Demand Forecasting using Machine Learning Methods: An Empirical Study on Walmart Retail Sales Forecast

Yongjai (Jay) Lee

Stockholm School of Economics

May 26, 2023

Motivation

- Economists conduct extensive research on demand estimation. Less is explored in a forecasting environment
- Business and Economics literature continues to adopt predictive models in diverse research areas
- Demand forecasting is crucial in practical settings and serves as a basis for supply chain management
- Motivated to find practical tools that economists and time series forecasters can utilize

Question

- The study maintains an exploratory purpose. The goal is a comparison of the chosen modeling approaches for forecasting Walmart's retail demand.
- Emphasis is put on the comparison of the benchmark model to the machine learning models. Prediction error, benefits, and disadvantages of each model are discussed

Literature

- Abbasimehr et al. (2020) propose a demand forecasting method based on multi-layer LSTM networks and finds a strong ability to capture nonlinear patterns in time series data
- Mirakyan et al. (2017) model and forecast electricity price using support vector regression, artificial neural networks, and ridge regression methods
- Wang and Guo (2020) propose a hybrid model with greater predictive performance than a single ARIMA model or a single XGBoost model in predicting stock prices
- Sagheer and Kotb (2019) compare petroleum production forecast prediction from deep LSTM recurrent networks to a benchmark ARIMA from two oilfields

Background

- Supply Chain Management can be viewed as a value chain: identifying and creating value for customers via upstream or downstream links in different processes and activities
- Value chain creates a competitive advantage by optimizing each activity and maximizing the overall value created for customers
- SCM is driven by demand and demand forecasting is the starting point for understanding the supply chain
- Demand planning in SCM is essential and in many cases uses econometric methods, such as ARIMA, for quantitative analysis.

Background

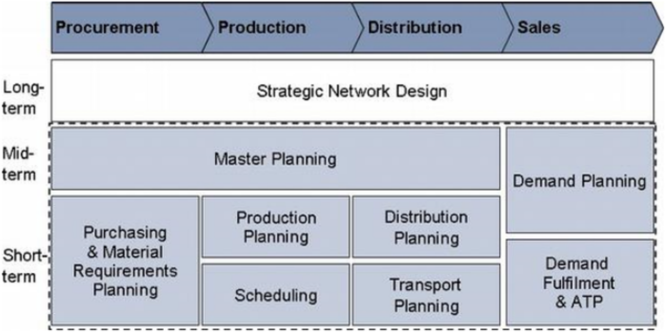

Figure 1: Supply Chain Planning Matrix (Fleischmann et al., 2002)

Background

- Demand estimation in general has its origin within economic literature
- Industrial Organization (IO) literature uses structural methods to estimate a causal relationship between price and quantity
- It uses panel datasets for causal inference. A popular empirical method is the BLP method of random-coefficients logit model which uses IV
- Growing amount of applied research in IO includes more computional methods, similar to the trends within the business literature.

Data

- Hierarchical sales data from Walmart which covers stores in three US States, California, Texas, and Wisconsin.
- Data is publicly released for the M5 forecasting competition which the Makridakis Open Forecasting Center (MOFC) hosts.
- The dataset consists of many household and food products with daily sales ranging from 2011 to 2016.
- 1913 daily observations. A large number of observations allow for a fine time series modeling, and is long enough to account for annual seasonality.

Data

- Top 10 selling products were chosen for forecasting as products in high demand typically need effective demand planning
- Daily sale of a product is the output variable. Explanatory variables such as price, promotions, day of the week, and special events were incorporated
- Data limitation: Many of the time series experienced non-linear properties such as zero demands for periods of time

Data

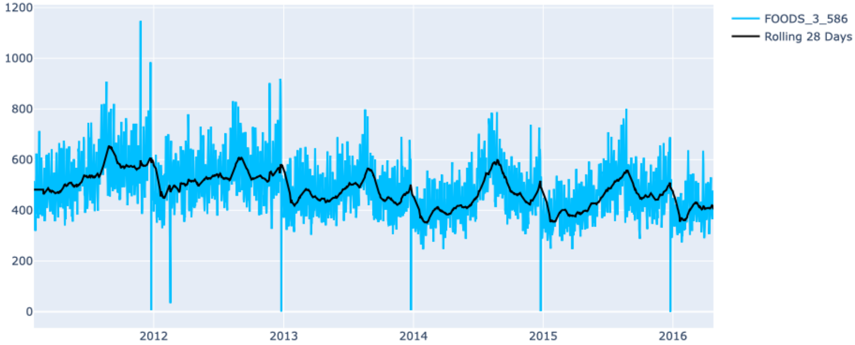

Figure 2: Example plot of time series data

Empirical Strategy

- ARIMA

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

- Combines Autoregression (AR), Stationarity (I), and Moving Average (MA)
- Where p represents the order of the autoregressive part, d the degree of differencing, and q the order of the moving average part.
- The parameters (p, d, q) of the ARIMA model are estimated using statistical techniques, and the model is then used to forecast future values based on the observed past data.
- Assumption: Stationarity in the data, linearity in the data, parametric

