International Conference on Business Research
University of Moratuwa, Sri Lanka
December 04, 2023

https://doi.org/10.31705/ICBR.2023.14

PREDICTIVE ANALYTICS FOR INVENTORY OPTIMIZATION

K.F.R.F. Comester, P.G.M.P. Gunathilake, S.K. Godakumbura, M.M.F.M. Begum, D.D.R.R. Amaranath, I. Mahakalanda, and T. de Silva*

Department of Decision Sciences, University of Moratuwa, Sri Lanka
*rochellecomester@gmail.com**

ABSTRACT

In 2021, the Sri Lankan apparel manufacturing industry faced a severe downturn due to the COVID-19 pandemic and economic crises, highlighting the need for accurate sales predictions amid global supply chain disruptions. Traditional statistical models struggle to handle such crises, necessitating the exploration of machine learning methods for forecasting sales. This study aimed to identify the most effective predictive models for finished apparel goods sales, addressing data complexities like seasonality, trend, and stationarity, with a focus on enhancing decision-making in the industry. The dataset consisted of 128 weekly records of point-of-sale (POS) data for three specific apparel items sold in the US and manufactured in Sri Lanka, spanning from January 2021 to June 2023. Also, the inflation rate in the USA is used as an exogenous variable. Data preprocessing began with rationalization, followed by splitting it into training and testing sets. Two models, ARIMA and SARIMAX, were constructed using the training data to analyze the time series. Model performance was assessed using Mean Square Error (MSE), with the goal of generating future sales predictions. The results indicated that the ARIMA model outperformed SARIMAX, exhibiting significantly lower MSE values. This outcome suggests that ARIMA is the superior model for forecasting sales in this context. Future research aims to validate this result by incorporating additional datasets, ensuring the continued effectiveness of the ARIMA model in predicting apparel sales. In conclusion, this study highlights the critical role of advanced machine learning techniques, in improving sales predictions for the Sri Lankan apparel manufacturing industry. By addressing data complexities and employing robust validation methods, this research contributes to more precise planning and decision-making, essential for navigating disruptions in the global supply chain and economic uncertainties.

Key Words: Apparel manufacturing industry, Machine learning, Model comparison, Sales prediction, Time series models

1. Introduction

Sri Lanka, a country that is well known for its export apparel industry, exports apparel worth over $2,400 million to the USA per year. Notably, a significant portion of this export comprises athletic wear and lingerie. In the context of the US apparel sector, it is undeniable that it exerts a direct influence on the Sri Lankan apparel manufacturing industry.

The apparel industry, both in the United States and Sri Lanka, encountered substantial challenges during the COVID-19 pandemic. In the United States, the sales of apparel demonstrated an upward trend leading up to 2019 but witnessed a notable drop in 2020 due to the pandemic. Subsequently, sales remained depressed in 2021 and 2022. However, according to statistics, there has been a renewal in apparel sales since 2023, and this upward trend is predicted to continue. These effects were reflected in Sri Lankan apparel exports as well, with a substantial decline in 2020. Consequently, the development of advanced mechanisms for sales prediction and decision-making became crucial in order to mitigate these challenges effectively.

The research focuses on an apparel company in Sri Lanka that manufactures solely for export. The company needs to plan ahead and maintain a sufficient finished goods inventory to meet the demand for clothing items in the USA. Holding inventory, the task of storing and maintaining a certain level of inventory is a crucial business area. Maintaining finished goods is vital, as it will ensure that customer demands are always met. However since inventory holding is costly, a company must strike a balance between the two and maintain the most optimal level of inventory that minimizes the cost while guaranteeing sufficient stock to meet customer demand.

With recent uncertain economic events, anticipating future sales levels to decide the inventory levels that need to be maintained has become more crucial than ever before to reduce the associated costs, especially for the apparel manufacturers in Sri Lanka. With time series forecasting, the technique of using past data by itself to predict future values is the approach for predicting future data related to price levels or sales can be utilized. According to Pavlyshenko (2019), there are ideal time series models that can be used for predicting sales, like ARIMA and SARIMAX and research evaluated the contribution of an exogenous variable, the inflation rate in the US. After evaluating the two models based on their MSE, it was concluded that ARIMA reported a comparatively lower MSE.

