Pontifícia Universidade Católica
do Rio de Janeiro

Erick Meira de Oliveira

Getting the most out of the wisdom of the crowds: improving forecasting performance through ensemble methods and variable selection techniques

Tese de Doutorado

Thesis presented to the Programa de Pós-Graduação em Engenharia de Produção of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Engenharia de Produção.

Advisor: Prof. Fernando Luiz Cyrino Oliveira

Rio de Janeiro
February 2020

Pontifícia Universidade Católica
do Rio de Janeiro

Erick Meira de Oliveira

Getting the most out of the wisdom of the crowds: improving forecasting performance through ensemble methods and variable selection techniques

Thesis presented to the Programa de Pós-Graduação em **Engenharia de Produção** of PUC-Rio in partial fulfillment of the requirements for the degree of **Doutor em Engenharia de Produção**. Approved by the Examination Committee.

Prof. Fernando Luiz Cyrino Oliveira
Advisor
Departamento de Engenharia Industrial – PUC-Rio

Prof. Lilian Manoel de Menezes Willenbockel
University of London - UL

Prof. Reinaldo Castro Souza
Departamento de Engenharia Industrial – PUC-Rio

Prof. Hélio Côrtes Vieira Lopes
Departamento de Informática – PUC-Rio

Prof. José Francisco Moreira Pessanha
CEPEL

Prof. Marcelo Cabus Klotzle
IAG Escola de Negócios – PUC-Rio

Rio de Janeiro, February 6th, 2020

Erick Meira de Oliveira

Erick Meira holds a B. Eng. (*cum laude*) in Oil Engineering from the Federal University of Rio de Janeiro (UFRJ) and an MSc degree in Industrial Engineering from the Pontifical Catholic University of Rio de Janeiro (PUC-Rio).

He is also a permanent employee of the Brazilian Agency for Research and Innovation (Finep), a federal, state-owned company linked to the Ministry of Science, Technology, Innovations and Communications (MCTIC).

Bibliographic Data

de Oliveira, Erick Meira

Getting the most out of the wisdom of the crowds: improving forecasting performance through ensemble methods and variable selection techniques / Erick Meira de Oliveira; advisor: Fernando Luiz Cyrino Oliveira. – 2020.

113 f. : il. color. ; 30 cm

Tese (doutorado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Industrial, 2020.

Inclui bibliografia.

1. Engenharia Industrial – Teses. 2. Previsão. 3. Séries temporais. 4. Seleção de modelos. 5. Combinação de previsões. 6. Métodos de ensemble. I. Oliveira, Fernando Luiz Cyrino. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Industrial. III. Título.

CDD: 658.5

Acknowledgments

First and foremost, I thank the Almighty God for giving me the wisdom, strength and power to persevere throughout this journey. Without His blessings, none of this would have been possible.

I am eternally grateful to my family for their undying support and unconditional love. Thank you for believing in me.

I feel great pleasure in expressing my profound gratitude and venerable regards to my advisor, Prof. Fernando Luiz Cyrino Oliveira, for his meticulous guidance, scientific supervision and constant encouragement during my research. I am also deeply indebted to all my beloved teachers and collaborators from other universities for their co-operation and timely helps. Special thanks in this regard are due to Prof. Marcelo Klotzle, Prof. Lilian de Menezes, Dr. Jooyoung Jeon and Dr. Fotios Petropoulos.

There are also many colleagues who have been supportive and helpful during my PhD period, and I am very grateful for that. There are some people I would like to thank in particular: Felipe Fogliano, Katie Halet and Lucas Bastos. I also appreciate the support of the university staff throughout the course of my PhD.

My warm thanks go to Dr. André Assis de Salles, my undergraduate supervisor, who played an important role in my academic growth and professional development.

Lastly, I gratefully acknowledge the support provided by Finep's Graduate Incentive Program (PIPG) during the second half of my doctoral studies (partial leave of absence) and during my stay as a Visiting Postgraduate Scholar (VPS) at the School of Management, University of Bath (UK) (full leave of absence).

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Abstract

De Oliveira, Erick Meira; Cyrino Oliveira, Fernando Luiz (Advisor). **Getting the most out of the wisdom of the crowds: improving forecasting performance through ensemble methods and variable selection techniques.** Rio de Janeiro, 2020. 113p. Tese de Doutorado – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

This research focuses on the development of hybrid approaches that combine ensemble-based supervised machine learning techniques and time series methods to obtain accurate forecasts for a wide range of variables and processes. It also includes the development of smart selection heuristics, i.e., procedures that can select, among the pool of forecasts originated via ensemble methods, those with the greatest potential of delivering accurate forecasts after aggregation. Such combinatorial approaches allow the forecasting practitioner to deal with different stylized facts that may be present in time series, such as nonlinearities, stochastic components, heteroscedasticity, structural breaks, among others, and deliver satisfactory forecasting results, outperforming benchmarks on many occasions.

The thesis is divided into a series of essays. The first endeavor proposed an alternative method to generate ensemble forecasts which delivered satisfactory forecasting results for certain types of electricity consumption time series. In a second effort, a novel forecasting approach combining Bootstrap aggregating (Bagging) algorithms, time series methods and regularization techniques was introduced to obtain accurate forecasts of natural gas consumption and energy supplied series across different countries. A new variant of Bagging, in which the set of classifiers is built by means of a Maximum Entropy Bootstrap routine, was also put forth. The third contribution brought a series of innovations to model selection and model combination in forecasting routines. Gains in accuracy for both point forecasts and prediction intervals were demonstrated by means of an extensive empirical experiment conducted on a wide range of series from the M- Competitions.

Keywords

Forecasting; Time Series; Model Selection; Forecast combinations; Ensemble methods; Bagging; Regularization techniques

Resumo

De Oliveira, Erick Meira; Cyrino Oliveira, Fernando Luiz (Orientador). **Tirando o máximo proveito da sabedoria das massas: aprimorando previsões por meio de métodos de ensemble e técnicas de seleção de variáveis**. Rio de Janeiro, 2020. 113p. Tese de Doutorado – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

A presente pesquisa tem como foco o desenvolvimento de abordagens híbridas que combinam algoritmos de aprendizado de máquina baseados em conjuntos (*ensembles*) e técnicas de modelagem e previsão de séries temporais. A pesquisa também inclui o desenvolvimento de heurísticas inteligentes de seleção, isto é, procedimentos capazes de selecionar, dentre o *pool* de preditores originados por meio dos métodos de conjunto, aqueles com os maiores potenciais de originar previsões agregadas mais acuradas. A agregação de funcionalidades de diferentes métodos visa à obtenção de previsões mais acuradas sobre o comportamento de uma vasta gama de eventos/séries temporais.

A tese está dividida em uma sequência de ensaios. Como primeiro esforço, propôs-se um método alternativo de geração de conjunto de previsões, o que resultou em previsões satisfatórias para certos tipos de séries temporais de consumo de energia elétrica. A segunda iniciativa consistiu na proposição de uma nova abordagem de previsão combinando algoritmos de *Bootstrap Aggregation* (*Bagging*) e técnicas de regularização para se obter previsões acuradas de consumo de gás natural e de abastecimento de energia em diferentes países. Uma nova variante de *Bagging*, na qual a construção do conjunto de classificadores é feita por meio de uma reamostragem de máxima entropia, também foi proposta. A terceira contribuição trouxe uma série de inovações na maneira pela qual são conduzidas as rotinas de seleção e combinação de modelos de previsão. Os ganhos em acurácia oriundos dos procedimentos propostos são demonstrados por meio de um experimento extensivo utilizando séries das Competições M1, M3 e M4.

Palavras-chave

Previsão; Séries Temporais; Seleção de modelos; Combinação de previsões; Métodos de ensemble; Bagging; Técnicas de regularização

Contents

List of Figures

List of Tables

1
Introduction

In light of the rapid economic development and to respond to an ever-growing competitive environment, businesses and organizations have become increasingly complex. As a result, decision makers find it increasingly more difficult to weigh all the factors in a given situation without some explicit, systematic support.

A major concern that is common to most decision-making circumstances is the uncertainty of future outcomes. Every day, corporate leaders, planners and policymakers are faced with the challenge of making decisions without knowing what will happen in the future. For instance, inventory is ordered without certainty as to what sales will be; new equipment is purchased despite uncertainty about demand for products, and investments are made without knowing what profits will be. In this context, the availability of tools which can correctly recognize emerging changes in the business environment and accurately predict future ones has become a key factor for effectively planning and eventually succeeding in business. This brings forecasting methods to the forefront of management practice in organizations.

A general approach to forecasting is the use of quantitative methods, particularly the ones which resort to time series information (data collected at regular intervals over time). In this connection, a wide range of quantitative forecasting methods have been proposed throughout the last decades – see DE GOOIJER & HYNDMAN (2006) for a comprehensive review.

Contemporary evidence from international forecasting competitions points toward the use of combinatorial approaches as the state-of-the-art in forecasting time series, particularly for long time horizons – see, for instance, the results from the M4 competition (MAKRIDAKIS et al., 2018, 2020) and the Global Energy Forecasting Competition (GEFCom) (HONG et al., 2019). Furthermore, the use of

hybrid approaches, i.e. procedures that utilize both statistical and Machine Learning (ML) features, ranked the best in terms of accuracy in both competitions.

A particular class of machine learning algorithms that has received considerable attention in time series forecasting applications are the so-called ensemble methods. In brief terms, these supervised learning algorithms construct a set of classifiers and then classify new data points by taking a (weighted) vote of their predictions. That way, these methods are capable of including different forms of uncertainty which may be present when building a predictive model from data, namely data uncertainty, model uncertainty, and parameter uncertainty (PETROPOULOS et al., 2018).

In spite of their potential, a caveat common to all combinatorial approaches (including ensemble methods) is that they assume that the forecasts to be combined are reasonable. To overcome this, an additional step must be considered: the correct treatment of the sources of uncertainty. Surprisingly, this step has been mostly overlooked in the literature (PETROPOULOS et al., 2018; KOURENTZES et al., 2019).

All things considered, this thesis focuses on the development of hybrid strategies, combining ensemble-based machine learning algorithms (mainly Bootstrap Aggregating – Bagging routines), time series methods, and variable selection/weighting techniques to obtain accurate forecasts for a wide range of variables and processes. The underlying rationale is that by combining different sets of features from carefully selected models, one can substantially improve the performance (accuracy) of forecasting methods.

To demonstrate its contributions, this thesis was divided into a series of essays. The first contribution was the proposal of an alternative method to generate ensemble forecasts which delivered satisfactory forecasting results for certain types of electricity consumption time series (DE OLIVEIRA & CYRINO OLIVEIRA, 2018). In a second effort, a novel forecasting approach combining Bagging algorithms, time series methods and regularization techniques was introduced to obtain accurate forecasts of natural gas consumption and energy supplied time series across different countries. A new variant of Bagging, in which the set of classifiers is built by means of a Maximum Entropy Bootstrap routine, was also put forth. The third contribution brought a series of innovations to how model selection and model combination routines can be conducted. The gains in forecasting

accuracy for both point forecasts and prediction intervals were demonstrated by means of an extensive empirical experiment conducted on a wide range of series from the M- Competitions (98,830 in total). It should be highlighted that the methods/procedures presented in this third work were developed in partnership with field researchers from the School of Management, University of Bath, where the author stayed as a Visiting Postgraduate Scholar during the months of March to July 2019.

The rest of the thesis unfolds as follows. Chapter 2 provides a brief overview on the use of combinatorial approaches and ensemble-based methods for forecasting, with a special focus on *Bootstrap Aggregation* (*Bagging*) algorithms. Chapter 3 delves into the details of the most up-to-date techniques involving the use of Bagging for forecasting. It proceeds by proposing a framework for forecasting ensembles in which four main stages/tasks can be identified: (i) an (optional) data treatment or decomposition; (ii) resampling; (iii) forecasting; and (iv) combination. This framework provides a common ground for the discussion of the contributions comprising this thesis. The first alternative method to generate ensemble forecasts and its applications on electric energy consumption time series are presented in Chapter 4. Chapter 5 describes the new forecasting approach applied to natural gas consumption and energy supply forecasts. Chapter 6 introduces the concepts of treating and pruning and demonstrates how they can improve the accuracy of both point forecasts and prediction intervals in **any forecasting approach** involving model selection or combination. Finally, Chapter 7 summarizes the findings of the essays, emphasizing their main take-away messages, and indicates possible avenues for future research.

2
Literature Overview

2.1
Combining forecasts

The combination of different forecasting methods is a well-established procedure in the literature of time series forecasting. Since the seminal works of BATES & GRANGER (1969) and NEWBOLD & GRANGER (1974), there have been nearly five decades of research and empirical evidence in favor of forecast combination over the selection of a single forecasting model (CLEMEN & WINKLER, 1986; AKSU & GUNTER, 1992; MACDONALD & MARSH, 1994; DE MENEZES et al., 2000; ELLIOTT & TIMMERMANN, 2004; STOCK & WATSON, 2004; DEKKER et al., 2004; JOSE & WINKLER, 2008; GUIDOLIN & TIMMERMANN, 2009; ANDRAWIS et al., 2011; KOLASSA, 2011; KOURENTZES et al., 2014; AYE et al., 2015; ELLIOTT & TIMMERMANN, 2016; BARROW & KOURENTZES, 2016; KOURENTZES et al., 2019). Results from global forecasting competitions, such as the M, M-3 and M-4 Competitions (MAKRIDAKIS et al., 1982; MAKRIDAKIS & HIBON, 2000; MAKRIDAKIS et al., 2018) have also been virtually unanimous in concluding that combining multiple forecasts leads to increased forecast accuracy. In many cases one can make dramatic performance improvements by simply averaging the forecasts. The two key reported advantages are the reduction of forecast error variance and not having to rely on a single forecast method (CLEMEN & WINKLER, 1986; TIMMERMANN, 2006).

Even though empirical evidence suggests potential gains in accuracy when more than one model/method is taken into consideration when building the final

forecasts, there is still no consensus as to what the best approach to forecast combination is (DEBNATH & MOURSHED, 2018).

A crucial (and intuitive) point which it is often overlooked in the empirical literature of forecast combination is the proper treatment of the uncertainties associated with the identification of the "best model" (KOURENTZES et al., 2019). In most applied studies, the issue of forecast quality is subsumed in the task of using multiple alternative forecasting models or methods and picking the ones that are identified as most appropriate, given the data at hand. Even though each empirical study has its own merits (usually by being the first application of a particular combination of methods to a specific set of time series), the real additionality of such contributions is of little relevance to the state-of-the-art in time series forecasting methods.

In general, according to the taxonomy of BREIMAN (1996), three sources of uncertainty are present in time series forecasting: the one inherent to the information available (the available data sample); that related to model selection; and another one originating from the estimation of the involved parameters in each selected model. It is worth noting that these uncertainties are interlinked, in the sense that different sample sizes will result in different parameter estimates, which in turn may result in different model forms (KOURENTZES et al., 2019). Parameter estimation uncertainty may originate from the estimation algorithm and setup; for instance, different initial values may result in different estimates. Different model structures may impose specific restrictions in parameters, simplifying, or not, the estimation problem, and so on.

Recent works have sought to address the issue of uncertainty reduction by employing specific model selection metrics in the training or validation phases, such as the AKAIKE (1974) information criterion (AIC) – see, for instance, BILLAH et al. (2006) and KOLASSA (2011) – or by means of cross-validation techniques, as in FILDES & PETROPOULOS (2015) and BARROW & CRONE (2016). Naturally, the use of different criteria can lead to different selections of allegedly optimal forecasts. That way, such criteria are still subject to the above-mentioned sources of uncertainty. This is added to the fact that most selection criteria, particularly those which are based on likelihood values or one-step ahead in-sample fit, suffer from an additional limitation: implicitly they assume that the

postulated forecasting model is true. Otherwise, the likelihood function is not appropriate for any multi-step forecast that we require from the model (CHATFIELD, 2000; XIA & TONG, 2011). FILDES & PETROPOULOS (2015) provide empirical evidence of the disadvantage of one-step ahead forecast based selection criteria.

2.2
Hybrid, ensemble-based approaches to forecasting

A recently emerged strand of literature dedicated to combinatorial approaches for forecasting is the use of hybrid techniques, i.e., approaches that incorporate functionalities from traditional statistical methods and from machine learning techniques. Such approaches allow the forecasting practitioner to deal with different stylized facts that may be present in time series – such as nonlinearities, stochastic components, heteroscedasticity, structural breaks, among others – and, at the same time, deliver satisfactory forecasting results, outperforming benchmarks on several occasions.

A particular class of computationally intensive methods that has demonstrated satisfactory forecasting results when combined with classical time series methods are the so-called ensemble-based methods. They are based on the concept of *Decision committee learning*: "committee members" are applied to a classification/forecasting task and their individual outputs are combined to create a single classification/forecast from the committee as a whole (WEBB, 2000). Examples of such methods include: classification ensembles formed by stacked generalization (WOLPERT, 1992) or by stochastic search (ALI et al., 1994); NOCK & GASCUEL's (1995) decision committees; averaged decision trees (OLIVER & HAND, 1995); Bootstrap Aggregation (*Bagging*) (BREIMAN, 1996) algorithms; Weight Aggregation (*Wagging*) (BAUER & KOHAVI, 1999) routines; and the Boosting (FREUND, 1995) algorithm and its variants, such as the AdaBoost (FREUND & SCHAPIRE, 1997), the Arc-X4 (BREIMAN, 1998) and the MultiBoosting (WEBB, 2000).

Among decision committee learning approaches, certain algorithms have received particular attention in the forecasting literature, due to their remarkable consistency in reducing the final forecasting error. These are methods which operate by selectively resampling from the training data to generate derived training sets to which the base learner is applied. Popular examples of such procedures are the *Bagging* (BREIMAN, 1996) and the *Boosting* (FREUND, 1995) algorithms.

A notable feature common to both *Bagging* and *Boosting* routines is that, on average, the error reduces as the committee size increases, but the marginal error reduction associated with each additional committee member tends to decrease. In other words, each additional member, on average, has less impact on a committee's prediction error than any one of its predecessors (SCHAPIRE et al., 1998). On the other hand, the operating modes of the two algorithms differ substantially. The *Boosting* algorithm generates the classifiers sequentially, while *Bagging* generates them in parallel. *Boosting* also changes the weights of the training instances provided as input to each inducer based on classifiers that were previously built. As a result, *Boosting* appears to have greater average effect, leading to substantially larger error reductions than bagging on average. Much of the benefit realized by *Boosting*, however, seems to be due to overfitting (QUINLAN, 1996). This explains the failure of *Boosting* routines on some datasets, particularly when the interest is in predictive accuracy. The empirical results from BAUER & KOHAVI (1999), for instance, suggest that certain Boosting algorithms, such as the AdaBoost (FREUND & SCHAPIRE, 1997), do not deal well noisy data. Some authors also argue that *Bagging* is more consistent, in the sense that it increases the error of the base learner less frequently than *Boosting* does (WEBB, 2000). Finally, another important feature of *Bagging* algorithms, one that is particularly useful in forecasting approaches, is the possibility of selecting the predictors originated from the forecasting ensemble by means of user-defined techniques, i.e., the practitioner is not restricted to the pre-defined weight schemes of *Boosting* approaches.

2.3
Bagging applications to time series forecasting

In light of aforementioned, this thesis focuses on the use of Bagging algorithms to generate forecast ensembles, and the subsequent selection and weighting of the most relevant predictors in the ensemble by means of smart selection heuristics. To understand its roots, this subsection provides a brief chronological review of relevant works using Bagging in time series forecasting contexts.

INOUE & KILIAN (2004) are likely to have pioneered the use of Bagging in time series forecasting. Using a multiple regression approach, the authors demonstrated that Bagging consistently led to more accurate forecasts than dynamic factor models when the number of predictors is large, but smaller than the sample size. LEE & YANG (2006) showed that Bagging may improve the binary and quantile predictions in small samples using asymmetric loss functions. INOUE & KILIAN (2008), in turn, proposed three variants of the original Bagging algorithm (BREIMAN, 1996) to investigate whether including indicators of real economic activity when forecasting U.S. consumer price inflation led to lower Mean Squared Forecast Error (MSFE) estimates. They demonstrated that Bagging could reduce the MSFE, although they argued that the method was not the only capable of doing so.

Another strand of literature arose from the work of CORDEIRO & NEVES (2009), who first proposed combining Bagging and exponential smoothing methods, and tested it using series from the M3 competition. Their so-called *Boot.EXPOS* approach could be viewed as a variant of the Sieve bootstrap approach (BÜHLMANN, 1997) and had some success in forecasting series with marked seasonal and trendy components (mainly quarterly and monthly data). HILLEBRAND & MEDEIROS (2010) showed that Bagging led to accuracy improvements on two types of models when forecasting realized volatility of several stocks from the Dow Jones Industrial Average: the log-linear model and a nonlinear model for the realized kernel estimator of integrated volatility. RAPACH & STRAUSS (2010) combined Bagging with a dynamic linear regression model for forecasting U.S. employment growth. They compared it with several forecast combination methods and showed that the use of Bagging often delivered the lowest MSFE values. WANG et al. (2012) proposed the combined use of Bagging with Support Vector Machines (SVM) and Artificial Neural Networks (ANN). They

showed that their approach generated more accurate results than single-model SVM and ANN models and other ensemble methods.

ZONTUL et al. (2013) combined Bagging with an algorithm called REPTree to produce forecasts of wind speed in Kirklareli (Turkey) and showed that their method provided better results compared to competing machine learning methods. JIN et al. (2014) proposed a revised version of Bagging to investigate the dependency in time series data. The method was found to outperform the one-step-ahead linear, local constant and local linear models when forecasting several financial time series. MAÇAIRA et al. (2015), in turn, proposed a variant of Bagging which involved generating bootstraps of the noise obtained from a Multi-channel Singular Spectrum Analysis (MSSA) decomposition (HASSANI et al., 2015). The method was used to forecast up to 60 months ahead natural inflow energy series in Brazil and outperformed some forecasting benchmarks.

Inspired by the *Boot.EXPOS* approach of CORDEIRO & NEVES (2009), BERGMEIR et al. (2016) proposed a novel forecasting approach combining Bagging with exponential smoothing methods. In brief terms, it involved first pre-treating and decomposing the original series into three additive components (trend, seasonal and remainder). Replicas for the remainder would then be generated by means of a slightly different version of the original Moving Blocks Bootstrap (MBB) algorithm (KÜNSCH, 1989). Once the desired number of replicas was achieved, the series were reconstructed from their structural components (by adding again the trend and seasonal components to the remainder bootstraps). That way, multiple new series (bootstraps) were created. An exponential smoothing forecasting model was built for the original data and each of the bootstraps separately. Finally, the point forecasts originating from each model were aggregated using the median. The authors demonstrated that their approach, which became known as Bagged.BLD.MBB.ETS[1], outperformed *Boot.EXPOS* and other simple benchmarks, particularly for monthly series from the M3 Competition.

[1] BLD is an acronym for Box–Cox and loess-based decomposition (BLD), MBB stands for Moving Blocks Bootstrap and ETS stands both for ExponenTial Smoothing and for Error, Trend, and Seasonality, which are the three components that define a model within the ETS state space modelling framework proposed by HYNDMAN et al. (2002) – see Section 3.1.3 for details.

Motivated by the findings from BERGMEIR et al. (2016), DANTAS et al. (2017) applied the Bagged.BLD.MBB.ETS forecasting approach in the context of air transportation demand time series, and the results outperformed the benchmarks methods. DE OLIVEIRA & CYRINO OLIVEIRA (2018), in turn, proposed an alternative method to generate ensemble forecasts – a variant of the KREISS (1988) / BÜHLMANN (1997) sieve bootstrap method applied to the remainder component of an STL decomposition (CLEVELAND et al., 1990). This new variant of Bagging delivered satisfactory forecasting results for certain types of electricity consumption time series, outperforming BERGMEIR's et al. (2016) approach on several occasions. The work, entitled "Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing methods", was published as a full-length article in *Energy* in 2018 (v. 144, p. 776–788) and relates to the first contribution of the thesis, described in detail in Chapter 4.

PETROPOULOS et al. (2018) explored the sources of uncertainty (model, data and parameter) in Bagging procedures and demonstrated that the benefits of Bagging originate predominantly from model uncertainty. They then proposed a more sophisticated combination strategy that specifically tackled this source of uncertainty: The Bootstrap Model Combination (BMC). Considering all series from the M (MAKRIDAKIS et al., 1982) and M3 (MAKRIDAKIS & HIBON, 2000) competitions, the BMC delivered better forecasts when compared to Bagged.BLD.MBB.ETS of BERGMEIR et al. (2016).

DANTAS & CYRINO OLIVEIRA (2018), in turn, developed a new forecasting approach combining Bagging, Exponential Smoothing and Clustering. In brief terms, Partitioning Around the Medoids (PAM) (KAUFMAN & ROUSSEEUW, 1987) is initially used to identify clusters of similar bagged forecasts. Then, forecasts from each cluster are selected in order to create a smaller subset of forecasts with reduced error-variance to be combined using the median. The proposed approach was evaluated using series from the M3 and CIF 2016 competitions and led to more accurate forecasts than several benchmarks from both competitions, including previous existing Bagging approaches.

3
How Bagging works for time series forecasting

To illustrate how Bagging for time series forecasting works in practice and, at the same time, provide a starting point for the discussion of the contributions in this thesis, the present chapter describes the most up-to-date techniques involving the use of Bagging in forecasting routines.

We start by delving into the details of the Bagged.BLD.MBB.ETS method proposed by BERGMEIR et al. (2016), since it provides a sound base of comparison with other recently developed Bagging routines for forecasting. Then, we explore new features brought by the Bootstrap Model Combination (BMC) approach devised by PETROPOULOS et al. (2018) and the Bagged.Cluster.ETS approach depicted in DANTAS & CYRINO OLIVEIRA (2018).

3.1
Bagged.BLD.MBB.ETS

As foreshadowed in Section 2.3, The Bagged.BLD.MBB.ETS / Bagged ETS procedure proposed by BERGMEIR et al. (2016) involves combining an ensemble algorithm, namely the *Bootstrap Aggregation* (*Bagging*), with exponential smoothing formulations.

The Bootstrap was first devised by EFRON (1979) following an earlier work on the jackknife procedure by QUENOUILLE (1949). In its original form, the technique consisted of re-sampling the underlying data, in order to get an approximation of the sampling distribution of some statistic of interest. Adaptations have been developed for time series, since the data are typically autocorrelated. The Bootstrap Aggregation (Bagging), in turn, is a supervised machine learning technique, proposed by BREIMAN (1996).

The underlying idea behind Bagging for time series forecasting is to use predictors that are built on bootstrapped versions of the original data. That way, a random pool (ensemble) of forecasts is formed, and then combined into one single output, by weight-averaging for instance. Hence, Bagging allows the practitioner to include different types of uncertainty that may arise when building a predictive model from data, namely data uncertainty, model uncertainty, and parameter uncertainty (PETROPOULOS et al., 2018). Approaches can differ, however, in many aspects/steps of the methodology (such as training data pre-treatment, prior decomposition for isolation of key features, selection of which components are going to be bootstrapped, choice of bootstrapping methods, among others).

