Inventory Optimization with Predictive and Time Series Modeling

Yuanshan Zhang, Leonardo Trucios, Shubham Mishra
Questrom School of Business
Boston University

1. Introduction

1.1. Objective

Spinnaker Analytics is helping an online retailer to optimize its inventory by forecasting order quantities in the next year for different products so that these products will not run into shortage when orders are placed

1.2. Dataset

The dataset consists of over 1 million online sales records of different products over 5 years

2. Methodology

Quantitative forecasting approaches can be statistical or machine learning based. It is usually better to apply simple statistical models when the underlying mechanisms are not known or too complicated [1]. Therefore, SARIMA and Exponential Smoothing are the two essential models in this project, and 'Quantity Ordered' is used as the time series.

3. EDA - Time Series Analysis

3.1. Stationarity

60% of products have stationary series by ADF testing

3.2. Time Dependency

Time dependencies are examined by ACF and PACF (Chart 1 and Chart 2), showing significant spikes mostly occur at lags 1 and 12

3.3. Trend & Seasonality

Seasonalities are examined using seasonal decomposition, showing non-linear trends and repetitive seasonal patterns every 12 month

Chart 1: Autocorrelation Function (ACF)

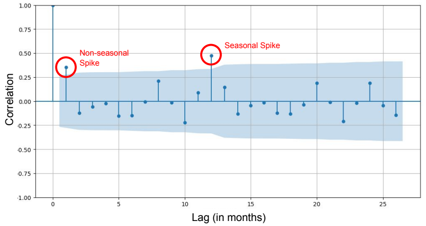


Chart 2: Partial Autocorrelation Function (PACF)

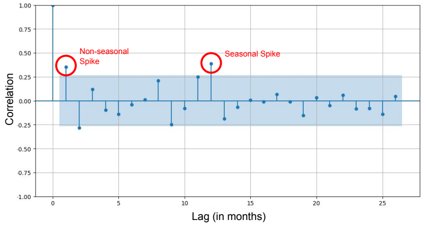
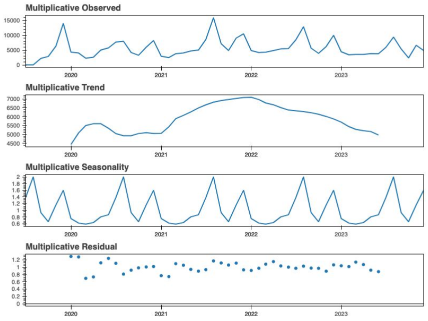


Chart 3: Trend & Seasonality



4. Preprocessing

- Data cleaning
- Data integrity check
- Identify and remove discontinued products
- Aggregation
- Deal with discontinuities in time series using linear interpolation
- Prepare data for modeling

5. Models

ARIMA, SARIMA, SARIMAX, Double Exponential Smoothing (DES), Triple Exponential Smoothing (TES), and LSTM were built for 341 products and evaluated by:

$$\text{Percentage Error} = \frac{\sum_t^n |\hat{y}_t - y_t|}{\sum_t^n y_t} \times 100\%$$

This formula represents the error for predicting the 12-month total inventory

6. Results and Findings

- SARIMA and TES gives better predictions when seasonality is obvious as shown by Chart 4
- ARIMA and DES gives better predictions when seasonality is not obvious as shown by Chart 5
- LSTM significantly outperforms other models for some products, but it should be used if sufficient data and when interpretability is not required [2]
- Stacking these models will constraint the errors within 30% for 78% products and 10% for 45% products

Chart 4: A prediction by Triple Exponential Smoothing with 2.77% error

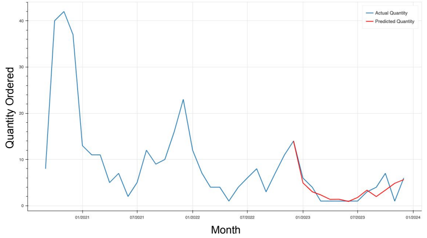
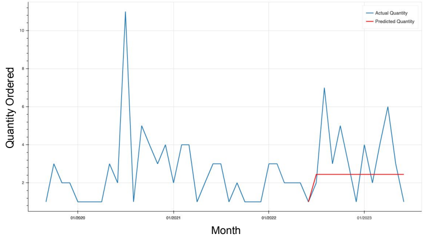


Chart 5: A Prediction by ARIMA with 28.55% error



7. Next Steps

- Train the models on the entire dataset
- Apply cross-validation for model selection
- Stack the models

Acknowledgements
This capstone project is supported by Spinnaker Analytics and Boston University

References

[1] Neptune.ai. (n.d.). (August 18, 2023) ARIMA & SARIMA: Real-World Time Series Forecasting. Retrieved from https://neptune.ai/blog/arima-sarima-real-world-time-series-forecasting-guide
[2] Uber Technologies Inc. (September 6, 2018). Forecasting at Uber: An Introduction. Retrieved from https://www.uber.com/blog/forecasting-introduction/