2. Literature Review

In the rapidly evolving landscape of forecasting methodologies, it is essential to examine the effectiveness of various approaches in accurately predicting sales dynamics. This literature review delves into the scope of sales forecasting, with a focus on the application of traditional and advanced forecasting methods in the context of diverse industries. By exploring various forecasting techniques, this review aims to provide a comprehensive understanding of the differences and challenges associated with forecasting methodologies, thereby facilitating informed decision-making and strategic planning in the dynamic landscape of sales forecasting.

Among the numerous forecasting methods, ARIMA is a basic time series forecasting technique yet simpler to implement and well-suitable for analyses of data that have trends and/or seasonal patterns. (Wang & Aviles, 2023; Bansal, 2020). Based on many studies, ARIMA consistently underscores as a successful application in various industries, including food and financial markets where time series analysis is presented by emphasizing its ability to adaptability and providing accurate sales forecasts (Wang & Aviles, 2023; Ariyo et al., 2014; Thomassey, 2010). According to Ariyo et al., (2014), the robustness and efficiency of ARIMA, especially in short-term prediction is a known fact and has been proven by using the model for financial time series forecasting to predict stock prices. The studies state that ARIMA could reasonably compete with any emerging forecasting method, which proves that ARIMA is one of the better approaches for short-term prediction in sales forecasting as well. However, ARIMA also has some limitations when using the model in the clothing industry, it requires a dataset that covers characteristics of product sales and other business environments (Wang & Aviles, 2023) since fashion trends, seasonality, and consumer preferences play a vital role in the apparel industry.

In addition to ARIMA, the implementation of SARIMAX (Seasonal ARIMA with Exogenous Variables) emerges as a superior forecasting methodology, particularly in industries where external variables significantly influence sales dynamics. Research by Doshi et al. (2023) emphasizes the critical role of incorporating exogenous factors in the forecasting process, highlighting SARIMAX's capacity to manage inventory efficiently, thereby enhancing its applicability in industries such as apparel, where external variables play a pivotal role in sales forecasting. The empirical findings of Arunraj and Ahrens (2016) further strengthen the superiority of SARIMAX over traditional models, such as SARIMA and naive approaches, as evidenced in the successful prediction of daily sales in a German discount retail store. This superiority is not limited to the retail sector alone, as shown by the studies conducted by Alharabi and Csala (2022) in the domain of long-term electricity forecasting. The study underlines SARIMAX's ability to outperform complex alternatives, indicating its versatility and adaptability across diverse industries. Moreover, the dynamic nature of SARIMAX, as highlighted by Arunraj et al. (2016) and Alharbi & Csala (2022), enables the model to effectively account for unforeseen changes in future demand, thus enhancing the accuracy and dependability of sales forecasts. The model's capability to integrate external variables and capture the impact of various dynamic factors contributes to its robustness and reliability in addressing forecasting challenges. While acknowledging the concerns raised by Athapaththu, Grero, and Fernando (2020) regarding potential errors in dealing with continuous datasets, it is important to emphasize that with proper data preprocessing and meticulous consideration of the specific related factors, SARIMAX remains a powerful and adaptable forecasting tool, well-suited for capturing the complexities of external variables and improving the precision of sales predictions.

Further, it is a must to acknowledge the significant role played by the limitations of advanced forecasting models when choosing the appropriate approach for research. While the utilization of deep learning, particularly through recurrent neural network (RNN) algorithms such as the Long Short-Term Memory (LSTM) model, has gained grip in various time series analysis tasks, the limitations associated with its application become increasingly apparent, especially when dealing with small datasets. The research conducted by Xue et al. (2019) underscores the growing prominence of these advanced techniques, yet the study by M. J. Hamayel and A. Y. Owda (2021) reveals the constraints arising from the narrow focus of LSTM models, thereby emphasizing the potential challenges in concluding findings to broader forecasting environments. Furthermore, the critical role of parameter setting, as highlighted by Yurtsever & Gökay (n.d.), is pivotal in mitigating issues like insufficient training and overfitting, which become more evident with smaller datasets. Accordingly, these limitations underscore the necessity for a comprehensive understanding of the consequences associated with the utilization of advanced deep-learning techniques, especially in the context of sales forecasting.