To provide a common framework for further discussions and ease the interpretation of the selected state-of-the-art Bagging techniques, we argue that most Bagging routines for forecasting can be summarized in four main stages: (i) an (optional) data treatment or decomposition; (ii) resampling; (iii) forecasting; and (iv) combination, in which the outputs from all members are averaged and sometimes removed, when they are unlikely to improve the forecast. In the next subsections, we explain in detail each stage according to BERGMEIR's et al. (2016) approach.

We finally clarify that practical implementation of every stage is conducted using the R programming language (R CORE TEAM, 2019) and its related packages. The overall procedure devised by BERGMEIR et al. (2016) can be implemented in R through the baggedModel( ) function from the *forecast* package (HYNDMAN & KHANDAKAR, 2008; HYNDMAN et al., 2019) using appropriate arguments, or the baggedETS( ) wrapper function.

3.1.1
Pretreatment and decomposition

BERGMEIR's et al. (2016) approach involves first generating replicas for the remainder component of a Seasonal-Trend decomposition using *Loess* (STL decomposition) (CLEVELAND et al., 1990) applied to a Box–Cox (BC) (BOX & COX, 1964) transformed time series.

The BC transformation aims at stabilizing the variance of a time series. It is also capable of making highly skewed distributions less skewed. It is defined as follows:

$$\omega_t = \begin{cases} \log\,(y_t)\,, & \lambda = 0; \\ \left(y_t^{\lambda} - 1\right)/\lambda\,, & \lambda \neq 0 \end{cases} \qquad (1)$$

where y_t represents the original time series, ω_t its transformed version and λ is the transformation parameter. It is worth noting that there is still no consensus on the method of choosing λ. BERGMEIR's et al. (2016) choice was to restrict λ to lie in the interval [0, 1] and use the method of GUERRERO (1993) to choose its value. In short, the chosen method partitions the original data into subseries of length equal to the seasonality (or length two, if the series is non-seasonal). Then, the sample mean m and the standard deviation s are calculated for each of the subseries, and λ is chosen in such a way that the coefficient of variation of $s/m^{(1-\lambda)}$ across the subseries is minimized. The Box-Cox transformation can be implemented by means of the BoxCox( ) function from the *forecast* package in R.

The STL, in turn, is a sequence of six smoothing operations that employ locally-weighted regression (*Loess*) on the (pretreated) series, dividing it into three additive components: trend, seasonal and remainder. STL offers major advantages when compared to other decomposition methods, such as: the possibility to handle any type of seasonality (regardless of the frequency) and to change the seasonal component over time; the possibility to control the smoothness of the trend-cycle; and its robustness to outliers when estimating the trend-cycle and seasonal components (HYNDMAN & ATHANASOPOULOS, 2013).

In *Loess*, a neighborhood is first defined for each data point and the points in that neighborhood are subsequently weighted according to their distances from the respective data point. A polynomial of degree d is then fitted to these points - usually d = 1 or d = 2. Higher degrees do not improve much the fit. Indeed, CLEVELAND et al. (1990) argue that taking d = 1 is reasonable if the underlying pattern in the data has gentle curvature. The trend component is equal to the value of the polynomial at each data point.

In summary, the steps performed during the STL decomposition are: (i) detrending; (ii) cycle-subseries smoothing, in which series are built for each seasonal component, and smoothed separately; (iii) low-pass filtering of smoothed cycle-subseries, when the subseries are put together again, and then smoothed; (iv) detrending of the seasonal series; (v) deseasonalizing the original series using the seasonal component calculated in the previous steps; and (vi) smoothing the deseasonalized series to get the trend component.

In R, STL can be applied by means of the stl( ) function from the *stats* package (R CORE TEAM, 2019). Essential parameters are "*periodic"* for s.window and default values for the polynomial degrees: $d = 0$ in step (ii) and $d = 1$ in steps (iii) and (iv).

3.1.2 Resampling

BERGMEIR's et al. (2016) approach to resampling involves generating replicas for the remainder component of the STL decomposition. To that end, they put forth a slightly different version of the original Moving Blocks Bootstrap (MBB) algorithm – see next paragraph for details. After obtaining the desired number of remainder bootstraps, the trend and seasonal components are added to each replica, and the BC transformation is inverted. That way, multiple new series (bootstraps) were created.

The Moving Blocks Bootstrap (MBB) approach was first introduced by KÜNSCH (1989), who proposed drawing data blocks of equal size from the series until the desired length was achieved. That way, for a series of length n, with a block size of l, $n - l + 1$ (overlapping) possible blocks exist. However, bootstrap procedures for time series replicates must take into account both stationarity and autocorrelation in the data. To meet that end, BERGMEIR's et al. (2016) proposed drawing $\lfloor n/l \rfloor + 2$ blocks[2] from the remainder series of a STL Decomposition, and discarding a random number of values, between zero and $l - 1$, from the beginning

[2] $\lfloor n/l \rfloor$ stands for the "floor" of n/l division, i.e., the largest integer less than or equal to n/l.

of the bootstrapped series. Then, to obtain a series with the same length as the (original) remainder series, they further discarded as many values as necessary to obtain the required length. This process ensures that the bootstrapped series do not begin or end on a block boundary. Finally, the trend and seasonality are combined with the bootstrapped remainder to get the final bootstrapped sample.

The above procedure can be implemented in R by means of the bld.mbb.bootstrap( ) function from the forecast *package*. The method requires, however, the pre-definition of the block size parameter. Although not a consensus in the related literature, BERGMEIR et al. (2016) recommended the use of block sizes of $l = 24$ and $l = 8$ for monthly and quarterly series, respectively. These correspond to two full years of observations, to ensure that any remaining seasonality is captured. For yearly data, a block size of $l = 8$ was employed, even though no explanations were provided for such choice. It is worth noting that the same guidelines in terms of block size is followed in subsequent works using the MBB algorithm for bootstrapping in time series forecasting contexts – see, for instance, DANTAS et al. (2017); DE OLIVEIRA & CYRINO OLIVEIRA (2018); PETROPOULOS et al. (2018); and DANTAS & CYRINO OLIVEIRA (2018).

The flowchart of Figure 3.1 illustrates the first part of the BERGMEIR's et al. (2016) approach (Section 3.1.1 and the current section), whilst Figure 3.2 depicts how the MBB procedure works in practice.

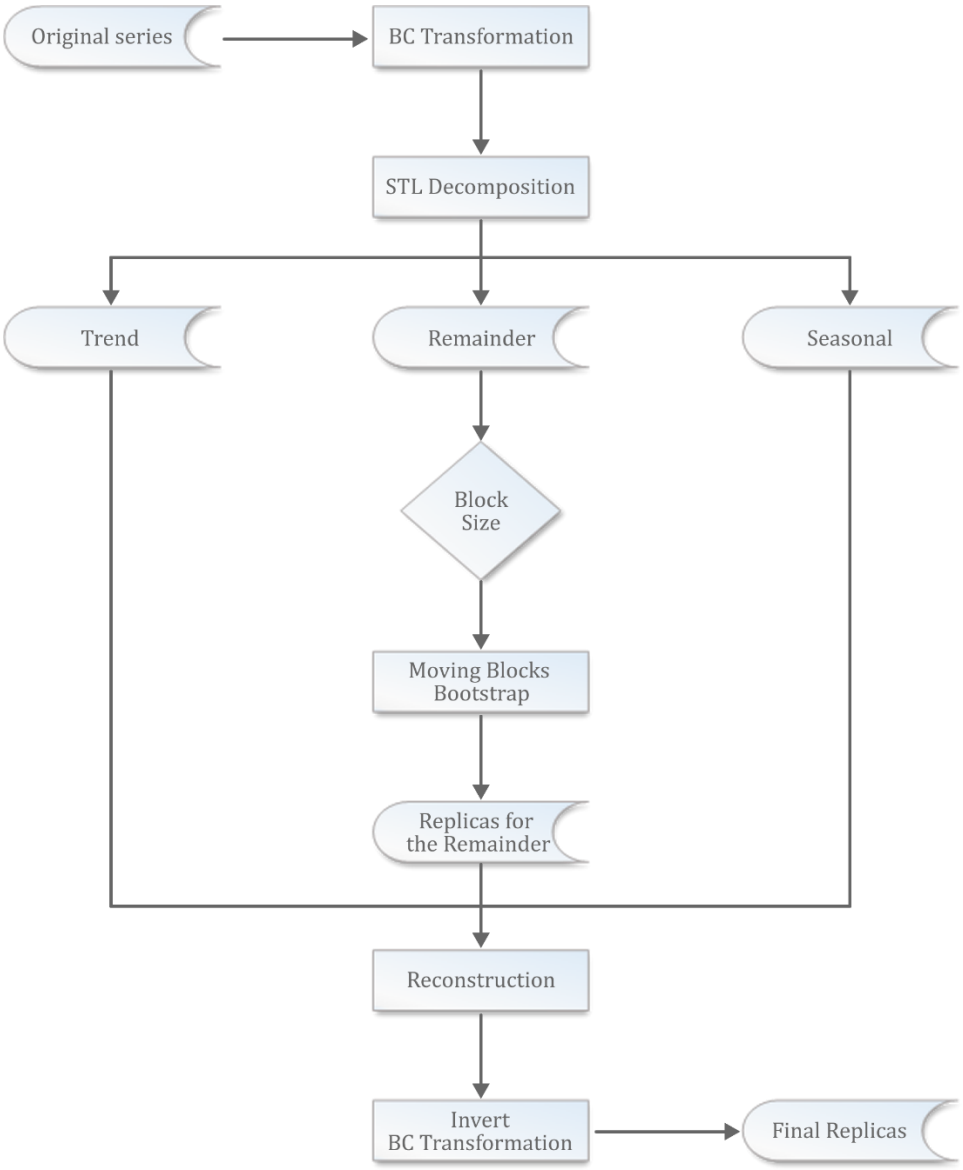


Figure 3.1 Bagged.BLD.MBB.ETS algorithm – First Part – Flowchart.
Source: Adapted from BERGMEIR et al. (2016).

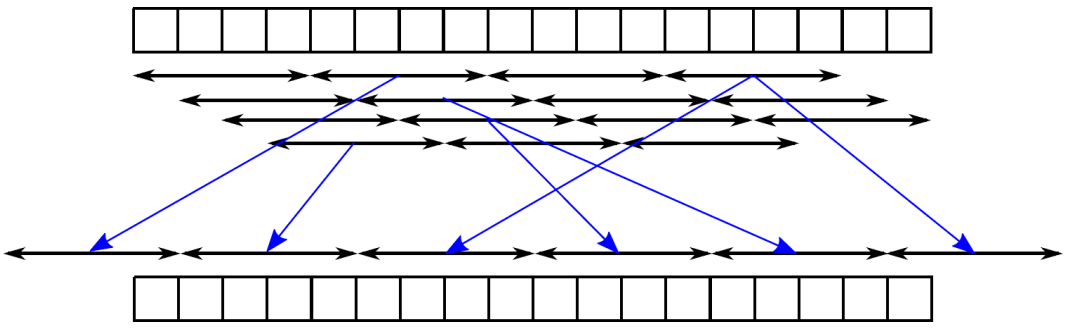


Figure 3.2 Illustration of the MBB algorithm.
Source: Adapted from PETROPOULOS et al. (2018).

3.1.3
Forecasting with ETS

The second part of BERGMEIR's et al. (2016) Bagged.BLD.MBB.ETS algorithm involves building a forecasting model for the original data and each of the bootstraps separately. In their work, the ETS family of models was selected to that end.

ETS (an acronym standing for ExponenTial Smoothing or, alternatively, Error, Trend and Seasonality) stands for a finite set state space based exponential smoothing formulations, which can be obtained by considering variations in the combination of the error, trend and seasonal components of a time series. Exponential smoothing, in turn, consists of traditional procedures that attribute exponentially decreasing weights for past data so that recent observations are given relatively more weight in forecasting than older ones.

Even though the basic structures were provided a long time ago, with the seminal works of HOLT (1957, reprinted 2004) and WINTERS (1960), exponential smoothing methods are still widely applied, mainly due its simplicity and its ability to adapt to many different situations (GOODWIN, 2010). In addition, ETS formulations have a theoretical foundation in state space modelling, allowing for straightforward implementation in many statistical packages (HYNDMAN et al., 2002, 2008; HYNDMAN & ATHANASOPOULOS, 2013).

There are several different approaches to exponential smoothing. For an extensive list of the most commonly used exponential smoothing methods in the literature, the interested reader is referred to the compiling works of GARDNER Jr. (1985, 2006). As for ETS in particular, according to the taxonomy proposed by PEGELS (1969) and extended by GARDNER Jr. (1985), the possibilities for the trend and seasonal components are depicted in Table 3.1. In addition, the error term can also vary between additive or multiplicative. That way, a total of 30 different formulations can be achieved.

Table 3.1 Possible variations for the trend and seasonal components of ETS formulations under a state space-based approach

	Seasonal Component		
Trend Component	**None (N)**	**Additive (A)**	**Multiplicative (M)**
None (N)	N, N	N, A	N, M
Additive (A)	A, N	A, A	A, M
Additive Damped (A_d)	A_d, N	A_d, A	A_d, M
Multiplicative (M)	M, N	M, A	M, M
Multiplicative Damped (M_d)	M_d, N	M_d, A	M_d, M

Each model in a state space based formulation consists of two sets of equations: (i) a measurement equation that describes the observed data; (ii) and some transition equations that describe how the unobserved components or states (level, trend, seasonal) change over time. Let's consider one possible combination as an example: an additive error, multiplicative trend, multiplicative season model, or AMM, according to the above-mentioned notation. First, we consider a *p*-dimensional state vector $x_t = (a_t, b_t, s_t, s_{t-1}, \dots, s_{t-m})'$, with a_t and b_t being the contemporaneous estimates of the level and linear trend parameters and s representing the included seasonal terms. We also let $\hat{y}_t = a_{t-1}\, b_{t-1}\, s_{t-m}$ be the one-period ahead forecast of y_t. Then, the prediction error decomposition is

$$y_t = \hat{y}_t + e_t = a_{t-1}\, b_{t-1}\, s_{t-m} + e_t \qquad (2)$$

Following ORD et al. (1997), we may write a nonlinear dynamic model representation of the exponential smoothing equations using a state space model with a common error term:

$$\begin{aligned} y_t &= h\,(x_{t-1}, \theta) + k\,(x_{t-1}, \theta)\, e_t \\ x_t &= f\,(x_{t-1}, \theta) + g\,(x_{t-1}, \theta)\, e_t \end{aligned} \qquad (3)$$

where h and k are known continuous scalar functions, f and g are known continuous functions with continuous derivatives from $\Re^p \to \Re^p$ and $e_t \sim iid\,(0, \sigma^2)$ are the independent past realizations of y and x.

Conceptually, the y_t equation represents how the various state variable components $(a_{t-1}, b_{t-1}, s_{t-m})$ are combined to express the series in terms of a smoothed forecast $\hat{y}_t = h(x_{t-1}, \theta)$ and the prediction error (e_t). The x_t equations, in turn, outline the process by which the component estimates are updated using the previous period's estimates and the current prediction error - e_t. The multiple functions are a notational device for writing the additive and multiplicative errors in compact form. With additive errors, we have $k \equiv 1$, so that $y_t = h\,(x_{t-1}, \theta) + e_t$. In short, we may think of the updating smoothing equations as being weighted averages of a term which depends on the current prediction error (and prior states), and one which depends on the prior states. The resulting state space equations for an additive error, multiplicative trend and multiplicative season model are:

$$\begin{aligned} \hat{y}_t &= a_{t-1}\, b_{t-1}\, s_{t-m} \\ a_t &= a_{t-1}\, b_{t-1} + \alpha\, {e_t}/{s_{t-m}} \\ b_t &= b_{t-1} + \alpha\beta\, {e_t}/{(s_{t-m}\, a_{t-1})} \\ s_t &= s_{t-m} + \gamma\, {e_t}/{(a_{t-1}\, b_{t-1})} \end{aligned} \tag{4}$$

To conserve space, further details on the ETS state space specification are not presented here. For a thorough overview on the subject, the interested reader is referred to ORD et al. (1997). For information regarding the alternatives on state space estimation methods, the work of HYNDMAN et al. (2008) is indicated.

Concerning the practical implementation of ETS, an optimal model and set of parameters is identified for each series using the function ets( ) from the *forecast* package for the R statistical software. The input to the function is a vector formed by the original data values organized in a time series format. The output of ets( ) is a model form (together with the optimal parameters) consisting of three terms: error, trend and seasonality. Model selection/parameter optimization is often performed by minimizing one (or more) information criterion. The default is to select the ETS combination offering the lowest Akaike Information Criteria with corrections (AIC_c) (SUGIURA, 1978), a commonly adopted practice in empirical literature. Finally, forecasts for each optimal model can then be computed for a desired number of steps-ahead using the forecast( ) function (or forecast.ets( ) wrapper function), available from the R *forecast* package.

It is interesting to note that certain combinations from Table 3.1 give birth to well-known models in the forecasting literature. These models are frequently used as benchmarks to compare the forecasting performance of competing methods in a out-of-sample evaluation. Two widely referenced methods are (i) the three parameter Holt-Winters additive model and (ii) the three parameter Holt-Winters multiplicative model. The former can be obtained using the ets( ) function with default parameters, with the exception of the model selection, which is set to model="AAA". An alternative is to use the wrapper function hw( ) and setting *seasonal* to "additive". The latter, in turn, can be called upon by simply adjusting the *seasonal* setting to "multiplicative" in the hw( ) wrapper function[3].

3.1.4
Combination

The last step in BERGMEIR's et al. (2016) Bagged.BLD.MBB.ETS approach is to aggregate the forecasts obtained for each bootstrapped time series to generate the final output. To that end, the authors opt to use the simple median, given that it is less sensitive to outliers than other averaging approaches, reducing the effects of occasional poor forecasts.

3.1.5
Overall procedure

The flowchart of Figure 3.3 summarizes the stages described in Sections 3.1.1–3.1.4 and highlights the fundamental choices made by BERGMEIR et al. (2016) under the proposed framework. Following a Box-Cox (BC) transformation, each series is decomposed into three components (trend, seasonal and remainder) using the STL Decomposition. The remainder is then resampled using the Moving

---

[3] Alternatively, the ets( ) function can also be used, this time with the following arguments:
ets(X, "MAM", alpha = NULL, beta = NULL, gamma = NULL, phi = NULL, damped = FALSE, opt.crit = "mse", lambda = NULL, biasadj = FALSE)
where X is the underlying train series.

Blocks Bootstrap (MBB) approach (with the desired block size for each replication). Finally, the components are added together again, and the BC transformation is inverted. The overall procedure is repeated J times, J being the number of desired replications. A forecasting model is subsequently built for the original data and each of the bootstraps separately: in the present procedure, the 30 possible combinations of the ETS family of models described in Section 3.1.3 are considered as competing models for each series. The $J + 1$ forecasts (forecasts of the original data and the J bootstraps generated) are finally aggregated/combined using the simple median.

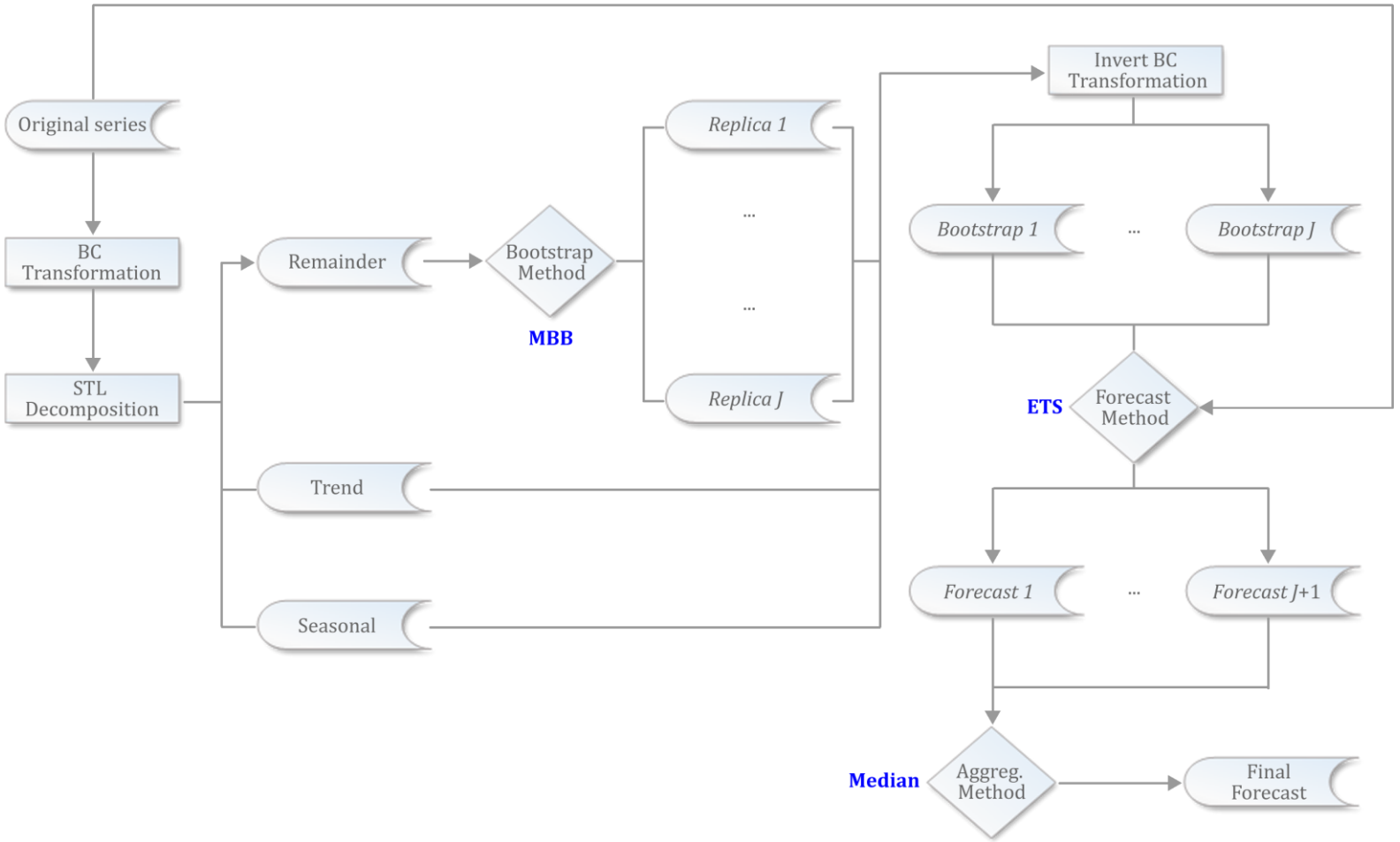


Figure 3.3 A usual Bagging routine for forecasting.

Texts in blue represent the fundamental choices adopted in BERGMEIR's et al. (2016) in each stage of the algorithm. *Source:* The author.

3.2
The Bootstrap Model Combination (BMC)

As previously outlined in Section 2.3, the Bootstrap Model Combination (BMC) was devised by PETROPOULOS et al. (2018) as an alternative strategy to tackle the isolated effect the model uncertainty arising in Bagging strategies, i.e., the fact that different models might be selected as optimal for the bootstrapped series.

The BMC is quite similar to BERGMEIR's et al. (2016) Bagged.BLD.MBB.ETS approach in that the bootstraps are originated by resampling the remainder from an STL decomposition and independently predicted using exponential smoothing formulations. However, the bootstraps are not directly used for forecasting, but rather to drive the selection of exponential smoothing model forms, which are then applied to the original data. The forecasts originating from this last step are then combined, with weights reflecting the frequency that the selected formulations were identified as optimal for the bootstraps.

The fundamental differences between Bagged.BLD.MBB.ETS and BMC are illustrated in the flowchart of Figure 3.4.

In further details, Bagged.BLD.MBB.ETS aggregates the $J + 1$ Point Forecasts (PFs) using their medians. BMC, in turn, identifies from the pool of $J + 1$ forecasts, the K unique ETS model forms and apply them to the original series. Then, it combines the results from K PFs using as weights the frequency with which the unique forms were identified as optimal, i.e., the amount of times they were selected divided by $J + 1$. Considering all series from the M (MAKRIDAKIS et al., 1982) and M3 (MAKRIDAKIS & HIBON, 2000) competitions, the BMC delivered better forecasts when compared to Bagged.BLD.MBB.ETS.

Referring once again to the stages framework suggested in the last section, one could note that the procedure underlying the BMC is akin to the one depicted in Figure 3.3, with two exceptions: (i) the forecasting methods are now the unique model forms identified when applying ETS on the original series and its bootstraps; and (ii) the combination strategy is the aggregation by frequencies, in lieu of the simple median.

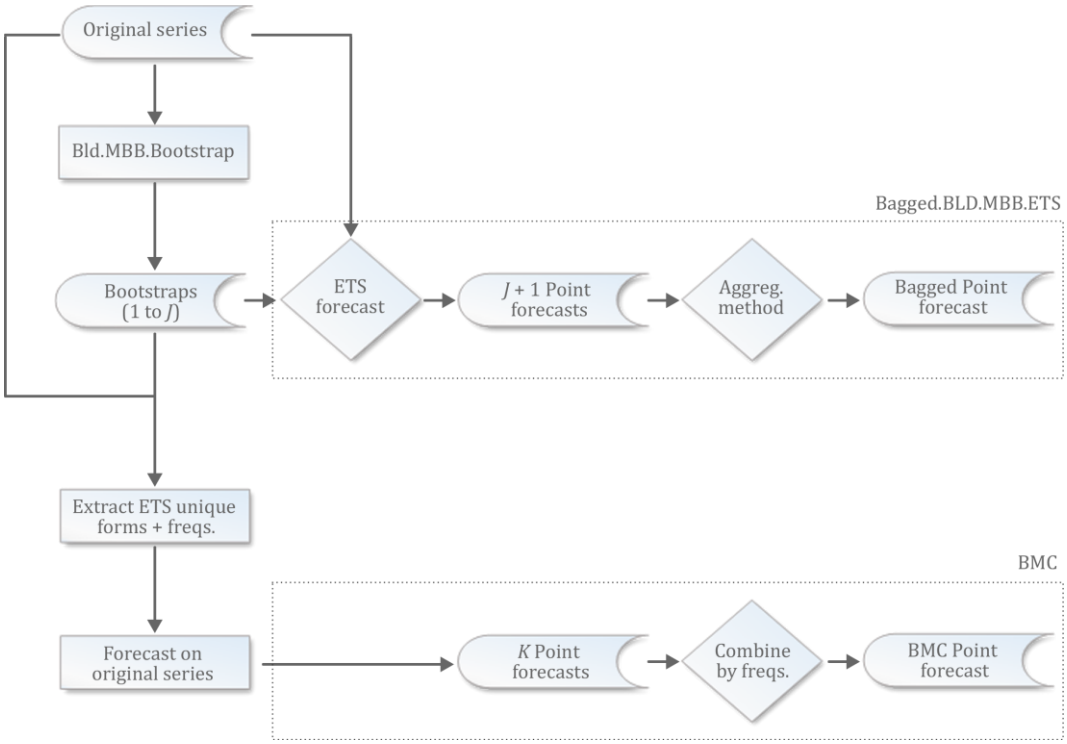


Figure 3.4 Bagged.BLD.MBB.ETS and BMC.

