The Application of Machine Learning Methods to Time Series Forecasting

Improving Forecasting Techniques for Smart City Planning

Authors:

Manuel Alexander Schreiber
(676991)

Conor Hasselgard Cavanaugh
(82502)

Supervisor: Prof. Lisbeth La Cour, PhD
(Department of Economics)

Co-Supervisor: Prof. Raghava Rao Mukkamala, PhD
(Department of Digitalization
and Centre for Business Data Analytics)

A thesis presented for the degrees of a

Master of Science (M.Sc.)
in Advanced Economics and Finance

Cand.Merc.Mat (M.Sc.)
in Erhvervsøkonomi og Matematik

Number of pages (characters):
Copenhagen Business School
Copenhagen, June 30, 2020

Abstract

Contents

1 Introduction **6**

1.1 Preliminary Overview of Relevant Literature 6

1.2 Planned Methodology and Research Design 6

2 Research Questions **8**

3 Smart City **9**

4 Forecasting Methods **10**

4.1 Traditional Statistical models 10

4.1.1 ARIMA models 10

4.1.2 Exponential Smoothing models 10

4.1.3 Holt-Winters models 10

4.2 Machine Learning Methods 10

4.2.1 Bagging for Time Series 10

4.2.2 Tree-based models for Time Series 12

4.2.3 Neural Network Models 12

5 Data **13**

6 Analyses and Results **14**

7 Discussion **15**

8 Conclusion **16**

A Appendix **17**

List of Figures

4.1 Bootstrapping procedure for a time series (adapted from Petropoulos, Hyndman, and Bergmeir (2018)) 11
4.2 Dynamic ensemble model (adapted from Galicia, Talavera-Llames, Troncoso, Koprinska, and Martınez-Álvarez (2019)) 12

List of Tables

Acknowledgements

I want to thank my supervisor/parents/classmates...

Chapter 1

Introduction

Machine learning methods have been developed and refined to the point that they now pose a serious challenge to classic statistical models in the area of forecasting. For example, Ahmed, Atiya, Gayar, and El-Shishiny (2010) came up with a large-scale comparison study based on the M3 competition dataset to compare the major machine learning models for time series forecasting. Exponential smoothing models are among the most frequently used forecasting methods. However, Bergmeir, Hyndman, and Benıtez (2016) argue that these models can be substantially outperformed for the purpose of forecasting by applying the bagging algorithm to a bootstrapped component of the initial time series to estimate an ensemble of exponential smoothing models and by combining the resulting point forecasts.

1.1 Preliminary Overview of Relevant Literature

Januschowski et al. (2020) provide an overview of the spectrum of Machine learning and statistical methods and the boundaries and intersections in the context of forecasting. This distinction is a common way of classifying forecasting methods in the forecasting literature. However, the authors contend that there is no added value from such a classification of the methods and they encourage the scientific communities to collaborate more. Maasoumi and Medeiros (2010) present applications in Economics and Finance at the intersection of Statistical Learning and Econometrics. Inter alia, the authors analyze bagging and combination forecasts and find that bagging forecasts often deliver the lowest mean squared forecast errors.

1.2 Planned Methodology and Research Design

As suggested by Bergmeir et al. (2016), a Box-Cox transformation can be applied to the series in order to stabilize the variance of the time series and subsequently an STL decomposition is employed to break the time series down into the trend, the seasonal part and the remainder. Bootstrapping the remainder of the series is done to achieve stationarity. Furthermore, Galicia, Talavera-Llames, Troncoso, Koprinska, and Martınez-Álvarez (2019) suggest the use of an ensemble time series forecasting model consisting of three components, which are decision trees, gradient boosted trees and Random Forests. For this model, the ensemble weights are computed by weighted least squares. They found that the ensemble outperforms its individual components. The three ensemble components can thus be employed for the purpose of this paper.

Chapter 2

Research Questions

This paper is concerned about the following reseach questions:

(Q1) How can forecasting methods be employed to improve city infrastructure planning in a smart city?

(Q2) What are the best forecasting models for the three Smart City usecases traffic, healthcare and security?

(Q3) Can the forecasting performance of traditional statistical models for time series forecasting be outperformed or complemented by Machine Learning methods?