There is a lack of studies specifically examining how forecasting models, like the ones discussed, are used for predicting sales in the Sri Lankan apparel industry, especially during uncertain economic times. Despite many research efforts in various sectors, there is a clear need for focused research into how these forecasting methods can be effectively applied within the unique context of Sri Lankan apparel manufacturing. Understanding this gap is crucial for developing tailored and reliable forecasting techniques that can help the industry make informed decisions, particularly in the face of the current economic uncertainties.

3. Methodology

3.1. Data set

The experimental dataset utilized in this study was taken from a leading apparel manufacturing organization in Sri Lanka, where sales data for three distinct lingerie items were displayed. The dataset spanned a duration of nearly two and a half years, capturing sales information from January 2021 to June 2023. Each entry in the dataset represented the weekly sales revenue, accompanied by the corresponding year and week of observation. The sales values were denominated in USD.

3.2. Data preprocessing

3.2.1. Data transformation and cleaning

Prior to conducting the analysis, the dataset undergoes meticulous preprocessing procedures to ensure data integrity. Feature selection shrinks the dataset to just the first three columns, which reflect item sales, in order to concentrate on the information that is crucial for inventory optimization. Thus, three lingerie items; Item 01, Item 02, and Item 03 are selected. To maintain uniformity, column names are cleaned while checking for any duplicate rows and examining for any missing values, and it is found that the dataset has none of these issues.

3.2.2. Check for stationarity

In addition to the above measures, the stationarity of the dataset is a critical consideration. To evaluate the stationarity of the time series data, the Augmented Dickey-Fuller (ADF) test was used. If any series within the dataset was found to be non-stationary, differencing techniques were applied to transform the series into a stationary form. The results of the stationarity checks on the chosen items provide crucial information about their time-series properties. To begin with, the Augmented Dickey-Fuller (ADF) test gives strong evidence that Items 2 and 3 are stationary with no need for correction. However, Item 1 is found to be non-stationary, and as a corrective measure, we perform differencing on 'Item 01,' which leads to the formation of an additional stationary time series called 'Item 01_diff.'

3.2.3. Splitting the data

To assist in creating and assessing models, the dataset was divided into training and testing subsets. The data set was split according to the traditional 80:20 ratio where 80% of the data set was split as the training data while the rest was allocated to the testing data.

3.3. Model fitting

To assess the predictive capabilities of different models, including ARIMA and SARIMAX, we followed a systematic methodology using the dataset. After splitting the dataset into training and testing subsets, all models were trained and evaluated using the testing data, with the mean square error (MSE) used as the validation criterion for model selection. All three models that were undertaken had the same preprocessing techniques before model fitting for each model.

3.3.1. ARIMA model

For the ARIMA model, the AutoARIMA algorithm is employed for the three chosen goods, namely, "Item 01," "Item 02," and "Item 03." The dataset undergoes extensive time series modeling using AutoARIMA to estimate future sales and assist in inventory optimization. For each model, the AutoARIMA function is used to automatically determine the best orders (p, d, and q).

3.3.2. SARIMAX model

In our SARIMAX model, the first step involves preprocessing the exogenous variable, the US inflation rate, obtained from the reliable OECD website. To align it with our weekly sales data, which has a different time resolution, we needed to convert this monthly inflation data into a weekly frequency. We accomplished this by employing linear interpolation, which bridged the gaps between the monthly observations by creating a continuous connection and estimating values between two monthly data points with a straight-line relationship.

The inflation rate is calculated monthly, reflecting the rate for each specific month. When transitioning from a monthly to a weekly perspective, we opted for a linear interpolation method. This entailed taking the average of the inflation rate values, effectively producing a value that represents the inflation rate for a given week. Considering that Week 2's sales depend on Week 1's inflation rate, the data set was aligned to the corresponding weekly sales data, ensuring a seamless alignment between the two.

