1,1,0)12 and ARIMA(1,1,6)(1,1,0)12 models to the red wine data and SARIMA models demonstrate the most superior performance.

Keywords: Forecast, SARIMA, ANNs, MAPE.

1. Introduction

The wine industry plays a crucial role in the economy of Australia. Since the 1820s, the commercial wine industry has been flourishing in Australia due to the country's superior geographic location. Currently Australia is the sixth largest producer of wine in the world, producing 1.37 billion liters of wine per year and occupying 4.9% of the total wine production in the world. With such a large proportion of production in the world market, the wine industry of Australia has a significant effect on its people. A booming wine industry increases employment and exports, and contributes to the economic growth of Australia. The wine estates become a popular tourist attraction and stimulate the development of tourism. Since the wine industry is an important pillar of Australia's economy, accurately forecasting the demand for wine is of great importance.

2. Literature

In this paper we will forecast the demand for the Red wine and compare the performance of traditional statistical forecasting models, such as Auto Regressive Integrated Moving Average (ARIMA) to that of artificial neural networks (ANNs). ARIMA models have been used extensively in a variety of fields, from medical, to supply chain to fish production. Liu et al. (2011) used ARIMA models to forecast the incidence of hemorrhagic fever with renal syndrome. Tsitsika et al. (2007) utilized ARIMA models to forecast the sting pelagic fish production. Contreras et al. (2003) employed ARIMA models to predict the next day electricity prices. Kumar et al. (2004) forecasted daily maximum surface ozone concentrations using ARIMA. These univariate ARIMA models provided excellent results in terms of explained variability and predictive power.

More recently, neural networks gained popularity as far as time series forecasting is concerned. According to Hill et al. (1996), artificial neural networks (ANNs) offer good predictive performance and become a reliable alternative approach to traditional time series forecasting. Zhang&Berardi (2001) applied them to predict the exchange rate.

Foster et al. (1991) using M-competition data found that ANNs underperform compared to Brown's, Holt's and the least squares statistical models for annual time series; they did not run the models on monthly data. Tang et al. (1990) show that ANNs and Box-Jenkins models produce similar results for time series with long history. Kang (1991) found that Box-Jenkins technique produces superior MAPE to that found with 18 different ANNs. Kang also pointed out that ANNs perform better given a longer forecast horizon.

Selecting the best ANN architecture (the number of hidden layers and hidden nodes in each layer) and choosing the right type of a learning algorithm is a challenging task that is crucial to successful ANN modeling. With ANNs there is a sizeable possibility of model over-fitting since there are no prior theoretical relationships to function as constraints. Typically ANN models that require fewer

Published by CSP © 2020 the Authors

parameters to be estimated tend to perform significantly better. However, ANNs are more robust to missing and inaccurate data and can handle interaction between several independent variables easily.

The remainder of this article is organized as follows. The next section describes the relationship between feed-forward ANNs and statistics and includes a discussion of some of the limitations of ANNs in terms of statistics. The following section provides some examples of neural networks that address specific applications. The article closes with conclusions.

3. Data

Monthly sales data on Australian red wine from January 1980 to December of 1993 are used in this study for detailed examination. The data is from the source: R. J. Hyndman, Time Series Data Library, www.robjhyndman.com/TSDL. It contains monthly sales of six types of Australian wines (red, rose, sweet white, dry white, sparkling, and fortified) for the period from January 1980 to December of 1993, and the sales of the Red wine are chosen to be studied.

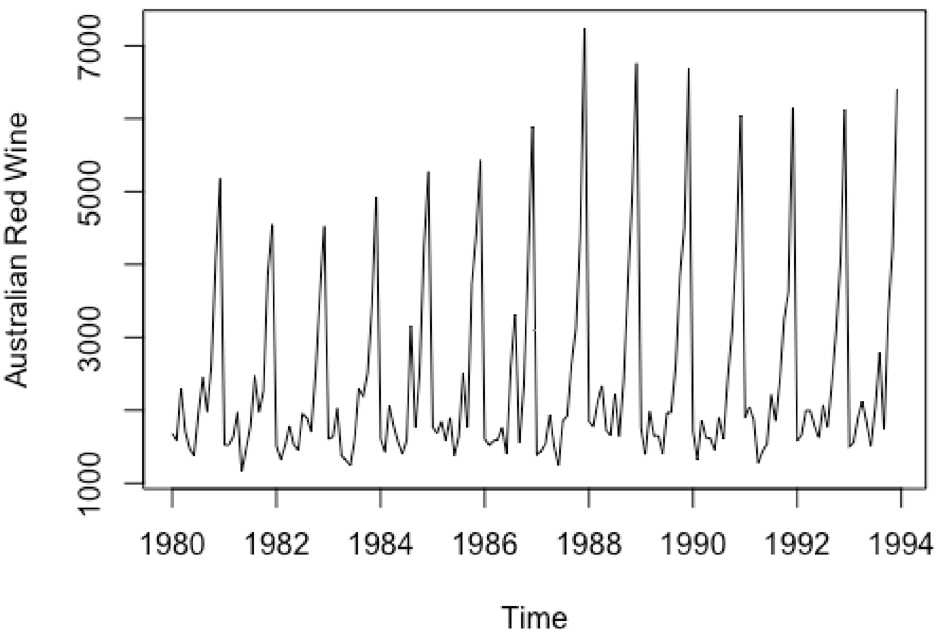

Fig 1. The Monthly sales data on Australian red wine from January 1980 to December of 1993

