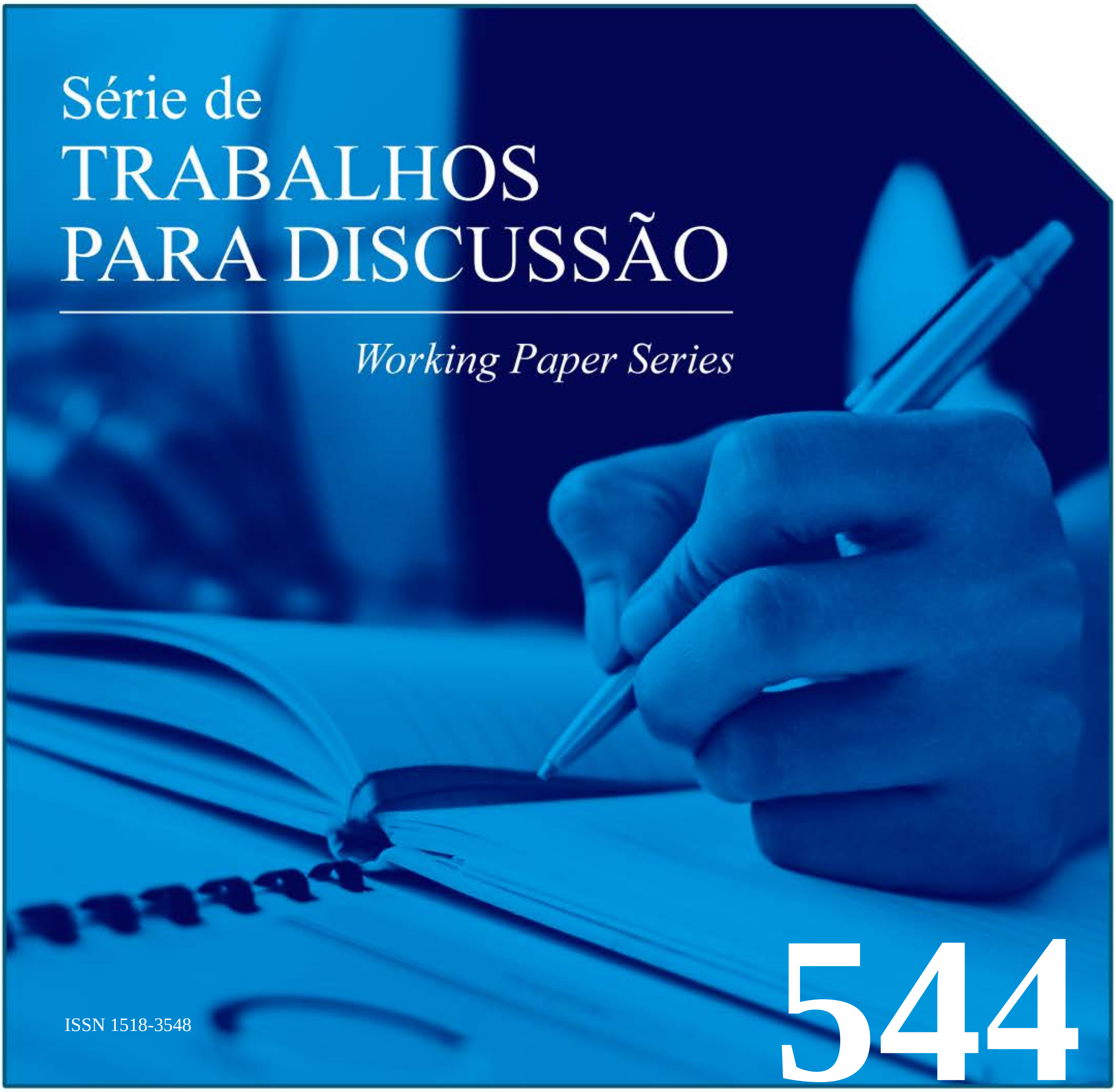


FEBRUARY 2021

Machine Learning and Oil Price Point and Density Forecasting

Alexandre Bonnet R. Costa, Pedro Cavalcanti G. Ferreira, Wagner P. Gaglianone,
Osmani Teixeira C. Guillén, João Victor Issler, Yihao Lin

BANCO CENTRAL DO BRASIL

ISSN 1518-3548
CGC 00.038.166/0001-05

Working Paper Series	Brasília	no. 544	February	2021	p. 1-100

Working Paper Series

Edited by the Research Department (Depep) – E-mail: workingpaper@bcb.gov.br

Editor: Francisco Marcos Rodrigues Figueiredo

Co-editor: José Valentim Machado Vicente

Head of the Research Department: André Minella

Deputy Governor for Economic Policy: Fabio Kanczuk

Citizen Service Division

Banco Central do Brasil

Deati/Diate

SBS – Quadra 3 – Bloco B – Edifício-Sede – 2º subsolo

70074-900 Brasília – DF – Brazil

Toll Free: 0800 9792345

Fax: +55 (61) 3414-2553

Internet: http//www.bcb.gov.br/?CONTACTUS

Non-technical Summary

This paper studies *machine learning* techniques to forecast the oil price. Given the importance of crude oil to the global economy, constructing reliable forecasts of the oil price is a relevant issue in applied macroeconomics, since large and unexpected fluctuations of this commodity impact the global economy, affecting the welfare of countries that are oil exporters as well as those that import this commodity.

In the era of *big data*, recent automated tools can potentially improve the oil price forecast accuracy over traditional approaches. The goal of this paper is to build oil price forecasts from 22 methods, including several new machine learning techniques, based on regression trees or regularization procedures, as well as standard econometric models and forecast combinations, besides the structural factor model of Schwartz and Smith (2000), which is a model of reference in the field.

To evaluate the predictive power of each method, an extensive out-of-sample forecasting exercise is conducted in both monthly and quarterly frequencies. The database contains 315 macroeconomic and financial variables. The sample covers the period from January 1991 to June 2020, and forecast horizons vary from one month up to five years.

Overall, the empirical results reveal a good performance of the machine learning methods in the short and medium horizons. Future oil prices and the Schwartz-Smith model also provide forecasts with comparable accuracy in such horizons. At longer horizons, forecast combinations become relevant too.

In several cases, the accuracy gains in respect to the random walk (benchmark) forecast are statistically significant and reach two-digit figures, in percentage terms, using the R^2 out-of-sample statistic. This is an expressive improvement vis-a-vis the previous literature, thus confirming that machine learning tools can indeed contribute to the standard statistical toolkit used in macroeconomic forecasting.

Sumário Não Técnico

Este artigo investiga técnicas de aprendizado de máquina para previsão do preço do petróleo. Dada a importância do petróleo para a economia global, a construção de previsões confiáveis do preço do petróleo é uma questão relevante em macroeconomia aplicada, uma vez que grandes ou inesperadas flutuações dessa *commodity* têm um impacto na economia global, afetando tanto o bem-estar de países exportadores de petróleo como daqueles que importam essa *commodity*.

Na atual era de *big data*, novas ferramentas automatizadas podem potencialmente melhorar a precisão da previsão do preço do petróleo em relação às abordagens tradicionais. O objetivo deste artigo é construir previsões do preço do petróleo a partir de 22 métodos, incluindo diversas novas técnicas de aprendizado de máquina, baseadas em árvores de regressão ou técnicas de regularização, bem como modelos econométricos usuais e combinações de previsões, além do modelo estrutural de fatores de Schwartz e Smith (2000), que é um modelo de referência na área.

Para avaliar a capacidade preditiva de cada método, um amplo exercício de previsão fora da amostra é realizado nas frequências mensal e trimestral. A base de dados contém 315 variáveis macroeconômicas e financeiras. A amostra considerada abrange o período de janeiro de 1991 a junho de 2020, e os horizontes de previsão variam de um mês até cinco anos.

De maneira geral, os resultados empíricos revelam um bom desempenho dos métodos de aprendizado de máquina nos horizontes de curto e médio prazos. Os preços futuros do petróleo e o modelo de Schwartz-Smith fornecem previsões com equivalente grau de precisão em tais horizontes. Em horizontes mais longos, as combinações de previsão também se tornam relevantes em termos de capacidade preditiva.

Em vários casos, os ganhos de precisão em relação à previsão do passeio aleatório (modelo *benchmark*) são estatisticamente significativos e atingem valores de dois dígitos, em termos percentuais, usando a estatística R^2 fora da amostra. Trata-se de uma melhoria expressiva em relação à literatura anterior, confirmando dessa forma que ferramentas de aprendizado de máquina podem, de fato, contribuir para o conjunto de ferramentas estatísticas utilizadas em previsões macroeconômicas.

Machine Learning and Oil Price Point and Density Forecasting*

Alexandre Bonnet R. Costa† Pedro Cavalcanti G. Ferreira‡
Wagner P. Gaglianone§ Osmani Teixeira C. Guillén¶
João Victor Issler‖ Yihao Lin**

Abstract

The purpose of this paper is to explore machine learning techniques to forecast the oil price. In the era of *big data*, we investigate whether new automated tools can improve over traditional approaches in terms of forecast accuracy. Oil price point and density forecasts are built from 22 methods, including regression trees (random forest, quantile regression forest, xgboost), regularization procedures (elastic net, lasso, ridge), standard econometric models and forecast combinations, besides the structural factor model of Schwartz and Smith (2000). The database contains 315 macroeconomic and financial variables, used to build high-dimensional models. To evaluate the predictive power of each method, an extensive pseudo out-of-sample forecasting exercise is built, in monthly and quarterly frequencies, with horizons from one month up to five years. Overall, the results indicate a good performance of the machine learning methods in the short run. Up to six months, the lasso-based models, oil future prices, and the Schwartz-Smith model provide the best forecasts. At longer horizons, forecast combinations also become relevant. In several cases, the accuracy gains in respect to the random walk forecast are statistically significant and reach two-digit figures, in percentage terms, using the R^2 out-of-sample statistic; an expressive achievement compared to the previous literature.

Keywords: Machine Learning; Commodity Prices; Forecasting.

JEL Classification: C14; C15; C22; C53; C55; E17; E31.