Source: The author.

3.3
The Bagged.Cluster.ETS

Another variant of Bagging for time series forecasting was recently put forth by DANTAS & CYRINO OLIVEIRA (2018). Their proposal was to combine Bagging, exponential smoothing and clustering algorithms. In brief terms, the approach, named after Bagged.Cluster.ETS, aimed at reducing the covariance effect of bagged forecasts by using Partitioning Around Medoids (PAM) to produce clusters of similar forecasts, then selecting several forecasts from each cluster to create a group with a reduced variance. The approach was tested on different sets of time series from the M3 (MAKRIDAKIS & HIBON, 2000) and CIF 2016 competitions (ŠTĚPNIČKA & BURDA, 2017) and proved itself as a tough competitor, with its forecasts being more accurate than those from 25 benchmarks in the first competition and 23 in the second, including BERGMEIR et al. (2016) Bagged.BLD.MBB.ETS in both cases. Reporting one more time to Figure 3.3, the difference for Bagged.Cluster.ETS lies in the last stage (combination method), where a clustering and subsetting phase precedes the median aggregation.

4
First essay: A new variant of Bagging applied to mid/long term electric energy consumption forecasting

This chapter refers to the first contribution of the thesis. In a nutshell, it proposes an alternative method to generate ensembles of forecasts and applies it to a range of electricity consumption time series across different developed and developing economies. The proposed approach is compared with several time series and machine learning methods and with the Bagging approach of BERGMEIR et al. (2016) (Bagged.BLD.MBB.ETS). The results show that the former outperforms the latter on several occasions. The work, entitled "*Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing methods*", was published as a full-length article in *Energy* in 2018 (v. 144, p. 776–788).

Before delving into the details of the involved methodology, the next section provides a brief insight on the issues involved in mid/long term energy forecasting. The methodology is then introduced in Section 4.2. Section 4.3 summarizes the selected data whilst Section 4.4 describes and discusses the results of the quantitative analysis. Finally, Section 4.5 concludes the findings of the study and presents directions for future research.

4.1
Introduction to energy demand planning and its challenges

Accurate load forecasting is of utmost concern in decision-making within the electric sector, as the consequences of overestimation or underestimation can be costly. For instance, when delivered power is higher than the actual demand, the provider not only wastes resources but may also bear expensive costs due to strong

spot market regulation in several countries. On the other hand, underestimation naturally results in failures and shortages, which in turn translates into a loss of productive time and quality and subjects the provider to sanctions and penalties.

Over the past decades, a large number of approaches have been proposed to estimate and forecast electric energy consumption. In short, these approaches can be divided into two main categories: short-term and mid-/long-term. Whilst the first is concerned with time frames of minutes to hours, the horizon for mid/long-term forecasting ranges from a few weeks to several years (AL-HAMADI & SOLIMAN, 2005). Forecasting electric energy demand over the latter period is often regarded as a challenging task, due to the nonlinear, multidimensional nature of this variable (SHAO et al., 2016). In addition, many unpredictable factors affect electricity demand modeling, such as structural breaks and transitory effects from external variables. Nevertheless, mid/long-term load forecasting assumes a particular importance for electric power utility planning. Even though short-term forecasting forms the basis of the electrical energy trade and spot price calculation (CASTELLI et al., 2015), several decisions are made on the basis of mid/long-term energy demand forecasting, such as the construction of new generation facilities, the purchase of existing generating units, the development of transmission and distribution systems, among others.

In light of the aforementioned, despite the drawbacks in terms of complexity and uncertainty, the pursuit of models that may enhance prediction of energy demand has been a prominent issue in the fields of energy/environmental policy and economic development. To contribute in this regard, this essay proposes an alternative Bagging approach to forecast two-year ahead (medium to long-term) forecasts for monthly electric energy demand in different parts of the world, including both developing and developed countries.

As of the date of publication of this first essay, Bagging techniques are yet to be fully explored in the context of total electric energy consumption. Furthermore, a different variation of Bagging is here introduced leading to satisfactory forecasting results in terms of accuracy for several countries.

4.2
Methods

The Bagging strategy developed in this first essay is built using the same core ideas from BERGMEIR's et al. (2016) Bagged.BLD.MBB.ETS approach – see Sections 3.1.1–3.1.4 for details. However, in lieu of resorting to the Moving Blocks Bootstrap (MBB) algorithm to generate the replicas of the original time series in Bagging, we propose a variant of the Sieve Bootstrap method applied to the remainder of an STL decomposition. This approach, henceforth addressed as Remainder Sieve Bootstrap (RSB), is described in details in the next subsection.

4.2.1
Remainder Sieve Bootstrap

The sieve bootstrap approach has its roots in the works of KREISS (1988) and BÜHLMANN (1997). Their methodology was based on the idea of fitting parametric models first and then resampling from the residuals. In brief terms, given a sample $X_1, \dots, X_n$, from a stationary process, the method involved: (i) selecting the order $p = p(n)$ of an autoregressive (AR) approximation using the Akaike Information Criterion (AIC, AKAIKE, 1974); (ii) using the AR(p) model to filter the residuals series, obtaining centered residuals and their empirical cumulative distribution function; (iii) resampling the (supposed) i.i.d. centered residuals; (iv) using the AR(p) for obtaining a new series X_t^* by recursion. Finally, given $X_1^*, \dots, X_T^*$, the method computes the estimation of the AR coefficients and then obtain future bootstrap observations by recursion from the new series.

In CORDEIRO & NEVES (2009), a different approach was proposed: the authors first suggested fitting an EXPOS model to the data and then proceeding like BÜHLMANN's (1997) method over the residuals. The EXPOS model is named after the best fit model from a set of exponential smoothing forecasting methods. The *BOOT.EXPOS* procedure, as the method became known, demonstrated promising results in forecasting series with seasonal and trendy components.

In this essay, we set forth a variant of the above-mentioned procedure. In lieu of fitting an exponential smoothing model to the data, we first decompose the original time series into its trend, seasonal and remainder components, following the STL approach on the Box-Cox pretreated series, akin to the initial steps proposed by BERGMEIR et al. (2016). Then, provided that the last component is already stationary, we estimate the best fit Autoregressive-Moving Average – ARMA(p, q) – model for the remainder, using AIC with corrections (AIC_c) (SUGIURA, 1978) or the most parsimonious formulation which ensures that there are no autocorrelation issues in the residuals. Finally, we resample the centered residuals and use the ARMA(p, q) to obtain new series for the remainder.

4.2.2
Forecasting with ETS and ARIMA

After having obtained all the bootstrapped time series, we estimated and subsequently forecasted each generated series using two major classes of time series methods: the ETS state space formulations and the Seasonal Autoregressive Integrated Moving Average (SARIMA) family of models (BOX & JENKINS, 1970). Concerning ETS, their fundamentals were already presented in details in Section 3.1.3 of this thesis. We clarify, however, that we employed three different exponential smoothing approaches as forecasting methods for the bootstrapped series, i.e., in the third stage of the Bagging framework depicted in Figure 3.3. These methods are briefly described in the following lines.

(i) The **Holt-Winters additive model** (WINTERS, 1960), appropriate for series with a linear time trend and additive seasonal variation. Its component form can be written as follows:

$$\hat{y}_{t+k} = a + bk + s_{t+k} \qquad (5)$$

where a and b are the permanent component (base signal) and the linear trend parameters, respectively, and s_t are the additive seasonal factors. These parameters, in turn, are defined by the following recursive expressions:

$$a(t) = \alpha\,[y_t - s_t(t-m)] + (1-\alpha)\,[a(t-1) + b(t-1)]$$
$$b(t) = \beta\,[a(t) - a(t-1)] + (1-\beta)\,b(t-1) \quad (6)$$
$$s_t(t) = \gamma\,[y_t - a(t)] + (1-\gamma)\,s_t(t-m)$$

where $0 < \alpha,\ \beta,\ \gamma < 1$ are the hyperparameters and m is the seasonal frequency (supposedly monthly). In these terms, forecasts are computed by:

$$\hat{y}_{T+k} = a(T) + b(T)\,k + s_{T+k-m} \quad (7)$$

where the seasonal factors are used from the last s estimates.

(ii) The **Holt-Winters multiplicative model** (WINTERS, 1960), adequate to series with a linear time trend and multiplicative seasonal variation. The smoothed series $\hat{y}_t$ in this case is given by:

$$\hat{y}_{t+k} = (a + bk)\,s_{t+k} \quad (8)$$

The three coefficients are now defined by the following recursions:

$$a(t) = \alpha\left[\frac{y_t}{s_t(t-m)}\right] + (1-\alpha)\,[a(t-1) + b(t-1)]$$
$$b(t) = \beta\,[a(t) - a(t-1)] + (1-\beta)\,b(t-1) \quad (9)$$
$$\rho_t(t) = \gamma\left[\frac{y_t}{a(t)}\right] + (1-\gamma)\,s_t(t-m)$$

Forecasts are then computed by:

$$\hat{y}_{T+k} = (a(T) + b(T)\,k)\,s_{T+k-m} \quad (10)$$

(iii) **State space based (exponential smoothing) formulations.** As thoroughly explored in Section 3.1.3, this approach consists of a set of 30 different formulations which can be obtained by considering variations in the combination of the error, trend and seasonal components of a time series. As usual in automated ETS routines, the default here was to select the ETS combination offering the lowest Akaike Information Criteria with corrections (AIC_c) (SUGIURA, 1978).

The SARIMA models, first proposed by BOX & JENKINS (1970), consist of an alternative approach to exponential smoothing. In brief terms, SARIMA

models are similar to exponential smoothing methods inasmuch as they are adaptive, can model trends and seasonal patterns, and can be automated. They differ, however, in that they are based on autocorrelations (patterns in time) rather than a structural view of level, trend and seasonality. It is argued that SARIMA formulations tend to succeed better than exponential smoothing methods for longer, more stable data sets and not as well for noisier, more volatile data (MAKRIDAKIS et al., 1982).

Non-seasonal ARIMA models are generally denoted by ARIMA(*p,d,q*) where parameters *p, d*, and *q* are non-negative integers, *p* being the order of the autoregressive model, *d* the degree of differencing, and *q* the order of the moving-average model. Seasonal ARIMA models, in turn, are usually denoted by SARIMA$(p,d,q)_x(P,D,Q)_S$ and can written as follows:

$$\nabla_S^D \nabla^d \phi(B)\, \Phi(B^S)\, Z_t = \theta(B)\, \Theta(B^S)\, a_t \qquad (11)$$

where:

- *S* refers to the number of periods in each season;
- the uppercase *P,D,Q* refer to the autoregressive, differencing, and moving average terms for the seasonal part of the ARIMA model;
- a_t is the error term;
- B is the backward shift operator (eg. $By_t = y_{t-1}$);
- $\phi(B)$ and $\Phi(B^S)$ are the non-seasonal and seasonal autoregressive polynomials, respectively;
- $\theta(B)$ and $\Theta(B^S)$ are the non-seasonal and seasonal moving-average polynomials, respectively;
- ∇^d and ∇_S^D are the non-seasonal and seasonal differencing operators, respectively.

Concerning practical implementation, a SARIMA model can be automatically selected for a given time series by means of the auto.arima( ) function from the *forecast* package in R. The function implements an algorithm which combines unit root tests, minimization of the AICc and Maximum Likelihood Estimation (MLE) to select the best fit SARIMA model.

4.2.3
Aggregation using the mean and the median

In the last step, we aggregate the forecasts obtained for each bootstrapped time series to generate the final output. To that end, we use two different methods: the simple mean (or equal weights combination) and the median. Besides its simplicity, the simple mean has proved to be a tough benchmark for forecasting combinations, see for example STOCK & WATSON (2004). The median, in turn, is less sensitive to outliers, reducing the effects of occasional poor forecasts.

4.3
Data and overall procedure

The empirical analysis was based upon monthly data of total electric energy consumption (GWh) in different developed and developing countries: Canada, France, Italy and Japan for the former case; and Brazil, Mexico and Turkey for the latter. For the Brazilian electric energy consumption, we referred to the data provided by the major Brazilian electric utilities company, Eletrobras, available at the Brazilian Central Bank time series database ELETROBRAS (2017). All other data were collected from the International Energy Agency (IEA) Monthly Electricity Statistics report, which provides electricity production and trade data for all OECD Member Countries (IEA, 2017). The time period of the analysis spanned from July 2006 (the first date available for OECD countries) to December 2016. Months from July 2006 to December 2014 were considered as training set whilst the observations from January 2015 to December 2016 comprised the test set for the out-of-sample experiment.

Recalling the overall procedure proposed in this essay, a transformed version of each original time series was generated by means of a BOX & COX (1964) transformation. For the transformation parameter (λ), we followed BERGMEIR et al. (2016) and restricted it to lie in the interval [0,1], then used the method of GUERRERO (1993) to choose its value. Following the BC

transformation, each series was decomposed into three components (trend, seasonal and remainder), using the STL Decomposition (CLEVELAND et al., 1990). The remainder, was then either: (i) estimated by means of an ARMA(p, q) process and its residuals resampled according to the Remainder Sieve Bootstrap (RSB) procedure proposed in Section 4.2.1; or (ii) directly bootstrapped using the MBB approach, as proposed in BERGMEIR et al. (2016). For each bootstrap approach, a total of 100 new series were generated. Finally, the components were added and the BC transformation was inverted. The overall procedures are illustrated in the flowchart of Figure 4.1.

Several models are proposed to estimate and subsequently forecast the original and bootstrapped versions of the total electric energy consumption time series. In this essay, we restrict our analysis to four methods:

(i) an auto ARIMA approach, implemented via the auto.arima( ) function in R – see Section 4.3.2 for details;

(ii) a three parameter Holt-Winters additive model;

(iii) a three parameter Holt-Winters multiplicative model; and

(iv) an auto state space ETS approach. For this case, we use the ets( ) function in R and let it decide which Error, Trend and Seasonal (ETS) combination best suits the data – see Section 3.1.3 for details.

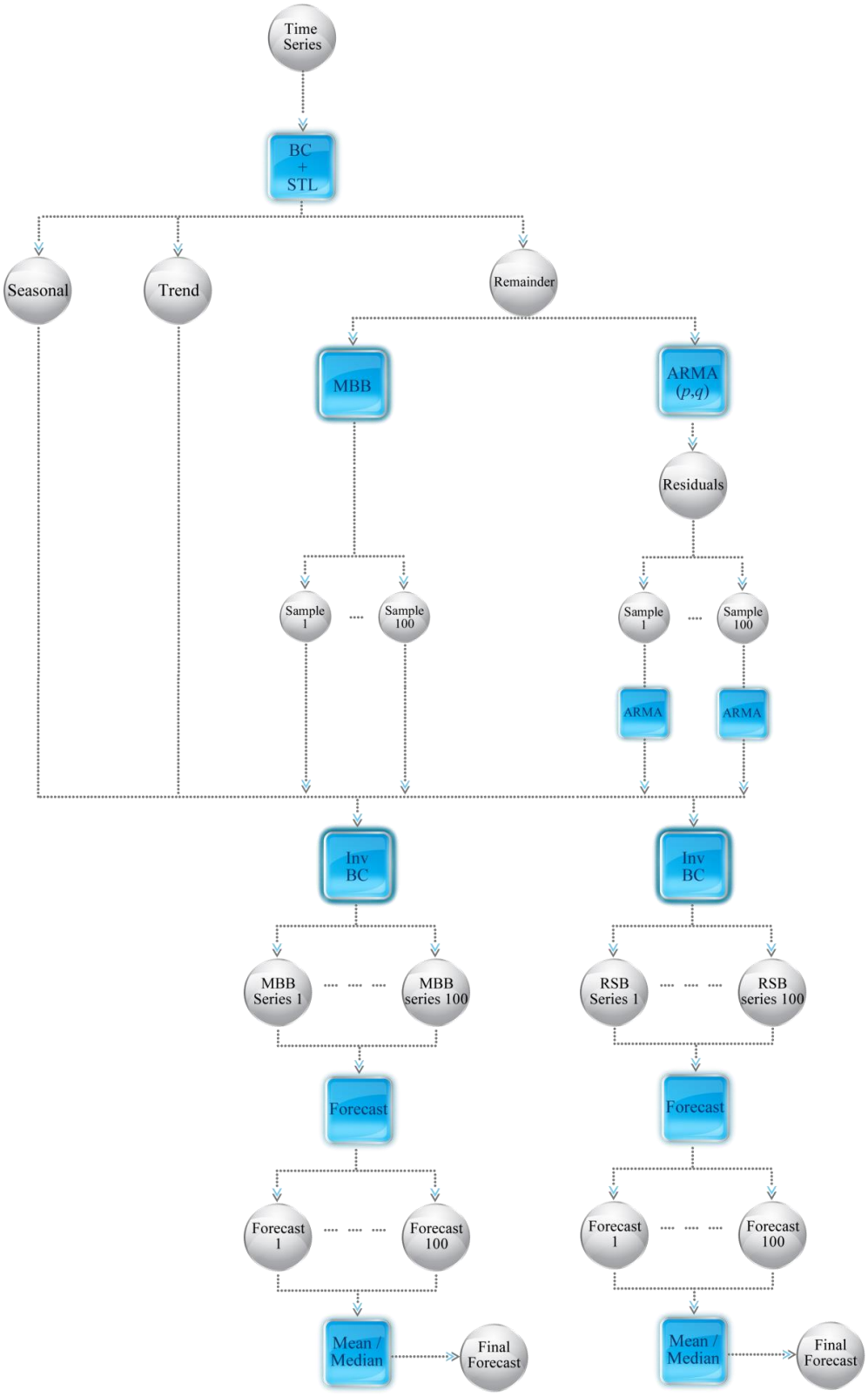


Figure 4.1 MBB and RSB-based Bagging approaches.

Source: DE OLIVEIRA & CYRINO OLIVEIRA (2018).

Finally, the forecasts were combined using either the simple mean or median and the predictive power was assessed by means of an out-of-sample experiment using the test set (January 2015 to December 2016). The following measures were used to evaluate the accuracy of the forecasts:

Mean Absolute Percentage Error (MAPE):

$$MAPE = \left(\sum_{t=T+1}^{T+h} \left| \frac{\hat{y}_t - y_t}{y_t} \right| / h \right) \times 100\% \qquad (12)$$

Symmetric Mean Absolute Percentage Error (sMAPE):

$$sMAPE = \left\{ \sum_{t=T+1}^{T+h} \left[\frac{|\hat{y}_t - y_t|}{(|\hat{y}_t| + |y_t|)/2} \right] / h \right\} \times 100\% \qquad (13)$$

Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\sum_{t=T+1}^{T+h} (\hat{y}_t - y_t)^2 / h} \qquad (14)$$

Theil Inequality Coefficient (TIC):

$$TIC = \frac{\sqrt{\sum_{t=T+1}^{T+h} (\hat{y}_t - y_t)^2 / h}}{\sqrt{\sum_{t=T+1}^{T+h} \hat{y}_t^{\,2} / h} + \sqrt{\sum_{t=T+1}^{T+h} y_t^{\,2} / h}} \qquad (15)$$

In the above formulae, $\hat{y}_t$ is the predicted (forecasted) value whereas y_t is the real (observed) value. h, in turn, is the number of forecasting steps ahead.

4.4
Empirical Findings and Discussion

4.4.1
Performance gains from Bagging

The empirical results (best highlighted in bold) for the developed and developing countries are summarized in Tables 4.1 and 4.2, respectively. A visualization of the forecasts generated by the best bagging approaches plotted against the actual values (for each country) can be seen in Figure 4.2.

With the exception of the Japanese case, where the best forecast in terms of MAPE and sMAPE was achieved using a auto ETS formulation, the Bagging approaches led to considerably superior results in terms of accuracy. In several cases the gains were noteworthy when compared with single forecasts on the real data. For the Italian electricity consumption, for instance, the sMAPE and the RMSE obtained using a Remainder Sieve Bootstrap (RSB) ETS approach were almost 30% and 58% lower than the ones obtained using the auto ETS method.

It is worth noting that, for developed countries, the bagged forecasts that used the RSB approach performed better, in terms of MAPE and sMAPE, than the ones that resorted to the MBB counterpart. In terms of RMSE and TIC, the only case where the MBB outperformed the RSB was for the French monthly electricity consumption. Even so, the difference between the error metrics was not too significant. As for the developing countries, the MBB approach provided slightly better results on two of the three involved countries (Mexico and Turkey). This is a substantial improvement over previous bagging methods, as the MBB-based Bagging approach proposed by BERGMEIR et al. (2016) has been regarded as a benchmark for forecasting monthly data.

Another interesting feature is the fact that the mean and the median of the forecasted values differed considerably in nearly every occasion (the Brazilian case is the sole exception). The results obtained using the simple median aggregation approach were considerably superior to the ones obtained by pooling the forecasts using equal weights (simple mean). Considering that the median is less sensitive to outliers, this suggests that the outliers (whether in the original or the generated time series) exert a considerable effect on the overall results.

Table 4.1 Forecast evaluation – developed countries (best results in **bold**)

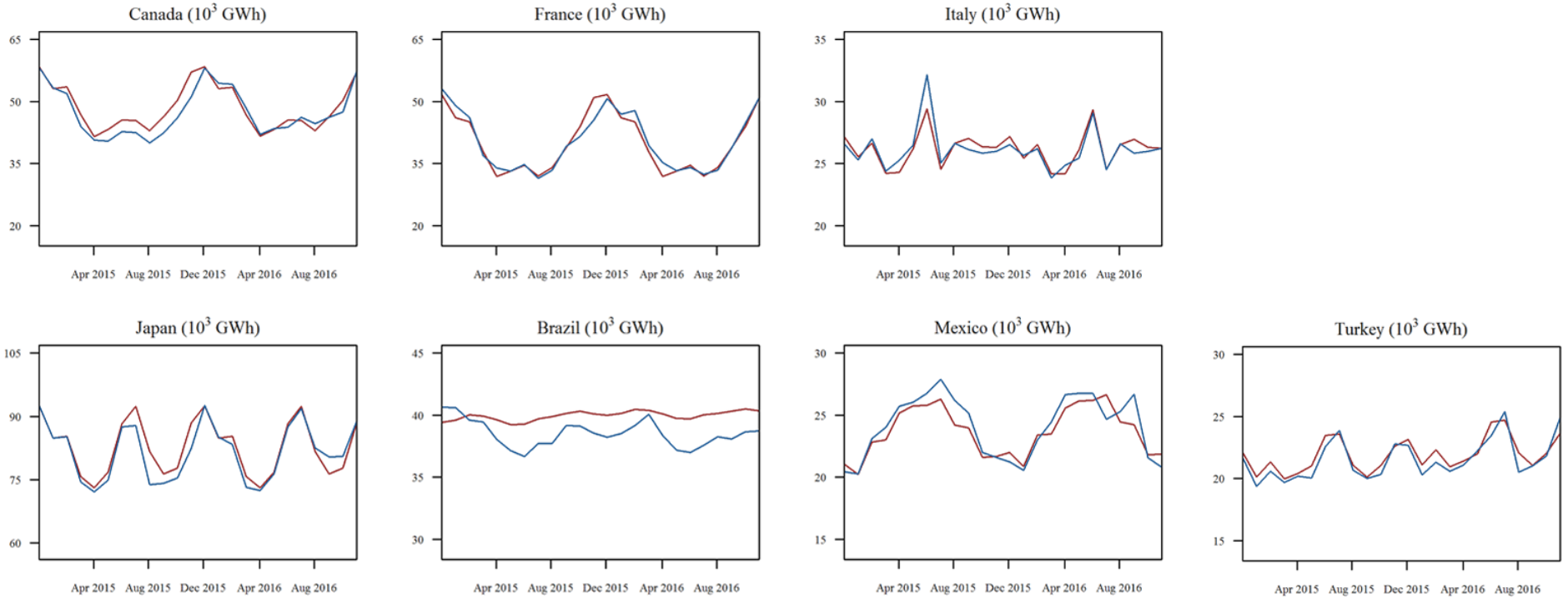
Forecast Approach	Statistic	MAPE (%)	SMAPE (%)	RMSE (GWh)	TIC	MAPE (%)	SMAPE (%)	RMSE (GWh)	TIC
		Canada				France			
MBB.Arima	Mean	4.271	4.187	2275.225	0.024	4.056	4.067	1981.013	0.024
	Median	3.881	3.958	1960.548	0.021	3.684	3.617	1442.878	0.019
RSB.Arima	Mean	4.986	4.899	2629.646	0.027	5.116	5.132	2497.840	0.031
	Median	5.004	5.004	2503.508	0.027	4.892	4.903	1784.197	0.023
Auto Arima	Single	5.140	5.050	2718.240	0.028	5.946	6.014	2865.827	0.036
MBB.Add.	Mean	4.023	3.966	2146.523	0.022	4.157	4.239	2012.938	0.025
H-W	Median	3.433	3.424	1748.092	0.019	3.520	3.583	1618.959	0.021
RSB.Add.	Mean	3.911	3.817	2251.833	0.023	5.248	5.388	2413.163	0.030
H-W	Median	**3.174**	**3.176**	1589.605	0.017	4.320	4.416	2019.551	0.027
Add H-W	Single	4.206	4.078	2474.736	0.026	5.634	5.792	2463.391	0.031
MBB.Mult.	Mean	4.011	3.948	2149.314	0.022	3.807	3.880	1944.163	0.024
H-W	Median	3.294	3.347	1743.229	0.019	3.722	3.793	1582.883	0.021
RSB.Mult	Mean	3.884	3.793	2239.347	0.023	4.834	4.959	2351.984	0.029
.H-W	Median	3.296	3.253	**1575.038**	**0.017**	4.558	4.664	1722.974	0.023
Mult H-W	Single	4.141	4.023	2427.717	0.025	5.033	5.158	2319.074	0.029
MBB.ETS	Mean	3.855	3.787	2141.358	0.022	2.781	2.784	1663.936	0.020
	Median	3.385	3.389	1681.856	0.018	2.098	2.100	**815.629**	**0.011**
RSB.ETS	Mean	4.234	4.152	2306.102	0.024	2.994	3.004	1785.948	0.022
	Median	3.879	3.956	1794.504	0.019	**1.955**	**1.954**	867.615	0.011
Auto ETS	Single	4.040	3.944	2268.954	0.023	2.489	2.479	1534.811	0.019
		Italy				Japan			
MBB.Arima	Mean	2.533	2.595	1024.809	0.020	3.494	3.570	3507.538	0.022
	Median	2.527	2.559	625.712	0.012	3.426	3.486	2957.626	0.018
RSB.Arima	Mean	2.562	2.619	1243.208	0.024	3.575	3.645	3516.507	0.022
	Median	1.455	1.456	377.912	0.007	3.585	3.609	2922.618	0.018
Auto Arima	Single	3.314	3.407	1221.386	0.024	3.229	3.288	3267.591	0.020
MBB.Add.	Mean	2.502	2.556	917.959	0.018	3.494	3.570	3507.538	0.022
H-W	Median	2.409	2.439	607.370	0.012	3.426	3.486	2957.626	0.018
RSB.Add.	Mean	2.126	2.163	848.341	0.016	3.575	3.645	3516.507	0.022
H-W	Median	1.583	1.583	402.300	0.008	3.585	3.609	2922.618	0.018
Add H-W	Single	1.904	1.943	803.252	0.015	3.229	3.288	3267.591	0.020
MBB.Mult.	Mean	2.458	2.512	928.555	0.018	3.909	3.994	3638.494	0.023
H-W	Median	2.317	2.344	635.191	0.012	3.997	4.058	3053.883	0.019
RSB.Mult.	Mean	2.012	2.049	856.750	0.016	5.624	5.805	5008.852	0.031
H-W	Median	1.419	1.409	371.320	0.007	6.128	6.322	5071.051	0.032
Mult H-W	Single	1.829	1.868	818.579	0.016	3.735	3.770	3345.802	0.021
MBB.ETS	Mean	1.745	1.773	755.870	0.015	3.526	3.588	3255.967	0.020
	Median	1.609	1.596	402.184	0.008	3.594	3.660	2818.551	0.017
RSB.ETS	Mean	1.855	1.860	748.582	0.014	3.711	3.781	3490.502	0.022
	Median	**1.305**	**1.296**	**327.862**	**0.006**	3.044	3.057	**2635.629**	**0.016**
Auto ETS	Single	1.806	1.838	768.508	0.015	**2.274**	**2.233**	2687.012	0.016