4. Research Methodology

We compare three forecasting models and select the best one based on the Mean Absolute Percentage Error (MAPE). This metric is widely used and is highly robust (Madrakis 1993). We use a rolling window forecast method with 36 windows and 12 month-ahead forecasts for improved accuracy. The first window contains the data from January 1980 to January 1990 with 167 observations. The second window uses the data from January 1980 to February 1990 with 168 observations. Finally, the last window employs the data from January 1980 to February 1990 with 168 observations. For each of the 36 forecasts we compute the corresponding MAPE and later on we compute the average of the 36 MAPEs.

We have used the *mlp* (multi-layer perceptron) function in R developed by Kourentzes to forecast the demand for the Australian red wine. The *mlp* function trains 20 single hidden layer ANNs and produces an ensemble forecast by computing by default the median of all forecasts. The ensemble operator for the *mlp* function could be also set to either the “mode” or the “mean” regime, so that the researcher can control how forecasts are combined. The size of the ANN ensemble could also be specified within the *mlp* function. Furthermore, the *mlp* function automatically specifies the autoregressive lags and seasonal dummy variables to be used in the model. Kourentzes recommends both controlling for the size of the ensemble and using either the “median” or the “mode” operator to combine forecasts since the “mean” is highly sensitive to outliers.

A single hidden layer (SHL) network with two nodes and up to 48 autoregressive lags was employed. We found that increasing the size of the ensemble from 20 to 120 ANNs improves the MAPE by about two percentage points, so we changed the default size of the ensemble to 120 ANNs. We have also tried all three possible ways of combining ensemble forecasts and came to the conclusion that the “median” operator performs the best, while the “mean” operator’s performance is the worst. Therefore, we kept the default way of combining forecasts. We also allowed the *mlp* function to automatically select inputs from the first 48 autoregressive lags of the dependent variable.

The architecture of the network we used is displayed in Fig. 1 and the results are shown in Table 1. We compare the forecasting results produced by the mlp function to those of Seasonal Auto Regressive Moving Average (SARIMA) models. We fit ARIMA(1,1,5)(1,1,0)12 and ARIMA (1,1,6) (1,1,0)12 models to the red wine data and find that SARIMA models demonstrate far superior performance as compared to the MLP ANNs (see Table 1).

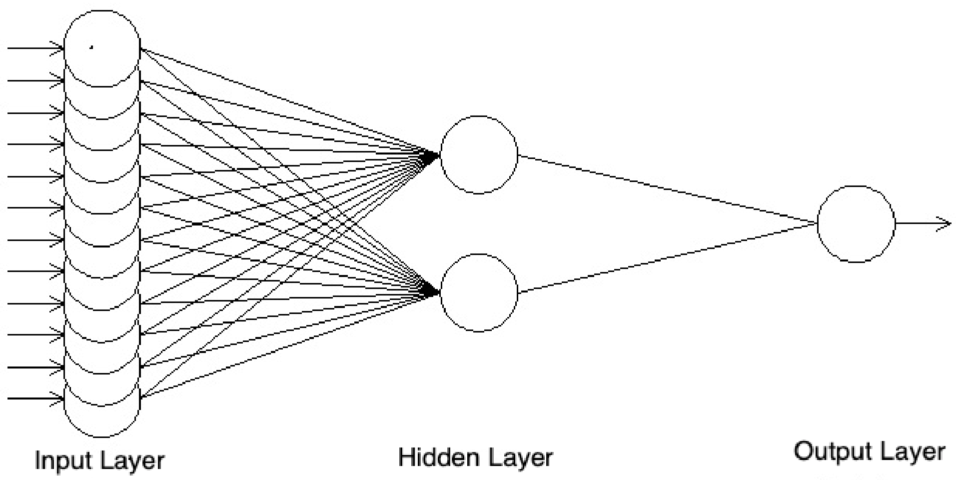

Fig 2. Single Hidden Layer ANN with Two Neurons

Table 1. Results of Forecasting Models

Forecasting Models	Naive Forecast	Neural Network	SARIMA (1) Forecast	SARIMA (2) Forecast
Average of MAPE	67.7	18.6	11.7	12.1

5. Conclusion

It can be seen from the results that Naive forecast is the worst in terms of seasonal forecasting. Neural Network is the second worst although in other cases it may be the best. SARIMA models outperform other two models and between the two SARIMA models, SARIMA (1) Forecast does slightly better then SARIMA (2) Forecasting.

