Demand Forecasting using Data Analytics and Stochastics

Shivam M Dave, Aditya Patel, Forum Bhutaiya and Prof. Mohammed Awwad
Industrial and Systems Engineering Department
University at Buffalo, The State University of New York
Buffalo, NY 14260, USA

Abstract

Every Industry and Organization is leaning towards extensive use of data due to its availability and usefulness, including supply chain engineering. This has given a rise in the development of demand forecasting models and techniques for accurate predictions. Implementing big data to the supply chain has gained a lot of attention from various organizations to attain accuracy and dependability. Using Markov Chains and Bayesian Statistics for short term predictions are widely studied due to their simplicity along with other statistical techniques. The supply chain has a huge amount of data available ranging from transportation to consumer demand. This historical data can be used to model the probabilities to forecast the demands or uncertainties in the future. This paper focuses on understanding and implementing various analysis techniques and multiple ordered Markov chains to forecast the demand and see the plausible applications of this technique in the field of supply chains.

Keywords

Markov Chains, Supply Chain, Forecasting, Bayesian Statistics.

1. Literature Review

In the field of probability, there are two approaches, Frequentist approach and Bayesian approach. Both of these concepts are used widely and studied in other fields. In the field of the supply chain, forecasting is a crucial part of the system. There are a lot of sophisticated algorithms that can forecast with good accuracy. Forecasting can be categorized into 2 parts, Now-casting: Short term forecasting and Long term forecasting. These methods of forecasting can help make decisions based on the forecast. These decisions can be tactical or strategic based on prediction. Stochastic Process is a method of forecasting done using mathematical abstraction and modeling. The stochastic process has found its applications in many fields around and particularly, Markov chains are the best sources of forecasting models in order to capture dynamic behavior having a large stochastic component (AT 1960). (Zhihang Peng 2010) showed that Markov chains can be used even in the field of diseases to forecast infectious diseases. It compared the historical methods of forecasting such as Monte-Carlo simulations, multivariate statistics analysis which greatly relies on the historical data to an improved Markov theory. They determined the scientific classification of the weighted Markov chain by finding the system's initial state and ensuring the state transition probability matrix. (C. Catal 2019) predicted the sales using various different regression techniques such as Bayesian regression, Linear regression, Neural Network regression, Decision Forest regression and various time series ARIMA models. As per the analysis, Boosted Decision Tree regression provided a new decision support system that gave the best performance on the sales data of Walmart. (Kalekar 2004) represented that when the data has trend and seasonality, Holt-Winters exponential smoothing method is used which constantly updates the most recent observations of the data to calculate the forecast. This helps in accurate forecasting and avoids poor performance of the forecast model.(Denis Tikunov 2004) evaluated a technique to predict traffic on a cell of radio access network using Holt-Winter's exponential smoothing. The Erlang method proposed, was used to evaluate GSM/GPRS network for a 3G mobile network which helped in estimating future network traffic value. (OCZKI 2014) shows that forecasting method is used even in human resource planning which helps in reducing HR costs and improves the organization's performance. Markov chain analysis is used to forecast internal labor in a company or retail store. It helps in analyzing labor supply required in each job category and also the number of future vacancies that need to be filled. Another effective forecasting method, the Hidden Markov Model (HMM) is used in fields of Natural Language Processing, Robotics, and Bio-genetics. HMM helps in predicting the hidden variables that are unknown variables from a set of observed variables in the data set (Dorairaj 2018). It considers joint probabilities of the sequence of hidden states to determine the best possible sequence of hidden states. These are a few works in the field of Supply Chain and forecasting. In the subsequent sections, we present a comparison between the forecasting methods used in the field of supply chain considering Walmart's sales data.