Stationarity checks are conducted using the Augmented Dickey-Fuller (ADF) test, and measures are taken to make the data stationary. Finally, the data is divided into training and testing subsets. Subsequently, SARIMAX models are applied to the sales data, incorporating the Inflation rate as an exogenous variable influencing sales fluctuations over time. The initial step involves identifying the best SARIMA model by assessing the optimal values for p, q, d, P, Q, D, and S through the AIC (Akaike Information Criterion). Following the selection of the optimal SARIMA models, the exogenous variable is introduced as the X factor. The performance of these newly augmented SARIMA models is then evaluated using the Mean Squared Error (MSE).

4. Results and Discussion

In this section, the study presents its findings regarding the development of accurate machine learning and deep learning models with the capability to forecast future sales, extending the lead time to 4 weeks.

4.1. ARIMA

The results obtained from the ARIMA model can be summarized as follows:

Table 1. Summary of the results of the ARIMA.

Item	Model	AIC
Item 01	ARIMA (0, 0, 1)	1909.587
Item 02	ARIMA (1, 0, 0)	1932.261
Item 03	ARIMA (1, 0, 0)	1906.104

Based on the derived ARIMA models for each item, it is evident that Item 01 has an ARIMA model of (0,0,1) while Item 02 and 03 share the same model structure, which is ARIMA (1,0,0) model, which is also known as AR (1) model. This signifies that the future values of the items depend solely on their past values, along with an error term. Additionally, considering the Box-Jenkins test values for these models, it can be concluded that the process is characterized as "white," indicating the absence of significant autocorrelation in the residuals. The ARIMA Forecasts can be visually shown as in Figure 1.

4.2. SARIMAX

In the SARIMAX model, which is an extension of ARIMA, the inflation rate of the US, which is the exogenous variable, is taken into account. Consequently, the SARIMAX models for each item are derived as follows:

Table 2. Summary of the results of the SARIMAX.

Item	Model	AIC
Item 01	SARIMAX (1, 1, 2, 0, 1, 1)	1465.422
Item 02	SARIMAX (1, 2, 1, 0, 1, 1)	2026.361
Item 03	SARIMAX (0, 2, 1, 1, 1, 1)	2000.999

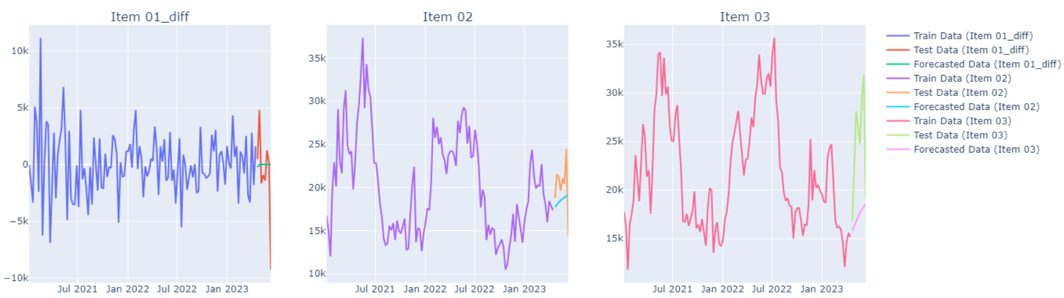

Figure 1. ARIMA forecast for the three items.

For Item 01, the derived SARIMAX model incorporates a seasonal autoregressive (AR) term with a lag of 1, and a seasonal moving average (MA) term with 1 lag. It also undergoes one seasonal differencing, which suggests a mild seasonal effect with a 12-month seasonality For Item 02, the non-seasonal component comprises an autoregressive (AR) term with a lag of 1 and one lag in the moving average (MA) term. In contrast, the seasonal component encompasses a seasonal autoregressive term (AR) with 1 lag, a seasonal moving average term (MA), and one seasonal differencing. This model accounts for both non-seasonal and seasonal dynamics. Regarding Item 03, the non-seasonal component includes both a non-seasonal AR term and a non-seasonal MA term, with two differencing to achieve stationarity. The seasonality component consists of a seasonal autoregressive term (SAR), a seasonal moving average term (SMA), and a seasonal differencing term. The seasonal component has a 12-month seasonality, and the model captures both non-seasonal and seasonal patterns for Item 03.