References

[1]. Liu, Q.; Liu, X.; Jiang, B. and Yang, W. " Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model", Biomed Central, 2011, pp. 1-7.

[2]. Tsitsika, E.V.; Maravelias, C.D & Haralatous, J. "Modeling & forecasting pelagic fish production using univariate and multivariate ARIMA models", Fisheries Science, Volume 73 (2007), pp 979-988.

[3]. Contreras, J.; Espinola, R.; Nogales, F.J. and Conejo, A.J. "ARIMA models to predict Next Day Electricity Prices," IFEE Transactions on power system, Vol. 18 (2003), No.3, pp 1014 - 1020.

[4]. Kumar, K.; Yadav, A.K. Singh, M.P.; Hassan, H. and Jain, V.K. "Forecasting Daily Maximum Surface Ozone Concentrations in Brunei Darussalam—An ARIMA Modeling Approach.", Journal of the Air & Waste Management Association, 54:7, 809-814, 2004.

[5]. Tim Hill, Marcus O'Connor and William Remus, "Neural Network Models for Time Series Forecasts", Management Science, Vol. 42 (1996), No. 7, pp 1082-1092.

[6]. G. P. Zhang and V. L. Berardi, "Time Series Forecasting with Neural Network Ensembles: An Application for Exchange Rate Prediction", The Journal of the Operational Research Society, Vol. 52 (2001), No. 6, pp 652-664.

[7]. Foster, B., F. Collopy, and L. Ungar, "Neural Network Forecasting of Short, Noisy Time Series", presented at the ORSA TIMS National Meeting, May, 1991.

[8]. Tang, Z., C. de Almeida, and P. Fishwick, "Time Series Forecasting Using Neural Networks vs. Box-Jenkins Methodology", Simulation, 57, 5, 303-310, 1990.

[9]. Kang, S., "An Investigation of the Use of Feed forward Neural Networks for Forecasting", Ph.D. Dissertation, Kent State University, The United States, 1991.

[10]. Airman, Edward I., Marco, Giancarlo, & Varetto, Franco. "Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)", Journal of Banking & Finance, 18 (3), 505-529., 1994.

[11]. Briesch, Richard A., Chintagunta, Pradeep, & Matzkin, Rosa, "Semiparametric estimation of brand choice behavior", Journal of American Statistical Association, 97(460), 973-982., Dec.2002.

[12]. Cheng, Bing, & Titterington, D. M, "Neural networks: A review from a statistical perspective", Statistical Science, 9(1), 2-54, 1994.

[13]. Curry, Bruce, & Moutinho, Luiz, "Neural networks in marketing:Modeling consumer responses to advertising stimuli", European Journal of Marketing, 27(1), 5-20., 1993.

[14]. Flynn, Leisa R., Eastman, Jacqueline K., & Newell, Stephen J, "An exploratory study of the application of neural networks to marketing: Predicting rock music shopping behavior", Journal of Marketing Theory & Practice, 3(2), 75-85., 1995.

[15]. Geman, Stuart, Bienenstock, Elie, & Doursat, Rene, "Neural networks and the bias/ variance dilemma", Neural Computation, 4(1), 1-58., 1992.

[16]. Gronhold, Lars, & Martensen, Anne, "Analysing customer satisfaction data: A comparison of regression and artificial neural networks" International Journal of Market Research, 47(2), 121-130., 2005.

[17]. Hinton, Geoffrey E., "How neural networks learn from experience", Scientific American, 145-172., Sep.1992.

[18]. Herrero, Javier, Valencia, Alfonso, & Dopazo, Joaquin, "A hierarchical unsupervised growing neural network for clustering gene expression patterns", Bioinformatics, 17(2), 126 – 136., 2001.

[19]. Kumar, Akhil, Rao, Vithala R., & Soni, Harsh, "An empirical comparison of neural network and logistic regression", Marketing Letters, 6 (4), 251-263., 1995.

[20]. Malthouse, E. C., "Limitations of nonlinear PCA as performed with generic neural networks", IEEE Transactions on Neural Networks, 9, 165-173., 1998.

[21]. Ripley, B. D., "Neural networks and related methods for classification", Journal of Royal Statistical Society B, 56(3), 409-456., 1994.

[22]. West, Patricia M., Brockett, Patrick L., & Golden, Linda L, "A comparative analysis of neural networks and statistical methods for predicting consumer choice", Marketing Science, 16 (4), 370-392., 1997.