These methods have also been adopted in other disciplines like stock market predictions and text predictions in smartphones. The bsts package in R original developed by (SCOTT 2017) for structural time series. According to

(Larsen 2016) the Bayesian approach to structural time series is more transparent as it explains seasonality and explaining uncertainty for insufficient historical data. In (Larsen 2016) approach, he used a typical time series for Number of Air passengers growing over a span of years in the USA, as in Bayesian approach we have three main components prior distribution or prior knowledge, the evidence or the actual data and updated knowledge of the event using evidence and prior distribution known as 'Posterior Distribution'. We define our prior knowledge before looking at the data then keep updating our knowledge with every dataset in time by creating a likelihood function and define our posterior distribution or beliefs after looking at the historical data in our case historical demand or sales. Another method we have approached this problem was using frequentist approach by calculating the number of possible outcomes over the total number of events as proposed by (Sirajudeen 2017), but in a broader sense by replacing binary outcomes with the possible probability of sales on that day. we can make a sequence of any random variable X behave like a Markov Chain, this method defines binary sequence of whether sale was made that day or not as 1 and 0, by such approach obtaining a long sequence of 0's and 1's he can calculate transition probabilities that whether a sale was made that or not. Thus citing these papers, we developed and implemented these methods in univariate time series data to obtain a forecast for sales. Another method that has been widely adopted in univariate time series forecasting is Exponential Smoothing State space method. Each exponential smoothing method can be described by two components trend 'T' and seasonality 'S' with multiple combinations of additive(A), additive damped(A_d), multiplicative(M), multiplicative damped(M_d) and none(N) the methods like simple exponential smoothing, Holt's linear method, Holt's Winter method can be defined by these combinations as shown in Figure 1. As (Hyndman 2018) stated "Alternative to improving these models by minimizing the sum of squared error is maximizing the probability of the data from specified model i.e. finding the best possible combination of the E, T, S parameters by combination of components A, A_d, M, M_d and N and maximizing the likelihood." This is shown in Figure 1.

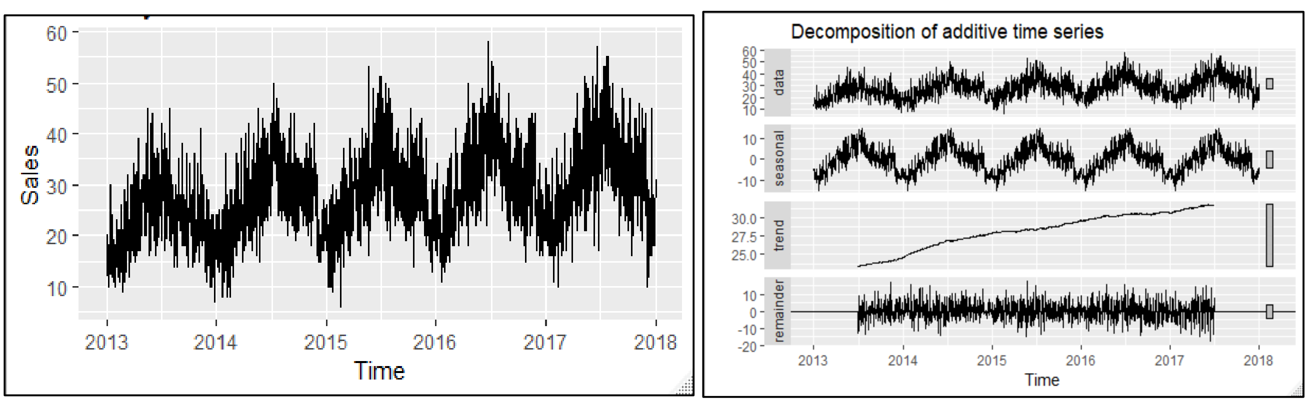
Short hand	Method
(N,N)	Simple exponential smoothing
(A,N)	Holt's linear method
(A_d,N)	Additive damped trend method
(A,A)	Additive Holt-Winters' method
(A,M)	Multiplicative Holt-Winters' method
(A_d,M)	Holt-Winters' damped method

Figure 1. Classification Developed by (Pegels 1969) and further developed by Gardner(1985) from (Hyndman 2018)

2. Methodology and Data

We obtained two datasets and used R software to implement the traditional and probabilistic forecasting techniques as follows:

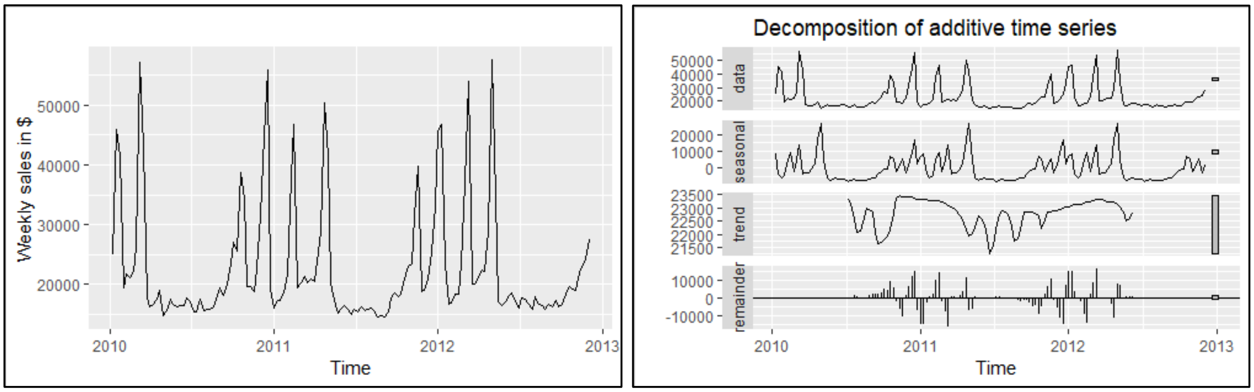
Dataset 1 from kaggle.com which is univariate daily data from multiple stores and multiples items, however actual location and store names are kept anonymous. This data set has daily data from 2013 to 2018. With seasonality, trend and noise shown in **Error! Reference source not found.**2. For store 2 item 2



(a) (b)

Figure 2. Figures (a) and (b) represent the Time series plot for Store 2 item 1

Dataset 2 we have used is multivariate data with weekly frequency also https://www.kaggle.com/gcarra/walmart with following trend and seasonality as shown in Figure . For a specific store and item



(a) (b)

Figure 3. Figures (a) and (b) represent Time series plot with Trend and Seasonality for Weekly data

Both datasets are separated into training and test sets. The models were implemented on the test set to compare of RMSE value between forecasted vs observed.

2.1 Frequentist Approach using Markov Chains.

Markov chain is a concept studied in probability and statistics that defines the events and its outcomes in states and predict the probability of event transitioning from one state to another. The transition probability can be defined as $P(X_t = x_t | X_{t-1} = x_{t-1})$. This is an important property of Markov chain that probability of transition or hoping to the current state is only dependent on the previous state and not on other past values. Similarly, the probability of transitioning into the next state only depends on the current state and not the past state. In term of sales, we can define this probability as:

$P(X_t = sales_t \ | sales_{t-1}) \ for \ sales_t, sales_{t-1}, \ldots \in S.$

$where \ S \ is \ state \ space \ which \ will \ define \ the \ range \ of \ sales \ i.e. sales \ will \ be \ between \ a \ range \ of \ values \ for \ example \ sales \ can \ be \ between \ (5,10], \ldots, (55,60] \ in \ terms \ of \ units \ or \ can \ be \ in \ terms \ of \ dollars.$

Thus by defining a specific range to sales, a long sequence of such range is generated based on the data and counting how specific sales range is followed by other ranges(states) for every state and we can predict the outcome of the next day. The transition matrix obtained by this method is shown in Figure 4.

A 11 - dimensional discrete Markov Chain defined by the following states:
(10,15], (15,20], (20,25], (25,30], (30,35], (35,40], (40,45], (45,50], (5,10], (50,55], (55,60]
The transition matrix (by rows) is defined as follows:

	(10,15]	(15,20]	(20,25]	(25,30]	(30,35]	(35,40]	(40,45]	(45,50]	(5,10]	(50,55]	(55,60]
(10,15]	0.23750000	0.37500000	0.26250000	0.0875000	0.00000000	0.012500000	0.000000000	0.000000000	0.02500000	0.000000000	0.00000000
(15,20]	0.14009662	0.21739130	0.37198068	0.1884058	0.04830918	0.009661836	0.004830918	0.000000000	0.01932367	0.000000000	0.00000000
(20,25]	0.03932584	0.17696629	0.28089888	0.2387640	0.14606742	0.064606742	0.025280899	0.002808989	0.02528090	0.000000000	0.00000000
(25,30]	0.02476780	0.10526316	0.24458204	0.2755418	0.19504644	0.092879257	0.055727554	0.006191950	0.00000000	0.000000000	0.00000000
(30,35]	0.01687764	0.07172996	0.15189873	0.2700422	0.21097046	0.168776371	0.088607595	0.012658228	0.00000000	0.008438819	0.00000000
(35,40]	0.00000000	0.07971014	0.21014493	0.1449275	0.23188406	0.188405797	0.094202899	0.043478261	0.00000000	0.007246377	0.00000000
(40,45]	0.00000000	0.03947368	0.13157895	0.1710526	0.26315789	0.157894737	0.118421053	0.092105263	0.00000000	0.013157895	0.01315789
(45,50]	0.00000000	0.00000000	0.04761905	0.1904762	0.33333333	0.142857143	0.142857143	0.095238095	0.00000000	0.047619048	0.00000000
(5,10]	0.31250000	0.25000000	0.25000000	0.1250000	0.00000000	0.000000000	0.000000000	0.000000000	0.06250000	0.000000000	0.00000000
(50,55]	0.00000000	0.00000000	0.00000000	0.0000000	0.60000000	0.000000000	0.400000000	0.000000000	0.00000000	0.000000000	0.00000000
(55,60]	0.00000000	0.00000000	0.00000000	0.0000000	0.00000000	1.000000000	0.000000000	0.000000000	0.00000000	0.000000000	0.00000000

Figure 4. Transition Matrix obtained by daily sales data

Thus from this matrix, if your sales today are in range (5,10] then the probability of going to state (10,15] is 31.25%. But the problem with this method is the transition probabilities are not high enough to give confidence to the user.

Thus, aggregating the daily data into weekly was sensible. After aggregating the data of daily sale into weekly by taking the sum of daily data. As shown in Figure 5 aggregating the data

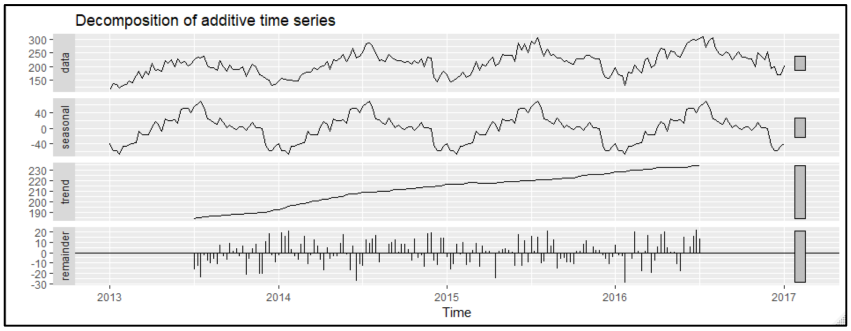


Figure 5. Effect of Aggregation on daily sales data.

Creating the Markov transition probability matrix yielded so better results when used on sales data is shown in Figure 6.

```
MLE Fit
 A  6 - dimensional discrete Markov Chain defined by the following states:
 (100,140], (140,180], (180,220], (220,260], (260,300], (300,340]
 The transition matrix  (by rows)  is defined as follows:
           (100,140]  (140,180]  (180,220]  (220,260]  (260,300] (300,340]
(100,140] 0.66666667 0.22222222 0.11111111 0.00000000 0.00000000 0.0000000
(140,180] 0.04347826 0.63043478 0.30434783 0.02173913 0.00000000 0.0000000
(180,220] 0.00000000 0.19444444 0.51388889 0.26388889 0.02777778 0.0000000
(220,260] 0.00000000 0.02325581 0.22093023 0.61627907 0.13953488 0.0000000
(260,300] 0.00000000 0.00000000 0.02631579 0.34210526 0.50000000 0.1315789
(300,340] 0.00000000 0.00000000 0.00000000 0.00000000 0.71428571 0.2857143
```

Figure 6. Transition Probability Matrix for Weekly aggregation

Thus, looking at the transition matrix knowing the present sales for the week is say (300,340] we can say with 71% confidence that next week the sales will be between (260,300] or by averaging $\frac{260+300}{2} = 280\sim approx.$(Giorgio Alfredo Spedicato)