*The views expressed in the paper are those of the authors and do not necessarily reflect those of the Banco Central do Brasil, Getulio Vargas Foundation or of Petrobras. We are especially grateful for the helpful comments and suggestions given by the participants at the 2019 Workshop on Energy Economics - Econometric Analysis of Energy Demand and Climate Change (Seoul, May 2019). The authors gratefully acknowledge the support from CNPq and INCT through different grants.

†Corporate Risks, Petrobras. E-mail: abrcosta@petrobras.com.br

‡Graduate School of Economics, Getulio Vargas Foundation. E-mail: Pedro.Ferreira@fgv.br

§Research Department, Banco Central do Brasil. E-mail: wagner.gaglianone@bcb.gov.br

¶Open Market Operations Department, Banco Central do Brasil, and Ibmec/RJ. E-mail: osmani.guillen@bcb.gov.br

‖Corresponding Author. Graduate School of Economics, Getulio Vargas Foundation, Praia de Botafogo 190, s.1104, Rio de Janeiro, RJ 22.253-900, Brazil. E-mail: Joao.Issler@fgv.br

**Graduate School of Economics, Getulio Vargas Foundation. E-mail: lin-291@hotmail.com

1 Introduction

Traditional forecasting methods often rely on fitting data to a pre-specified relationship between dependent and independent variables, thus assuming a specific functional and stochastic process. In contrast, a different approach to statistical analysis and forecasting, in particular, is offered by *machine learning* (ML), which is a narrow form of artificial intelligence, often described as the *art and science of pattern recognition.* Indeed, ML is to a great extent a data-driven framework, since it requires mild assumptions about the underlying statistical relationship in the data. According to Hansen (2019): *"The term 'machine learning' is a new and somewhat vague term, but typically is taken to mean procedures which are primarily used for point prediction in settings with unknown structure. Machine learning methods generally allow for large sample sizes, large number of variables, and unknown structural form."*

Although machine learning encompasses a wide variety of models, it generally comprises two core elements: a *learning method*, where data are used to determine the best fit for the input variables, and an *algorithm*, which models the relationship between the input and output. According to Jung et al. (2018), ML methods can be categorized into three types:

(i) *supervised learning*, where the dependent variables are clearly identified, even if the specific relationships in the data are not known (e.g., linear regression, logistic regression);

(ii) *unsupervised learning*, where there is no specific output defined beforehand, and the goal is to recognize data patterns and determine output classification categories (e.g., cluster analysis, principal components);

(iii) *reinforcement learning*, which iteratively search for an optimal location of the input variables that yields the highest reward, that is, optimizes a given "reward" function using no training set (e.g., dynamic programming models, sarsa, Q-learning).

According to Varian (2014), the growing amounts of data and ever complex-growing relationships warrant the usage of machine learning in economics. One of the advantages of ML over traditional approaches is to automate as many of the modeling choices as possible in a manner that is not subject to the discretion of the forecaster (Hall, 2018).

Producing accurate forecasts is not an easy task, since it requires an approach complex enough to incorporate relevant variables but also focused on excluding irrelevant data. ML methods, in general, are able to deal with large amounts of data (*big data*) and nonlinear patterns in the data, often hidden to standard linear models, thus offering an alternative and compelling approach to traditional econometric models.[1]

Given the importance of crude oil to the global economy, constructing reliable forecasts of the oil price is a relevant issue in applied macroeconomics, since large and unexpected fluctuations of this commodity will have an impact on the global economy, affecting the welfare of countries that are oil exporters as well as those that import this commodity.

[1] According to Hall (2018), it is crucial to control the model complexity by using an algorithm that yields a model complex enough to avoid underfitting the data, but not so complex as to overfit it.

According to Alquist et al. (2013), not only accurate oil price forecasts have the potential to improve the forecast-accuracy of relevant macro variables, but also some sectors of the economy directly depend on oil price forecasts for their business (e.g., the oil spot price is critical to investment decisions in the oil industry). Also, central banks and private sector agents quite often view the price of oil as one of the key elements in producing macroeconomic projections and in assessing risks.

The relationship between oil price dynamics and key macroeconomic variables is well documented in the literature; see, for instance, Hamilton and Herrera (2004), Baumeister and Kilian (2016), Kilian and Vigfusson (2017), Bjørnland, Larsen and Maih (2018), Bjørnland and Zhulanova (2018).

The literature on oil price forecasting is also vast. Just to mention a few papers, see Cologni and Manera (2008), Miller and Ni (2011), Ravazzolo and Rothman (2012), Hong and Yogo (2012), Gargano and Timmermann (2014), Baumeister and Kilian (2015), Mohaddes and Pesaran (2016), Gogolin et al. (2018) and Yu et al. (2019).

The objective of this paper is to forecast the real oil price (Brent crude) based on a large number of macroeconomic and financial variables. Our goal is also to assess whether machine learning techniques can offer real improvement to forecast-accuracy in applied macroeconomics, and thus make a contribution to the standard statistical toolkit used in macro forecasting. Our research contributes to the latter literature in two ways: The first original contribution is to *density forecast* the oil price using machine learning tools. The second contribution is to help "opening" the machine learning *black box*,[2] by providing a full set of auxiliary graphs to help investigating the forecasting exercise results.[3]

In sum, machine learning tools are used here to build Brent oil price forecasts based on 22 competing methods, including regularization[4] procedures that introduce penalties for *overfitting*[5] the data (e.g., LASSO and Elastic Nets), more recent supervised machine learning techniques (e.g., Quantile Regression Forest and XGBoost), which are nonparametric approaches based on the recursive binary partitioning of the covariate space, besides standard econometric models (e.g., ARIMA), the forecast combination methods discussed in Duarte

[2]The *black box* expression applied to ML has been around for years now. It is often used to critisize neural networks' lack of explainability. Here, we turn the *black box* into a *gray box* by providing complementary tools to analyze and further understand the ML results.

[3]For instance, (i) *word cloud* and *variable importance* plots to reveal the most important variables for oil price forecasting according to a given ML method of interest; (ii) decomposition of the mean-squared forecast error plots, which allows one to disentangle the effect of forecast bias from the variance of the forecast. This is particularly important in model selection and helps understanding why some methods display a better forecast accuracy compared to others; and (iii) time series plots of the differences between the cumulative squared prediction error, which complement the graphical analysis, by presenting the cumulative performance of a given forecasting method over time in respect to a selected benchmark.

[4]For example, the elastic net mixes two types of regularization, by penalizing the number of variables in the model and the extent to which any given variable contributes to the model's forecast. By applying such penalties, the elastic net model *learns* which variables are most important, thus eliminating the need for researchers to make discretionary choices about which variables to include in the model.

[5]In statistics, *overfitting* denotes the production of an analysis, which is assumed to be valid for the entire population (for instance, an estimated input-output relationship), that corresponds too closely to a particular set of data, but it may fail to fit additional data, or forecast future observations, reliably.

et al. (2019), and the two-factor model of Schwartz and Smith (2000).

To do so, we put together a set of 630 time series, coming from 315 macroeconomic and financial variables used to build high-dimensional models. In order to evaluate the forecast accuracy of each approach, an extensive pseudo out-of-sample forecasting exercise is conducted in monthly and quarterly frequencies. The sample covers the period from January 1991 to June 2020, and forecast horizons vary from one month up to five years.

Overall, the results corroborate recent findings in favor of the nonlinear automated procedures, indicating machine learning algorithms can indeed statistically surpass, in the short run, some traditional methods in terms of Root Mean Squared Error (RMSE). One of the reasons is the ability of some machine learning techniques in reducing the forecast variance while maintaining the forecast bias under control.[6] As result, forecast accuracy can be improved when compared to traditional oil price forecasting models.

In particular, the adaptive LASSO (or simply *adalasso*) exhibited the lowest RMSE at the one-month forecast horizon. The empirical exercise also revealed a good performance of other machine learning approaches (e.g., Random Forest and XGBoost) at short/medium horizons, providing forecasts that are statistically superior to the random walk for horizons up to three months. In the monthly frequency, other LASSO family models, the Brent future prices and the Schwartz-Smith model provided the best forecasts for horizons up to six months. At longer horizons, the forecast combination techniques discussed in Duarte et al. (AF and BCAF) gain importance, together with the Brent future prices and the Schwartz-Smith forecasts.

In both frequencies, and in several cases, the forecast accuracy gains over the benchmark model (random walk without drift) are statistically significant, and reach two-digit figures, in percentage terms: the R^2 out-of-sample statistics, for the best model in each horizon, range from 14% to 40% in monthly frequency, and between 9% to 49% in quarterly frequency; expressive results compared to the previous literature.

Regarding density forecasts, it is worth mentioning the good performance, in most part of the horizons considered at monthly frequency, of the Brent future prices, the forecast combination model AF (long horizons) and the Schwartz-Smith model. The excellent result of the Schwartz-Smith densities, generated from model simulations, in great part of forecast horizons at the quarterly frequency, should also be mentioned.

The outline of the paper is as follows. Section 2 presents the methodology comprising machine learning and traditional econometric models to forecast the oil price. Section 3 presents the forecasting exercise and Section 4 concludes. The Technical Appendix provides additional results.

[6] In the context of neural networks, Neal et al. (2018) find both bias and variance can decrease as the number of parameters grows (i.e., model complexity). The authors also discuss this outcome by introducing a new decomposition of the variance to disentangle the effects of model optimization and data sampling.

2 Methodology

2.1 Point Forecast

In this paper, oil price forecasts are constructed from 22 forecasting methods listed in Table 1. Besides some traditional approaches to forecast the oil prices, such as the random walk and the ARIMA models, this paper considers factor models, which are well-known in the macroeconometrics literature (e.g., Stock and Watson, 2002; Schwartz and Smith, 2000). The set of forecasting methods also includes several non-linear machine learning methods, based on regularization procedures (e.g., LASSO and elastic net) or regression trees (e.g., random forest and quantile regression forest).