4.4. MSE

After constructing the models and forecasting future sales, the Mean Squared Error (MSE) serves as the metric for evaluating the effectiveness of the models. MSE quantifies the difference between the predicted values and the actual values, making it a commonly utilized statistic for comparative analysis. It enables the measurement of how closely the model's predictions align with the actual data, with models performing better in terms of reducing prediction errors having lower MSE values. As a result, the following MSE values are obtained:

Table 3. MSE of each model for selected items.

Item	ARIMA	SARIMAX
Item 01	7.537911e+06	1.170565e+07
Item 02	6.785578e+06	2.152620e+08
Item 03	2.608944e+07	2.199547e+08

Table 3 presents the evaluation metrics for all two models selected for the chosen items. It is evident from the table that, as assessed by the MSE criterion, the ARIMA model performs admirably across all three items, boasting the lowest MSE values. To visually compare the performance of these models, Figure 2 showcases the MSE values for each model across the inventory items. Notably, the ARIMA model consistently outperforms SARIMAX, exhibiting the lowest MSE values.

Figure 2. Comparison of MSE of the models.

It should be noted that the higher Mean Squared Error (MSE) values observed here are due to the limited amount of available data points. Capturing the complexity and variations of lingerie sales, particularly at a period characterized by economic uncertainty such as the global crisis, is an immense task, given the available sales data spanning little over two years. Time series forecasting models, such as traditional approaches like ARIMA or more sophisticated techniques like SARIMAX, rely on previous data to identify patterns and make precise forecasts. Increasing the size of the dataset and extending the period will likely enhance the ability of these models to encompass and consider an increased number of variables that impact sales, hence resulting in more accurate and refined predictions.

However, the study aims to develop models capable of making precise sales predictions, particularly within the context of a major Sri Lankan apparel manufacturing firm facing uncertainties caused by the pandemic. These uncertainties involve fluctuating customer demands and increased holding costs. The insights derived from this research for sales prediction hold significant relevance.

The results obtained from the ARIMA model, which demonstrated the lowest MSE values across all items, underscore the importance of our findings. These results indicate that, within our specific scenario, the traditional ARIMA strategy outperformed the SARIMAX models in terms of prediction accuracy. Despite the incorporation of external factors in the SARIMAX models, their contribution did not yield the expected improvements in prediction accuracy.

5. Conclusion

According to the research conducted, two models have been analyzed to determine the most suitable model for predicting sales one month ahead while considering the impact of a global crisis. Between the two, ARIMA and SARIMAX, it has been observed that the most suitable model is ARIMA. This determination is because ARIMA exhibits a lower Mean Squared Error (MSE) comparatively. Though the results of the models have a high MSE, the comparative decision as to which model is comparatively suitable can be made. Consequently, ARIMA is considered the most suitable model for predicting sales during a crisis within the Sri Lankan apparel sector, for the organization under study. It is worth mentioning that the dataset employed in this study was derived from an economically uncertain period, and therefore, has a limited number of data points. Hence, it is suggested that future studies focus on collecting and integrating a greater amount of data points, encompassing multiple phases of the economy and diverse market conditions. This approach is expected to significantly enhance the ability to predict these models and broaden their practicality in real-world contexts.

References

Alharbi, F. R., & Csala, D. (2022). A Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) Forecasting Model-Based Time Series Approach. *Inventions*, *7*(4), 94. https://doi.org/10.3390/inventions7040094

Ariyo, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Stock Price Prediction Using the ARIMA Model. *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*. https://doi.org/10.1109/uksim.2014.67

Arunraj, N. S., Ahrens, D., & Fernandes, M. (2016). Application of SARIMAX Model to Forecast Daily Sales in Food Retail Industry. *International Journal of Operations Research and Information Systems*, *7*(2), 1–21. https://doi.org/10.4018/ijoris.2016040101

Athapaththu, A. M. K. B., Grero, E. H., Perera, S. M., & Fernando, M. G. N. A. S. (2020). Forecasting Better Prices for Trip Packages based on Historical Sales Data and Related Factors: In the context of Europe Railway Tourism. *2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer)*. https://doi.org/10.1109/icter51097.2020.9325480

Bansal, A. (2020). *Role of Machine Learning in Inventory Optimization using Time-series Forecasting ROLE OF MACHINE LEARNING IN INVENTORY OPTIMIZATION USING TIME-SERIES FORECASTING*.