2.2 Limitations and Advantages

For a small scale organization or bulk service provider with constant demand, this method can be useful or even with the slow trend by simply adding new values every week and can update their transition probability and creating a transition diagram as shown in Figure 7(a) and Figure 7(b). Another advantage of this method is keeping the matrix updated with new data. The transition probabilities get more and more robust every time and can now-cast better for the next week or any other required time frame defined by the user. We can also calculate the next week transition probabilities or further by simply raising the power of the matrix $P^n = \begin{bmatrix} P_{00} & \cdots & P_{0n} \\ \vdots & \ddots & \vdots \\ P_{n0} & \cdots & P_{nn} \end{bmatrix}^n$

A disadvantage of this method is it provides values of forecast within a specific range and not the actual value. If the trend has bigger slope new transition states will have to be added constantly to incorporate them in the transition matrix which can decrease the probability of transitioning as new states will be added.

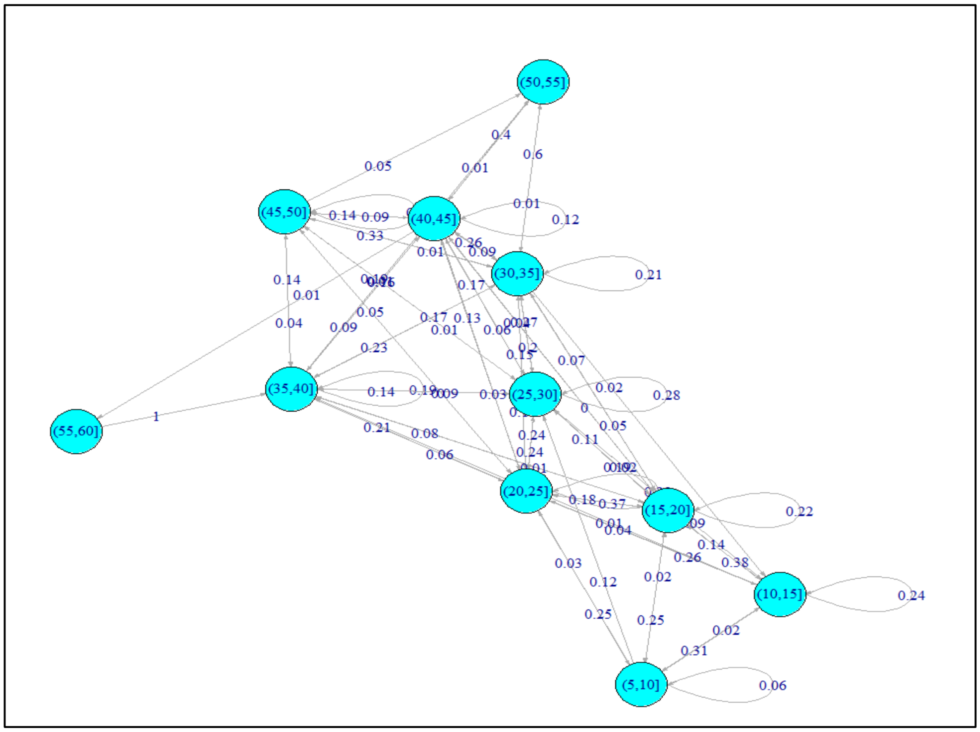


Figure 7 (a). Markov Chain Transition Diagrams for Daily data.