Table 1 - Models/methods used to forecast the oil prices

	Model	*References*
1	Random walk	-
2	Random walk with drift	-
3	Random walk with drift (last 5 years)	Alquist et al. (2013)
4	ARIMA	-
5	Factor model 1	Bai and Ng (2002, 2008)
6	Factor model 2	Bai and Ng (2002, 2008)
7	Elastic net	Zou and Hastie (2005)
8	LASSO	Tibshirani (1996)
9	Adaptive LASSO	Zou (2006)
10	Ridge regression	Hoerl and Kennard (1988)
11	Random forest	Breiman (2001)
12	Quantile regression forest	Meinshausen (2006)
13	XGBoost	Chen and Guestrin (2016)
14	AF	Issler and Lima (2009)
15	BCAF	Issler and Lima (2009)
16	Brent futures	-
17	Schwartz-Smith (mean)	Schwartz and Smith (2000)
18	Schwartz-Smith (median)	Schwartz and Smith (2000)
19	Mean (all models)	-
20	Median (all models)	-
21	Mean (selected models)	-
22	Median (selected models)	-

The list of models, of course, is far from an exhaustive list, since more complex models could be included. Although this extension would be valuable, the list presented here seems to be a reasonable starting point to compare the accuracy of traditional econometric approaches with competing machine learning techniques.

Our variable of interest is the Brent oil real price Y_t, and our goal is to forecast the $h-$period change of the logarithm of Y_t at period $t+h$, that is $(y_{t+h} - y_t)$, where $y_t = \ln(Y_t)$, using the information set available at period t. In this sense, the dependent variable $(y_{t+h} - y_t)$ is modeled as a function of a set of predictors $\widetilde{x}_t$, measured at time t, as follows:

$$(y_{t+h} - y_t) = \Upsilon_h(\widetilde{x}_t) + \varepsilon_{t+h}, \tag{1}$$

where $\Upsilon_h(\cdot)$ is a possibly nonlinear mapping of a set of predictors, ε_{t+h} is the forecasting error and $\widetilde{x}_t$ may include weakly exogenous predictors, lagged values of oil prices and a large number of potential covariates. Let $\widetilde{x}'_t \equiv \{\mathbf{1}_t, x_t, x_{t-1}, \ldots, x_{t-s}\}$, where $\mathbf{1}_t$ is a constant term, $x_t = \{x_{1,t}, \ldots, x_{n,t}\}$ is a set of n predictors and s is the maximum lag adopted for the set of variables x_t when forming the database $\widetilde{x}'_t$.

In order to build our forecasting exercise, the sample is divided into two periods: the first one $(t = 1, ..., T_1)$ is labeled as "training set", used to estimate the tuning parameters and model coefficients. The second period, also known as the "test set", comprising the last P observations $(t = T_1 + 1, ..., T)$, is used to confront the observations of $(y_{t+h} - y_t)$ with out-of-sample forecasts. This way, $P = T - T_1$ observations are used to compare different forecasts and compute forecast-accuracy measures.

In regularization methods (models 7-10), the mapping $\Upsilon_h(\cdot)$ is linear, such that:

$$(y_{t+h} - y_t) = \widetilde{x}'_t \beta_h + \varepsilon_{t+h}, \tag{2}$$

where β_h is a vector of unknown parameters, estimated using a sample of $t = 1, ..., T_1$ observations. Note that for these models, the *direct forecast* approach is adopted, where the oil price change $(y_{T_1+h} - y_{T_1})$ is modeled as a function of a set of predictors $\widetilde{x}'_{T_1}$ available at period T_1. In other words, for each horizon h, a different vector of unknown parameters β_h is estimated (in contrast to the iterated multistep approach; see Marcellino, Stock and Watson, 2006). This way, one avoids the necessity of estimating a model for the time-evolution of $\widetilde{x}_t$. The pseudo out-of-sample forecast of $(y_{T_1+h} - y_{T_1})$ from these ML approaches, labelled $f_{y_{T_1+h}}$, is given by:

$$f_{y_{T_1+h}} = \widetilde{x}'_{T_1} \widehat{\beta_h}, \quad \text{for } h = 1, ..., H. \tag{3}$$

To evaluate forecast-accuracy, the root mean-squared error (RMSE) is computed for all forecasts of the Brent oil real prices Y_t, generated from the models listed in Table 1. Next, the 22 forecasting methods considered in this paper are described in details.

Model 1 (RW): A natural *benchmark* for all competing methods to forecast the real price of oil is the canonical random walk (RW) model, which assumes here the $h-$period oil price change is an unforecastable martingale difference sequence (MDS), that is $E(y_{t+h} - y_t \mid \mathcal{F}_t) = 0$, for all $t = 1, ..., T_1$ and $h = 1, ..., H$. Thus, the RW forecast assumes the oil price remains unchanged along the out-of-sample period, that is, $f^{m1}_{y_{T_1+h}} = 0$ for all h.

Models 2 and 3 (RW with drift): These variants of the random walk approach assume $E\left(y_{t+h} - y_t \mid \mathcal{F}_t\right) = drift{*}h$, where the $drift$ parameter is estimated over the training sample (model 2) or over the last five years (model 3).[7] Thus, $f^{m2,m3}_{y_{T_1+h}} = \widehat{drift} * h$.

Model 4 (ARIMA): One of the most common statistical models used for time-series forecasting is the autoregressive moving average (ARMA) model, which assumes future observations are primarily driven by recent observations. Here, one considers the ARIMA (Autoregressive Integrated Moving Average) approach, which allows for integrated series. The logarithm of the real oil price, which often exhibits persistent behavior, seems to be consistent with this setup. Thus, one assumes in this approach $y_t = \ln(Y_t)$ follows an ARIMA(p, d, q) process, where p is the number of AR terms, d is the integration order of y_t, and q is the number of MA terms.

Model 5 (Factor model 1, direct forecast): The idea that time variations in a large number of variables can be summarized by a small number of factors is empirically attractive and it is employed in a large number of studies in economics and finance; see Forni et al. (2000) and Stock and Watson (2002). Zagaglia (2010) uses a factor model to forecast the nominal oil price along the 2003-2008 period. Here, one explores the use of factor models for forecasting the real price of oil. Let $x_{i,t}$ be the observed data for the $i-$th cross-section unit at time t, for $i = 1, ..., N$ and $t = 1, ..., T_1$, and consider the following factor representation of the data:

$$x_{i,t} = \lambda_i' F_t + e_{i,t}, \tag{4}$$

where F_t is a vector of common factors, λ_i is a vector of factor loadings associated with F_t and $e_{i,t}$ is the idiosyncratic component of $x_{i,t}$. Note that λ_i, F_t and $e_{i,t}$ are unknown since only $x_{i,t}$ is observable. Here, one estimates the factors and respective loadings using principal components analysis (PCA), which is a well-established technique for dimension-reduction in time series. The number of components is determined by the Bai and Ng (2002) criterion. After the PCA estimation of the common factors F_t, the *direct forecast* approach is used to model the oil price change at time $t + h$, as follows:

$$(y_{t+h} - y_t) = \beta_h F_t + \varepsilon_{t+h}. \tag{5}$$

The respective out-of-sample forecast is given by:

$$f^{m5}_{y_{T_1+h}} = \widehat{\beta_h}\widehat{F_{T_1}}, \quad \text{for } h = 1, ..., H. \tag{6}$$

It is worth mentioning this approach only uses here a subset of predictors, which are pre-selected by taking into account our variable of interest is the oil price change. Bai and

[7] This "local" drift model assumes, for instance, oil traders extrapolate from the recent behavior of the spot price when they form expectations about the future prices. According to Alquist et al. (2013), the local drift model is designed to capture "short-term forecastability" that arises from local trends in the oil price data.

Ng (2008) shows the factor model forecasting performance could be improved by previously selecting (or targeting) the predictors. The core idea is that irrelevant predictors employed to build a factor model only add noise into the analysis, and thus produce factors with a poor predictive performance.

In this sense, it is adopted a pre-selection of variables (*target predictors*) to be included in the factor analysis, as follows: (i) regress $(y_{t+h} - y_t)$ on the intercept and the candidate variable $\widetilde{x}'_{i,t} \in \widetilde{x}'_t$, for all $i = 1, ..., N$; (ii) compute the t-statistic for the coefficient associated to $\widetilde{x}'_{i,t}$; and (iii) include $\widetilde{x}'_{i,t}$ in the set of predictors (used to extract the factors) only if it is statistically significant at a 5% level.

Model 6 (Factor model 2, iterated forecast): This approach is a variant of the previous one, but using an iterated method instead of the direct forecast approach. The idea is again to employ common factors, but to model the oil price change in a *contemporaneous* way in respect to the factors, that is:

$$(y_{t+h} - y_t) = \gamma F_{t+h} + v_{t+h}. \tag{7}$$

Following Bańbura et al. (2013), the factors are assumed to follow a VAR process, that is, $F_t = \Phi(L)F_t + u_t$. The out-of-sample forecast from this factor model is given by:

$$f^{m6}_{y_{T_1+h}} = \widehat{\gamma}\widehat{F_{T_1+h|T_1}}, \quad \text{for } h = 1, ..., H, \tag{8}$$

where $\widehat{F_{T_1+h|T_1}}$ is the $h-$step ahead forecast of the vector of common factors using a VAR model for F_t, estimated in a recursive scheme.

Again, the factor model considers *target predictors*, as discussed in model 5.