Table 4.2 Forecast evaluation – developing countries (best results in **bold**)

Forecast Approach	Statistic	MAPE (%)	SMAPE (%)	RMSE (GWh)	TIC	MAPE (%)	SMAPE (%)	RMSE (GWh)	TIC
		Brazil				Mexico			
MBB.Arima	Mean	4.724	4.603	1933.379	0.025	3.531	3.585	1091.476	0.023
	Median	4.627	4.522	1784.111	0.023	**3.041**	**3.046**	**680.274**	**0.014**
RSB.Arima	Mean	4.368	**4.264**	1805.707	0.023	3.503	3.559	1100.538	0.023
	Median	**4.359**	4.266	**1675.724**	**0.021**	3.280	3.276	750.486	0.016
Auto Arima	Single	4.677	4.550	1943.011	0.025	3.092	3.122	968.572	0.020
MBB.Add.H-W	Mean	6.789	6.536	2803.963	0.035	4.554	4.674	1375.443	0.029
	Median	6.700	6.483	2574.868	0.032	4.241	4.333	1083.106	0.023
RSB.Add.H-W	Mean	6.447	6.222	2658.016	0.034	4.947	5.058	1446.868	0.030
	Median	6.250	6.061	2383.104	0.030	4.257	4.350	1106.579	0.023
Add H-W	Single	7.170	6.884	2961.887	0.037	5.128	5.298	1558.574	0.033
MBB.Mult.H-W	Mean	6.647	6.405	2745.442	0.035	4.608	4.728	1364.842	0.029
	Median	6.588	6.378	2495.533	0.031	4.156	4.162	995.764	0.021
RSB.Mult.H-W	Mean	6.381	6.162	2629.106	0.033	4.566	4.657	1345.790	0.028
	Median	6.122	5.940	2335.423	0.030	3.832	3.908	952.716	0.020
Mult H-W	Single	7.180	6.891	2973.818	0.037	4.779	4.911	1428.301	0.030
MBB.ETS	Mean	6.471	6.242	2661.628	0.034	6.192	6.441	1780.398	0.038
	Median	6.570	6.361	2502.552	0.032	6.086	6.278	1442.600	0.031
RSB.ETS	Mean	6.411	6.188	2649.286	0.033	6.341	6.610	1853.388	0.040
	Median	6.195	6.009	2366.104	0.030	6.046	6.234	1463.423	0.031
Auto ETS	Single	7.214	6.927	2965.903	0.037	6.921	7.228	1953.420	0.042
		Turkey							
MBB.Arima	Mean	2.644	2.632	712.075	0.016				
	Median	2.151	2.138	490.369	0.012				
RSB.Arima	Mean	2.744	2.729	724.887	0.017				
	Median	2.507	2.511	556.709	0.013				
Auto Arima	Single	2.277	2.279	681.329	0.016				
MBB.Add.H-W	Mean	3.079	3.038	755.223	0.017				
	Median	2.756	2.718	595.679	0.014				
RSB.Add.H-W	Mean	3.326	3.275	807.492	0.018				
	Median	2.993	2.949	642.197	0.015				
Add H-W	Single	2.623	2.594	707.484	0.016				
MBB.Mult.H-W	Mean	2.740	2.701	707.057	0.016				
	Median	**2.035**	**2.015**	**452.767**	**0.011**				
RSB.Mult.H-W	Mean	2.995	2.938	795.979	0.018				
	Median	2.392	2.364	495.507	0.012				
Mult H-W	Single	2.421	2.383	698.838	0.016				
MBB.ETS	Mean	2.512	2.527	724.285	0.017				
	Median	2.489	2.459	547.975	0.013				
RSB.ETS	Mean	2.686	2.728	830.497	0.019				
	Median	2.224	2.224	460.625	0.011				
Auto ETS	Single	2.913	2.984	932.234	0.022				

Figure 4.2 Electricity demand by country.

Best forecasts in red, actual values in blue. *Source*: DE OLIVEIRA & CYRINO OLIVEIRA (2018).



There was no consensus concerning the superiority of either ETS or ARIMA when combined with the Bagging algorithms employed in this essay. The (auto) ARIMA approach seemed to perform better for the Brazilian and Mexican cases, whereas the exponential smoothing methods adapted well for the monthly consumption in developed countries.

4.4.2
Comparison with other methods

For robustness checks, we compared the developed approaches with other univariate methods established in the literature. Care was taken to choose models that dealt with different stylized facts in electricity demand time series, such as nonlinearities, stochastic components (trend, seasonality, residuals), heteroscedasticity, among others. Particularly, we selected the following methods for comparison:

- a feedforward Artificial Neural Network (ANN) model (RUMELHART et al., 1985; AUER et al., 2008), to address complex nonlinear behavior;
- a feedforward ANN model with prior Box-Cox transformation (BC-ANN), in an attempt to ensure that residuals will be roughly homoscedastic;
- a univariate Support Vector Regression (SVR), an advanced machine learning algorithm, able to learn from training data and form complex non-linear decision boundaries (SMOLA & SCHÖLKOPF, 2004). To select the best subset of variables for prediction (in our case, the lagged values of the electricity demand with the most predictive power), the Correlation-based feature selection (CFS) algorithm (HALL, 1999) was used in each country's training set;
- the Theta method (ASSIMAKOPOULOS & NIKOLOPOULOS, 2000), a technique equivalent to a simple exponential smoothing with drift (with a particular restriction for this last component). The technique has performed particularly well in the M3-competition (MAKRIDAKIS & HIBON, 2000) for monthly series and for microeconomic data;

- two variations of the univariate Singular Spectrum Analysis (SSA) technique - a decomposition-reconstruction method that seeks to filter the noise and forecast the signal of an underlying time series using multiple steps (Embedding, Singular Value Decomposition, Grouping and Diagonal Averaging). In this work, we employ both the Recurrent SSA (RSSA) and the Vector SSA (VSSA) variations (GOLYANDINA et al., 2001).

The results obtained using the above methods are presented in Tables 4.3 and 4.4, for the developed and developing countries, respectively. For each country in the tables, the first and second rows refer to the two best (most accurate) forecasting methods from Table 4.1 (developed countries) or Table 4.2 (developing countries). For ANN formulations, the selected model is given in the form ANN (p,P,k) $_{[m]}$, where p is the number of lagged inputs (autoregressive terms), P is the number of autoregressive terms for the seasonal part of the time series, k is the number of nodes in the hidden layer and m is the seasonal frequency. For SVRs, numbers in parenthesis are the lag variables selected by the CFS algorithm for the training set. For the SSA models, the parameters refer to the window length (L) and the number of eigenvalues / eigentriples (r), in that order. The selection was made on the basis of the lowest Root Mean Squared Error (RMSE) for the calibration period (24 months before the out-of-sample period), i.e. the L and r parameters are the same from the model which demonstrated the lowest RMSE when forecasting for the period January 2013-December 2014.

The results outlined in Tables 4.3 and 4.4 endorse the superiority of the proposed bagging methods. The Japanese case remains the only exception, but now results are not conclusive in terms of the best forecasting technique for the 2015-2016 period. The auto ETS approach performed better in terms of MAPE and sMAPE for the Japanese electricity demand, whilst the Theta forecasts were slightly more accurate in terms of RMSE and TIC.

Table 4.3 Comparison with other methods – developed countries (best in **bold**)

Forecast Approach	**MAPE** (%)	**SMAPE** (%)	**RMSE** (GWh)	**TIC**	**Forecast Approach**	**MAPE** (%)	**SMAPE** (%)	**RMSE** (GWh)	**TIC**
	Canada					France			
RSB.Add.H-W - Median	**3.174**	**3.176**	1589.605	0.017	RSB.ETS - Median	**1.955**	**1.954**	867.615	0.011
RSB.Mult.H-W - Median	3.296	3.253	**1575.038**	**0.017**	MBB.ETS - Median	2.098	2.100	**815.629**	**0.011**
ANN (1, 1, 2) [12]	4.137	4.078	2334.346	0.024	ANN (2, 1, 2) [12]	3.406	3.440	1957.902	0.024
BC-ANN (1, 1, 2) [12]	4.138	4.085	2307.694	0.024	BC-ANN (1, 1, 2) [12]	3.303	3.370	2002.234	0.025
SVR (6, 20, 27, 84, 96)	8.509	8.550	4977.226	0.052	SVR (8, 96)	8.199	8.074	3881.359	0.048
Thetha	4.137	4.022	2385.858	0.025	Thetha	2.846	2.861	1666.439	0.021
RSSA (35, 33)	4.746	4.697	2661.387	0.028	RSSA (35, 7)	4.413	4.422	1980.674	0.024
VSSA (27, 26)	5.857	5.672	3245.611	0.033	VSSA (34, 11)	4.289	4.322	1897.870	0.023
	Italy					Japan			
RSB.ETS - Median	**1.305**	**1.296**	**327.862**	**0.006**	RSB.ETS - Median	3.044	3.057	2635.629	0.016
RSB.Mult.H-W - Median	1.419	1.409	371.320	0.007	Single ETS	**2.274**	**2.233**	2687.012	0.016
ANN (12, 1, 6) [12]	3.863	3.960	1405.311	0.027	ANN (7, 1, 4) [12]	4.318	4.310	4637.648	0.028
BC-ANN (12, 1, 6) [12]	3.393	3.520	1498.755	0.029	BC-ANN (6, 1, 4) [12]	5.804	5.856	5782.212	0.036
SVR (34, 99)	3.516	3.590	1613.364	0.031	SVR (2, 100)	7.368	7.243	6883.180	0.042
Thetha	2.065	2.107	839.800	0.016	Thetha	2.392	2.370	**2615.827**	**0.016**
RSSA (29, 21)	3.078	3.053	1107.605	0.021	RSSA (21, 7)	5.276	5.443	5254.590	0.033
VSSA (30, 16)	2.791	2.866	1127.547	0.022	VSSA (23, 11)	4.424	4.591	4869.373	0.030

Notes on the parameters for each method: See main text from Section 4.4.2.

Table 4.4 Comparison with other methods – developing countries (best in **bold**)

Forecast Approach	**MAPE** (%)	**SMAPE** (%)	**RMSE** (GWh)	**TIC**	**Forecast Approach**	**MAPE** (%)	**SMAPE** (%)	**RMSE** (GWh)	**TIC**
	Brazil					Mexico			
RSB.ARIMA - Mean	4.368	**4.264**	1805.707	0.023	MBB.ARIMA - Median	**3.041**	**3.046**	**680.274**	**0.014**
RSB.ARIMA - Median	**4.359**	4.266	**1675.724**	**0.021**	Single ARIMA	3.092	3.122	968.572	0.020
ANN (1, 1, 2) [12]	5.531	5.360	2259.505	0.029	ANN (1, 1, 2) [12]	6.762	7.019	2042.153	0.043
BC-ANN (1, 1, 2) [12]	5.415	5.251	2212.039	0.028	BC-ANN (2, 1, 2) [12]	5.810	5.998	1750.736	0.037
SVR (2, 3, 4, 100)	5.069	4.910	2160.605	0.027	SVR (2, 100)	15.040	16.749	4538.936	0.102
Thetha	5.153	5.012	2086.524	0.026	Thetha	6.545	6.820	1860.834	0.040
RSSA (26, 13)	6.318	6.072	2751.797	0.035	RSSA (35, 7)	3.557	3.567	1080.091	0.022
VSSA (20, 11)	7.033	6.722	3059.962	0.038	VSSA (24, 10)	3.303	3.341	1042.185	0.022
	Turkey								
MBB.Mult.H-W - Median	**2.035**	**2.015**	**452.767**	**0.011**					
MBB.ARIMA - Median	2.151	2.138	490.369	0.012					
ANN (1, 1, 2) [12]	3.604	3.490	1320.487	0.030					
BC-ANN (2, 1, 2) [12]	3.387	3.463	1028.813	0.024					
SVR (2, 100)	5.723	6.022	1887.737	0.045					
Thetha	2.914	2.975	898.173	0.021					
RSSA (25, 7)	3.559	3.503	884.791	0.020					
VSSA (39, 24)	6.174	5.941	1457.826	0.033					

Notes on the parameters for each method: See main text from Section 4.4.2.

4.4.3
Discussion

The performance gains demonstrated by the Bagging approaches are remarkable as accurate forecasts are decisive for assertive profit/cost management and investment decisions, as well as for the definition of sectoral policies in a local or national scale. For the energy sector, particularly, precise mid/long-term demand forecasting is of the utmost importance for several decision-making processes, such as the construction of new generation facilities, the purchase of existing generating units, the development of transmission and distribution systems, among others. In a more general sense, accurate forecast results are also paramount to reach agreements between different stakeholders (generators, transmitters, distributors, traders, consumers, investors, government and national regulation institutes).

It should be noted that a considerable amount of the variation in the monthly electric energy consumption is due to external factors, which cannot be captured by univariate forecasting methods. Some remarkable examples are the influences of the electric energy generation and, particularly, Industrial Output in several countries. For the Brazilian case, for instance, the Industrial Sector accounted for almost 43% of the total electric energy consumption between the years of 2006 and 2014 (train period). Another point worth noting is the important role that the Gross Domestic Product (GDP) plays in electric energy consumption behavior (KUCUKALI & BARIS, 2010; BURKE & CSEREKLYEI, 2016). By quickly glancing the GDP data in Brazil, one may notice substantial falls in its common trend, reflecting the recent political and economic turmoil in the country. Along with the energy rationing and the lower industrial output, this might have been an important factor for the substantial decline in the energy demand in Brazil in the last years.

Notwithstanding the above, formulations that consider external influences on the variable of interest usually yield satisfactory results when simulating historical data but fail to perform well in forecasting several steps ahead (more than 12 steps, as in our case). On these grounds, the combination of Bagging approaches and univariate forecast methods emerges as a promising alternative to predict mid-/long-term behavior for a broad variety of time series in different economic sectors.

4.5
Main conclusions from the first essay

This essay proposed an alternative method, here addressed as Remainder Sieve Bootstrap (RSB), to generate the ensemble of forecasts prior to final aggregation in Bagging routines. It also represented the first endeavor to consider the use of Bagging in the context of electric energy demand forecasting. The obtained results attested that the developed method could improve over a range of benchmarks, such as univariate time series methods and machine learning techniques. The method also performed equally well (being superior in some cases) to alternative Bagging approaches to forecasting, such as BERGMEIR's et al. (2016) Bagged.BLD.MBB.ETS method. This constituted an important contribution to the field of forecasting at the time of the publication.

A range of suggestions for future research were made by the time of the publication such as the use of alternative decomposition and resampling schemes. A top-down disaggregation approach before proceeding to the forecasting routines was also indicated. This entailed applying the proposed approach for each class of consumption (Industrial, Commercial, Residential and Other Sectors) and then pooling together the forecasts to obtain the estimates of the total demand. Such sector-specific studies would provide a more in-depth understanding of the demand for electric energy across different countries and could further enhance the accuracy of forecasts.

5
Second essay: Ensemble approaches and regularization techniques to natural gas consumption and energy supplied forecasts

In this second contribution, a novel forecasting approach combining Bootstrap aggregating (Bagging) algorithms, time series methods and regularization techniques was introduced. A new variant of Bagging, in which the set of classifiers is built by means of a Maximum Entropy Bootstrap (MEB) routine, was also put forth. The approach was evaluated on two types of monthly energy demand across a wide range of European/OECD-European economies: energy supplied and gross inland natural gas consumption. These series were chosen because of their relevance to real world operational problems and decision making in energy and environmental policy.

It should be emphasized that the use of a proper variable weighting technique in the aggregation of forecasting ensembles (in our case, the regularization routines) was an issue that had not yet been addressed for Bagging approaches when the work was put forth. To the best of our knowledge, there had been only one work partially addressing this issue (DANTAS & CYRINO OLIVEIRA, 2018), but the technique developed was only capable of feature selection, i.e., the selected variables were still given equal weights in the final (aggregation) phase.

The empirical experiments and robustness checks conducted throughout the work demonstrated the superiority, in terms of forecasting accuracy, of the proposed approach over traditional forecasting methods and over recently developed Bagging routines for forecasting. The paper originating from this essay, entitled “A novel approach to ensembles applied to energy consumption time series”, is currently being considered for publication.

Despite the importance of the literature overview in forecasting energy supplied and gross inland natural gas consumption across different countries/regions, in this essay we focused only on the involved methodology from the second paper/contribution, its results and the overall discussion. The rest of the paper can be made available upon request.

5.1
Proposed methodology

In spite of the substantial improvements in forecasting performance brought by ensemble methods, as illustrated in the last essay, resampling and variable weighting schemes appear to have only been partially addressed in the literature. Resampling in most Bagging routines for forecasting, for instance, has been mostly conducted via the modified Moving Blocks Bootstrap (MBB) algorithm (BERGMEIR et al., 2016; DANTAS et al., 2017; PETROPOULOS et al., 2018; DANTAS & CYRINO OLIVEIRA, 2018). However, the MBB is very sensitive to the choice of the block size, for which there is currently no consensus in the literature on what would be optimal for different types of series. In addition, MBB, like most bootstrapping approaches, repeats original values while not using many others, and values that are in the neighbourhood of observed points in the time series cannot be included in a replica.

Concerning current limitations in variable weighting schemes, the BMC approach of PETROPOULOS et al. (2018), presented in details in Section 3.2, is restricted to exponential smoothing. The Bagged.Cluster.ETS method of DANTAS & CYRINO OLIVEIRA (2018) (Section 3.3), in turn, is also not problem-free since: (i) the number of clusters needs to be defined (an automatic procedure is offered by the authors, but it does not guarantee the best results); (ii) the method is very computational intensive, when compared with other ensemble-based routines; and (iii) only feature selection is achieved, as the selected variables are still equally weighted in the aggregation phase. Furthermore, as GUO & LUH (2004) highlighted, the weights in ensembles can reflect the overall historical prediction performance, but are unlikely to exploit the information in current input data.

Hence, traditional ensembles are generally unable to make the best use of all the available information at the time of forecasting.

Given the current state of the literature, this essay proposes an ensemble approach that includes (i) a resampling algorithm that expands the range of values in replicas and is not conditional on pre-selection of key parameters and (ii) regularization while combining forecasts. This proposal therefore develops a hybrid approach that draws on knowledge from statistics, machine learning and forecasting, which are fields that until recently had been developed separately, as observed by WERON (2014) in his review of the state-of-the-art in forecasting electricity prices.

5.1.1
Resampling via the Maximum Entropy Bootstrap

The first part of the proposed approach is akin to the Bagged.BLD.MBB.ETS procedure proposed by BERGMEIR et al. (2016), since it involves generating replicas for the remainder component of an STL decomposition applied to a Box–Cox (BC) transformed time series. However, instead of using the Moving Blocks Bootstrap (MBB) algorithm to replicate the remainder, a Maximum Entropy Bootstrap (MEB) routine is adopted, so that ensembles are created from a density distribution that satisfies the maximum entropy principle (VINOD & LÓPEZ-DE-LACALLE, 2009). To the best of our knowledge, this method has not been adopted in this context. After bootstrapping the remainder, the trend and seasonal components are added, and the BC transformation is inverted. The procedure is repeated J times, J being the number of desired replicas. Subsequently, forecasts are generated for each series in the pool, using ETS and ARIMA formulations. The proposed resampling and forecasting procedures are summarized in Figure 5.1.

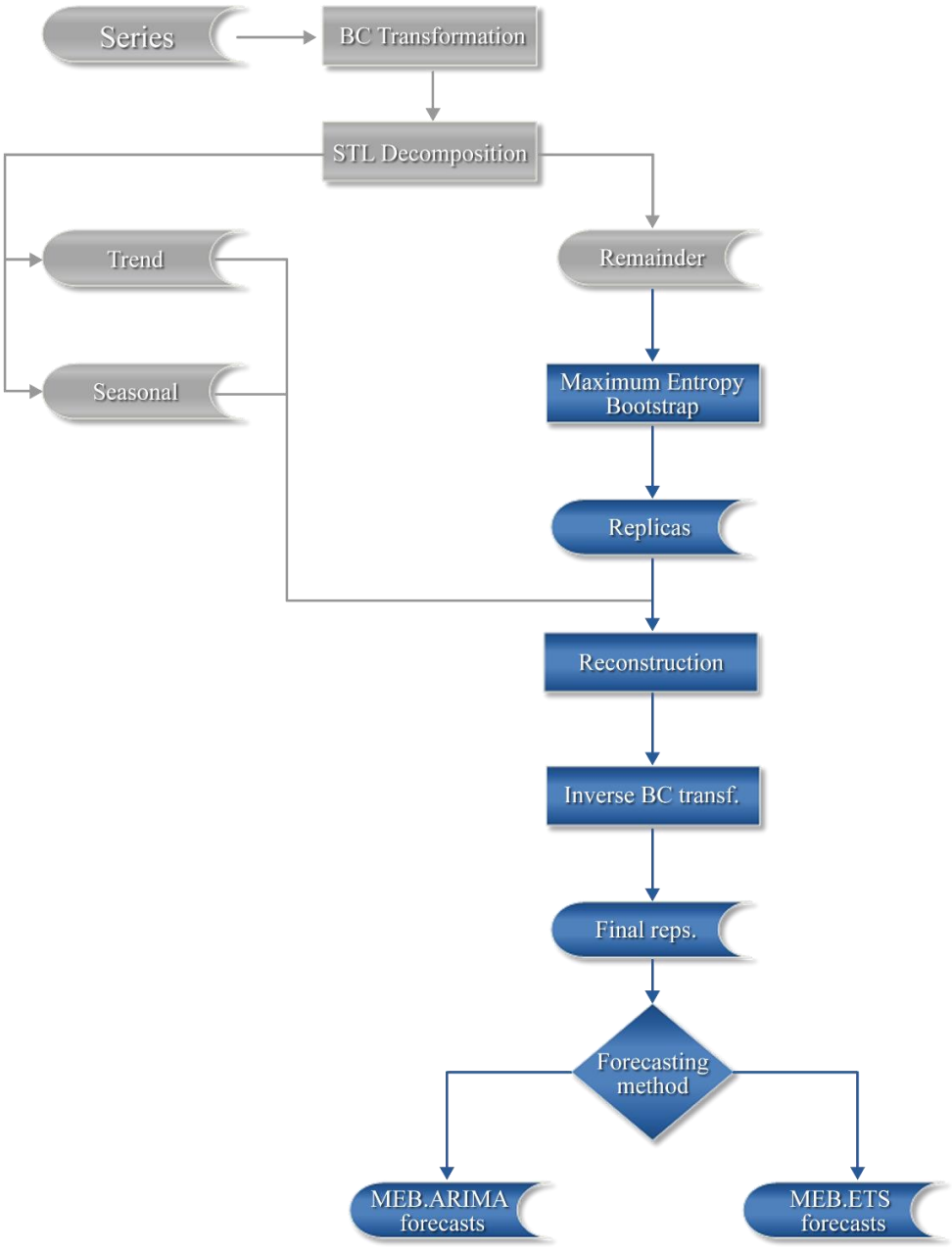


Figure 5.1 MEB – Data treatment, resampling and forecasting stages.
Source: The authors.

The Maximum Entropy Bootstrap (MEB) approach was devised by VINOD (2004) as an alternative resampling procedure for non-stationary time series or in which the stationarity hypothesis is difficult to ascertain. It involves constructing replicates of an original series by means of a seven-step algorithm designed to satisfy the ergodic theorem, ensuring that the grand mean of all ensembles is close

to the original sample mean. In brief terms, for a x_t time series of size T, the following steps are proposed:

1) Sorting of the original data in increasing order to create order statistics $x_{(t)}$ and storing of the ordering index vector;

2) Computation of the intermediate points from the order statistics: $z_{(t)} = x_{(t)} - x_{(t-1)}/2, t = 2,\ 3, \dots, T-1$;

3) Calculation of the trimmed mean (m_{trm}) of the deviations $x_{(t)} - x_{(t-1)}$ among all consecutive observations. In addition, computation of the lower and upper limits of the density, $z_0 = x_{(1)} - m_{trm}$ and $z_T = x_{(T)} + m_{trm}$, respectively;

4) Construction of a maximum entropy density function with the *z* values as limiting points. The density is built by joining uniform distribution intervals of equal probability. The uniform densities are also designed to satisfy the mean-preserving constraint (and eventually the ergodic theorem). To that end, the interval means for the uniform density, m_t, must satisfy the following relations:

$$\begin{gathered} m_1 = 0.75x_{(1)} + 0.25x_{(2)} \\ m_k = 0.25x_{(k-1)} + 0.50x_{(k)} + 0.25x_{(k+1)}, \\ k = 2, \dots, T-1 \\ m_T = 0.25x_{(T-1)} + 0.75x_{(T)} \end{gathered} \tag{16}$$

5) Inverse transforming sampling: generation of T random numbers from the [0, 1] uniform interval, computation of sample quantiles of the ME density at those points and sorting in increasing order;

6) Reordering of the sorted sample quantiles by using the ordering index of step 1. This recovers the time dependence relationships of the original data;

7) Finally, steps 1–6 are repeated until the desired number of replicas is achieved.

As noted throughout the steps, the MEB procedure offers appealing characteristics, such as the retention of the basic shape and time dependence structure of the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) of the original series for its bootstrapped versions without having to resort to shape-destroying transformations like detrending or differencing to achieve stationarity (VINOD & LÓPEZ-DE-LACALLE, 2009). Furthermore, the

results are not sensitive/conditional on pre-selection of key parameters, as in other bootstrap procedures[4]. Besides avoiding stationarity, VINOD (2006) adds that the MEB procedure circumvents three other limiting properties of the traditional independent identically distributed (*iid*) bootstrap:

i) The traditional resample obtained from shuffling with replacement repeats some original values while not using as many others. It never admits nearby data values in a resample. *A priori*, there is no reason to believe that values near the observed x_t are impossible;

ii) Traditional resamples must lie in the closed interval $[\min(x_t), \max(x_t)]$. Since the observed range is random, we cannot rule out somewhat smaller or larger x_t. Note that the third step of the MEB algorithm implies a less restrictive/wider range;

iii) Traditional bootstrap routines involve shuffling x_t in a way that any dependence information in the time series sequence $(x_1, \dots, x_t, x_{t+1}, \dots, x_T)$ is lost. If one tries to restore the original order to the shuffled resample, he/she ends up with essentially the original set x_t, except that some dropped values are replaced by the repeats of adjacent values. Hence, it is impossible to generate a large number of sensibly distinct resamples in a traditional bootstrap shuffle without admitting nearby values.