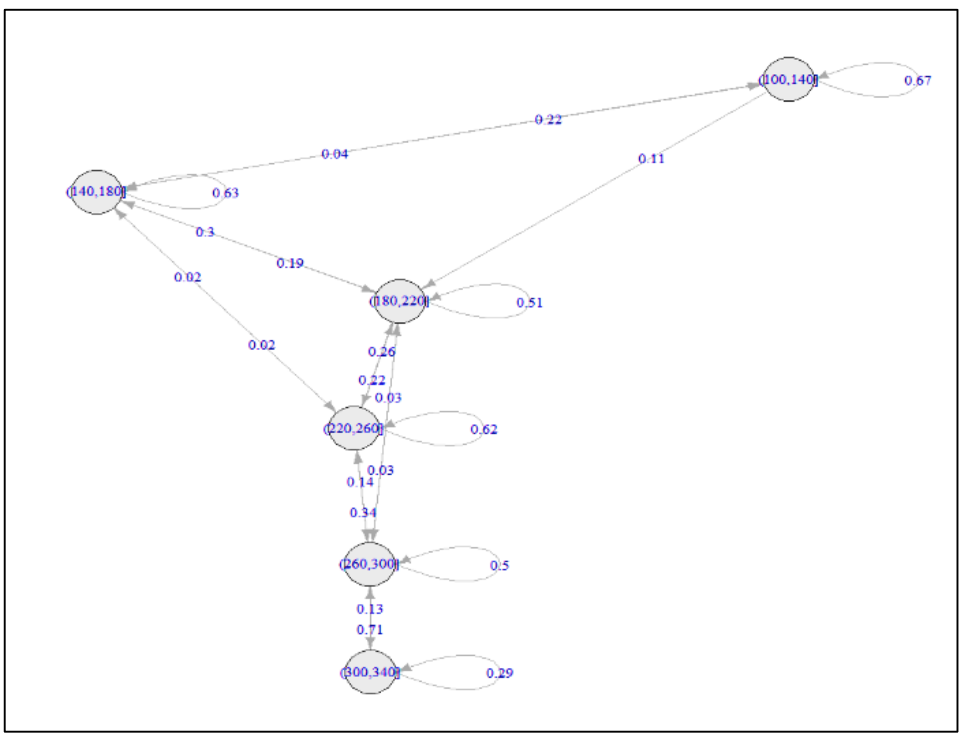
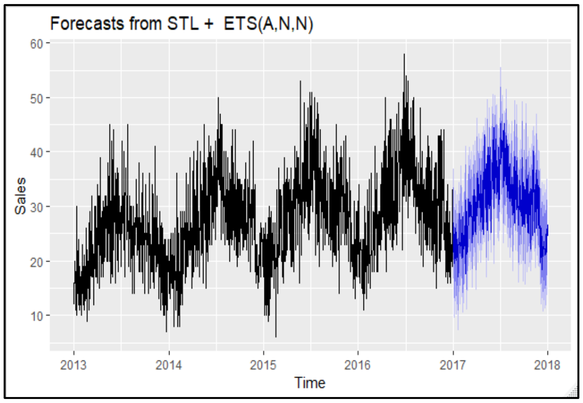


Figure 7 (b). Markov Chain Transition Diagrams for Aggregated data.

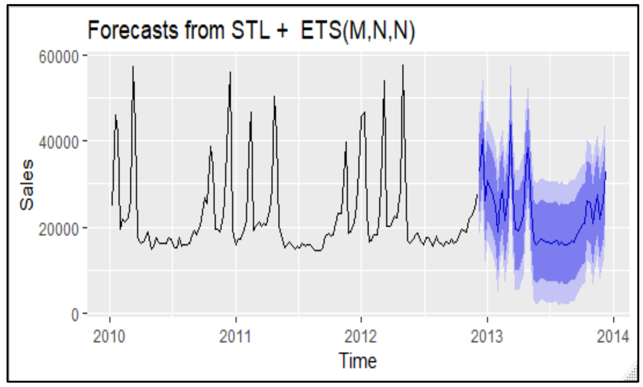
3. Error Trend Seasonality(ETS) Models.

The ETS model developed using *'forecast'* package in R software as per (Hyndman 2018) implemented on univariate Walmart data we obtained following results the models selected was ANN which is simple exponential smoothing with additive errors with seasonal adjustments.




```
                    ME     RMSE      MAE       MPE
Training set 0.2700754 5.152951 4.082466 -3.067862
                 MAPE      MASE       ACF1
Training set 16.42218 0.5942865 0.02788956
```

(a)




```
                    ME     RMSE      MAE       MPE
Training set -5.258838 7107.701 4723.962 -6.358619
                 MAPE      MASE      ACF1
Training set 19.13393 0.5762592 0.4946025
```

(b)

Figure 8.2 Figure (a) and (b) shows the forecast obtained implementing ETS model

3.1 Bayesian Structural Time Series

This approach revolves around Bayesian statistics. The three components of Bayesian Statistics namely:
1) Prior Distribution bases on prior knowledge of data assumed normal in *bsts* package.
2) Likelihood Function developed on data or (evidence) sequence of time series we provide.
3) Posterior Distribution of the each and every point and the mean of the distribution defined as forecast.