Chapter 3

Smart City

Chapter 4

Forecasting Methods

4.1 Traditional Statistical models

4.1.1 ARIMA models

4.1.2 Exponential Smoothing models

4.1.3 Holt-Winters models

4.2 Machine Learning Methods

4.2.1 Bagging for Time Series

According to Hastie, Tibshirani, and Friedman (2009) the bagging estimate is defined by

$$\hat{f}_{bag}(x) = \frac{1}{B}\sum_{b=1}^{B}\hat{f}^{*b}(x) \tag{4.1}$$

The bagging estimate is obtained by averaging the predictions $\hat{f}^{*b}(x)$ for $b = 1, .., B$ from the statistical learning models fitted to a collection of B bootstrap samples obtained from the single training data set via taking repeated samples with replacement, i.e. bootstrapping.
Bagging is therefore a general-purpose technique for reducing the variance of a statistical learning method, which also increases that method's prediction accuracy (James, Witten, Hastie, & Tibshirani, 2013).

The well-established bagging method was first introduced by Breiman (1996) but it was not successfully applied in a time series forecasting context until 2016.
According to Petropoulos, Hyndman, and Bergmeir (2018) the complication with time series consists in accounting for non-stationarity and autocorrelation in the bootstrapping procedure in order to produce bootstrapped samples that resemble the original data.
Bergmeir et al. (2016) propose a bootstrapping procedure for time series as illustrated in Figure 4.1 that includes a Box-Cox transformation in order to stabilize the variance and a decomposition either in form of the loess method to extract trend and remainder in case of a non-seasonal time series or an STL decomposition in

order to break the series down into the trend, seasonal and remainder components. The Box-Cox transformation, which was first introduced by Box and Cox (1964), is defined as follows:

$$w_t = \begin{cases} log(y_t), & \lambda = 0 \\ (y_t^\lambda - 1)/\lambda, & \lambda \neq 0 \end{cases} \tag{4.2}$$

The optimal $\lambda \in [0, 1]$ is chosen by dividing the series into subseries of length equal to the seasonality and minimzing the coefficient of variation $s/m^{(1-\lambda)}$ across those subseries, where s stands for the standard deviation and m for the sample mean of the subseries. The authors then apply a bootstrapping method that allows for autocorrelation, the moving block bootstrap (MBB) as suggested by Künsch (1989), to the extracted remainder of the series. In the next step, the series is reconstructed from its structural components, i.e. trend, seasonality, and the bootstrapped remainder. Finally, the Box-Cox transformation is inverted. The whole process is then repeated in order to obtain the bootstrapped series.

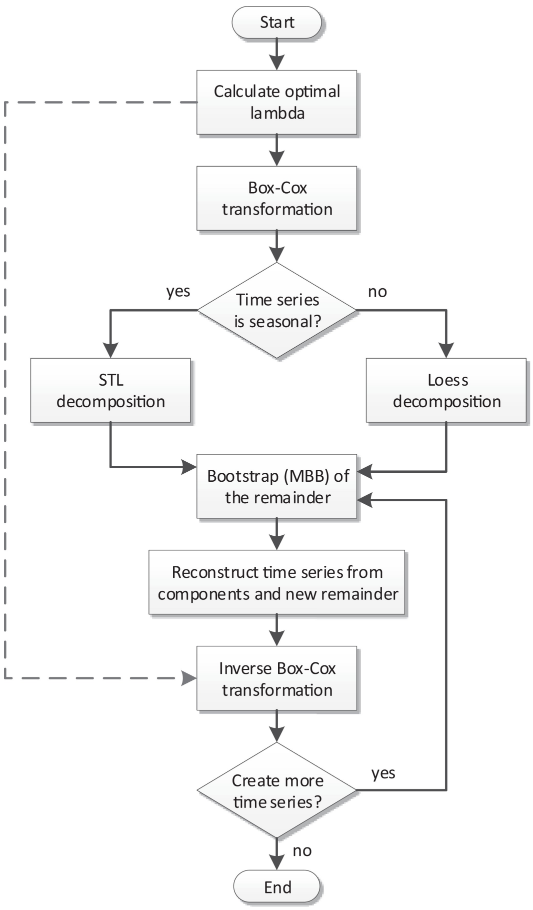

Figure 4.1: Bootstrapping procedure for a time series (adapted from Petropoulos, Hyndman, and Bergmeir (2018))

Bergmeir et al. (2016) find that the ensemble of bagged exponential smoothing models outperforms the regular exponential smoothing model consistently for monthly data on the M3 forecasting competition dataset, which is a common medium of comparison of newly introduced forecasting methods with existing state of the art models.