In addition to the above, the MEB is of straightforward implementation and is available in different statistical packages[5]. For such reasons, the procedure has been effectively employed in different empirical applications, such as: time series inferences related to the Asian economy (VINOD, 2006); investigation of associations between energy consumption and economic health in Turkey (YALTA, 2011); estimation of air temperature quantiles in certain regions of Central Europe (BARBOSA et al., 2011). Notwithstanding its increased popularity,

[4] To be fair, replicates originating from MEB may vary according to the definition of the trimming parameter for the computation of the limiting intermediate points – see the algorithm steps for further details. This is more of a practitioner choice (most practical implementations are taken using a 10% trimmed mean for the deviations) rather than a decisive factor, such as the block size definition in the MBB procedure.

[5] In R, the MEB can be implemented using the meboot( ) function of the *meboot* package (VINOD & LÓPEZ-DE-LACALLE, 2009). Following a commmon practice, we set the trimming proportion to 10% by adding trim = 0.10.

we are unaware of previous applications of the MEB procedure in the context of time series forecasting, especially as part of a Bagging approach.

5.1.2
Combination via Regularization

The second main difference between our proposed methodology and previous Bagging routines for forecasting lies in the way the pool of forecasts are combined. Instead of using the mean or the median, regularization routines assign weights for each forecast in the selected ensemble under a multiple regression framework. The idea is to substantially reduce the variance of the final forecasting model at the cost of introducing some bias, an approach which has proven to be very beneficial for the predictive performance of the model when (i) there are many predictors; and/or (ii) the predictors are highly correlated with each other. Both situations are both clearly present in Bagging routines for forecasting.

For illustrative purposes, Figure 5.2 demonstrates how the bias-variance trade-off works under a multiple regression setting and how the search for the optimal model complexity is conducted in such case. In brief terms, a model's error can be decomposed into three parts: the error resulting from a large variance, the error resulting from significant bias, and a remainder (unexplainable part). As the model complexity (in our case, the number of forecasted bootstraps) increases, the bias decreases. An unbiased ordinary least squares (OLS) estimate, for instance, would deliver a result on the right-hand side of the picture, which is far from optimal. The main rationale of regularization is thus to lower the variance at the cost of some bias, moving left on the plot, towards the optimum.

There are basically two types of regularization techniques: The Ridge regression (HOERL & KENNARD, 1970) and the Least Absolute Shrinkage and Selection Operator (LASSO)[6] (TIBSHIRANI, 1996, 2011). In both cases, the

[6] There are also the so-called elastic-net models, which are something between the RIDGE and the LASSO formulations, obtained by varying the α, the elastic-net penalty parameter over the range of 0 (Ridge) – 1 (LASSO) – see FRIEDMAN et al. (2010) for further details. As a side note, we have

traditional OLS loss function is augmented in such a way that one not only minimizes the sum of squared residuals but also penalizes the size of parameter estimates.

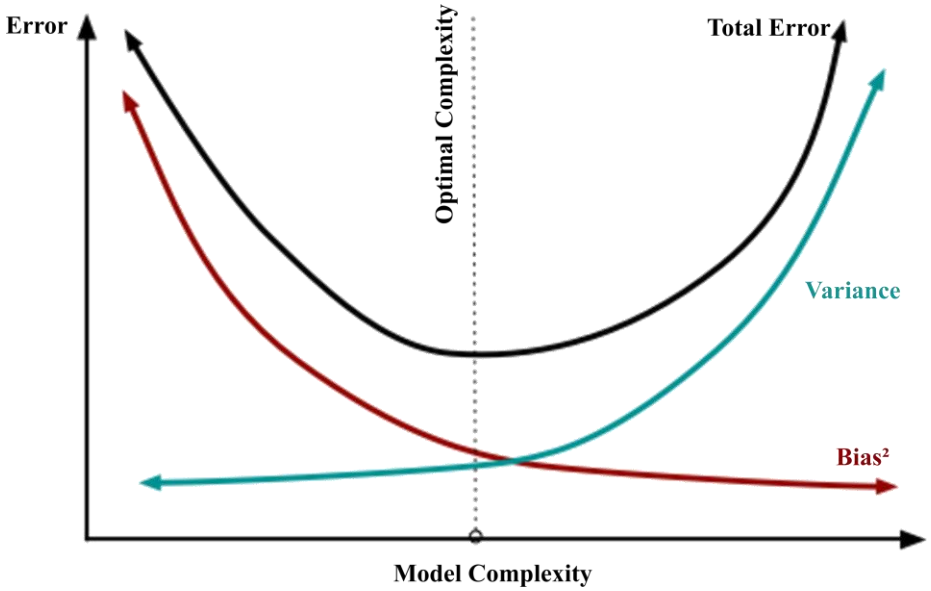


Figure 5.2 Bias-Variance trade-off.
Source: Adapted from FORTMANN-ROE (2012).

Supposing that we are aiming at predicting *n* observations of the response variable, *Y*, with a linear combination of *m* predictor variables, *X*, and a normally distributed error term with σ^2 variance. In this case, under Ridge, the loss function is defined as:

$$L_{Ridge}(\hat{\beta}) = \sum_{i=1}^{n}\left(y_i - x'_i\hat{\beta}\right)^2 + \lambda\sum_{j=1}^{m}\hat{\beta}_j^2 \qquad (17)$$

where λ is the regularization penalty parameter. Minimizing the above formula gives the Ridge regression estimates $\hat{\beta}_{Ridge} = (X'X + \lambda I)^{-1}(X'Y)$, where *I* stands for the identity matrix. One can easily note that as $\lambda \to 0, \hat{\beta}_{Ridge} \to \hat{\beta}_{OLS}$ and as $\lambda \to \infty, \hat{\beta}_{Ridge} \to 0$.

considered several versions of these models, but they did not offer substantial improvements over Ridge and LASSO for the series considered.

By incorporating the regularization coefficient in the formulas for bias and variance we obtain:

$$Bias\left(\hat{\beta}_{Ridge}\right) = -\lambda(X'X + \lambda I)^{-1}\beta$$
$$Var\left(\hat{\beta}_{Ridge}\right) = \sigma^2(X'X + \lambda I)^{-1}X'X\,(X'X + \lambda I)^{-1} \quad (18)$$

From the above equation, we observe that as λ becomes larger, the variance decreases, and the bias increases. This leaves us with the following question: how much bias are we willing to accept in order to decrease the variance?

There are basically two strategies to tackle this issue. A more traditional approach would be to choose λ in a way that some information criterion is the smallest. An alternative is to perform cross-validation and select the value of λ that minimizes the cross-validated sum of squared residuals (or some other measure). The former approach emphasizes the model's fit to the data and the relative impact of exogenous inputs in the variable of interest, while the latter is more focused on its predictive performance. In this work, we follow this second strategy. Basically, we choose a set of *P* values of λ to test, split the dataset into *K* folds, and select the optimal λ according to the following algorithm:

Algorithm 1 Choice of lambda

procedure cross-validation(*P* = nlambda, *K* = nfolds)
  for p in 1 to *P* **do**
    for k in 1 to *K* **do**
      keep fold k as hold-out data
      use the remaining folds and $\lambda = \lambda_p$ to estimate $\hat{\beta}_{Ridge}$
      predict hold-out data: $y_{test,k} = X_{test,k}\,\hat{\beta}_{Ridge}$
      compute the sum of squared residuals: $SSR_k = \left\|y - y_{test,k}\right\|^2$
    end for k
    average SSR over the folds: $SSR_p = 1/k \sum_{k=1}^{K} SSR_k$
  end for p
  choose optimal λ value: $\lambda_{opt} = \underset{p}{argmin}\, SSR_p$
end procedure

where $\|\ \ \|^2$ is the quadratic norm.

For practical implementations, we use the cv.glmnet( ) function from the *glmnet* package in R (FRIEDMAN et al., 2010) and consider $K = 10$ cross-validation folds and $P = 1000$ possible lambda values, whose sequence is defined by the own function. The choice of λ_{opt}, i.e., the value of lambda which minimizes the averaged sum of squared residuals, is conducted using a validation set of the same size of the test set.

An important feature of our approach is that forecasts are generated only once for the period comprising both the validation and combining phases. For example, consider the case of forecasting a monthly series 12 steps (1 year) ahead via a regularized ensemble. In our approach, forecasts for each replica are computed up to 24 steps-ahead: the first twelve steps comprise the validation set and are used to find the optimal weights of the regularized model; the last half (steps 13 to 24) is then used for the combination, using the optimal weights obtained in the validation set (first half). Other alternatives, which were also considered, are: (i) generate first forecasts for the validation set only; conduct cross-validation to obtain the optimal weights; generate the forecasts for the combination period, but considering the validation set as part of the train set; and then use the weights obtained in the validation phase to combine these last forecasts; or (ii) conduct the cross-validation using the replicas (not their forecasts) as predictors; generate forecasts for the combination period; and combine then using the optimal weights obtained for the replicas. The rationale behind our approach is that, by conducting validation and combination in the same set of forecasts, we do not modify the data generating process of the forecasts. This is a subtle difference that can significantly improve the accuracy of the final forecast.

Finally, we also conduct the LASSO regularization technique following the same guidelines as depicted above. Under LASSO, the loss function is defined as:

$$L_{LASSO}(\hat{\beta}) = \sum_{i=1}^{n} \left(y_i - x'_i\hat{\beta}\right)^2 + \lambda \sum_{j=1}^{m} |\hat{\beta}_j| \qquad (19)$$

As in Ridge, LASSO also adds a penalty for non-zero coefficients, but unlike the former, which penalizes the sum of squared coefficients (L2 penalty), LASSO penalizes the sum of their absolute values (L1 penalty). As a result, for high

values of λ, many coefficients are exactly zeroed under LASSO. Ridge regularization, by contrast, always keeps some information from all the predictors.

5.2
Applications

The empirical analysis in the second essay included two sets of series across several European economies: (i) Gross Inland Natural Gas Consumption (in terajoules, TJ); and (ii) Energy Supplied (in gigawatt-hour, GWh). Data for the former were collected from the Statistical Office of the European Union database (EUROSTAT, 2019). Data for the latter were compiled from the International Energy Agency (IEA) Monthly Electricity Statistics reports, which provide information on energy production and trade for all OECD Member Countries (IEA, 2019).

The analysis spanned from January 2008 to May 2019 (the last date available for all involved European countries). Data from January 2008 to May 2018 was considered as training set for models using the median for aggregation. When employing regularization, a validation set was included between June 2017 and May 2018, in which weights were assigned to each of the forecasts in the selected ensemble (MEB.ETS or MEB.ARIMA). The test set comprised the last 12 observations (June 2018 – May 2019). A total of 18 countries were selected in each dataset, including the main consumers, namely France, Germany, Italy, Netherlands, Spain and United Kingdom. Figure 5.3 depicts the train set of selected countries, to provide a basis for comparison. As can be noted, the series differ considerably, highlighting the challenge faced by forecasters.

We compared forecasts of the proposed approaches with those from several forecasting methods, ranging from traditional benchmarks to the most recent Bagging routines for forecasting. They are summarized in Table 5.1. We further clarify that implementation was conducted using the R programming language (R CORE TEAM, 2019) and related packages. More specifically, we used R version 3.5.0 (2018-04-23) and *forecast* version 8.8 for ETS and ARIMA modelling.

Furthermore, a parallel implementation was adopted, where the following packages were used: *doSNOW* (1.0.16), *foreach* (1.4.4) and *snow* (0.4–3).

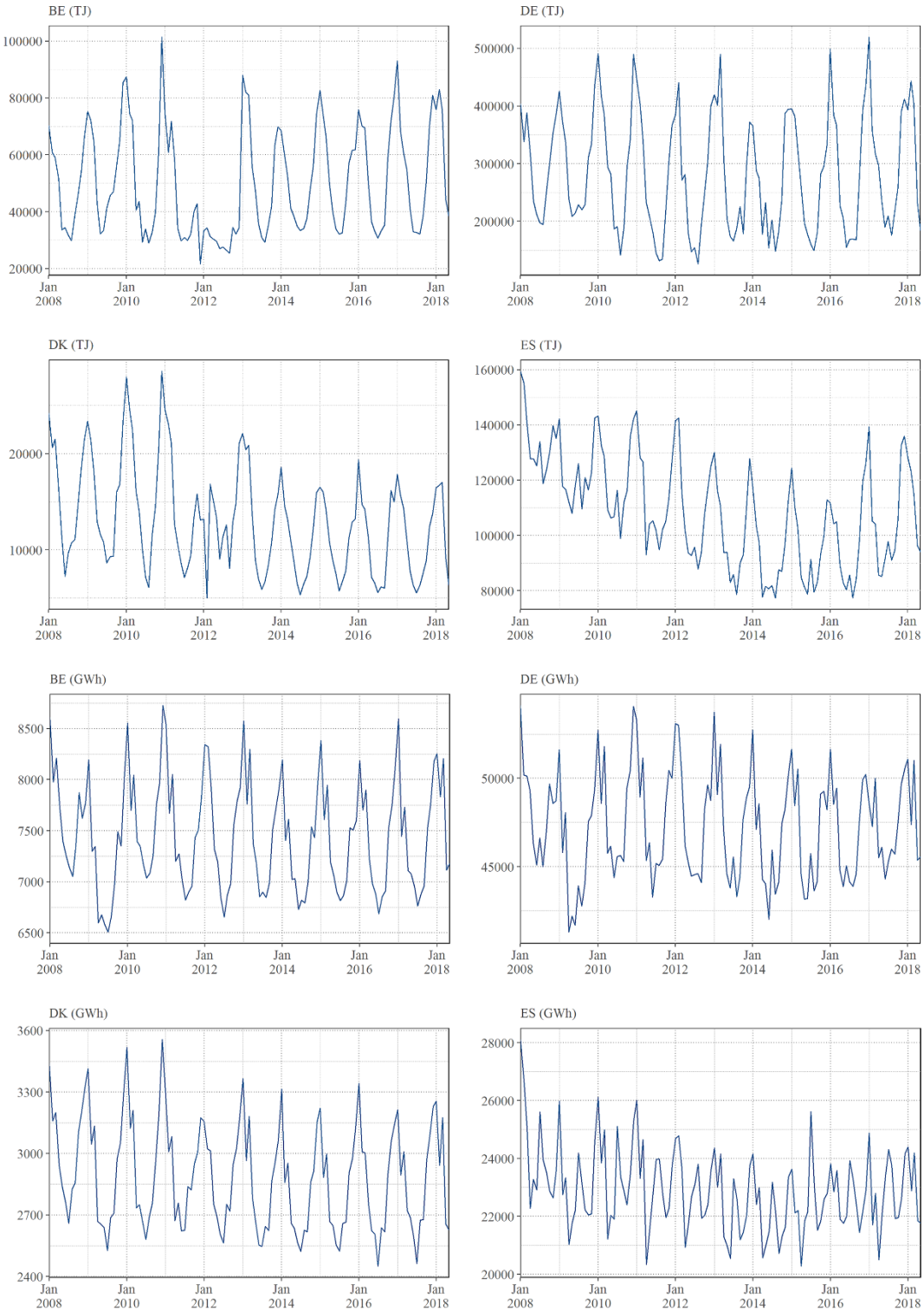


Figure 5.3 Gross inland natural gas consumption in terajoules (TJ) and energy supplied in gigawatt-hour (GWh).

Train set, sample 4 countries - Belgium (BE); Germany (DE); Denmark (DK); and Spain (ES). *Sources:* EUROSTAT (2019) and IEA (2019).

Table 5.1 Selected methods for comparison

Method	Implementation / Source	Short description
Traditional Benchmarks		
ETS	R *forecast* package ets( ) function	auto Error, Trend and Seasonality specification
ARIMA	R *forecast* package auto.arima( ) function	automatically-selected (S)ARIMA model
Additive HW	R *forecast* package hw( ) function[1]	three parameter Additive Holt-Winters method
Multiplicative HW	R *forecast* package hw( ) function[2]	three parameter Multip. Holt-Winters method
Competing Bagging approaches		
Bagged.BLD.MBB.ETS	BERGMEIR et al. (2016)	see Section 3.1 for details
BMC	PETROPOULOS et al. (2018)	see Section 3.2 for details
Bagged.Cluster.ETS	DANTAS & CYRINO OLIVEIRA (2018)	see Section 3.3 for details

Notes: ets( ) and auto.arima( ) are used for model selection. The forecast( ) function must be used on the output to generate the forecasts. [1]Set seasonal argument to "additive"; [2]Set seasonal argument to "multiplicative".

For this particular experiment, we opted to generate 99 replicas for each ensemble. To facilitate the replication of our results, all resampling procedures were conducted using the same random seed, which was set to 123 using the set.seed( ) function in R before bootstrapping. To gauge the overall accuracy of the forecasts, we summarized the results according to the mean across all involved series of the set of metrics specified in Table 5.2. The choice of metrics (specially sMAPE) was mainly to allow comparability with published results, thus providing a common ground for discussion. It should be noted that sMAPE and MASE were the official evaluation metrics for point forecasts in the M4 Competition (MAKRIDAKIS et al., 2018). The MASE is a scale-free metric devised by HYNDMAN & KOEHLER (2006) as a generally applicable measurement of forecast accuracy. As for RMSE,

although averaging its values across multiple series is unusual, it provides an estimate of how much energy (in TJ or GWh) can be "saved" by opting for a more accurate forecasting approach in comparison with other methods.

Table 5.2 Evaluation metrics

Metric	Formula	Unit of measurement
Root Mean Squared Error (RMSE)	$\sqrt{\frac{\sum_{t=1}^{h}\left(Y_t - \widehat{Y}_t\right)^2}{h}}$	Same as the original series
Symmetric Mean Absolute Percentage Error (sMAPE)	$\frac{1}{h}\sum_{t=1}^{h}\frac{2\left\|Y_t - \widehat{Y}_t\right\|}{\left\|Y_t\right\| + \left\|\widehat{Y}_t\right\|} \times 100$	Percentage points (%)
Mean Absolute Scaled Error (MASE)	$\frac{1}{h}\frac{\sum_{t=1}^{h}\left\|Y_t - \widehat{Y}_t\right\|}{\frac{1}{n-m}\sum_{t=m+1}^{n}\left\|Y_t - Y_{t-m}\right\|}$	Dimensionless

Notes: Y_t and $\widehat{Y}_t$ are the real (actual) and forecasted values of the underlying series, respectively; h is the forecasting horizon (number of forecasting steps ahead); m is the seasonal period.

5.3
Results and Discussion

5.3.1
Results

The results are summarized in Tables 5.3 and 5.4 (where best performance is highlighted in **bold**), for the cases of total natural gas consumption and total energy supplied, respectively. Averages of performance metrics across all selected countries are provided.

Overall, the most accurate forecasts in terms of RMSE and MASE were delivered by combining the MEB algorithm for resampling and the Ridge regularization routine as aggregation method. The same was observed for energy supplied time series according to sMAPE. However, based on sMAPE, the most accurate forecats for natural gas consumption were obtained via a combination of

MEB resampling, ETS forecasts and LASSO regularization. The performance of this combination, however, was similar to the one delivered by the MEB.ETS.Ridge approach.

Table 5.3 Forecast evaluation: Natural gas consumption

Resampling algorithm	Forecasting approach	Combining method	Average RMSE (TJ)	Average sMAPE (%)	Average MASE
Proposed approaches					
MEB	ETS	Ridge	7625.495	8.193	0.566
MEB	ETS	LASSO	7719.702	**8.154**	0.567
MEB	ARIMA	Ridge	**6800.515**	8.678	**0.552**
MEB	ARIMA	LASSO	7158.693	9.026	0.567
Median aggregation					
MEB	ETS	Median	7839.111	10.147	0.626
MEB	ARIMA	Median	7267.769	10.921	0.622
Alternative Bagging approaches					
MBB	ETS	Median[a]	8092.746	9.230	0.608
MBB	ETS	BaggedCluster[b]	8017.247	9.135	0.605
MBB	ETS	BMC[c]	7926.985	9.615	0.615
Traditional Benchmarks					
None	ETS	Single	8084.310	10.409	0.641
None	ARIMA	Single	7295.507	10.908	0.621
None	Add HW	Single	8056.783	11.048	0.644
None	Multip HW	Single	7906.997	9.463	0.621

Notes: Overall results (average of the evaluation metrics across all countries) considering 12 steps ahead forecasts (best in **bold**). MBB and MEB stand for Moving Blocks Bootstrap and Maximum Entropy Bootstrap, respectively. [a, b, c] stand for the methods proposed in BERGMEIR et al. (2016), DANTAS & CYRINO OLIVEIRA (2018) and PETROPOULOS et al. (2018), respectively. Block size for the MBB algorithm in these methods comprises 24 observations, following the same guidelines as the authors in their original papers. BMC is the abbreviation for Bootstrap Model Combination. HW is the Holt-Winters Method. Pretreatment for all ensemble methods involves using BC transformation and STL decomposition prior to resampling.

Table 5.4 Forecast evaluation: Energy supplied

Resampling algorithm	Forecasting approach	Combining method	Average RMSE (GWh)	Average sMAPE (%)	Average MASE
Proposed approaches					
MEB	ETS	Ridge	411.960	2.911	0.830
MEB	ETS	LASSO	425.832	2.933	0.840
MEB	ARIMA	Ridge	**371.904**	**2.759**	**0.778**
MEB	ARIMA	LASSO	377.937	2.814	0.804
Median aggregation					
MEB	ETS	Median	475.540	3.315	0.976
MEB	ARIMA	Median	380.166	2.832	0.804
Alternative Bagging approaches					
MBB	ETS	Median[a]	451.480	3.170	0.917
MBB	ETS	BaggedCluster[b]	448.410	3.150	0.915
MBB	ETS	BMC[c]	474.504	3.330	0.982
Traditional Benchmarks					
None	ETS	Single	478.981	3.336	0.983
None	ARIMA	Single	389.915	2.889	0.820
None	Add HW	Single	453.338	3.400	0.992
None	Multip HW	Single	453.246	3.264	0.957

Notes: See Table 5.3.

Concerning the choice of the forecasting method, regularized ensembles seem to benefit from the use of ARIMA models during forecasting. However, if we consider ensembles aggregated using the median, MEB.ARIMA results are usually poorer than MEB.ETS. That is, forecasting each series in the artificial ensemble with ARIMA models may initially bring more variance to the ensemble, but this variance is usually handled well by regularization routines. In light of this fact, greater gains from regularization techniques are expected in ensembles whose components were generated using Neural Networks (NNs), Support Vector

Regressions (SVRs), and other methods that have a large number of parameters and can result in high variance between committee members in the ensemble. We leave this as a direction for future research, since it is beyond the scope of the present study.

5.3.2
Robustness checks

In this section we considered forecasting performance under alternative settings, such as different resampling methods, forecasting horizons and ensemble sizes (number of series to be combined). We started by comparing the proposed methods depicted from Tables 5.3 and 5.4 with similar approaches, with the exception that the MBB was this time used as an alternative algorithm in the resampling phase, to assess the potential differences between MEB and MBB in ensemble generation. The results for natural gas consumption and energy supplied forecasts are summarized in Table 5.5. They show that ensembles that considered the MEB for resampling provided more accurate forecasts than the ones based on MBB for resampling. A possible explanation lies in the way ensembles are created according to these two algorithms: MEB-generated ensembles are more diversified since MEB admits values near the original time series observations, as opposed to MBB.

Table 5.5 Robustness checks: comparisons with the MBB algorithm

Resampling algorithm	Forecasting approach	Combining method	Average RMSE (TJ)	Average sMAPE (%)	Average MASE
I. Natural gas consumption forecasts					
Proposed approaches using MEB for resampling					
MEB	ETS	Ridge	7625.495	8.193	0.566
MEB	ETS	LASSO	7719.702	**8.154**	0.567
MEB	ETS	Median	7839.111	10.147	0.626
MEB	ARIMA	Ridge	**6800.515**	8.678	**0.552**
MEB	ARIMA	LASSO	7158.693	9.026	0.567
MEB	ARIMA	Median	7267.769	10.921	0.622
Proposed approaches using MBB for resampling					
MBB	ETS	Ridge	7854.193	8.396	0.574
MBB	ETS	LASSO	8703.086	8.968	0.622
MBB	ETS	Median	8092.746	9.230	0.608
MBB	ARIMA	Ridge	7943.413	9.077	0.603
MBB	ARIMA	LASSO	9394.249	10.824	0.692
MBB	ARIMA	Median	8174.486	11.394	0.671
II. Energy supplied forecasts					
Proposed approaches using MEB for resampling					
MEB	ETS	Ridge	411.960	2.911	0.830
MEB	ETS	LASSO	425.832	2.933	0.840
MEB	ETS	Median	475.540	3.315	0.976
MEB	ARIMA	Ridge	**371.904**	**2.759**	**0.778**
MEB	ARIMA	LASSO	377.937	2.814	0.804
MEB	ARIMA	Median	380.166	2.832	0.804
Proposed approaches using MBB for resampling					
MBB	ETS	Ridge	411.232	2.894	0.824
MBB	ETS	LASSO	434.314	3.069	0.893
MBB	ETS	Median	451.480	3.170	0.917
MBB	ARIMA	Ridge	403.509	2.891	0.823
MBB	ARIMA	LASSO	469.441	3.107	0.893
MBB	ARIMA	Median	416.715	3.059	0.891

Notes: Block size for MBB comprises 24 observations.

We further examined the likely gains from including or excluding replicas in the ensemble pool. To this end, we conduct the same empirical experiment depicted in Section 5.3.1 using different ensemble sizes (50, 200, 500 and 1000). We used the same random seed from the previous exercise before bootstrapping. In a nutshell, results were not very sensitive to the number of replicas involved in the Bootstrap Aggregation. Figure 5.4 illustrates, for the same sample of 4 countries depicted in Figure 5.3, the differences in the final forecasts obtained by conducting Ridge regularization in an ensemble comprised of 100 and 1000 MEB.ARIMA forecasts. Some changes can be noticed in the overall forecasting performance with only 50 replicas, with regularization routines performing poorer in some countries. Even so, in the majority of cases, the regularization approaches are still considerably superior to traditional benchmarks and recently developed Bagging routines for forecasting[7].

Finally, results were also assessed in different forecasting horizons. Table 5.6[8], for instance, depicts the values for the average MASE computed at three different forecasting horizons: steps 1–4; 5–8; and 9–12. LASSO regularized forecasts seem to offer optimal results in short forecasting horizons. This is in line with the "more prone to overfit" behaviour of LASSO routines, since they usually "throw away" predictors (by making their corresponding coefficients equal to zero) which are considered of limited use in the validation set. The same predictors, however, may hold important information when forecasting in longer horizons. Ridge, in turn, always keeps information from all the predictors (coefficients are never exactly zeroed) and, for this reason, it is favoured in the long run.

[7] The full results for different ensemble sizes are available upon request.