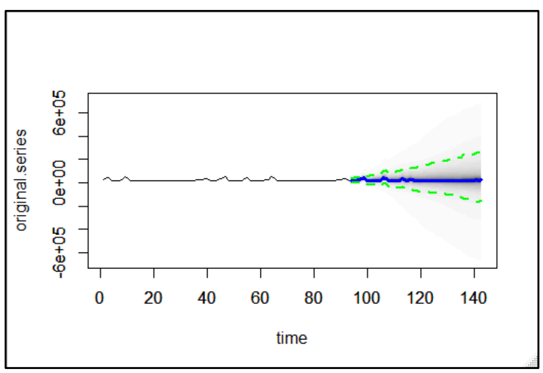


Figure 9. Predictions using BSTS model for Data 2

We can write Bayesian structural model as (Larsen 2016) :
$Y_t = \mu_t + x_t\beta + S_t + e_t, e_t \sim N(0, \sigma_e^2)$
$\mu_{t+1} = \mu_t + v_t, v_t \sim N(0, \sigma_v^2)$
$x_t = univariate\ data$
$S_t = seasonality\ , \mu_t = \ local\ level/trend$

We add local level trend, seasonality and number of seasons and seasonality duration to develop a likelihood function as state specification further passed to bsts algorithm. MCMC iterations are performed on this state specification. Another component of Bayesian Structural Time Series is MCMC iterations known as Markov Chain Monte Carlo iterations are used for probability distribution sampling. We can create a required distribution by observing the chain of our parameter and its sequence along time. Basically, MCMC iterations are performed on the number of samples repeatedly and gradually increasing the number of samples and bringing the desired output as close as the distribution of the output or forecasted parameter. We obtain a posterior distribution of prediction with 2 values of a range taking the mean of the prediction we obtain the final forecast.

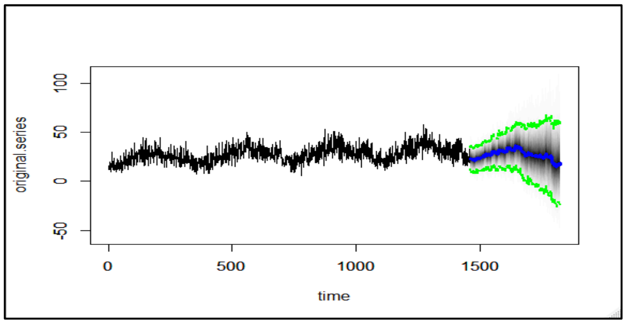


Figure 30. Prediction of dataset 1

The RMSE value for the univariate Walmart data was obtained 4.41, compared to ETS's error which was 5.152.

Conclusion

We have presented the use of Bayesian statistics and Time series forecasting in the field of the supply chain using short term forecasting stochastic methods viz. frequentist approach - Markov chain and ETS model - Bayesian structural time series. It is observed that a Markov chain with a countably infinite state space can be stationary which means that the process can converge to a steady state. They are used in a broad variety of fields, including the supply chain. Markov chain can be used when the process satisfies the Markov property viz. the future state of a stochastic variable is only dependent on its present state. This means that knowing the previous history of the process will not improve future predictions. By analyzing the historical data of a market, it is possible to distinguish certain patterns in its past movements. From these patterns, Markov diagrams can then be formed and used to predict future market trends as well as the risks associated with them. In case of the ETS, the taxonomy is based on characterizing each model against error, trend and seasonality. Bayesian structural time series model is one of the ETS models that applies probabilities to statistical problems. It provides tools to update people's beliefs in the evidence of new data and concludes that forecasting by increasing likelihood of an accurate forecast is sensible when compared to reducing error. Using the bsts function in R, the RMSE of the forecast was obtained as 4.41.