Model 7 (Elastic net): The elastic net is a regularization and variable selection method proposed by Zou and Hastie (2005) as a generalization of the LASSO. Similarly to the LASSO, the elastic net simultaneously does automatic variable selection and continuous shrinkage, and it can select groups of correlated variables. Simulation studies show the elastic net often outperforms the LASSO, in terms of predictive power, while enjoying a similar sparsity representation. The elastic net encourages a grouping-effect, where highly correlated regressors tend to be jointly included (or excluded) from the model, and it can be particularly useful when the number of predictors k is high when compared to the number of observations T. For a nonnegative shrinkage parameter λ, and a combination parameter α strictly between 0 and 1, the elastic net solves the following problem:

$$\widehat{\beta} = \underset{\{\beta_1,...,\beta_k\}}{\arg\min} \left(\frac{1}{T}\sum_{t=1}^{T} \left((y_{t+h} - y_t) - \sum_{j=1}^{k} x'_{j,t}\beta_j \right)^2 + \lambda P_\alpha(\beta) \right), \tag{9}$$

where

$$P_{\alpha}(\beta) = \sum_{j=1}^{k} \alpha \left|\beta_j\right| + \frac{(1-\alpha)}{2}\beta_j^2. \tag{10}$$

Note that the elastic net becomes the LASSO when $\alpha = 1$. As α shrinks toward 0, the elastic net approaches the ridge regression. For other values of α, the penalty term $P_{\alpha}(\beta)$ interpolates between the l_1-norm of β and the squared l_2-norm of β. Once again, the tuning parameter λ controls the overall strength of the penalty. Note the objective function is convex and so can be minimized using any convex optimization method such as gradient or coordinate descent.

Although the elastic net is defined here by using (λ, α), this is not the only choice as the tuning parameters; see Zou and Hastie (2005) for further details. For example, one could use the l_1-norm of the coefficients or the fraction of the l_1-norm to parameterize the elastic net. There are well-established methods for choosing the tuning parameters (λ, α). For instance, K-fold *cross-validation* (CV) is a popular method for computing the prediction error and comparing different models using training data. The loss often used is the mean squared error (MSE) and the goal is to produce the "cross-validation curve", which computes the MSE as a function of the tuning parameter λ over a pre-selected grid.[8]

In the elastic net, since there are two tuning parameters, one needs to cross-validate the model on a two-dimensional surface. The minimum MSE, thus, provides the pair (λ, α) to be used in the final model estimation. In this paper, however, the Bayesian Information Criterion (BIC) is adopted, instead of cross-validation, to choose the tuning parameters.[9] Finally, the vector of parameters β can be estimated using the penalized maximum likelihood, in which the regularization path (i.e., the path of each coefficient β_j against, for instance, the l_1-norm of the whole coefficient vector as λ varies) can be computed.

Model 8 (LASSO): The least absolute shrinkage and selection operator (LASSO) was originally proposed by Tibshirani (1996). The core idea is to shrink to zero the irrelevant coefficients. The LASSO is a penalized least squares method imposing an l_1-penalty on the regression coefficients, as follows:

$$\widehat{\beta} = \underset{\{\beta_1,\ldots,\beta_k\}}{\arg\min} \left(\frac{1}{T}\sum_{t=1}^{T} \left((y_{t+h} - y_t) - \sum_{j=1}^{k} x'_{j,t}\beta_j \right)^2 + \lambda \sum_{j=1}^{k} \left|\beta_j\right| \right), \tag{11}$$

[8] To do so, for each fold, the algorithm splits the training set of observations in two parts: *training folds* (used for the estimation of parameters) and *test fold* (based on the remaining observations, used for model predictions). Then, forecast errors are computed and used to calculate the MSE over the entire set of predictions using all K-folds.

[9] Zou et al. (2007) show one can consistently estimate the degrees of freedom of the LASSO model using *information criteria* as alternative to the CV approach. An advantage of such procedure is that selecting the model using information criterion is faster than using cross-validation. More importantly, performing CV in a time-series context may be complicated in cases where the data are not independent and identically distributed (i.i.d.); see Medeiros et al. (2016).

where β is the vector of parameters and λ is the shrinkage parameter. Due to the nature of the l_1-norm, the LASSO approach is able to do continuous shrinkage and automatic *variable selection* simultaneously, whereas the ridge regression only shrinks the coefficients close to zero (but does not exclude them from the model). Again, setting $\lambda = 0$ leads to the OLS estimation. According to Cheng et al. (2019), LASSO is "*the most intensively studied statistical method in the past 15 years*". Indeed, it has shown success in many practical situations, since it can handle more variables than observations. Nonetheless, it has some limitations and might even become an inappropriate variable selection method in some cases. Zou and Hastie (2005) list a few examples: (i) when the number of predictors k is greater than the number of observations T, the LASSO selects at most T variables before it saturates, due to the nature of the convex optimization problem; (ii) in the case of *grouping effect*[10], the LASSO tends to select only one variable from the group (and does not care which one is selected); (iii) in the case of $T > k$ and in the presence of highly correlated predictors, it has been empirically observed that ridge regression tends to perform better than LASSO.

Model 9 (Adaptive LASSO): Zou (2006) shows the LASSO estimator is inconsistent for variable selection under certain circumstances. This way, the author proposes a new version of the LASSO, called the adaptive LASSO (or simply *adalasso*), where adaptive weights are used for penalizing different coefficients in the l_1-penalty. According to the author, the adaptive LASSO enjoys the oracle properties (i.e., it performs as well as if the true underlying model were known) and does not select useless variables (which may damage the forecasting accuracy). The core idea behind the model is to use some previously known information to select the variables more efficiently. In practice, it consists of a two-step estimation that uses a first model to generate different weights w_j for each candidate variable $x_{j,t}$. These weights are used in the second-step as additional information. The *adalasso* estimator is, thus, defined as:

$$\widehat{\beta} = \underset{\{\beta_1,\ldots,\beta_k\}}{\arg\min} \left(\frac{1}{T}\sum_{t=1}^{T} \left((y_{t+h} - y_t) - \sum_{j=1}^{k} x'_{j,t}\beta_j \right)^2 + \lambda \sum_{j=1}^{k} w_j \left|\beta_j\right| \right), \tag{12}$$

where $w_j = \left|\widehat{\beta}_j^*\right|^{-\tau}$ represents the weights; $\widehat{\beta}_j^*$is a parameter estimated in the first-step, and $\tau > 0$ is an additional tuning parameter (which can be chosen by using the same criterion as λ) that determines how much one wants to emphasize the difference in the weights. In general, τ is set to unity and $\widehat{\beta}_j^*$ is estimated in the first-step using LASSO. According to Medeiros and Mendes (2016), the conditions required by the *adalasso* estimator are very general and the model works even when the errors are non-Gaussian, heteroskedastic and the number of variables increases faster than the number of observations.

[10] The grouping effect occurs if the regression coefficients of a group of highly correlated variables tend to be equal (up to a change of sign if negatively correlated).

Model 10 (Ridge regression): It is well known OLS often does poorly in prediction on future data (e.g., due to overfitting). In this sense, penalization techniques have been proposed in the literature to improve OLS accuracy. For instance, the ridge regression (see Hoerl and Kennard, 1970) minimizes the squared sum of the residuals subject to a bound on the l_2-norm of the parameters, as follows:

$$\widehat{\beta} = \underset{\{\beta_1,\ldots,\beta_k\}}{\arg\min} \left(\frac{1}{T} \sum_{t=1}^{T} \left((y_{t+h} - y_t) - \sum_{j=1}^{k} x'_{j,t} \beta_j \right)^2 + \lambda \sum_{j=1}^{k} \beta_j^2 \right), \tag{13}$$

where β is the $k \times 1$ vector of parameters, $(y_{t+h} - y_t)$ is the dependent variable, $\{x_{1,t}, \ldots, x'_{k,t}\}$ is the $k \times 1$ vector of regressors and λ is the shrinkage parameter, which controls the magnitude of the shrinkage penalty. The optimal value of λ can be determined by *cross-validation* (i.e., splitting the data into K folds and iteratively re-estimating the model for each fold) or using information criteria. Choosing a higher λ leads to a stronger shrinkage of the coefficients, whereas setting $\lambda = 0$ produces the same results of the ordinary least squares (OLS) regression. Also, because ridge regression is a continuous shrinkage method, it can achieve a better out-of-sample performance through a *bias-variance* trade-off (i.e., use regularization to balance the forecast errors due to bias and variance). In particular, the ridge regression is good at improving the OLS counterpart when multicollinearity is present. However, ridge cannot produce a parsimonious model, since it always keeps all the predictors in the model.

Model 11 (Random forest): Random Forest (RF) was introduced as a machine learning tool in Breiman (2001) and have since proven to be very popular and powerful for high-dimensional regression and classification. A random forest is a collection of regression trees, designed to reduce the prediction variance by using bootstrap aggregation (*bagging*) of randomly constructed regression trees. A *regression tree* is a nonparametric model based on the recursive *binary* partitioning of the covariate space X.[11] The main idea is that if a sufficiently large number of step functions are used, then a step function can be a good approximation to any functional form.[12] The model is often represented as a binary decision tree, with P parent nodes (also called "split nodes") and L terminal nodes (also called "leaves"; which represent different partitions of X).

In practice, one major problem with *regression trees* is their high forecast variance. Usually, a small change in the data lead to a very different sequences of splits. The main reason for such instability is the hierarchical nature of the algorithm: the effect of a big error in the top split is propagated down to all of the splits below it. To overcome this issue, one can

[11] Rather than splitting each node into just two groups, one might consider multiple splits into more than two groups at each stage. However, according to Hastie et al. (2009, p.311), while this can sometimes be useful, it is not a good general strategy, since multiple splits fragment the data too quickly, leaving insufficient data at the next level down.

[12] According to Hansen (2019): *"The literature on regression trees has developed some colorful language to describe the tools, based on the metaphor of a living tree. 1. A split point is node. 2. A subsample is a branch. 3. Increasing the set of nodes is growing a tree. 4. Decreasing the set of nodes is pruning a tree."*

employ the *bagging* technique (i.e., bootstrap aggregation), which consists on fitting the same regression tree several times to bootstrap-sampled versions of the training data and average the result. This bootstrapping approach often leads to better model performance because it decreases the forecast variance, without increasing too much the bias.[13]

The *random forest* approach uses a modified bagging algorithm (*random subspace projection*) that selects, at each candidate split in the learning process, a random subset of covariates. The reason for doing this is the correlation of the trees in an ordinary bootstrap sample: if one or a few covariates are very strong predictors for the dependent variable, these covariates will be selected in many of the K bootstraped trees, causing them to become correlated. According to Hansen (2019), the modification proposed by RF is to *decorrelate* the bootstrap regression trees by introducing extra randomness. The random forest algorithm can be summarized as follows:[14]

Given a training set (Y_i, X_i), for $i = 1, ..., n$, where Y is the dependent (response) variable and X represents a set of covariates, bagging repeatedly (K times) selects a random sample with replacement of the training set and fits regression trees to these bootstraped samples, that is, for $k = 1, ..., K$:

(i) sample with replacement n training observations from (X, Y); calling them (X_k, Y_k);

(ii) train a regression tree $T_k(\cdot)$ on (X_k, Y_k);

(iii) build the random forest prediction of Y conditioned on the test set (unseen samples x') by averaging the predictions from all the individual regression trees on x', as follows:

$$E_{\text{random forest}}(Y \mid X = x') = \frac{1}{K}\sum_{k=1}^{K} T_k(x'), \tag{14}$$

where $T_k(x')$ is the conditional forecast of Y from the k-th regression tree.