- RNN is a type of Artificial neural network with loops making information persist. LSTM is a type of Recurrent Neural Network (RNN) designed to better store and remember information for long periods of time

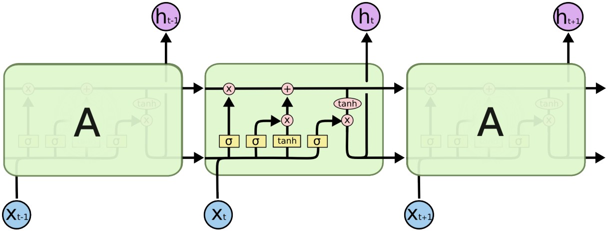

Figure 3: Structure of modules in an LSTM

LSTM

- RNN imitates structure of neural networks in the human brain and how it functions. Human brains have persistent thoughts which affect understanding of next reasoning.
- LSTM has the structure of repeating modules of neural network with four neural network layers. The horizontal line that goes through upper side of modules is called cell state. Cell state allows information to flow from each neural network module with information being added or removed from structures called gates.
- Gates are composed out of a sigmoid neural net layer lets additional information to cell state. Gates consist of input gate, cell, and forget gate

XGBoost

- Extreme Gradient Boosting is a supervised learning algorithm frequently used for regressions and classification problems.

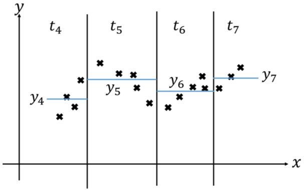

Figure 4: Example of regression tree with four terminal nodes

- XGBoost is a gradient boosted trees algorithm which combines hundreds of decision tree models. It iteratively trains decision trees, where each subsequent tree improves on the mistakes made by the previous trees

Results

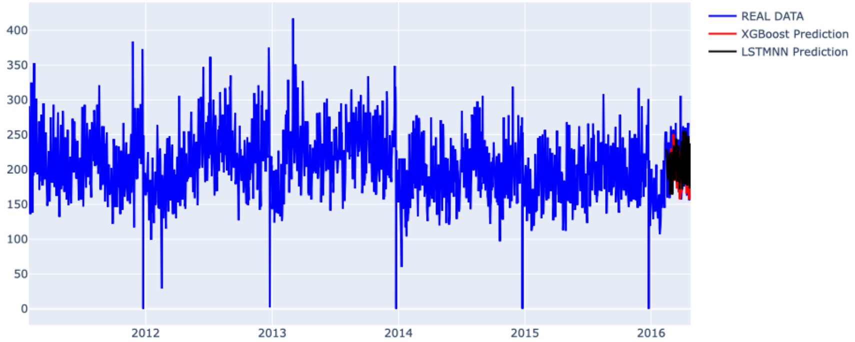

Figure 5: Predictions from LSTM and XGBoost

Results

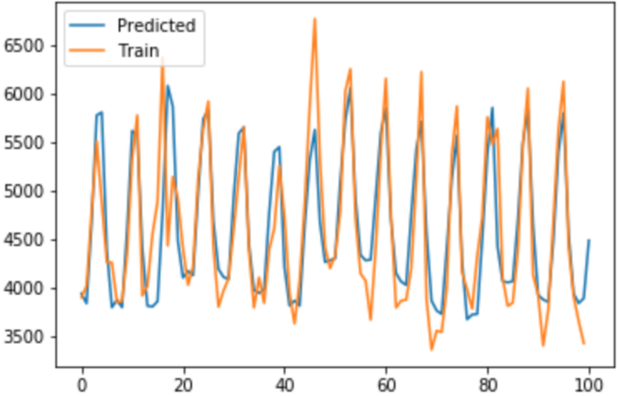

Figure 6: Predicted forecasts on test data for ARIMA

Results

- Last three months, 100 days, were split to be assigned to the test set. The rest of the data points were assigned to a training set
- The results suggest that machine learning methods provide more accurate prediction capabilities than the more traditional econometric ARIMA approach
- The machine learning models provided similar accuracy results in comparison to each other with LSTM performing better in more product series than XGBoost
- Machine learning methods bypass the assumption of stationary and linearity providing leniency in modeling and predicting diverse time series data.
- For retail forecasting where sales observe high fluctuations, including zero sales, machine learning methods are more useful for the use of forecasting