There are much more efficient ways to forecast the sale of a company. The future scope of this project is to implement the hidden Markov model which the system is modeled using unobservable states in Markov chain. We would also like to improve the Markov chain model by adding more memory and incorporating trend function.

References

C. Catal, K. E., B. Arslan and A. Akbulut (2019). "Benchmarking of Regression and Time Series Analysis Techniques for Sales Forecasting." Balkan Journal of Electrical & Computer Engineering, **7**(1).

Denis Tikunov, T. N. (2004). "Time series Forecasting using Holt-Winters Exponential Smoothing." Kanwal Rekhi School of Information Technology.

Dorairaj, S. (2018). "Hidden Markov Models Simplified." from https://medium.com/@postsanjay/hidden-markov-models-simplified-c3f58728caab.

Giorgio Alfredo Spedicato, T. S. K., Sai Bhargav Yalamanchi, Deepak Yadav. "The markovchain Package: A Package for Easily

Handling Discrete Markov Chains in R." from https://cran.r-project.org/web/packages/markovchain/vignettes/an_introduction_to_markovchain_package.pdf.

Hyndman, R. J., & Athanasopoulos, G. (2018). "Forecasting: principles and practice, 2nd edition." Retrieved 3/5/2019, 2019, from https://otexts.com/fpp2/.

Kalekar, P. S. (2004). "Time series Forecasting using Holt-Winters Exponential Smoothing." Kanwal Rekhi School of Information Technology.

Larsen, K. (2016) Sorry ARIMA, but I'm Going Bayesian.

OCZKI, J. (2014). "Forecasting Internal Labour Supply with a Use of Markov Chain Analysis." International Journal of Knowledge, Innovation and Entrepreneurship **Volume 2 No. 2**: 11.

Pegels, C. C. (1969). "Exponential forecasting: Some new variations." Management Science **15(5)**: 311-315.

SCOTT, S. L. (2017). "Fitting Bayesian structural time series with the bsts R package ". from http://www.unofficialgoogledatascience.com/2017/07/fitting-bayesian-structural-time-series.html.

Sirajudeen, A. (2017). "How to Predict Sales Using Markov Chain." from https://blog.arkieva.com/markov-chain-sales-prediction/.

Zhihang Peng, C. B., Yang Zhao, Honggang Yi, Letian Xia, Hao Yu, Hongbing Shen, Feng Chen (2010). "Weighted Markov chains for forecasting and analysis in Incidence of infectious diseases in Jiangsu Province, China." **24**: 207-214.

https://cran.r-project.org/web/packages/bsts/bsts.pdf

Biographies

Shivam M Dave is a master's student at University at Buffalo, New York, USA currently pursuing Industrial and System Engineering. With an internship experience in field of production planning and control and courses like Advance Data, Transportation Analytics and Applied Stochastic Process, has successfully implemented the process of forecasting in the field of Supply chain engineering. He has interest in studying and implementing the process in field of Supply chain, Transportation and Quality Manufacturing.

Aditya M Patel is a student pursuing master of science in Industrial and Systems engineering at University at Buffalo (State University at New York). In this project, he has implemented concepts learned from subjects such as Statistical Data Mining, Programming for Analysis, Advanced Data Analytics and Predictive Modeling and Supply Chain Engineering

Forum D Bhutaiya is a student pursuing masters of science in Industrial Engineering at University at Buffalo, New York, USA. She earned Bachelors in Engineering degree in Mechanical Engineering from Mumbai University, India. She has published conference papers. She is a member of IEEE, SME and INFORMS.

Mohamed Awwad is a Teaching Assistant Professor in the Department of Industrial and Systems Engineering at the University at Buffalo, The State University of New York, Buffalo, NY, USA. He received his Ph.D. and M.S. degrees in Industrial Engineering from the University of Central Florida, Orlando, FL, USA. Additionally, he holds M.S. and B.S. degrees in Mechanical Engineering from Cairo University, Egypt. Before his tenure with the University at Buffalo, Dr. Awwad held several teaching and research positions at the University of Missouri, Florida Polytechnic University, and the University of Central Florida. His research interests include applying operations research methods in the fields of logistics & supply chain, distribution center design, unconventional logistics systems design, healthcare and the military. He is a member of IISE, INFORMS and ASME.