Beheshti-Kashi, S., Karimi, H. R., Thoben, K.-D., Lütjen, M., & Teucke, M. (2014). A survey on retail sales forecasting and prediction in fashion markets. *Systems Science & Control Engineering*, *3*(1), 154–161. https://doi.org/10.1080/21642583.2014.999389

Doshi, M., Pandey, P., & Ramrakhiani, N. (2023, August 31). *Optimization of Demand Forecasting & Inventory Management Using SARIMAX*. Conference: BASIC-23 (Business Analytics and Social Innovation Conference 2023), Mumbai, Maharashtra, India. https://www.researchgate.net/publication/373523277_Optimization_of_Demand_Forecasting_Inventory_Management_Using_SARIMAX

Hamayel, M. J., & Owda, A. Y. (2021). A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms. *AI*, *2*(4), 477–496. https://doi.org/10.3390/ai2040030

Hong, J.-K. (2021). LSTM-based Sales Forecasting Model. *KSII Transactions on Internet and Information Systems*, *15*(4). https://doi.org/10.3837/tiis.2021.04.003

Kaneko, Y., & Yada, K. (2016, December 1). *A Deep Learning Approach for the Prediction of Retail Store Sales*. IEEE Xplore. https://doi.org/10.1109/ICDMW.2016.0082

Lakshmanan, B., Vivek Raja, P. S. N., & Kalathiappan, V. (2020). Sales Demand Forecasting Using LSTM Network. *Advances in Intelligent Systems and Computing*, 125–132. https://doi.org/10.1007/978-981-15-0199-9_11

Li, Y., Yang, Y., Zhu, K., & Zhang, J. (2021). Clothing Sale Forecasting by a Composite GRU–Prophet Model with an Attention Mechanism. *IEEE Transactions on Industrial Informatics*, *17*(12), 8335–8344. https://doi.org/10.1109/tii.2021.3057922

Pavlyshenko, B. (2019). Machine-Learning Models for Sales Time Series Forecasting. *Data*, *4*(1), 15. https://doi.org/10.3390/data4010015

Punam, K., Pamula, R., & Jain, P. K. (2018). *A Two-Level Statistical Model for Big Mart Sales Prediction*. https://doi.org/10.1109/GUCON.2018.8675060

Song, X., Liu, Y., Xue, L., Wang, J., Zhang, J., Wang, J., Jiang, L., & Cheng, Z. (2020). Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model. *Journal of Petroleum Science and Engineering*, *186*, 106682. https://doi.org/10.1016/j.petrol.2019.106682

Thomassey, S. (2010). Sales forecasts in clothing industry: The key success factor of the supply chain management. *International Journal of Production Economics*, *128*(2), 470–483. https://doi.org/10.1016/j.ijpe.2010.07.018

Wang, F., & Aviles, J. (2023). Contrasting Univariate and Multivariate Time Series Forecasting Methods for Sales: A Comparative Analysis. *Applied Science and*

Innovative Research, *7*(2), p127. https://doi.org/10.22158/asir.v7n2p127

Xue, N., Triguero, I., Figueredo, G. P., & Landa-Silva, D. (2019). Evolving Deep CNN-LSTMs for Inventory Time Series Prediction. *2019 IEEE Congress on Evolutionary Computation (CEC)*. https://doi.org/10.1109/cec.2019.8789957

Yurtsever, M., & Gökay, U. (n.d.). *A NOVEL HYBRID MODEL USING ARIMA AND LSTM FOR SALES FORECASTING*.