[8] To conserve space, results for the traditional benchmarks and alternative Bagging approaches were not depicted in Table 5.6 (available upon request). We clarify that they did not provide superior (more accurate) results than the best method (highlighted in **bold**) and than most regularized approaches.

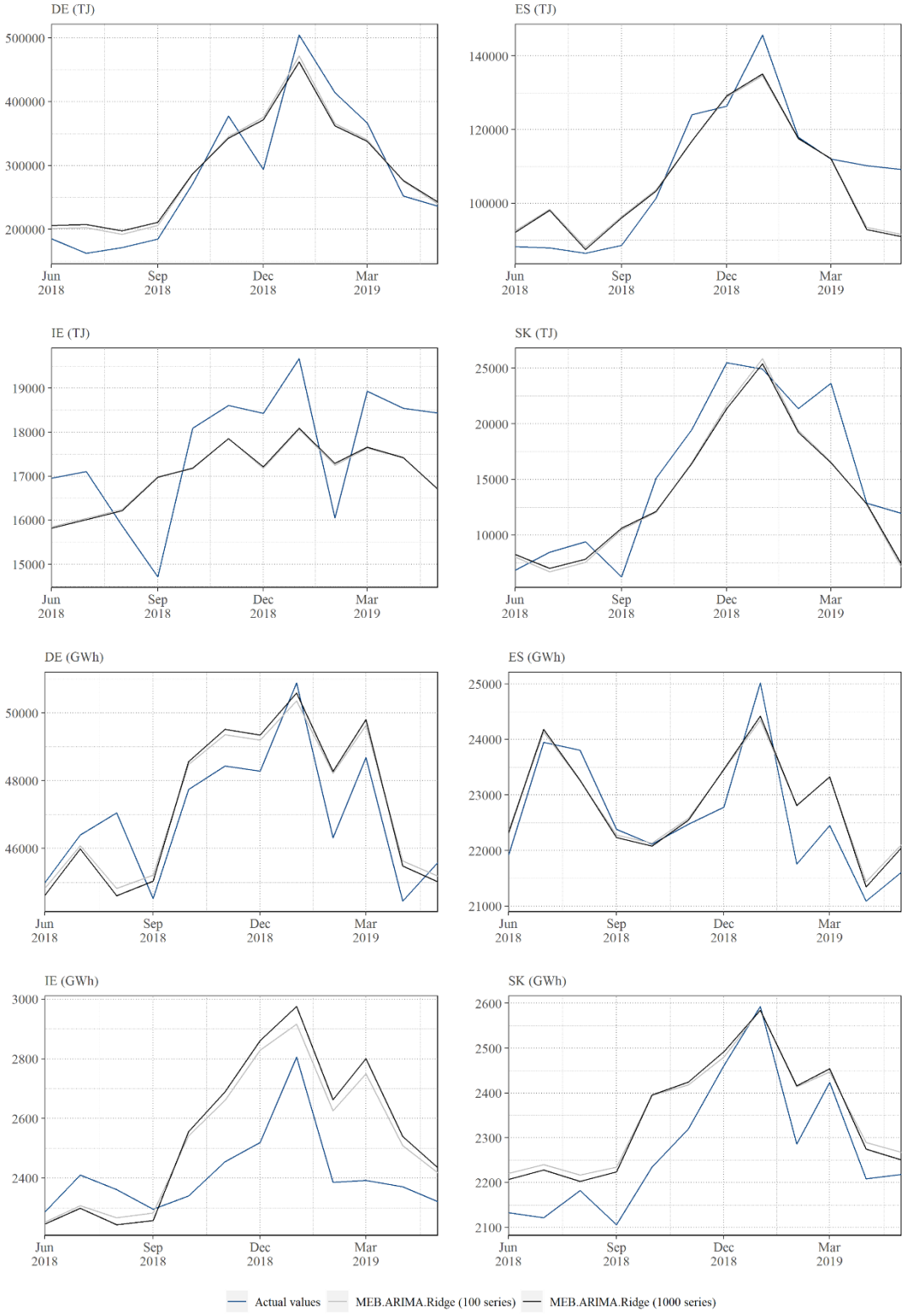


Figure 5.4 Robustness checks: Different ensemble sizes.

Aggregation with 100 forecasts in gray and with 1000 forecasts in black. Actual values in blue. Sample of 4 countries in each case.

Table 5.6 Robustness checks: average of MASEs at different horizons

Resampling algorithm	Forecasting approach	Combining method	Average MASE (steps 1-4)	Average MASE (steps 5-8)	Average MASE (steps 9-12)
I. Natural gas consumption forecasts					
Proposed approaches using MEB for resampling					
MEB	ETS	Ridge	0.393	0.625	0.679
MEB	ETS	LASSO	**0.364**	0.575	0.761
MEB	ETS	Median	0.488	0.645	0.744
MEB	ARIMA	Ridge	0.422	0.592	**0.641**
MEB	ARIMA	LASSO	0.417	**0.544**	0.741
MEB	ARIMA	Median	0.458	0.639	0.769
Proposed approaches using MBB for resampling					
MBB	ETS	Ridge	0.405	0.629	0.689
MBB	ETS	LASSO	0.377	0.621	0.869
MBB	ETS	Median	0.483	0.606	0.736
MBB	ARIMA	Ridge	0.469	0.615	0.724
MBB	ARIMA	LASSO	0.463	0.688	0.925
MBB	ARIMA	Median	0.570	0.657	0.786
II. Energy supplied forecasts					
Proposed approaches using MEB for resampling					
MEB	ETS	Ridge	0.785	0.752	0.952
MEB	ETS	LASSO	0.736	0.775	1.010
MEB	ETS	Median	0.820	0.963	1.145
MEB	ARIMA	Ridge	0.722	**0.667**	**0.944**
MEB	ARIMA	LASSO	0.703	0.716	0.993
MEB	ARIMA	Median	0.729	0.683	0.999
Proposed approaches using MBB for resampling					
MBB	ETS	Ridge	0.739	0.761	0.971
MBB	ETS	LASSO	**0.640**	0.864	1.177
MBB	ETS	Median	0.739	0.913	1.100
MBB	ARIMA	Ridge	0.708	0.750	1.011
MBB	ARIMA	LASSO	0.677	0.813	1.190
MBB	ARIMA	Median	0.713	0.880	1.079

5.3.3
Discussion and implications

The results outlined in Sections 5.3.1 and 5.3.2 endorsed the superiority of the proposed approaches over traditional forecasting methods and recently developed Bagging routines for forecasting. The performance gains are noteworthy since accurate forecasts are paramount for profit/cost optimization and assertive investment strategies, as well as for the definition of sectoral policies, whether in a regional or national scale. It should be noted that, for many countries, a considerable amount of the variation in natural gas demand may be due to external factors, which cannot be captured by univariate forecasting methods, as for example gas on gas competition, market liberalization expanding third-party access to key infrastructure, uncertainties over medium-term and long-term carbon pricing and emissions taxes inhibiting investment in gas infrastructure. In such cases, future predictions could also benefit from judgmental forecasts, possibly combining its outputs with the results from quantitative methods. This leaves a question for future research: how to include experts' judgements into the ensemble approaches for forecasting?

Another avenue for future research is to consider a multivariate setting, including the influence of external variables that may contain predictive information on natural gas demand or energy supplied across economies. We hasten to add, however, that multivariate formulations usually fail to perform well when forecasting several steps ahead (as in our case). On these grounds, the combination of ensemble methods and univariate forecasting techniques is a promising alternative for a wide range of time series in different industries/sectors.

5.4
Conclusions and future directions

This second essay proposed an ensemble-based forecasting approach combining Bootstrap aggregating (Bagging) algorithms, time series methods and regularization techniques. In doing so, this work integrated research from

combining forecasts, statistics and committee machines. A Maximum Entropy Bootstrap (MEB) routine was adopted and the use of regularization allowed for feature selection and variable weighting schemes in the combination of forecasts.

The results and robustness checks demonstrated that ensemble approaches, when combined with regularization techniques, offer accurate forecasts and are capable of dealing with different complex structures that are inherent to real world time series. Moreover, the MEB procedure was shown to be competitive when compared to the frequently used Moving Block Bootstrap (MBB) approach, outperforming the latter in most cases. This is a contribution to the literature, as the MBB has been the main benchmark for resampling monthly data under Bagging.

As previously outlined the first essay, further studies in this field of application may also benefit from a hierarchical disaggregation approach. For the natural gas sector, for instance, this would imply using the Decomposition and Bagging methods for each subsystem of the total consumption (Industrial, Electric Power, Residential, Transportation and Commercial). Such sector-tailored analysis may contribute to a more in-depth understanding of the demand for natural gas across different countries, as well as improve the forecasts of natural gas consumption in several countries. Finally, as methodological extensions of this research, investigations of other decomposition, bootstrap and forecasting methods constitute a future research agenda.

6
Third essay: new approaches to model selection and combination

The third and last essay introduces the concept of *treating*, a new way of selecting among model forms in automated forecasting routines. The procedure operates by selectively subsetting the ensemble of competing models based on information from their prediction intervals. An application to exponential smoothing formulations gives rise to an alternative forecasting method, the '*Treated ETS*'. By the same token, a *pruning* strategy that is capable of feature selection in combined forecasting methods is proposed. The benefits arising from pruning are demonstrated by applying it to different Bagging algorithms. To do so, the essay first proposes two different ways that Bagging routines can be extended to deliver prediction intervals for the point forecasts, another important contribution in the related field of knowledge.

The present essay can be considered the most significant contribution of the thesis, both in terms of theory and practice. First, because it demonstrates that model selection via traditional information criteria minimization may lead to inaccurate forecasts and unstable prediction intervals on certain occasions. Second, because it shows that prediction intervals contain important information that can be used to compare different forecasting methods. Third, based on the two previous findings, this essay sets forth strategies that can be used to improve the accuracy of both point forecasts and prediction intervals in **any forecasting approach** involving model selection or combination.

The paper originating from this essay, entitled "Treating and Pruning: new approaches to model selection and combination", is also being currently considered for publication. The essay starts with a brief introduction highlighting the fact that prediction intervals have often been by overlooked to the detriment of point

forecasts in the main stream forecasting literature. Section 6.2, in turn, provides an overview on how model selection is usually conducted in most exponential smoothing routines and the limitations arising from it. It also provides a chronological review of relevant works using Bagging in time series forecasting contexts. The proposed approaches are presented in details in Section 6.3. Section 6.4 introduces the selected data for the empirical analysis and summarizes the results in terms of both point forecasts and prediction intervals. Finally, Section 6.5 concludes and suggests directions for future works.

6.1 Introduction

It is nearly six decades since the basic structures of exponential smoothing methods were first proposed (HOLT, 1957; WINTERS, 1960). Still, thanks to their ease of use and adaptation to many different situations, exponential smoothing methods are not only widely applied but also considered competitive in many cases. For instance, automatic selection among exponential smoothing model forms ranked fourth best overall in terms of delivering accurate prediction intervals in the most recent M- Competition (MAKRIDAKIS et al., 2018; 2020). In spite of their widespread use, recent literature has demonstrated that it is possible to improve upon exponential smoothing formulations (HYNDMAN et al., 2002; TAYLOR, 2003; HYNDMAN et al., 2008; HYNDMAN & ATHANASOPOULOS, 2013).

Concurrently, the literature on forecast combination has now progressed to the point of considering the effect of subsetting the pool of available forecasts before aggregation (DE MENEZES et al., 2000; HENDRY & CLEMENTS, 2004; AIOLFI & TIMMERMANN, 2006; ELLIOTT, 2011; MATSYPURA et al., 2018; KOURENTZES et al., 2019; DIEBOLD & SHIN, 2019). The rationale behind subsetting has also been recently raised when forecasting using Bootstrap Aggregation (Bagging) routines by DANTAS & CYRINO OLIVEIRA (2018), who advocated the use of clustering methods to create a subset with a reduced variance.

In spite of the undeniable achievements on exponential smoothing formulations and on subsetting routines for forecast combination methods, no work

has considered looking at the information delivered by Prediction Intervals (PIs) when conducting model selection and/or combination. In fact, it was not until recently that PIs were considered in most forecasting works. For instance, the M4 Competition was the first of its kind to explicitly ask participants to deliver prediction intervals for their point forecasts, and ended with only 20 forecasters providing valid PIs (MAKRIDAKIS et al., 2018; 2020).

This essay demonstrates that prediction intervals, apart from providing practitioners with a convenient way to estimate the uncertainty of a point forecast, contain important information that can be used to improve the accuracy of forecasting methods involving model selection and/or combination. Concerning the former, we introduce a new way of selecting among competing formulations that involve 'treating' – discarding specific model forms from the set of models – before proceeding to selection via traditional methods, e.g. via information criteria minimization. Regarding the latter, we set forth a 'pruning' strategy that can be used to enhance the forecasts arising from any combining method.

Both treating and pruning are conducted based on the information retrieved from the prediction intervals of the forecasts. We explore the potential gains of these two strategies through an extensive empirical experiment on a wide range of monthly, quarterly and yearly time series from the M, M3 and M4 Competitions (MAKRIDAKIS et al., 1982; MAKRIDAKIS & HIBON, 2000; MAKRIDAKIS et al., 2018). To demonstrate how treating can be used to improve upon model selection, we apply this strategy to the automated exponential smoothing routine implemented in the forecast package for the R statistical software (HYNDMAN & KHANDAKAR, 2008; HYNDMAN et al., 2019). With regards to pruning, we explore its potential to improve upon combining methods on two recently developed Bagging routines for forecasting, presented in the works of BERGMEIR et al. (2016) and PETROPOULOS et al. (2018). These combining methods were selected in light of their promising results in the M3 Competition. Finally, we also propose different ways that Bagging routines can be extended to deliver prediction intervals for the point forecasts, another important development of this essay. Foreshadowing our results, we demonstrate that, apart from their simplicity and ease of use, treating and pruning require practically no additional

computation cost and can substantially improve the quality of both point forecasts and prediction intervals for a wide range of time series.

6.2
Exponential smoothing and Bagging for forecasting - state of the art

6.2.1
Exponential Smoothing and current limitations

There are several different approaches to exponential smoothing. As outlined in Section 3.1.3, HYNDMAN et al. (2002) provided a solid theoretical foundation for exponential smoothing in state space modelling, allowing for straightforward implementation in many statistical packages (HYNDMAN et al., 2008; HYNDMAN & ATHANASOPOULOS, 2013). Model selection under the framework of HYNDMAN et al. (2002) is based on the minimization of one or more information criteria. For instance, by default, the ets( ) function from the *forecast* package for the R statistical software uses the Akaike's Information Criterion corrected for small sample bias (AICc, SUGIURA, 1978) to select an appropriate model. Other information criteria, such as AKAIKE (1974) or SCHWARZ (1978) can also be used. A similar procedure is also conducted in the EViews statistical software (IHS GLOBAL INC., 2015).

Selecting models based on information criteria minimization may seem compelling to practitioners who believe that searching for the 'true' model may not make sense for empirical data, since the optimal model for the real data generating process will not usually be among the candidate models considered in any case (KOLASSA, 2011). Nevertheless, selecting a single best model out of a number of competing candidates may also be misleading. Multiple models may explain the data equally well, and selecting a single model discards the information that could be gauged from alternative models with high explanatory power (BUCKLAND et al., 1997). Another point that is often overlooked is that even if one relies on criteria that partially addresses overfitting (such as information criteria), the selected model(s) may still lead to inaccurate and/or unstable forecasts. Conducting some

sort of cross-validation routine may circumvent this problem in some cases, but this is not always guaranteed.

In light of the above, we propose looking at the outputs originating from competing ETS formulations and let them dictate which models are actually more likely to produce the best forecasts for a given time series and, accordingly, which should be discarded. More specifically, we aim to gather the prediction intervals delivered from competing models and check for deviant behaviors in the ensemble. This 'wisdom of the crowds' approach builds on the same argument of the previous paragraph: since multiple models may explain the data almost equally well, they will usually produce forecasts that are not very distant from one another. However, for models presenting 'hard to estimate' stylized facts such as structural breaks, nonlinear patterns and/or periods with large range of values, a best model may be identified on the basis of traditional criteria, but its forecasts can still be very inaccurate and sometimes display explosive behavior in long forecast lead times. On the other hand, competing models which also delivered low values for most information criteria but were not selected as best during estimation phase may produce better forecasts than the selected model. Under such circumstances, pre-treating the set of candidate models may contribute to reduce the odds of selecting an unstable model and hence improve the accuracy of the forecasting method. We demonstrate the usefulness of 'pre-treatment' in Section 6.4, using a wide range of time series (more than 100,000 series from the M-Competitions, split into monthly, quarterly and yearly frequencies). We also note that the additional computational cost is negligible, especially when compared with the time ETS routines take to estimate all competing model forms and collect their corresponding information criteria values.

6.2.2
Bagging in time series forecasting

This section intended to provide a brief overview of recent studies employing Bagging as a combining method for forecasting. Since the content of this section has already been previously explored in the thesis (Section 2.3), we opted to skip it here and promptly proceed to the methodology part.

6.3
Methods

6.3.1
Treating in model selection

The rationale behind 'treating' is to compare the prediction intervals originating from competing model forms in ETS, and discard the ones showing deviant behaviors from the majority in the ensemble. More specifically, it collects the upper limits of the prediction intervals and considers as outliers any values lying outside the range of $\pm 1.5\ IQR$, where $IQR = Q_3 - Q_1$ is the Inter-Quartile Range (difference between the 3rd and 1st quartiles). We recall that we use as a benchmark the automated ETS procedure implemented in the ets( ) function from the forecast package for the R statistical software (HYNDMAN & KHANDAKAR, 2008; HYNDMAN et al., 2019). According to this algorithm, not all the 30 ETS state space formulations are considered by default in model selection. Model forms involving multiplicative trends and combinations of additive errors and multiplicative seasonality are not estimated by default. Thus, at the end, there are 15 competing model forms out of the 30 different possibilities[9]. Provided that no transformations were conducted before estimating the model, the upper and lower limits of the prediction intervals delivered by the default ETS model forms are symmetric relative to the corresponding point forecast, so there are no differences in conducting treating by looking at one limit or another. The symmetry in prediction intervals may not hold for the other 15 model forms which are not considered by default in ets( ), since their prediction intervals are computed by simulation – see HYNDMAN et al. (2008) for further details. We also note that the

[9] The number of competing model forms is even smaller for yearly data (just six) since combinations for this frequency do not take into account any seasonality. In addition, for quarterly series with training sets comprising 13 observations or less only, there is insufficient amount of data to estimate models with damped trends. In such cases, we are left with only 10 out 15 competing model forms. By the same token, for yearly series with training sets comprising 9 observations or less only, the set of candidate models decreases from 6 to 4.

use of the Inter-Quartile Range for outlier detection is a well-established procedure in descriptive statistics (VINOD, 2014) and has been used in a vast number of applications, including subsetting pools of forecasts (KOURENTZES et al., 2019).

The choice of using prediction intervals in lieu of point forecasts to compare and occasionally discard model forms from the pool of ETS formulations is twofold: first, prediction intervals are quicker, in the sense that they require fewer forecasting steps, to indicate explosive patterns in forecasts; second, it can be quite challenging to identify deviant behaviors just by looking at point forecasts: differences, in relative terms, may not be so big (hampering the task of looking for outliers, for instance); and it is not uncommon to observe forecasts that deviate considerably from the ensemble at specific forecasting steps, usually due to the model from which they were originated, but are still competitive at large forecast lead times.

The outlier detection procedure in our 'Treated ETS' approach is conducted for every step in the forecast lead time and considers all competing model forms, regardless of whether a model form has already been identified as an outlier in the first forecasting step, for instance. At the end, every model identified as an outlier (even if just once) throughout the forecast lead time is discarded from the set of competing models. After this treatment, final model selection proceeds as usual: by finding, among the remaining models, the one offering the lowest value for AICc. Albeit unlikely to occur, it may be the case that all competing model forms are identified as outliers at least once during the forecast lead time. Under such circumstances, only the model forms which were the most frequently identified as outliers would be discarded from the set of competing models.

6.3.2
Pruning in model combination

As previously outlined, the rationale behind pruning is quite similar to treating. The main difference is that now we aim at subsetting the pool of forecasts to be combined, since some of them may deviate considerably from the rest of the ensemble. Therefore, pruning can be viewed as a feature selection strategy to improve the quality of prediction intervals and point forecasts of any forecast

combination method. To demonstrate its potential, in this paper we opt for using pruning on some benchmark Bagging strategies, given their flexibility to encompass different forecast methods for the ensemble of bootstraps. Particularly, we aim at improving point forecasts and prediction intervals originating from two different approaches discussed in Chapter 3: the Bagged.BLD.MBB.ETS method proposed by BERGMEIR et al. (2016); and the Bootstrap Model Combination (BMC) devised by PETROPOULOS et al. (2018). It should be noted that both approaches were developed with the focus of improving the accuracy of point forecasts. Therefore, extending their fields of application to generate prediction intervals can also be viewed as a novelty in this essay. In the next subsections, we demonstrate how the two selected Bagging strategies can be used to generate prediction intervals and how pruning can be applied to such cases.

6.3.3
Prediction intervals in Bagging

The strategies proposed to generate the prediction intervals in Bagging are built using the same core ideas for the point forecasts. As previously outlined, two Bagging strategies were here considered: BERGMEIR's et al. (2016) Bagged.BLD.MBB.ETS (henceforth 'Bagged ETS' to conserve space) and PETROPOULOS et al. (2018). These methods were selected in light of their promising results in the M3 Competition. In the case of Bagged ETS, besides aggregating the point forecasts, we also combine their corresponding prediction intervals using the median. This is possible because, besides the point forecasts, the forecast( ) function from the *forecast* package, when applied to an ETS model, also generates their corresponding prediction intervals, with a theoretical coverage level set by the practitioner. For instance, if a 95% coverage is aimed for, a prediction interval is generated using the 2.5% quantile as lower limit and the 97.5% quantile for the upper limit. The reader is referred to HYNDMAN et al. (2008) for details on how the quantiles are computed in ETS formulations.

Let J be the number of forecasts involved in the ensemble (forecasts of the original data and the $J-1$ bootstraps generated). That way, the upper and lower limits in Bagged ETS are obtained as follows:

$$U_{t,BaggedETS} = \text{median}\left[U_{t,1}, \dots, U_{t,J}\right]$$
$$L_{t,BaggedETS} = \text{median}\left[L_{t,1}, \dots, L_{t,J}\right] \quad (20)$$

where $U_{t,1}, \dots, U_{t,J}$ and $L_{t,1}, \dots, L_{t,J}$ are respectively the upper and lower limits of the J point forecasts in the ensemble. The above equation is applied for every step in the forecast lead time, i.e., $t = 1, \dots, h$, h being the total number of steps required.

As for BMC, we take a weighted average of the prediction intervals generated from applying the 'unique' ETS model forms on the original data, with weights defined by the frequency that the unique models were identified as optimal. Let K be the number of unique ETS model forms identified among the ensemble of J forecasts. Hence, the limits of the BMC prediction interval can be obtained according to the following equation:

$$U_{t,BMC} = \sum_{i=1}^{K} w_i \, U_{t,i}$$
$$L_{t,BMC} = \sum_{i=1}^{K} w_i \, L_{t,i} \quad (21)$$

where $w_i = 1, \dots, K$ are the weights of the K unique model forms, and $U_{t,i}$ and $L_{t,i}$ are the upper and lower limits of their corresponding prediction intervals.

Figure 6.1 illustrates how Bagged ETS and BMC can be used to generate both Bagged Point Forecasts (PFs) and Prediction Intervals (PIs). The figure also foreshadows how pruning can be achieved in each of these strategies (see the next section for details). Bagged ETS aggregates the J Point Forecasts (PFs) and their J corresponding Prediction Intervals (PIs) using their medians. BMC, in turn, identifies from the J forecasts the K unique ETS model forms and apply them to the original series. Then, it combines the results from K PFs (and corresponding PIs) using as weights the frequency with which the unique forms were identified as optimal, i.e., the amount of times they were selected divided by J.

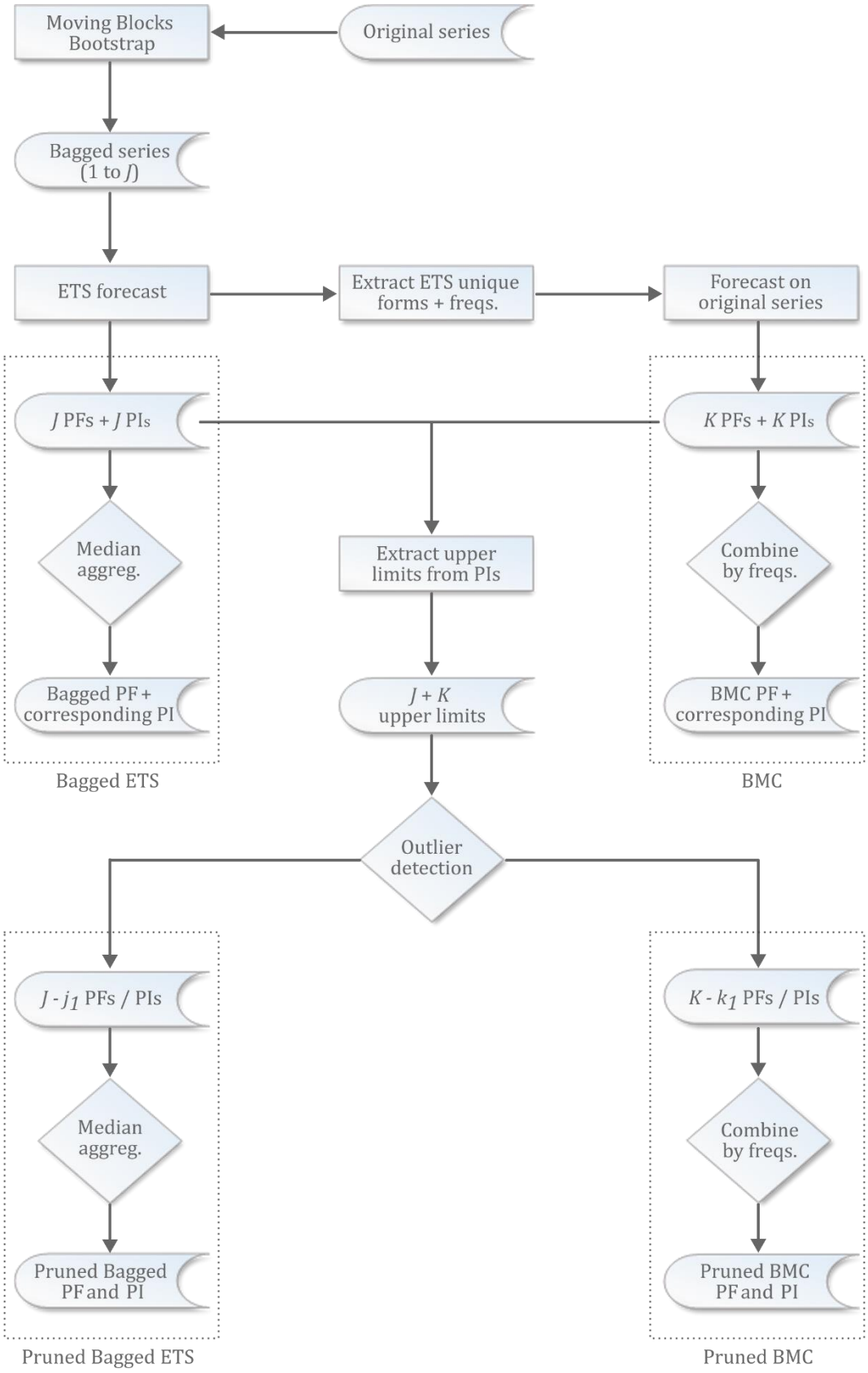


Figure 6.1 Bagged ETS and BMC and their pruned versions.