Model 12 (Quantile regression forest): Random forest approximates the conditional mean of Y by constructing a weighted average over the sample observations of Y. Nonetheless, the technique can also provide information about the full conditional distribution of the response variable, not only about the conditional mean. This information can be used, for instance, to build prediction intervals and account for outliers in the data. This way, conditional quantiles can be inferred with quantile regression forests (QRF), a generalization of random forests proposed by Meinshausen (2006).[15]

On the other hand, the conditional mean of Y can be approximated by a combination of conditional quantiles (i.e., integrating the conditional quantile function of Y over the entire

[13] While the predictions of a single tree are highly sensitive to noise in its training set, the average of many trees might be not, as long as the trees are not correlated. Besides, training many trees on a single training set would give strongly correlated trees, whereas bootstrap sampling helps de-correlating the trees by showing them different training sets.

[14] See the Technical Appendixes 3-4 and Hastie et al. (2009, chapters 9 and 15) for further details.

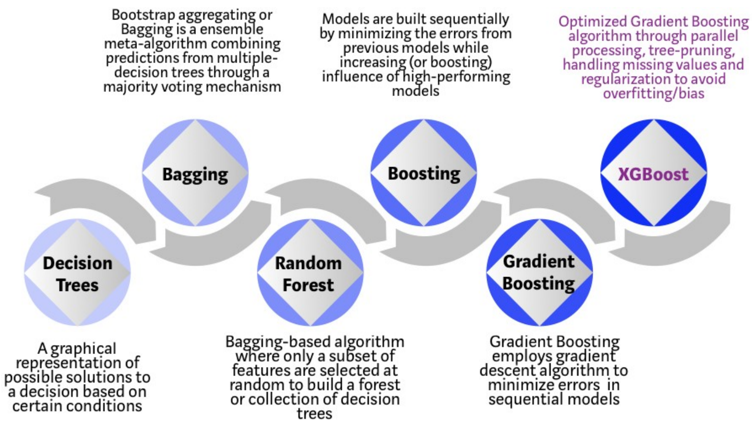
[15] The main difference between QRF and RF is that for each node (in each tree), RF keeps only the mean of the observations that fall into this node (and neglects all other information). In contrast, QRF keeps the value of all observations in this node (not just their mean) and assesses the conditional distribution based on this full information.

domain). In this sense, Araujo and Gaglianone (2020) proposed a quantile combination approach using QRF to build conditional mean forecasts of Y; see the Technical Appendix 3 for further details. The idea follows the averaging scheme of quantiles conditional on predictors selected by LASSO, as proposed by Lima and Meng (2017).[16] The advantage of both approaches relies on the fact that quantiles are robust to *outliers* (in our case, extreme unanticipated oil shocks), which potentially improves forecast-accuracy and likely impact the performance of standard models, which are usually designed to only account for average responses.

Model 13 (XGBoost): Extreme Gradient Boosting (or simply XGBoost) is a decision-tree-based ensemble algorithm that uses a gradient boosting setup proposed by Chen and Guestrin (2016). It improves upon the previous gradient boosting frameworks through systems optimization and algorithmic enhancements.[17]

According to Morde and Setty (2019), the XGBoost algorithm has the best combination of prediction performance and processing time compared to other algorithms. As result, it is widely used in many data science competitions (and there is a strong community of data scientists contributing to the XGBoost open source projects). Figure 1 shows a brief comparison of the most common decision tree algorithms.

Figure 1 - Algorithms for decision trees



Source: Morde and Setty (2019). Boosting is an *ensemble* technique (that is, makes an average of the predictions of a group of models) that constructs models sequentially, and each subsequent model corrects the errors of the previous one, whereas *bagging* constructs models independently.

In sum, XGBoost is a bagging-based algorithm with a key difference wherein only a subset of features is selected at random. Compared to Random Forest, XGBoost is normally used to

[16]According to the authors, the quantile combination method often results in a prediction model in which the coefficients of *fully weak* predictors (those that help predict no quantile at all) are not statistically significant, in contrast to statistically significant *strong* predictors (that help forecasting all quantiles), while the coefficients of *partially weak* predictors (useful to forecast some, but not all, conditional quantiles of Y) are adjusted to reflect the magnitude of their contribution to the conditional mean forecast. These methods potentially offers improvement in forecast accuracy compared to usual conditional mean models not designed to deal with *partial* and *fully weak* predictors across quantiles and over time.

[17]For instance: (i) the distributed *weighted quantile sketch* algorithm, to find the optimal split points among weighted datasets; (ii) sparsity awareness, that admits sparse features for inputs; (iii) cross-validation at each iteration; among others.

train gradient-boosted decision trees and other gradient boosted models, whereas RF uses the same model representation and inference (as gradient-boosted decision trees), but a different training algorithm. In addition, XGBoost supports *missing values* by default, since branch directions for missing values are learned during training.

In practice, XGBoost requires the right configuration of the algorithm for a dataset by tuning the *hyper parameters* (i.e., searching the parameter space for a set of values that optimizes the model architecture). Hyper parameter tuning is not automatic and must be fine-tuned manually. Most of hyper parameters in XGBoost are about the bias-variance trade-off. When one allows the model to get more complicated (e.g., more depth), the model has better ability to fit the training data (in-sample), resulting in a less biased model. However, such complicated model requires more data to fit. The best model should trade the model complexity with its predictive power carefully.[18][19] See Chen and Guestrin (2016) for further details.

Models 14 and 15 (AF and BCAF): Duarte et al. (2019) generate optimal oil price forecasts using forecast combination tools, in the context where the number of forecasts can grow without bounds, following the approach proposed in Issler and Lima (2009); see also Gaglianone and Issler (2019).[20] The main idea is to employ a bias-correction device on the cross-section average of individual forecasts. In this setup, the *Average Forecast* (AF) is a special case of the *Bias-Corrected Average Forecast* (BCAF), in which the bias term is statistically equal to zero. Such forecast combination setup works well in practice due to risk diversification: idiosyncratic forecast errors vanish, since the law of large numbers eliminates the uncertainty associated to them, as long as the number of combined forecasts increases with no bounds.

Here, the set of covariates used to forecast the oil price is, essentially, the same used in Duarte et al. (2019). Minor changes include the substitution of some FRED series without seasonal adjustment by the respective seasonally adjusted series, and the exclusion of the series from the Goyal and Welch (2008) database, due to infrequent data update.[21] On the other hand, in order to eliminate excessively high (or low) individual forecasts of the Brent oil

[18] One of the most important hyper parameters is the *max_depth*, which controls the model complexity. In general, the deeper a tree grows, the more complex the model will become, since there will be more splits to capture information about the data. Indeed, this is one of the key causes of overfitting in decision trees because the model can fit perfectly the training data (in-sample) but will not be able to generalize well on the test set (out-of-sample). Thus, reducing *max_depth* can avoid overfitting. Another key hyper parameter is the learning rate η, which scales the contribution of each tree by a factor of $0 < \eta < 1$. It is used to prevent overfitting by making the boosting process more conservative (lower values for η).

[19] Other way to tackle overfitting in XGBoost is to add randomness to make training robust to noise. This can be done by using hyper parameters *subsample* (ratio of the training instance. Setting it to 0.5 means that XGBoost randomly collects half of the data to grow trees, thus preventing overfitting) and *colsample_bytree* (ratio of features when constructing each tree). For more details, see: https://xgboost.readthedocs.io/en/latest/index.html

[20] Technical Appendix 1 provides further details on the referred forecast combination setup.

[21] The set of covariates used in this paper is the following: CONSPI; CRB; CRB_METALS; GPR_UKRAINE; HUNPROINDMISMEI; IPBUSEQ; IPG3311A2S; IPG3364T9S; IPN213111S; IPN3311A2RS; OIL_WTI; OIL_BRENT_REAL; PPICMM; S_P_PE_ratio; TB3SMFFM. See Technical Appendix 5 for further details on the description and source of the selected series.

prices from models AR, ARMA-X and VAR (which, in turn, impact the aggregate forecasts AF and BCAF), a *trimming* strategy is used. In other words, it is removed from the set of individual forecasts (used to build the AF or BCAF combined forecasts) those individual predictions of the Brent real oil price that are above US$ 400 or below US$ -30 (i.e., assumed here as *outliers*). Such approach can be used as long as the number of models diverges ($N \rightarrow \infty$), because even with *trimming*, the number of models used to compute the average of individual forecasts grows at the same rate, provided that it is proportional to N.

Model 16 (Brent futures): Contracts of future oil price, daily traded in global financial markets, naturally contain market expectations about the future prices of oil. Here, the Brent future prices from *ICE Brent Crude Futures* are considered, with maturities ranging from 1 up to 12, 24, 36, 48, 60 and 72 months.[22] This way, each contract maturity is considered as the respective forecast horizon,[23] and the Brent real oil price forecast as the contract nominal price of the Brent future (that is, assuming a neglible inflation along the considered horizon).[24]

Models 17 and 18 (Schwartz-Smith): Schwartz and Smith (2000) proposed a two-factor commodity price model, assuming the equilibrium price level, in continuous time, evolves according to a geometric Brownian motion with drift (equivalent to a random walk with drift in discrete time). This way, short-run deviations between the spot and equilibrium prices exhibit mean-reversion, following an Ornstein-Uhlenbeck process. From an econometrics point of view, the authors propose a decomposition of the oil price into two components: trend (long run, or fundamental price) and cycle (short-run variations around the trend). Although these two factors are not directly observable, they can be estimated by using a Kalman filter approach with spot and future prices.