Results

	ARIMA			LSTM			XGBoost		
item id	RMSE	MAPE	SMAPE	RMSE	MAPE	SMAPE	RMSE	MAPE	SMAPE
FOODS_3_694	30.89	2448.52	11.64	21.47	1731.47	8.31	24.86	1861.96	8.84
HOBBIES_1_354	18.19	1403.73	21.58	12.26	990.71	17.70	10.23	833.40	15.63
FOODS_3_444				0.71	70.51	200.00	0.16	833.40	200
FOODS_3_555	35.74	2741.97	14.81	26.81	2153.17	9.32	34.97	2666.27	11.46
FOODS_3_252	60.98	3970.46	18.23	42.35	3321.58	11.63	45.65	3540.68	12.86
FOODS_3_587				38.32	3034.70	11.81	37.59	3069.07	12.73
FOODS_3_202				30.92	2171.78	11.54	84.03	6274.30	44.51
FOODS_3_090				99.53	7059.44	12.77	163.17	13551.83	27.44
FOODS_3_120				85.90	6181.77	63.81	123.04	10756.21	79.17
FOODS_3_586	73.59	5034.80	14.84	49.62	3775.59	9.13	44.51	3493.33	8.45

Table 6.1: Prediction Accuracy of Top Demand Products - Baseline

Discussion

- A notable constraint in many of time series study is that the models are only tested on one series. Multiple products were examined in the study, allowing further exploration and validation of the findings
- Interpretation is a downside of machine learning methods compared to ARIMA. ARIMA provides more interpretable results with explicit relationships between variables based on statistical principles and assumptions
- Machine learning methods bypass the assumption of stationary and linearity providing leniency in modeling and predicting diverse time series data.
- For retail forecasting where sales observe high fluctuations, including zero sales, machine learning methods are more useful for use of forecasting

Discussion

- Multiple measures of accuracy, namely RMSE, MAPE, and SMAPE, are compared
- Further exploration is done with the LSTM model to compare prediction accuracy. Different numbers of days ahead for predictions are attempted and saw a loss of accuracy with longer-term predictions.
- The modeling process in the study gave precedence to a "fair assessment" of the models by following a general methodology
- A limitation of this study is that machine learning methods do not offer statistically stable confidence intervals for predictions
- Overall, the results are consistent with the anticipated outcomes derived from theory and previous findings.

Discussion

	LSTM		
item id	RMSE	MAPE	SMAPE
FOODS_3_694	21.78	1783.71	8.49
HOBBIES_1_354	11.07	909.55	16.53
FOODS_3_444	1.60	160.05	200.00
FOODS_3_555	27.08	2191.46	9.58
FOODS_3_252	42.67	3290.98	11.63
FOODS_3_587	40.73	3064.56	11.97
FOODS_3_202	30.73	2178.15	11.53
FOODS_3_090	126.09	10116.78	18.12
FOODS_3_120	86.96	6068.87	64.58
FOODS_3_586	49.06	3726.80	9.02

Table 6.2: LSTM model with 5 days ahead predictions

	LSTM		
item id	RMSE	MAPE	SMAPE
FOODS_3_694	21.99	1861.99	8.93
HOBBIES_1_354	10.72	894.56	15.66
FOODS_3_444	2.25	225.11	200.00
FOODS_3_555	33.32	2742.20	12.08
FOODS_3_252	47.79	3464.27	12.24
FOODS_3_587	42.44	3536.49	13.39
FOODS_3_202	31.17	2205.67	11.78
FOODS_3_090	117.53	8315.68	15.22
FOODS_3_120	99.23	6848.86	87.82
FOODS_3_586	53.54	4203.36	10.06

Table 6.3: LSTM model with 25 days ahead predictions

Conclusion

- The study provides additional evidence to the demand forecasting literature and compares machine learning methods with ARIMA
- Overall, the machine learning models showed higher forecast accuracy compared to the ARIMA model
- The valuable contribution to this research is the success of replication and extension on multiple time series
- Future research should continue to explore models in different demand contexts such as for nonphysical goods to gain more clarity.