4.2.2 Tree-based models for Time Series

Galicia et al. (2019) propose a dynamic ensemble model for big data time series forecasting purposes as illustrated in Figure 4.2 based on the three base models decision trees, random forests and gradient boosted trees. The ensemble weights are computed by weighted least squares assigning higher weights to more accurate ensemble members according to their past performance.

The authors found that the ensemble model outperformed the individual base models on a high sampling frequency dataset for the Spanish electricity market in terms of prediction accuracy as evidenced by lower mean absolute errors (MAE) and root mean squared errors (RMSE). Moreover, they found that the dynamic ensemble model could outperform Artificial Neural Network (ANN) and deep learning (DL) algorithms when evaluating forecast errors by yielding the lowest MAE and RMSE values on an Australia solar power dataset.

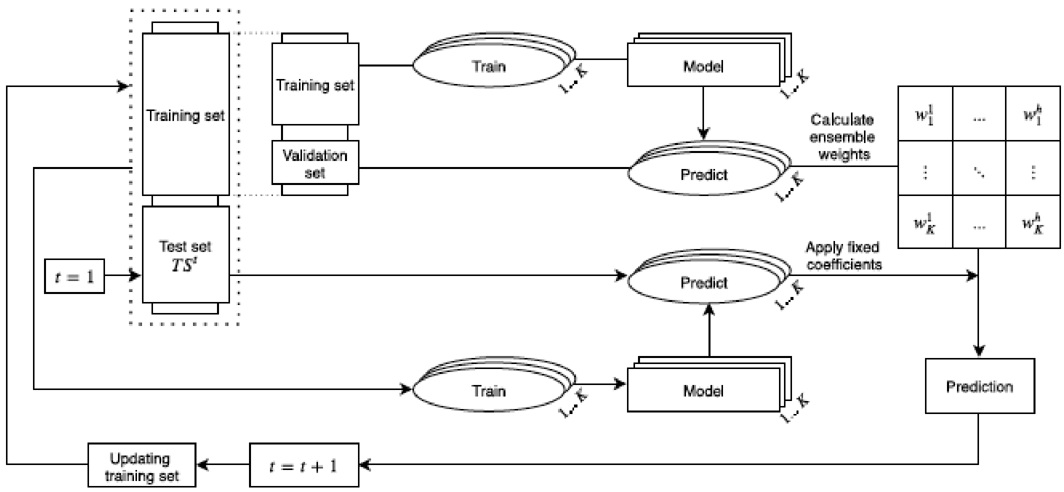

Figure 4.2: Dynamic ensemble model (adapted from Galicia, Talavera-Llames, Troncoso, Koprinska, and Martınez-Álvarez (2019))

4.2.3 Neural Network Models

Chapter 5

Data

Chapter 6

Analyses and Results

Chapter 7

Discussion

Chapter 8

Conclusion

Appendix A

Appendix

References

Ahmed, N. K., Atiya, A. F., Gayar, N. E., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, *29*(5-6), 594–621.

Bergmeir, C., Hyndman, R. J., & Benıtez, J. M. (2016). Bagging exponential smoothing methods using stl decomposition and box–cox transformation. *International journal of forecasting*, *32*(2), 303–312.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211–243.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*(2), 123–140.

Galicia, A., Talavera-Llames, R., Troncoso, A., Koprinska, I., & Martınez-Álvarez, F. (2019). Multi-step forecasting for big data time series based on ensemble learning. *Knowledge-Based Systems*, *163*, 830–841.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Science & Business Media.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.

Januschowski, T., Gasthaus, J., Wang, Y., Salinas, D., Flunkert, V., Bohlke-Schneider, M., & Callot, L. (2020). Criteria for classifying forecasting methods. *International Journal of Forecasting*, *36*(1), 167–177.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The annals of Statistics*, 1217–1241.

Maasoumi, E., & Medeiros, M. C. (2010). The link between statistical learning theory and econometrics: Applications in economics, finance, and marketing. *Econometric Reviews*, *29*(5-6), 470–475.

Petropoulos, F., Hyndman, R. J., & Bergmeir, C. (2018). Exploring the sources of uncertainty: Why does bagging for time series forecasting work? *European Journal of Operational Research*, *268*(2), 545–554.