Intuitively, price movements of future contracts at long maturities provide information about the equilibrium price level, whereas the differences between prices of short and long horizons give information about the short-run oil price variations. The authors argue that, although this model does not explicitly consider changes in convenience yields over time, this short-term/long-term model is equivalent to the stochastic convenience yield model developed in Gibson and Schwartz (1990); see also Cortazar and Naranjo (2006) and Cortazar et al. (2015) for further developments. Here, models 17 and 18 are, respectively, the mean and median of the Brent real oil price density forecast, based on a grid of quantiles $\tau = [0,01; 0,02; ...; 0,99]$, constructed with a numerical simulation of the Schwartz-Smith two-factor model.[25]

[22] We consider the contracts traded on the last workday of each month or quarter. For further details on Brent oil futures, see: https://www.theice.com/products/219

[23] A linear interpolation of contract future prices provides the oil price forecasts for those horizons in which there are no available maturities.

[24] Such assumption is justified by the order of magnitude of the variance of the monthly log difference of the Brent oil price, $ln(Y_t) - ln(Y_{t-1})$, which is roughly 100 times bigger than the variance of the monthly log difference of the U.S. producer price index (*PPI all commodities*), considering the sample period from January 1991 to June 2020.

[25] Technical Appendix 2 provides more details on the Schwartz-Smith factor model.

Models 19 and 20 (Mean and median, all models): The forecast combination literature (e.g., Palm and Zellner, 1992; and Timmermann, 2006) suggests that combining different models and/or forecasting methods, based on different information sets, might improve the out-of-sample forecast accuracy over individual models/methods. This exercise considers the simple average and the median of all models, respectively, on models 19-20.

Models 21 and 22 (Mean and median, selected models): Here, the mean and median of a subset of models is computed, only considering one method of each class of models. Thus, the following models are chosen (*ad hoc*): (1) *random walk*; (5) factor model 1; (9) adaptive LASSO (*adalasso*); (12) quantile regression forest; (14) AF (*average forecast*); (16) Brent futures; and (18) Schwartz-Smith median.

2.2 Density Forecast

Following the literature of commodity pricing models (e.g., Schwartz and Smith, 2000), it is assumed that the logarithm of the real oil price follows a normally distributed process.[26] In other words, the real oil price Y_t is assumed to follow a *log-normal* distribution. One of the key features of the log-normal distribution is that its support lies on the positive real line $\mathbb{R}^+$, that is $Y_t \in (0, +\infty)$. This feature is crucial to guarantee non-negative oil price forecasts. Let $y_t = \ln(Y_t) \sim N(\mu, \sigma^2)$. Then, $Y_t \sim$ log-normal(μ, σ^2). The main descriptive statistics of the log-normal distribution are the following:

$$\begin{aligned} mean(Y_t) &= \exp(\mu + \frac{\sigma^2}{2}), & (15) \\ median(Y_t) &= \exp(\mu), & (16) \\ mode(Y_t) &= \exp(\mu - \sigma^2), & (17) \\ variance(Y_t) &= \exp(2\mu + \sigma^2)\left(\exp(\sigma^2) - 1\right). & (18) \end{aligned}$$

The probability density function (pdf) of Y_t and its quantiles are given as follows:

$$\begin{aligned} pdf(Y_t) &= \frac{1}{Y_t \sigma \sqrt{2\pi}} \exp\left(-\frac{(\ln(Y_t) - \mu)^2}{2\sigma^2}\right), & (19) \\ quantile(Y_t, \tau) &= \exp(\mu + \sqrt{2\sigma^2}\, \mathrm{erf}^{-1}(2\tau - 1)), & (20) \end{aligned}$$

where erf (.) is the *error function*, defined as: $\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt.$

Using the forecasting methods/models described in the previous sections, once can build direct point forecasts of the $h-$period log variation of the real oil price at period $t + h$, that is, forecasts of $\Delta^h \ln(Y_{t+h}) \equiv \ln(Y_{t+h}) - \ln(Y_t) = (y_{t+h} - y_t)$, using the information set $\mathcal{F}_t$ available at period t.

In order to produce density forecasts of Y_{t+h}, one assumes here that the conditional distribution of $\Delta^h \ln(Y_{t+h})$ is Gaussian, with conditional mean $\mu_{t+h|t}$ and conditional vari-

[26] A positive random variable Y is log-normally distributed if the logarithm of Y is normally distributed.

ance $\sigma^2_{t+h|t}$, that is $\left(\Delta^h \ln(Y_{t+h}) \mid \mathcal{F}_t\right) \sim N(\mu_{t+h|t}, \sigma^2_{t+h|t})$ or, equivalently, $(\ln(Y_{t+h}) \mid \mathcal{F}_t) \sim N(\mu_{t+h|t} + y_t, \sigma^2_{t+h|t})$, since $y_t = \ln(Y_t) \in \mathcal{F}_t$.

Therefore, the conditional distribution of the real oil price Y_{t+h} is log-normal, with mean and variance given as follows:

$$E\left(Y_{t+h} \mid \mathcal{F}_t\right) = \exp(\mu_{t+h|t} + y_t + \frac{\sigma^2_{t+h|t}}{2}), \quad (21)$$
$$Var\left(Y_{t+h} \mid \mathcal{F}_t\right) = \exp(2\left(\mu_{t+h|t} + y_t\right) + \sigma^2_{t+h|t})\left(\exp(\sigma^2_{t+h|t}) - 1\right). \quad (22)$$

Similarly, the conditional quantile of Y_{t+h}, evaluated at quantile level $\tau_i \in (0,1)$, is computed as follows:

$$Q_{\tau_i}(Y_{t+h} \mid \mathcal{F}_t) = \exp(\mu_{t+h|t} + y_t + \sqrt{2\sigma^2_{t+h|t}}\, \mathrm{erf}^{-1}(2\tau_i - 1)). \quad (23)$$

Now, let $f^m_{t+h|t}$ be the model m estimate of the conditional mean of $\Delta^h \ln(Y_{t+h})$. Thus, $f^m_{t+h|t} = \widehat{\mu_{t+h|t}}$, where $\mu_{t+h|t} = E\left(\Delta^h \ln(Y_{t+h}) \mid \mathcal{F}_t\right)$. Also, let $\widehat{\sigma^2_{t+h|t}}$ be the model m estimate of the conditional variance of $\Delta^h \ln(Y_{t+h})$, that is $\sigma^2_{t+h|t}$, computed using the Newey and West (1987)'s HAC covariance matrix estimator, from a regression of the forecast error of $f^m_{t+h|t}$ on the intercept.[27]

Provided that $[\widehat{\mu_{t+h|t}}, \widehat{\sigma^2_{t+h|t}}]'$ are consistent estimates of $[\mu_{t+h|t}, \sigma^2_{t+h|t}]'$, one can obtain consistent estimates of the conditional quantiles of Y_{t+h}, along a grid of quantile levels $\tau \in [\tau_1, ..., \tau_n]'$, using equation (23). In particular, at the median ($\tau_i = 0.5$), it follows that $\widehat{Q_{\tau_i=0.5}}(Y_{t+h} \mid \mathcal{F}_t) = \exp(f^m_{t+h|t} + y_t)$, since $\mathrm{erf}^{-1}(0) = 0$.

Finally, the multi-step ahead density forecasts of Y_{t+h} are summarized by using a *fan chart* graph, based on the estimated conditional quantiles over the horizons $h = 1, ..., H$ and the considered grid of quantile levels. In order to obtain a *smooth* term-structure of conditional variances (i.e., across the considered horizons), one can also smooth out the estimated conditional variances using a *Spline* function.

2.2.1 Density Forecast Evaluation

The density forecasts are evaluated using three approaches: (i) coverage rate, (ii) log predictive density score, and (iii) interval score, next described.

Coverage Rate: According to Clark (2011, p.336): *"...a natural starting point for forecast density evaluation is interval forecasts - that is, coverage rates."* In this sense, a necessary (but not sufficient) condition for a "good" density model is to produce a conditional density with an adequate coverage rate.[28] The objective is to verify to which extent a given density

[27] The forecast error $\left(f^m_{t+h|t} - \Delta^h \ln(Y_{t+h})\right)$ is computed here along a pseudo out-of-sample forecasting exercise, that is, considering $t = T_1, ..., T_2$ and a given h.

[28] Coverage rates reveal the difference between the unconditional probability that realizations fall into the forecasted intervals and the respective nominal coverage. However, the main drawback is that coverage rates ignore time dependence and cluster behavior.

forecast departures from a selected nominal coverage rate.

In practice, one needs to compute the frequency of observations of Y_{t+h} that fall inside a selected forecast interval. In this paper, the 90% interval band is adopted, which leads to a forecast interval based on the conditional quantiles $\widehat{Q_{\tau,m}}(Y_{t+h} \mid \mathcal{F}_{t,m})$, estimated from model m, horizon h and quantile levels $\underline{\tau} = 0.05$ and $\overline{\tau} = 0.95$. The empirical coverage is, thus, defined as follows:

$$C_{m,h} = \frac{1}{(T_2 - T_1 + 1)} \sum_{t+h=T_1}^{T_2} \mathbf{1}_{\{\widehat{Q_{\underline{\tau},m}}(Y_{t+h}|\mathcal{F}_{t,m}) \leq Y_{t+h} \leq \widehat{Q_{\overline{\tau},m}}(Y_{t+h}|\mathcal{F}_{t,m})\}}. \tag{24}$$

The lower the distance between the nominal coverage $(\overline{\tau} - \underline{\tau})$ and the empirical coverage $C_{m,h}$, the better is the density forecast. In the case of $C_{m,h} >> (\overline{\tau} - \underline{\tau})$, the forecasted density is too wide, compared to data, whereas for $C_{m,h} << (\overline{\tau} - \underline{\tau})$ the density forecast is too narrow.