Source: The author.

6.3.4
Pruning for Bagging

Relating once again to Figure 6.1, pruning for Bagged ETS and BMC first considers merging the forecasts and prediction intervals from both ensembles, ending up with $J + K$ PFs (and corresponding PIs). This is recommended in light of the small number of forecasts comprising the BMC ensemble, which sometimes renders impossible the detection of outliers. By merging the PFs (and corresponding PIs) from both ensembles, it becomes easier to detect and remove unwanted outputs from the BMC ensemble using the $\pm 1.5\ IQR$ 'rule'. This may also prove beneficial to the Bagged ETS approach since, with the exception of the first forecast from both ensembles (which is the same since it is produced from the original data), the K added forecasts from the BMC ensemble can differ considerably from the J forecasts, bringing more diversity to the merged ensemble and ultimately making outlier detection more effective.

Even though pruning is conducted on the merged $(J + K)$ ensemble of forecasts, the final results are separated between Bagged ETS and BMC. In other words, after pruning, the resulting ensembles now encompass $J - j_1$ and $K - k_1$ forecasts (respectively for Bagged ETS and BMC), where j_1 and k_1 are the removed forecasts from each ensemble. Beyond this point, the Bagged ETS and BMC routines proceed as usual.

Pruning can be conducted as many times as desired, until no outliers can be identified in the resulting ensemble. Depending on the case, pruning twice leads to better results than pruning just once. The gains, however, were not too significant in our empirical tests with the M- Competitions and were usually detected for prediction intervals only, with a slight loss in accuracy for the point forecasts. We finally note that further pruning (three times or more) frequently led to poorer (less accurate) results, both in terms of point forecasts and prediction intervals.

6.4
Empirical investigation

6.4.1
Experiment settings

To assess the accuracy of our developed strategies and at the same time provide a common ground for discussion with previous related works, the empirical experiment was conducted using the databases from three well-known forecasting competitions, the M, M3 and M4 Competitions (MAKRIDAKIS et al., 1982; MAKRIDAKIS & HIBON, 2000; MAKRIDAKIS et al., 2018). We restrict our attention to yearly (181 + 645 + 23,000 = 23,826 series), quarterly (203 + 756 + 24,000 = 24,959 series) and monthly (617 + 1,428 + 48,000 = 50,045 series) data, which are the most used frequencies in practice and also in previous works concerning Bagging approaches (BERGMEIR et al., 2016; DE OLIVEIRA & CYRINO OLIVEIRA, 2018; PETROPOULOS et al., 2018; DANTAS & CYRINO OLIVEIRA, 2018). The predictive power of the proposed approaches was assessed using the same amount of out-of-sample data suggested in the competitions (6 observations for yearly series, 8 for quarterly and 18 for monthly), to allow comparability with published results. To gauge the accuracy of the developed strategies, we opted to summarize the results according to the following metrics:

- For Point Forecasts: Average of the Mean Absolute Scaled Errors (Mean of MASEs);
- For Prediction Intervals: Average of the Mean Scaled Interval Score (Mean of MSISs).

The MASE and MSIS are defined as follows:

$$MASE = \frac{1}{h}\frac{\sum_{t=1}^{h}\left|Y_t - \widehat{Y}_t\right|}{\frac{1}{n-m}\sum_{t=m+1}^{n}|Y_t - Y_{t-m}|}$$

$$MSIS = \frac{1}{h}\frac{\sum_{t=1}^{h}(U_t - L_t) + \frac{2}{\alpha}(L_t - Y_t)\mathbf{1}\{Y_t < L_t\} + \frac{2}{\alpha}(Y_t - U_t)\mathbf{1}\{Y_t > U_t\}}{\frac{1}{n-m}\sum_{t=m+1}^{n}|Y_t - Y_{t-m}|} \quad (22)$$

where Y_t and $\widehat{Y}_t$ are the actual and forecasted values of the underlying series, respectively; t is the forecast lead time from 1 to h steps ahead; m is the seasonal period; U_t and L_t are the upper and lower limits of the prediction interval produced

using the selected method; and $1 - \alpha$ is the desired (theoretical) coverage level. By introducing penalties for the width ($U_t - L_t$) and for the instances where the actual values are outside the specified bounds of the predicted interval, the MSIS offers a good balance between spread and coverage (hit rates).

The choice for the above-mentioned metrics was mainly to allow comparability with published results. It should also be noted that these are the official evaluation metrics for point forecasts and prediction intervals in the M4 Competition (MAKRIDAKIS et al., 2018). Apart from depicting the results in terms of mean and median of the above-mentioned metrics, we have also conducted the Multiple Comparisons with the Best (MCB) test (KONING et al., 2005) to assess whether the differences between the error measures were statistically significant.

6.4.2 Findings

Table 6.1 summarizes the average (across all series) MASE results for the point forecasts, whilst Table 6.2 compiles the average MSIS results for prediction intervals constructed with a desired coverage level (hit rate) of 95%, following the M4 Competition guidelines. For comparison purposes, we contrast the results obtained from the following methods:

(i) The auto state space exponential smoothing (ETS) approach, i.e. the forecasts obtained by selecting the best ETS specification for the original series and subsequently using it for forecasting. Despite its simplicity when compared with other forecasting methods, the ETS provides a sound base for comparison with the proposed Bagging approaches, since they also use ETS models to build the forecasts. In addition, it should be noted that ETS ranked third best overall in terms of closeness to an expected (desired) 95% hit rate and fourth best in terms of lowest MSIS, when all (100,000) series from the M4 Competition were considered (GRUSHKA-COCKAYNE & JOSE, 2020).

(ii) The Treated ETS approach presented in Section 6.3.1.

(iii) BERGMEIR's et al. (2016) Bagged.BLD.MBB.ETS method (here abbreviated to 'Bagged ETS').

(iv) The BMC approach, as proposed in PETROPOULOS et al. (2018).

(v) The selected Bagging strategies using, as forecasting method for the original data and the bootstraps, the Treated ETS in lieu of ETS; and,

(vi) The pruning strategy proposed in Section 6.3.4 applied to all of the above Bagging strategies.

Table 6.1 All competitions - Average MASE of the different forecasting methods

Method	Average MASE (Monthly)	Average MASE (Quarterly)	Average MASE (Yearly)
Exponential smoothing			
ETS	0.947	1.165	3.431
Treated ETS	0.939	1.161	3.395
Bagging Strategies			
Bagged ETS	0.955	1.180	3.286
Bagged Treated ETS	0.953	1.179	3.284
BMC ETS	0.925	1.146	3.323
BMC Treated ETS	0.922	1.145	3.318
Bagging with pruning			
Pruned Bagged ETS	0.955	1.181	3.288
Pruned Bagged Treated ETS	0.953	1.179	3.287
Pruned BMC ETS	*0.920*	*1.144*	*3.236*
Pruned BMC Treated ETS	**0.917**	**1.143**	**3.228**

Notes: Best (most accurate) approach in **bold**, second best in *italic*. BMC ETS stands for the BMC method devised in PETROPOULOS et al. (2018). We use the former notation to differentiate it from the BMC Treated ETS, which is the BMC applied to the forecasts generated by using the Treated ETS routine proposed in Section 6.3.1 on the bootstraps.

From the average results we note that Treated ETS provides more accurate results than ETS, with the former outperforming the latter in every case, regardless of the frequency of the time series, namely monthly, quarterly and yearly, and the

evaluation scenario, namely point forecasts and prediction intervals. This makes a new contribution to the literature, since the automatic ETS routine, as implemented in the ets( ) function from the *forecast* package for R, has been considered the benchmark for automatic model selection among competing ETS model forms and subsequent forecasting. Furthermore, as shown in Table 6.1 for point forecasts, Bagging routines deliver more accurate results when combined with Treated ETS rather than ETS.

Turning the attention to the MSIS values, as shown in Table 6.2, we note a major issue with using BMC to generate prediction intervals for monthly time series. When no pruning is conducted, regardless of the forecasting method selected for the bootstraps (ETS or Treated ETS), BMC generates very large prediction intervals for some series, resulting in very high overall MSIS values.

Table 6.2 All competitions - Average MSISs, computed at the 95% coverage level

Method	Average MSIS (Monthly)	Average MSIS (Quarterly)	Average MSIS (Yearly)
Exponential smoothing			
ETS	*8.258*	9.587	34.970
Treated ETS	**8.133**	9.513	34.466
Bagging Strategies			
Bagged ETS	8.700	9.746	36.948
Bagged Treated ETS	8.662	9.724	36.957
BMC ETS	3.301×10^{11}	10.355	32.825
BMC Treated ETS	6.603×10^{11}	10.374	32.633
Bagging with pruning			
Pruned Bagged ETS	8.727	9.780	37.276
Pruned Bagged Treated ETS	8.693	9.759	37.286
Pruned BMC ETS	8.342	**9.345**	*32.317*
Pruned BMC Treated ETS	8.370	*9.392*	**32.211**

Notes: Best (most accurate) approach in **bold**, second best in *italic*.

Upon closer inspection, we note that the issue with very high MSIS values arises in certain series from the M4 competition with notable structural breaks and/or outliers in the training set. As consequence, some bootstraps will be generated with extremely large values. ETS model forms for such bootstraps are not optimal for the original series, but they are applied to the latter according to how the BMC algorithm is designed. These model forms will usually generate very large prediction intervals since they contain multiplicative errors and are applied to the original series which already contains a large range of values. This is illustrated in Figure 6.2 and in Table 6.3. The former shows the training set ensemble (original series and its corresponding bootstraps) of the monthly series 41895 from the M4 competition. The latter reports on the selected model forms in BMC for the same series, along with the upper limits of the prediction intervals generated when such model forms are applied to the original series.

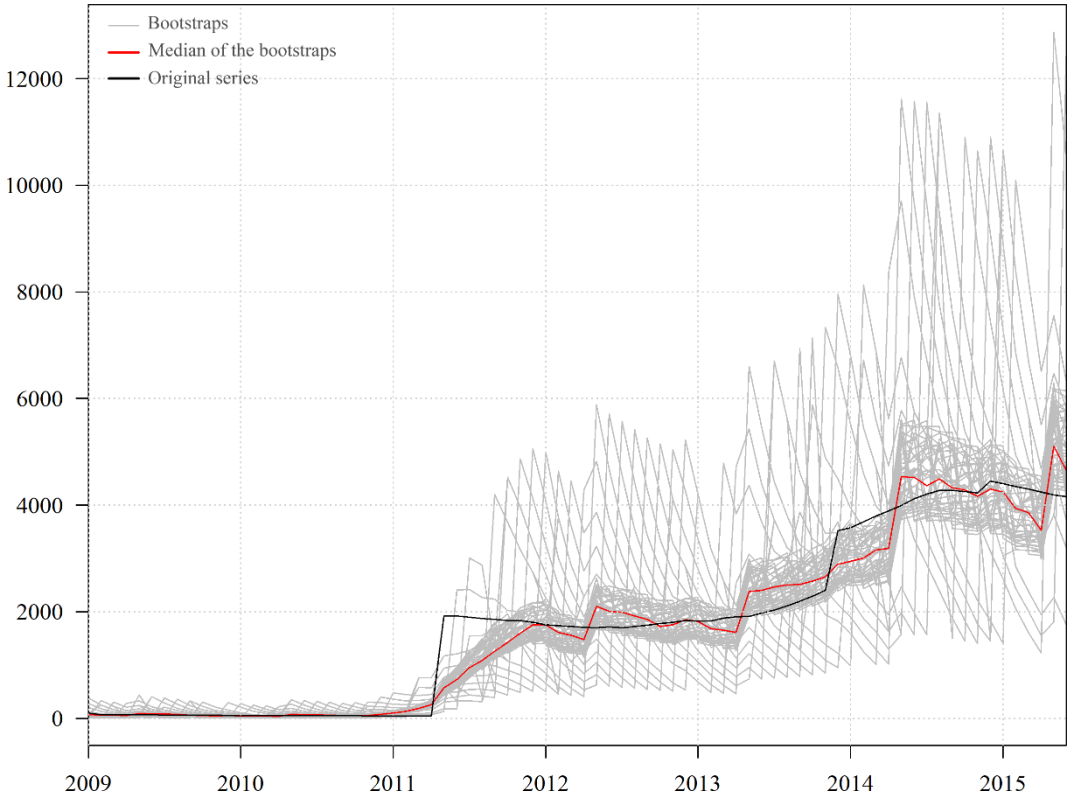


Figure 6.2 M4 competition monthly series 41895, training set. Original data in black, bootstraps in gray and median of the bootstraps in red. *Source*: The author.

Table 6.3 M4 competition monthly series 41895, test set
Actual values, selected ETS model forms and corresponding prediction interval upper limits, BMC ETS upper limits before and after pruning

Lead time	Step 1	Step 2	Step 3	...	Step 16	Step 17	Step 18
Actual values							
out-of-sample	4,119	4,074	4,032	...	3,483	3,443	3,399
ETS forms and upper limits							
A, N, N	4,662	4,870	5,029	...	6,164	6,225	6,285
A, N, A	4,607	4,779	4,900	...	6,017	6,034	6,261
A, A, N	4,711	4,969	5,179	...	6,991	7,105	7,217
M, A_d, M	19,094	16,407	16,815	...	16,745	16,054	14,830
M, A_d, N	14,714	14,750	14,784	...	15,131	15,156	15,180
M, A, M	16,476	17,197	15,089	...	14,853	15,388	16,470
M, N, M	13,114	23,203	17,954	...	67,102	60,490	61,591
M, A, N	41,749	363,072	3.38×10^6	...	1.51×10^{19}	1.42×10^{20}	1.33×10^{21}
BMC ETS upper PI limits							
No pruning	15,715	45,060	2.84×10^5	...	1.21×10^{18}	1.13×10^{19}	1.06×10^{20}
Pruning	13,669	13,681	12,662	...	12,876	13,069	13,481

The selected ETS form for the original M4 monthly series 41895[10] was (A, N, N), a combination of additive errors, no trend and no seasonality. Alternative ETS formulations with additive errors (A, A, N and A, N, A) produced relatively similar values for the upper limits of the prediction intervals, when compared with formulations with multiplicative errors. Irrationally high values for the prediction intervals were observed when the model form involved an additive trend and no seasonality (M, A, N), with the values for the upper limits being more than 10^{16} times higher than the actual (real) values at the last forecasting step. Such irrational behavior was considerably dampened when model forms considers a multiplicative seasonality (M) and/or an additive damped trend (A_d). By conducting pruning following the steps depicted in Section 6.3.4, we were able to discard the forecasts (and corresponding prediction intervals) from the last two ETS formulations –

[10] The one obtained by applying the default ets( ) function on the original series.

(M, N, M) and (M, A, N) – before proceeding to combination. As a result, the MSIS value decreased substantially, from 1.65×10^{16} with no pruning to 24.10 with pruning.

The same pattern observed in Figure 6.2 and in Table 6.3 is repeated in several cases: the BMC generates very large prediction intervals for at least 15 monthly series from the M4 competition, and relatively high values – when compared to ETS, for instance – for more than 100 series. In most cases, a substantial reduction in MSIS is achieved by conducting the proposed pruning strategy.

Turning once again to the overall MSIS results in Table 6.2, we note that the effect of pruning is substantial for BMC formulations but not for Bagged ETS. This is because the latter aggregates the bagged forecasts using the median, which diminishes the effect of the outliers in the ensemble. BMC, in turn, takes a weighted average of the forecasts and will thus always consider the effect of ETS formulations which generates very large prediction intervals. However, provided that proper pruning is conducted, BMC strategies are superior on average than Bagged ETS, both in terms of point forecasts and prediction intervals.

The benefits of treating for ETS model selection and pruning for BMC strategies are also shown in Figure 6.3, which depicts the average MSIS values computed at alternative prediction interval hit rates (85% to 99%) for four different methods. By contrasting the results delivered by ETS (in red) and Treated ETS (in yellow) in Figure 6.3, we note that the latter outperforms the former in every case scenario, regardless of the time series frequency or desired coverage level. We also compare the results obtained using BMC (in green) and its pruned version (in blue), illustrating the gains one can achieve by considering pruning in combined forecasting approaches.

We proceeded by exploring the results from the Multiple Comparisons with the Best (MCB) tests, illustrated in Figure 6.4. The results in terms of average MASE ranks were largely in line with the results from Table 6.1, with BMC Treated ETS and Pruned BMC Treated ETS depicted as the best methods and statistically significant from the others. The only exception was for quarterly series, where Pruned BMC Treated ETS ranked third and was considered statistically different from BMC ETS and BMC Treated ETS.

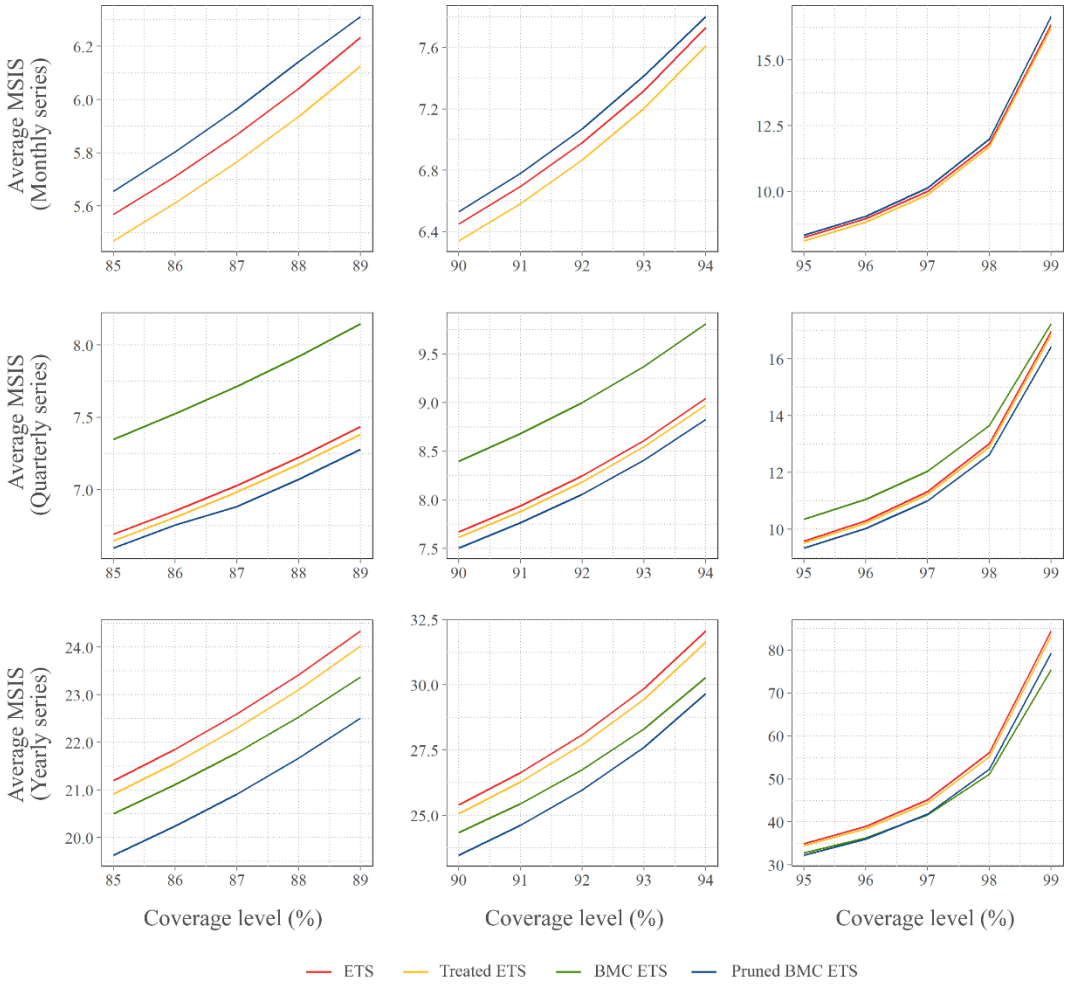


Figure 6.3 MSIS per different coverage levels (85–99%) – Four methods.

Average MSIS ranks, however, tell a slightly different history from the average MSIS values illustrated in Table 6.2. Pruned Bagged Treated ETS now ranks as the best overall in terms of average MSIS rank, in every case considered (monthly, quarterly or yearly series). An explanation lies in the fact that Pruned Bagged Treated ETS is usually the best method across the series, but when Bagging strategies fail in generating accurate and calibrated prediction intervals for their point forecasts, Pruned Bagged Treated ETS usually delivers worse results than Pruned BMC Treated ETS, the best method in terms of mean of MSISs, as depicted in Table 6.2. Even so, the results from Tables 6.1 and 6.2 and Figure 6.4 make it clear that Treated ETS consistently outperforms ETS, both in terms of Point Forecasts and Prediction Intervals, and that pruned Bagging strategies are usually more accurate than their traditional versions.

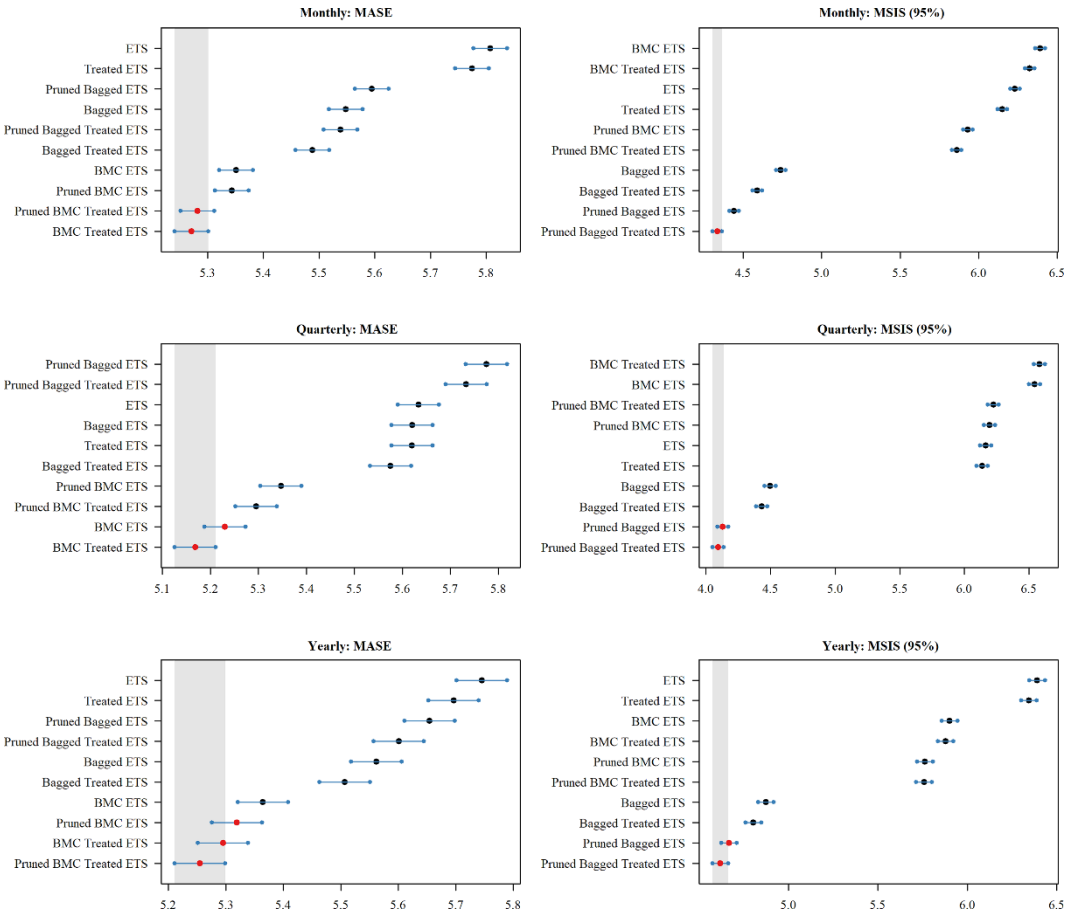


Figure 6.4 Multiple comparisons with the best for MASE and MSIS.

6.4.3
Relative performance on the M4 competition

As a final experiment, we compared the relative performance of the methods developed in this essay with the best methods from the M4 competition in terms of prediction intervals. Table 6.4 depicts the average MSIS values (at the 95% coverage level) for the automated exponential smoothing formulations, the two most accurate Bagging methods and the four best methods in M4.

It is interesting to note that in spite of their simplicity, as observed throughout the essay, treating and pruning led to very competitive results in terms of prediction intervals in the M4 competition. For monthly series, for instance, the Treated ETS and the two best Bagging routines developed with the aid of pruning – Pruned BMC ETS and Pruned BMC Treated ETS – ranked third among the best methods from the M4 competition. For quarterly series, pruned strategies ranked

second best overall. Finally, for yearly series, Treated ETS and Pruned BMC strategies ranked between the third and fourth best methods. Of course, one must take into account the fact that these are ex-post results, after the end of the competition. On the other hand, we argue that treating and pruning were not designed to beat benchmarks and/or rank among the best methods in the M4 competition, and yet provided very competitive results. It is also worth recalling that the overall accuracy of pruning in this case is restricted to how good Bagging strategies perform in practice. In other words, the results for pruning could have been even better if employed in alternative forecast combination methods and/or in conjunction with more sophisticated approaches, such as the ones presented in the M4 competition. This leaves a potential avenue for future research.

Table 6.4 M4 competition - Average MSIS, computed at the 95% desired coverage level, for the automated exponential smoothing formulations, the two most accurate Bagging methods and the four best methods from the competition

Method	Average MSIS (Monthly)	Average MSIS (Quarterly)	Average MSIS (Yearly)
Exponential smoothing			
ETS	8.30	9.49	34.90
Treated ETS	8.18	9.52	34.43
Best 2 Bagging methods			
Pruned BMC ETS	8.39	*9.25*	32.22
Pruned BMC Treated ETS	8.42	9.30	32.11
Best 4 methods from the M4 competition			
Submission 118	**7.20**	**8.55**	**23.90**
Submission 245	8.66	9.38	*27.48*
Submission 238	9.49	9.85	30.20
Submission 069	*8.03*	9.42	35.84

6.5
Conclusions and future directions

In this essay, a new way of selecting among model forms in automated forecasting routines was introduced. The approach, here addressed as *treating*,

operates by subsetting the pool of competing models based on the information delivered by their prediction intervals. An application to exponential smoothing formulations gave rise to an alternative forecasting method, the 'Treated ETS'. By the same token, we also proposed a pruning strategy that is capable of feature selection in combined forecasting methods.

The gains originating from treating and pruning were empirically demonstrated by means of an extensive experiment on a wide range of monthly, quarterly and yearly time series from the M- Competitions. We used as benchmarks for forecast combination two recently developed Bagging routines, which were originally developed with the focus of improving the accuracy of point forecasts. To demonstrate how the accuracy of these methods could be improved with the use of pruning, we first extended the fields of application of Bagging to generate prediction intervals, another important development of this work.