Log Predictive Density Score (LPDS): Another useful indicator to analyze density forecasts is the log predictive density score, or simply *logarithmic score* (e.g., Gneiting and Raftery, 2007, eq.54). This approach allows one to rank the investigated models $m = 1, ..., M$, for each forecast horizon $h = 1, ..., H$, according to their LPDS, as follows:

$$LPDS_{m,h} = \frac{1}{(T_2 - T_1 + 1)} \sum_{t+h=T_1}^{T_2} \ln\left(\widehat{d^m_{t+h|t}(Y_{t+h})}\right) \tag{25}$$

where $\widehat{d^m_{t+h|t}(Y_{t+h})}$ is the conditional density of Y_{t+h}, estimated from model m and horizon h, based on the information set available at period t. The referred density is evaluated at the observed value Y_{t+h} and (log) averaged along the pseudo out-of-sample observations $T_1, ..., T_2$. In our case, recall that Y_t follows a log-normal distribution, with conditional density given by equation (19). A higher score implies a better model (see Adolfson et al., 2005). According to Gneiting and Raftery (2007, p.374): *"The logarithmic score is strictly proper but involves a harsh penalty for low probability events and thus is highly sensitive to extreme cases."*

Interval Score: Scoring rules for intervals provide another way of checking how well-calibrated are density forecasts in respect to observed data. Given a central prediction interval forecast $[L, U]$, with associated probability $(1 - \alpha) \times 100\%$, where L and U represent the estimated conditional quantiles from model m, horizon h, quantile levels $\underline{\tau} = \frac{\alpha}{2}$ and $\overline{\tau} = \left(1 - \frac{\alpha}{2}\right)$, respectively, and $Y_{t+h} > 0$ is a realization of the variable of interest, one can define the following interval scoring rule, proposed by Gneiting and Raftery (2007, eqs. 43, 58):

$$S_{m,h} = \frac{1}{(T_2 - T_1 + 1)} \sum_{t+h=T_1}^{T_2} \left[(U - L) + \frac{2}{\alpha}(L - Y_{t+h}) \mathbf{1}_{\{Y_{t+h} < L\}} + \frac{2}{\alpha}(Y_{t+h} - U) \mathbf{1}_{\{Y_{t+h} > U\}} \right], \tag{26}$$

where $L = \widehat{Q_{\underline{\tau},m}}(Y_{t+h} \mid \mathcal{F}_{t,m})$ and $U = \widehat{Q_{\overline{\tau},m}}(Y_{t+h} \mid \mathcal{F}_{t,m})$. This is a *proper* scoring rule for intervals (Gneiting, 2011), constructed from two quantile losses at the $[\underline{\tau}; \overline{\tau}]$ quantile levels. Since this paper considers the 90% interval band, one should set $\alpha = 0.10$, $\underline{\tau} = 0.05$ and $\overline{\tau} = 0.95$. According to Gneiting and Raftery (2007, p.374): *"This scoring rule assesses both calibration and sharpness, by rewarding narrow prediction intervals and penalizing intervals missed by the observation."* Finally, note that this rule is negatively oriented, acting as a loss function. Thus, a lower score implies a better interval forecast.

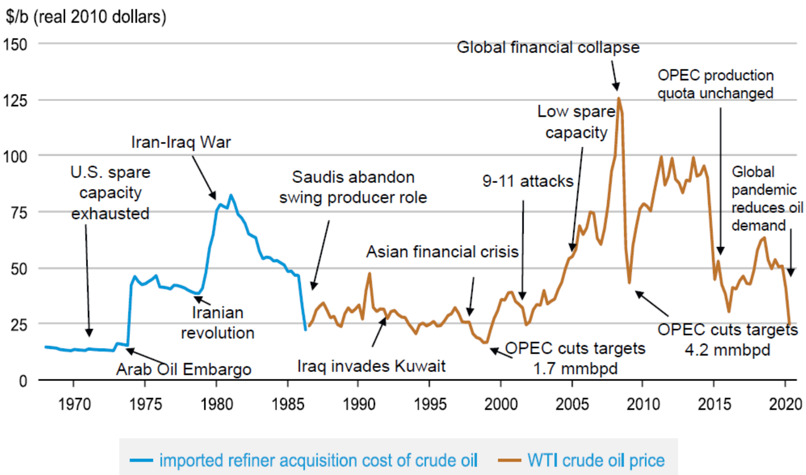
3 Empirical Exercise

3.1 Data

Although the nominal oil price receives great attention in the press, the relevant variable in terms of economic modeling is the *real price* of oil. The focus of the analysis is on the Brent oil price extracted from the International Financial Statistics (IFS) of the IMF. The nominal price data were deflated using the U.S. producer price index (PPI), obtained from the FRED database of the St. Louis FED.

Figure 2 shows that real oil prices over the past 50 years reacted to a variety of geopolitical and economic events.[29] To explain (and forecast) the real oil price dynamics, a quite diverse set of macroeconomic and financial variables drawn from a number of categories is used here. They came from a pool of $n = 315$ contemporaneous variables that are present in different databases: FRED-MD (McCracken and Ng, 2015), EPU (Economic Policy Uncertainty indexes of Baker, Bloom and Davis, 2015), GPR (Geopolitical Risk indexes of Caldara and Iacoviello, 2018) and Thomson Reuters Datastream, among others. The Technical Appendix 5 presents the full list of variables used as potential predictors for the real oil prices.

Figure 2 - Real oil prices



Source: U.S. Energy Information Administration (2020) report, available at:

https://www.eia.gov/finance/markets/crudeoil/reports_presentations/crude.pdf

[29] The real oil prices shown in Figure 2 are computed using the West Texas Intermediate (WTI) crude oil price, which is strongly correlated with the Brent oil price.

The relationship between the oil price dynamics and relevant macroeconomic variables is widely documented in the literature; see Hamilton and Herrera (2004), Kilian and Vigfusson (2013, 2017), Aastveit et al. (2015), Baumeister and Kilian (2012, 2016), Mohaddes and Pesaran (2016), Bjørnland, Larsen and Maih (2018), Bjørnland and Zhulanova (2018), among many others. The use of macro variables is motivated, for instance, by empirical evidence suggesting that measures of global real activity are useful for out-of-sample forecasting the real price of oil; see Alquist et al. (2013).[30] In this sense, the use of industrial production indexes from several countries, as well as U.S. industry-level and labor market indicators, within a high-dimensional context can be a promising route.

On the other hand, despite the fact that neither short-term interest rates nor trade-weighted exchange rates seem to have predictive power in the literature for the nominal price of oil, several financial market indicators are included in the set of predictors[31] (e.g., *Baltic Exchange Dry*[32] and indicators based on stock markets, money and credit, interest and exchange rates), relying on the usage of machine learning *nonlinear* approaches[33] to identify statistical relationships not captured by standard linear models.

Finally, several predictors not usually considered by economists are also included in the database, in order to potentially improve forecast accuracy, such as data from newspaper coverage used to build the economic policy uncertainty (EPU) and geopolitical risk (GPR) indexes; which nowadays are available freely for many countries.[34]

Our sample period covers roughly 30 years of data, ranging from January 1991 to June 2020 ($T = 354$ monthly observations). All variables are automatically tested for stationarity using the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test and first-differentiated when necessary.[35] The 315 variables are lagged one period[36] and considered in levels and first-differenced (or first- and second-differenced, in the case of $I(1)$ series), forming a final large data base containing 630 series. This way, $dim\,(\widetilde{x}_t') = 630$ variables used as potential predictors for the oil price variation in equation (1). All models are recursively estimated, considering both

[30] According to the authors, global real activity and changes in crude oil inventories can be viewed as leading indicators of the real price of oil. In addition, models based on the price index changes for non-oil industrial raw materials might capture the effect of persistent changes in the global business cycle on the (real) oil price, since shifts in the demand for industrial raw materials are also related to shifts in the demand for crude oil.

[31] See Miller and Ratti (2009).

[32] As proxy of shipping freight rates. According to Alquist et al. (2013), the idea of using fluctuations in shipping freight rates as indicators of changes in the global real activity is far from new and dates back to Isserlis (1938).

[33] Hamilton (2003) suggested a nonlinear relationship between oil prices and U.S. real GDP.

[34] The idea is to employ uncertainty proxies to capture oil shocks related to a speculative (or forward-looking) element in the real price of oil (see Kilian and Murphy, 2014).

[35] In factor models 1 and 2, all covariates are also standardized (i.e., considered with zero mean and unit variance), since such approach provided better results in terms of oil price forecast accuracy, compared to the use of covariates with their original mean and variance.

[36] Hamilton and Herrera (2004) point out that it is crucial to consider a rich lag structure in studying the dynamic relationship between the price of oil and the macro aggregates. However, previous empirical exercises (not reported) indicate that using more lags (2 or 3 lags) in our exercise generates oil price forecasts with higher RMSEs, especially at longer horizons, compared to the one-lag approach.

monthly and quarterly frequencies,[37] by using a growing window[38] (increasing sample size), as one incorporates every new time-series observation, one at a time.

In this context, each model is initially estimated using the first T_1 observations and the out-of-sample point forecasts are generated. One, then, adds an additional observation at the end of the *training set*, re-estimate the models and generate again out-of-sample forecasts. This process is repeated along the remaining data (*test set*). See Morales-Arias and Moura (2013) for a detailed discussion about recursive versus rolling window.

This paper uses data over the period from January 1991 to December 2005 $(T_1 = 180$ monthly observations) for model estimation (*training set*) and reserve the remaining data (*test set*) for the forecast comparison using $P = T - T_1 = 174$ observations, for $h = 1$. In this case, the evaluation period ranges from January 2006 to June 2020 (174 monthly forecasts). For $h = 24$ months, the evaluation period varies from December 2007 to June 2020 (151 forecasts). Thus, the first part of the sample is used to estimate the econometric models and train the machine learning approaches (selection of the tuning parameters and estimation of the β parameters), whereas the remaining observations are used for out-of-sample forecast comparison for horizons $h = 1, ..., 24$ months or $h = 1, ..., 20$ quarters.[39]

The empirical exercise is implemented using the R software (version 4.0.2, 64-bit). The ridge regression, LASSO and elastic net models are estimated using the R package *glmnet* (version 2.0-16), which fits a generalized linear model via penalized maximum likelihood. The adalasso model is implemented using the R package *HDeconometrics* (version of January 26, 2018), available at: https://github.com/gabrielrvsc/HDeconometrics. The same R package is used to compute the BIC information criterion. In turn, in order to implement the random forest and the quantile regression forest methods[40] the R package *ranger* (version 0.11.1) is employed, whereas the XGBoost approach is based on the R package *xgboost* (version 1.0.0.2).