The implications of the present study are significant in terms of both theory and practice. First, we demonstrate that model selection via traditional information criteria minimization may lead to inaccurate forecasts and unstable prediction intervals. Second, we show that prediction intervals, apart from providing practitioners with a convenient way to estimate the uncertainty of a point forecast, contain important information that can be used to improve the accuracy of forecasting methods without having to resort to procedures which are dependent on the choice of the practitioner, such as the use of a validation set, for instance. Third, based on these two previous findings, we set forth strategies that can be used to improve the accuracy of both point forecasts and prediction intervals **in any forecasting method involving model selection or combination**.

As methodological extensions of this research, future works may benefit from alternative schemes for subsetting the pool of competing model forms, in the case of treating, or the ensemble of forecasts to be combined, in the case of pruning. For the latter, for instance, we restricted our attention to demonstrate how subsetting could be achieved in Bagging routines. It would be interesting to see how the concept could be extended to other forecast combination methods. The use of alternative methods for outlier detection in ensembles, such as nonparametric methods, also constitute a future research agenda.

7
Summary of contributions and avenues for future research

This thesis comprised three main contributions involving the combined use of ensemble approaches and time series methods to the field of forecasting, summarized in Chapters 4, 5 and 6. The first of these efforts proposed an alternative method to generate the ensemble of forecasts prior to final aggregation, which delivered satisfactory results for total electricity consumption time series across different countries. The second endeavor put forth a novel forecasting approach through the combined use of Bagging algorithms, time series methods and regularization routines. The results from an empirical experiment involving different types of energy demand time series endorsed the superiority of the developed approach over traditional forecasting benchmarks and recently developed Bagging routines for forecasting. The last essay comprised the development of new ways of selecting among model forms in automated forecasting routines and conducting feature selection in combined forecasting methods. An important aspect of this essay, which differs considerably from the previous ones, is the validation of the proposed methodologies on a wide range of time series, such as the ones from the the M- Competitions (98,830 in total).

The main take-away message from the essays involved in the thesis is that ensemble approaches offer the forecasting practitioner the ability of properly addressing the many different complex structures that are inherent to real world time series, consequently improving the accuracy of forecasting methods in a wide range of contexts. In this connection, the thesis provides an alternative and challenging view to what has been considered so far as the Holy Grail of forecasting, namely the selection of a single method, from an ever-growing range of possibilities, which best extrapolates past historical data.

Another major advantage offered by the procedures developed in this thesis, when compared to alternative approaches, lies in their flexibility, in the sense that

the practictioner can easily adapt some stages to tackle particular needs and improve forecasting accuracy for a given phenomenon/situation. That way, a number of topics for future research can be suggested, depending on the intended application. A promising avenue in this regard is the development of alternative forecast selection heuristics, i.e., procedures that can select, among the pool of forecasts originated via ensemble methods, those with the greatest potential of delivering accurate final forecasts after aggregation. The first steps in this direction were taken in the second essay (Chapter 5), in which the use of regularization routines was proposed to select and/or assign weights to predictors in the forecast ensemble, to the detriment of traditional aggregation metrics (mean, median, among others). Another extension for future research, which was also initiated in the second essay, is the development of alternative approaches to generate replicas of the original data. This involves not only the proposition of alternative bootstrapping schemes, but also the proper use of simulation routines. Future studies can also benefit from alternative pre-treatment and decomposition schemes before resampling the original (or parts of the original) data.

Apart from the endless range of possibilities that exist when considering alternative methods in the different stages of forecasting ensemble approaches, another venue for future works includes the use of the information provided by forecasting ensembles to generate accurate prediction intervals to the point forecasts. The work presented in Chapter 6 represents a seminal effort in this regard. Finally, applications on datasets from other competitions, such as the Global Energy Forecasting Competition (GEFCom) (HONG et al., 2019) and the forthcoming M5 Competition, also constitute a future research agenda.

References

AIOLFI, M.; TIMMERMANN, A. Persistence in forecasting performance and conditional combination strategies. **Journal of Econometrics**, v. 135, n. 1-2, p. 31–53, 2006.

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716–723, 1974.

AKSU, C.; GUNTER, S. I. An empirical analysis of the accuracy of SA, OLS, ERLS and NRLS combination forecasts. **International Journal of Forecasting**, v. 8, n. 1, p. 27–43, 1992.

AL-HAMADI, H. M.; SOLIMAN, S. A. Long-term/mid-term electric load forecasting based on short-term correlation and annual growth. **Electric Power Systems Research**, v. 74, n. 3, p. 353–361, 2005.

ALI, K.; BRUNK, C.; PAZZANI, M. On learning multiple descriptions of a concept. In: ___ **Proceedings of Tools with Artificial Intelligence**. 6th ed. New Orleans, US: IEEE/TAI 94, 1994, p. 476–483.

ANDRAWIS, R. R.; ATIYA, A. F.; EL-SHISHINY, H. Combination of long term and short term forecasts, with application to tourism demand forecasting. **International Journal of Forecasting**, v. 27, n. 3, p. 870–886, 2011.

ASSIMAKOPOULOS, V.; NIKOLOPOULOS, K. The theta model: a decomposition approach to forecasting. **International Journal of Forecasting**, v. 16, n. 4, p. 521–530, 2000.

AUER, P.; BURGSTEINER, H.; MAASS, W. A learning rule for very simple universal approximators consisting of a single layer of perceptrons. **Neural Networks**, v. 21, n. 5, p. 786–795, 2008.

AYE, G. C.; BALCILAR, M.; GUPTA, R.; MAJUMDAR, A. Forecasting aggregate retail sales: the case of South Africa. **International Journal of Production Economics**, v. 160, p. 66–79, 2015.

BARBOSA, S. M.; SCOTTO, M. G.; ALONSO, A. M. Summarising changes in air temperature over Central Europe by quantile regression and clustering. **Natural Hazards and Earth System Sciences**, v. 11, n. 12, p. 3227–3233, 2011.

BARROW, D. K.; CRONE, S. F. Cross-validation aggregation for combining autoregressive neural network forecasts. **International Journal of Forecasting**, v. 32, n. 4, p. 1120–1137, 2016.

BARROW, D. K.; KOURENTZES, N. Distributions of forecasting errors of forecast combinations: implications for inventory management. **International Journal of Production Economics**, v. 177, p. 24–33, 2016.

BATES, J. M.; GRANGER, C. W. J. The Combination of Forecasts. **Operational Research Quarterly**, v. 20, n. 4, p. 451–468, 1969.

BAUER, E.; KOHAVI, R. An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants. **Machine Learning**, v. 36, n. 1–2, p. 105–139, 1999.

BERGMEIR, C.; HYNDMAN, R. J.; BENÍTEZ, J. M. Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation. **International Journal of Forecasting**, v. 32, n. 2, p. 303–312, 2016.

BILLAH, B.; KING, M. L.; SNYDER, R. D.; KOEHLER, A. B. Exponential smoothing model selection for forecasting. **International Journal of Forecasting**, v. 22, n. 2, 239–247, 2006.

BOX, G. E. P.; COX, D. R. An analysis of transformations. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 26, n. 2, p. 211–252, 1964.

BOX, G. E. P.; JENKINS, G. M. **Time series analysis: forecasting and control**. San Francisco: Holden Day, 1970. 1st ed.

BREIMAN, L. Bagging Predictors. **Machine Learning**, v. 24, n. 2, p. 123–140, 1996.

BREIMAN, L. Arcing Classifiers. **The Annals of Statistics**, v. 26, n. 3, p. 801–849, 1998.

BUCKLAND, S. T.; BURNHAM, K. P.; AUGUSTIN, N. H. Model Selection: An Integral Part of Inference. **Biometrics**, v. 53, n. 2, p. 603–618, 1997.

BÜHLMANN, P. Sieve bootstrap for time series. **Bernoulli**, v. 3, n. 2, p. 123–148, 1997.

BURKE, P. J.; CSEREKLYEI, Z. Understanding the energy-GDP elasticity: A sectoral approach. **Energy Economics**, v. 58, p. 199–210, 2016.

CASTELLI, M.; VANNESCHI, L.; DE FELICE, M. Forecasting short-term electricity consumption using a semantics-based genetic programming framework: The South Italy case. **Energy Economics**, v. 47, p. 37–41, 2015.

CHATFIELD, C. **Time-Series Forecasting**. Boca Raton: Chapman & Hall/CRC. 2000.

CLEMEN, R. T.; WINKLER, R. L. Combining Economic Forecasts. **Journal of Business and Economic Statistics**, v. 4, n. 1, p. 39–46, 1986.

CLEVELAND, R. B.; CLEVELAND, W. S.; McRAE, J. E.; TERPENNING, I. STL: A Seasonal-Trend Decomposition Procedure Based on Loess. **Journal of Official Statistics**, v. 6, n. 1, 3–73, 1990.

CORDEIRO, C.; NEVES, M. Forecasting time series with BOOT.EXPOS procedure. **REVSTAT-Statistical Journal**, v. 7, n. 2, p. 135-149, 2009.

DANTAS, T. M.; OLIVEIRA, F. L. C.; REPOLHO, H. M. V. Air transportation demand forecast through Bagging Holt winters methods. **Journal of Air Transportation Management**, v. 59, p. 116–123, 2017.

DANTAS, T. M.; CYRINO OLIVEIRA, F. L. Improving time series forecasting: an approach combining bootstrap aggregation, clusters and exponential smoothing, **International Journal of Forecasting**, v. 34, n. 4, p. 748–761, 2018.

DE GOOIJER, J. D.; HYNDMAN, R. J. 25 years of time series forecasting. **International Journal of Forecasting**, v. 22, n. 3, p. 443–473, 2006.

DE MENEZES, L. M.; BUNN, D. W.; TAYLOR, J. W. Review of guidelines for the use of combined forecasts. **European Journal of Operational Research**, v. 120, n. 1, p. 190–204, 2000.

DE OLIVEIRA, E. M.; CYRINO OLIVEIRA, F. L. Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing methods. **Energy**, v. 144, p. 776–788, 2018.

DEBNATH, K. B.; MOURSHED, M. Forecasting methods in energy planning models. **Renewable and Sustainable Energy Reviews**, v. 88, p. 297–325, 2018.

DEKKER, M.; VAN DONSELAAR, K.; OUWEHAND, P. How to use aggregation and combined forecasting to improve seasonal demand forecasts. **International Journal of Production Economics**, v. 90, n. 2, p. 151–167, 2004.

DIEBOLD, F. X.; SHIN, M. Machine learning for regularized survey forecast combination: Partially-egalitarian LASSO and its derivatives. **International Journal of Forecasting**, v. 35, n. 4, p. 1679–1691, 2019.

EFRON, B. Bootstrap Methods: Another Look at the Jackknife. **The Annals of Statistics**, v. 7, n. 1, p. 1-26, 1979.

ELETROBRAS (CENTRAIS ELÉTRICAS BRASILEIRAS S.A.). **Consumo de energia elétrica por classe de consumidor e por região**. Available at: https://www3.bcb.gov.br/sgspub/ (accessed 1 Feb 2017).

ELLIOTT, G. **Averaging and the Optimal Combination of Forecasts**. Working Paper. San Diego: University of California, 2011. 30 p.

ELLIOTT, G.; TIMMERMANN, A. Optimal forecast combinations under general loss functions and forecast error distributions. **Journal of Econometrics**, v. 122, n. 1, p. 47–79, 2004.

ELLIOTT, G.; TIMMERMANN, A. **Economic Forecasting**. New Jersey: Princeton University Press, 2016. 552 p.

EUROSTAT (European Statistics). **Supply of gas – gross inland consumption**. Available at: https://ec.europa.eu/eurostat/web/energy/data/database/ (accessed 2 Oct 2019).

FILDES, R.; PETROPOULOS, F. Simple versus complex selection rules for forecasting many time series. **Journal of Business Research**, v. 68, n. 8, 1692–1701, 2015.

FORTMANN-ROE, S. **Understanding the Bias-Variance Trade-off**. 2012. Available at: http://scott.fortmann-roe.com/docs/BiasVariance.html (accessed 3 January 2019).

FREUND, Y. Boosting a Weak Learning Algorithm by Majority. **Information and Computation**, v. 121, n. 2, p. 256–285, 1995.

FREUND, Y.; SCHAPIRE, R. E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. **Journal of Computer and System Sciences**, v. 55, n. 1, p. 119–139, 1997.

FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. **Journal of Statistical Software**, v. 33, n. 1, 2010.

GARDNER JR., E. S. Exponential smoothing: The state of the art. **Journal of Forecasting**, v. 4, n. 1, p. 1–28, 1985.

GARDNER JR., E. S. Exponential smoothing: The state of the art—Part II. **International Journal of Forecasting**, v. 22, n. 4, p. 637-666, 2006.

GOLYANDINA, N.; NEKRUTKIN, V.; ZHIGLJAVSKY, A. **Analysis of time series structure: SSA and related techniques**. Boca Raton: Chapman & Hall/CRC, 2001.

GOODWIN, P. The Holt–Winters approach to exponential smoothing: 50 years old and going strong. **Foresight: The International Journal of Applied Forecasting**, v. 19, p. 30–33, 2010. Available at: https://foresight.forecasters.org/shop/

GRUSHKA-COCKAYNE, Y.; JOSE, V. R. R. Combining prediction intervals in the M4 competition. **International Journal of Forecasting**, v. 36, n. 1, p. 178–185, 2020.

GUERRERO, M. V. Time-series analysis supported by power transformations. **Journal of Forecasting**, v. 12, n. 1, p. 37–48, 1993.

GUIDOLIN, M.; TIMMERMANN, A. Forecasts of US short-term interest rates: A flexible forecast combination approach. **Journal of Econometrics**, v. 150, n. 2, p. 297–311, 2009.

GUO, J.-J.; LUH, P. B. Improving Market Clearing Price Prediction by Using a Committee Machine of Neural Networks. **IEEE Transactions on Power Systems**, v. 19, n. 4, p. 1867–1876, 2004.

HALL, M. **Correlation-based feature selection for machine learning**. Hamilton (NZ): Waikato University, 1999.

HASSANI, H.; WEBSTER, A.; SILVA, E. S.; HERAVI, S. Forecasting U.S. tourist arrivals using optimal Singular Spectrum Analysis. **Tourism Management**, v. 46, p. 322–335, 2015.

HENDRY, D. F.; CLEMENTS, M. P. Pooling of forecasts. **The Econometrics Journal**, v. 7, p. 1–31, 2004.

HILLEBRAND, E.; MEDEIROS, M. C. The benefits of bagging for forecast models of realized volatility. **Econometric Reviews**, v. 29, n. 5, p. 571–593, 2010.

HOERL, A. E.; KENNARD, R. W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. **Technometrics**, v. 12, n. 1, p. 55–67, 1970.

HOLT, C. C. Forecasting seasonal and trends by exponentially weighted moving average. **Office of Naval Research Memorandum**, n. 52, 1957.

HOLT, C. C. Forecasting seasonals and trends by exponentially weighted moving averages. **International Journal of Forecasting**, v. 20, n. 1, p. 5–10, 2004.

HONG, T.; XIE, J.; BLACK, J. Global energy forecasting competition 2017: Hierarchical probabilistic load forecasting. **International Journal of Forecasting**, v. 35, n. 4, p. 1389–1399, 2019.

HYNDMAN, R.; ATHANASOPOULOS, G.; BERGMEIR, C.; CACERES, G.; CHHAY, L.; O'HARA-WILD, M.; PETROPOULOS, F.; RAZBASH, S., WANG, E.; YASMEEN, F. **forecast: Forecasting functions for time series and linear models**. 2019. Available at: http://pkg.robjhyndman.com/forecast

HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. **International Journal of Forecasting**, v. 22, p. 679–688, 2006.

HYNDMAN, R. J.; KOEHLER, A. B.; SNYDER, R. D.; GROSE, S. A state space framework for automatic forecasting using exponential smoothing methods. **International Journal of Forecasting**, v. 18, n. 3, p. 439-454, 2002.

HYNDMAN, R. J.; KOEHLER, A. B.; ORD, J. K.; GROSE, S. **Forecasting with exponential smoothing: The state space approach**. 1st ed. Berlin: Springer-Verlag; 2008.

HYNDMAN, R. J.; KHANDAKAR, Y. Automatic Time Series Forecasting: The forecast Package for R. **Journal of Statistical Software**, v. 27, n. 3, 1–22, 2008.

HYNDMAN, R.; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**. 1st ed. Perth: University of Western Australia, 2013. Available at: https://otexts.com/fpp2/

IEA (INTERNATIONAL ENERGY AGENCY). **Monthly electricity statistics**. Available at: https://www.iea.org/reports/monthly-oecd-electricity-statistics/ (accessed 4 Feb 2017 for the first essay; 2 Oct 2019 for the second essay).

IHS Global Inc. **EViews R Illustrated**. (9th ed.). IHS Global Inc, 2015. Available at: https://www.eviews.com/illustrated/EViewsIllustrated.pdf (accessed 10 Nov 2019).

INOUE, A.; KILIAN, L. **Bagging time series models**. Discussion paper. London: Centre for Economic Policy Research, 2004. 36 p.

INOUE, A.; KILIAN, L. How useful is bagging in forecasting economic time series? A case study of US consumer price inflation. **Journal of the American Statistical Association**, v. 103, n. 482, p. 511–522, 2008.

JIN, S.; SU, L.; ULLAH, A. Robustify financial time series forecasting with bagging. **Econometric Reviews**, v. 33, n. 5–6, p. 575–605, 2014.

JOSE, V. R. R.; WINKLER, R. L. Simple robust averages of forecasts: Some empirical results. **International Journal of Forecasting**, v. 24, n. 1, p. 163–169, 2008.

KAUFMAN, L.; ROUSSEEUW, P. J. Clustering by means of medoids. In: Y. Dodge, & editor (eds.). **Reports of the Faculty of Mathematics and Informatics**. Delft: North Holland (Elsevier), 1987. p. 405–416.

KOLASSA, S. Combining exponential smoothing forecasts using Akaike weights. **International Journal of Forecasting**, v. 27, n. 2, p. 238–251, 2011.

KONING, A. J.; FRANSES, P. H.; HIBON, M.; STEKLER, H. O. The M3 competition: Statistical tests of the results. **International Journal of Forecasting**, v. 21, n. 3, p. 397–409, 2005.

KOURENTZES, N.; BARROW, D. K.; CRONE, S. F. Neural network ensemble operators for time series forecasting. **Expert Systems with Applications**, v. 41, n. 9, p. 4235–4244, 2014.

KOURENTZES, N.; BARROW, D. K.; PETROPOULOS, F. Another look at forecast selection and combination: Evidence from forecast pooling, **International Journal of Production Economics**, v. 209, p. 226-235, 2019.

KREISS, J.-P. **Asymptotic Statistical Inference for a Class of Stochastic Processes**. Habilitationsschrift. Hamburg: Universitat Hamburg, 1988.

KUCUKALI, S.; BARIS, K. Turkey's short-term gross annual electricity demand forecast by fuzzy logic approach. **Energy Policy**, v. 38, n. 5, p. 2438–2445, 2010.

KÜNSCH, H. The Jackknife and the bootstrap for general stationary observations. **The Annals of Statistics**, v. 17, n. 3, p. 1217-1241, 1989.

LEE, T. H.; YANG, Y. Bagging binary and quantile predictors for time series. **Journal of Econometrics**, v. 135, n. 1, p. 465–497, 2006.

MACDONALD, R.; MARSH, I. W. Combining exchange rate forecasts: What is the optimal consensus measure? **Journal of Forecasting**, v. 13, n. 3, p. 313–332, 1994.

MAÇAIRA, P. M.; CYRINO OLIVEIRA, F. L.; SOUZA, R. C. Forecasting natural inflow energy series with multi-channel singular spectrum analysis and bootstrap techniques. **International Journal of Energy and Statistics**, v. 3, n. 1, p. 1550005-1–1550005-17, 2015.

MAKRIDAKIS, S.; ANDERSEN, A.; CARBONE, R.; FILDES, R.; HIBON, M.; LEWANDOWSKI, R.; NEWTON, J.; PARZEN, E.; WINKLER, R. The accuracy of extrapolation (time series) methods: Results of a forecasting competition. **Journal of Forecasting**, v. 1, n. 2, p. 111–153, 1982.

MAKRIDAKIS, S.; HIBON, M. The M3-Competition: results, conclusions and implications. **International Journal of Forecasting**, v. 16, n. 3, p. 451–476, 2000.

MAKRIDAKIS, S.; SPILIOTIS, E.; ASSIMAKOPOULOS, V. The M4 Competition: Results, findings, conclusion and way forward. **International Journal of Forecasting**, v. 34, n. 4, p. 802–808, 2018.

MAKRIDAKIS, S.; SPILIOTIS, E.; ASSIMAKOPOULOS, V. The M4 Competition: 100,000 time series and 61 forecasting methods. **International Journal of Forecasting**, v. 36, n. 1, p. 54–74, 2020.

MATSYPURA, D.; THOMPSON, R.; VASNEV, A. L. Optimal selection of expert forecasts with integer programming. **Omega**, v. 78, p. 165–175, 2018.

NEWBOLD, P.; GRANGER, C. W. Experience with forecasting univariate time series and the combination of forecasts. **Journal of the Royal Statistical Society - Series A (General)**, v. 137, n. 2, p. 131-165, 1974.

NOCK, R.; GASCUEL, O. On learning decision committees. In: ___ **International Conference on Machine Learning**. 12th ed. Tahoe City, US: ICML 95, 1995, p. 413–420.

OLIVER, J. J.; HAND, D. J. On Pruning and Averaging Decision Trees. In: ___ **International Conference on Machine Learning**. 12th ed. Tahoe City, US: ICML 95, 1995, p. 430–437.

ORD, J. K.; KOEHLER, A. B.; SNYDER, R. D. Estimation and prediction for a class of dynamic nonlinear statistical models. **Journal of the American Statistical Association**, v. 92, n. 440, 1621–1629, 1997.

PEGELS, C. C. Exponential forecasting: some new variations. **Management Science**, v. 15, n. 5, 311–315, 1969.

PETROPOULOS, F.; HYNDMAN, R. J.; BERGMEIR, C. Exploring the sources of uncertainty: Why does bagging for time series forecasting work? **European Journal of Operational Research**, v. 268, n. 2, p. 545-554, 2018.

QUENOUILLE, M. H. Problems in Plane Sampling. **The Annals of Mathematical Statistics**, v. 20, n. 3, p. 355–375, 1949.

QUINLAN, J. R. Bagging, boosting, and C4.5. In: ___ **National Conference on Artificial Intelligence**. 13th ed. Portland, US: AAAI 96, 1996, p. 725–730.

R CORE TEAM (2019) **R: A language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing. Available at: http://www.R-project.org/

RAPACH, D. E.; STRAUSS, J. K. Bagging or combining (or both)? An analysis based on forecasting US employment growth. **Econometric Reviews**, v. 29, n. 5–6, p. 511–533, 2010.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. **Learning internal representations by error propagation** (No. ICS-8506). San Diego: University of California, 1985.

SCHAPIRE, R. E.; FREUND, Y.; BARTLETT, P.; LEE, W. S. Boosting the margin: A new explanation for the effectiveness of voting methods. **The Annals of Statistics**, v. 26, n. 5, p. 1651–1686, 1998.

SCHWARZ, G. Estimating the Dimension of a Model. **The Annals of Statistics**, v. 6, n. 2, p. 461–464, 1978.

SHAO, Z.; CHAO, F.; YANG, S-L.; ZHOU, K-L. A review of the decomposition methodology for extracting and identifying the fluctuation characteristics in electricity demand forecasting. **Renewable and Sustainable Energy Review**s, v. 75, p.123-136, 2017.

SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. **Statistics and Computing**, v. 14, n. 3, p. 199–222, 2004.

ŠTĚPNIČKA, M.; BURDA, M. On the results and observations of the time series forecasting competition CIF 2016. In: ___ **IEEE international conference on fuzzy systems**. IEEE. 26th ed. Naples, Italy: FUZZ-IEEE 17, 2017, p. 1–6.

STOCK, J. H.; WATSON, M. W. Combination forecasts of output growth in a seven-country data set. **Journal of Forecasting**, v. 23, n. 6, p. 405–430, 2004.

SUGIURA, N. Further analysis of the data by Akaike's information criterion and the finite corrections. **Communications in Statistics - Theory and Methods**, v. 7, n. 1, p. 13–26, 1978.

TAYLOR, J. W. Exponential smoothing with a damped multiplicative trend. **International Journal of Forecasting**, v. 19, n. 4, p. 715–725, 2003.

TIBSHIRANI, R. Regression Shrinkage and Selection via the Lasso. **Journal of the Royal Statistical Society. Series B (Methodological)**, v. 58, n. 1, p. 267–288, 1996.

TIBSHIRANI, R. Regression Shrinkage and Selection via the lasso: a retrospective. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 73, n. 1, p. 273–282, 2011.

TIMMERMANN, A., 2006. Forecast combinations. In: ELLIOT, G.; GRANGER, C. W. J.; TIMMERMANN, A. (eds). **Handbook of Economic Forecasting, v. 1**. Oxford, UK: North Holland (Elsevier), 2006. p. 135–196.

VINOD, H. D. Ranking mutual funds using unconventional utility theory and stochastic dominance. **Journal of Empirical Finance**, v. 11, n. 3, p. 353–377, 2004.

VINOD, H. D. Maximum entropy ensembles for time series inference in economics. **Journal of Asian Economics**, v. 17, p. 955–978, 2006.

VINOD, H. D. **Matrix algebra topics in statistics and economics using R**. In Handbook of Statistics (pp. 143–176). Elsevier, 2014.

VINOD, H. D.; LÓPEZ-DE-LACALLE, J. Maximum Entropy Bootstrap for Time Series: The meboot R Package. **Journal of Statistical Software**, v. 29, n. 5, 2009.

WANG, Y.; XIAO, M.; ZHOU, Y. A hybrid ensemble approximation method for chaotic time series forecast. **Journal of Information and Computational Science**, v. 9, n. 18, p. 5849–5856, 2012.

WEBB, G. I. MultiBoosting: A Technique for Combining Boosting and Wagging. **Machine Learning**, v. 40, n. 2, p. 159–196, 2000.

WERON, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. **International Journal of Forecasting**, v. 30, n. 4, p. 1030–1081, 2014.

WINTERS, P. R. Forecasting sales by exponentially weighted moving averages. **Management Science**, n. 6, p. 324-342, 1960.

WOLPERT, D. H. Stacked generalization. **Neural Networks**, v. 5, n. 2, p. 241–259, 1992.

XIA, Y.; TONG, H. Feature Matching in Time Series Modeling. **Statistical Science**, v. 26, n. 1, 21-46, 2011.

YALTA, A. T. Analyzing energy consumption and GDP nexus using maximum entropy bootstrap: The case of Turkey. **Energy Economics**, v. 33, p. 453–460, 2011.

ZONTUL, M.; AYDIN, F.; DOAN, G.; SENER, S.; KAYNAR, O. Wind speed forecasting using REPTree and bagging methods in Kirklareli-Turkey. **Journal of Theoretical and Applied Information Technology**, v. 56, p. 17–29, 2013.