[37] At quarterly frequency, all covariates are aggregated using the quarterly average of monthly series, excepting the Brent oil prices from future contracts, which are considered at the last workday of each quarter.

[38] We adopt such an estimation scheme due to the greater efficiency, in general, of recursive regressions compared to rolling-window estimations. However, the latter approach could be justified under a framework with the possibility of structural changes.

[39] To avoid extra (and unnecessary) complications in the implementation of the forecasting exercise, we refrain to do a real-time analysis. Thus, a note of caution regarding the interpretation of results applies, mainly due to two concerns: (i) not all useful predictors may be available to the forecaster in real time; and (ii) several predictors are subject to data revisions (e.g., the CPI data become available only with a one-month delay). See Baumeister and Kilian (2012) for real-time forecasts of the real price of oil.

[40] We used $2,000$ trees in both the random forest and the quantile regression forest. In the latter method, we adopted the grid of quantile levels: $\tau \in (0.05, 0.10, 0.15, ..., 0.95)$.

3.2 Point Forecast Results

Figure 3 presents the out-of-sample forecasts (i.e., along the *pseudo* out-of-sample forecasting exercise) of selected models, in which each color represents a given *term-structure* of forecasts, formed at a given period t, for the following periods $t+h$, for $h = 1, ..., 24$ months. Figure 4 shows the log variation of the real price of oil, considering $h = 24$ months, plotted together with the respective h-period forecasts from the 22 models/methods listed on Table 1. See the Technical Appendixes 8 and 9 for several other results from the monthly and quarterly frequencies, respectively.

Figure 3 - Pseudo out-of-sample forecasts ($h = 1, ..., 24$ months, monthly freq.)

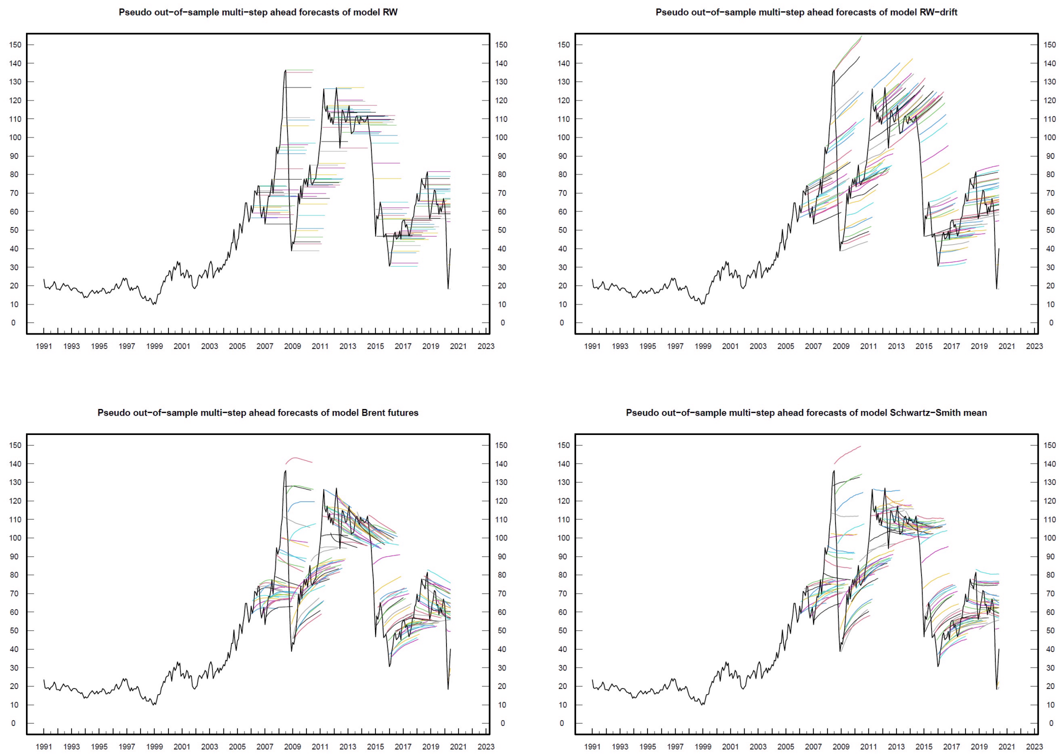
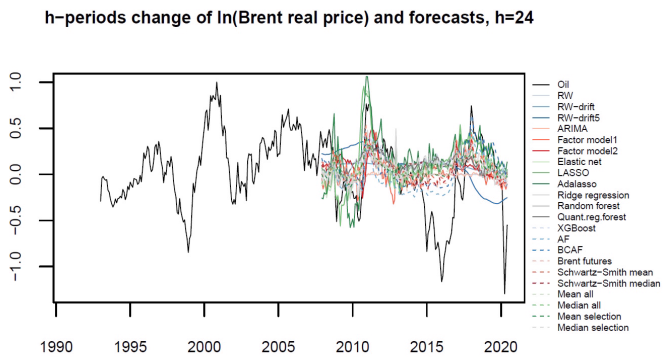


Figure 4 - Oil price variation and out-of-sample forecasts ($h = 24$ months, monthly freq.)



The individual forecast errors, for each horizon, are used to computed the Root Mean Squared Error (RMSE) from the out-of-sample evaluation period. In both model estimation and forecast evaluation, a real price of the Brent oil is computed at constant prices of the last sample observation used for model estimation (which, in turn, is time-varying along the pseudo out-of-sample forecasting exercise). The Clark and West (2007) approach[41] is used to statistically test the null hypothesis that a given forecasting method is as accurate as the random walk (*benchmark*), a usual forecast to be beaten in the oil price forecast literature, against the alternative that the competing method is more accurate than the no-change forecast.

Besides the RMSE, another way to present the results is to compute the R^2 out-of-sample statistics (or simply R^2oos), by comparing different forecast strategies with the *benchmark* model, which is an important benchmark to be beaten in the literature on oil price forecasting. For the Brent oil real price Y_{t+h}, the R^2oos-statistic is defined as follows (Rapach et al., 2010):

$$R^2oos = 100 \times \left[1 - \frac{\sum_{t=T_1+1}^{T} \left(Y_{t+h} - \widehat{f}^{i}_{t+h|t} \right)^2}{\sum_{t=T_1+1}^{T} \left(Y_{t+h} - \widehat{f}^{BMK}_{t+h|t} \right)^2} \right], \tag{27}$$

where $\widehat{f}^{i}_{t+h|t}$ is the forecast of Y_{t+h}, from method i, using information up to period t, and $\widehat{f}^{BMK}_{t+h|t}$ is the respective benchmark forecast. Positive (negative) values for the R^2oos statistic means that the forecast $\widehat{f}^{i}_{t+h|t}$ beats (is beaten by) $\widehat{f}^{BMK}_{t+h|t}$.

Table 2 presents the results of RMSE and R^2oos for the best model, in each horizon, in both frequencies; see the Technical Appendix 6 for the full results. The yellow cells reveal that the Adalasso, Elastic Net and BCAF are the best predictors with horizon up to six months, considering the exercise conducted in monthly frequency. In particular, note the good performance of the *machine learning* methods (e.g., Elastic Net, LASSO and Adalasso) in the short/medium term, providing forecasts statistically superior when compared to those from the random walk without drift in horizons from 1 to 3 months. For longer horizons, still considering the monthly frequency, the forecast combination techniques AF-BCAF gain importance, together with the Brent future prices and, to a lesser extent, the Schwartz-Smith forecasts.

In quarterly frequency, the best forecasts are those produced by the forecast combinations AF and BCAF, the Brent future prices, and the Schwartz-Smith model. Table 2 also reveals that, in both frequencies, the forecast accuracy gains in respect to the benchmark approach are statistically significant in several cases and reach two-digit figures, in percentage terms. Considering the random walk with no drift as benchmark, the R^2oos statistics for the best model, in each horizon, vary between 14% and 40% in monthly frequency, and between 9%

[41] The variances entering the test statistics use the Newey and West (1987) HAC covariance estimator.

and 49% in quarterly frequency; expressive results compared to the previous literature.[42]

Table 2 - Root Mean Squared Error (RMSE)

monthly frequency

	h = 1	h = 3	h = 6	h = 9	h = 12	h = 24
(1) RW	6.574	13.723	19.432	22.184	23.995	30.485
(2) RW-drift	6.618	13.972	20.146	23.394	25.689	35.333
(3) RW-drift5	6.696	14.474	21.423	25.240	27.716	38.260
(4) ARIMA	6.550	13.764	19.466	22.204	23.969	30.466
(5) Factor model1	6.006***	13.439**	19.119	21.540	22.800*	31.038
(6) Factor model2	5.625**	13.788	19.124	21.643	23.794	33.357
(7) Elastic net	5.192***	12.419*	18.035	22.400	26.486	40.171
(8) LASSO	5.221***	12.447*	17.929	22.455	26.535	40.037
(9) Adalasso	5.174***	12.454*	18.184	23.751	23.958	37.235
(10) Ridge regression	5.754***	12.935**	17.970	21.024	23.489	33.386
(11) Random forest	5.710***	13.258**	19.117	21.991	24.385	35.487
(12) Quant.reg.forest	5.742***	13.375**	19.078	21.948	24.278	35.373
(13) XGBoost	5.741***	13.593***	19.085**	21.899	24.394	35.174
(14) AF	9.489	13.897	17.814	19.828	20.946*	23.605*
(15) BCAF	9.376	13.723	17.665	19.800	21.060**	24.678**
(16) Brent futures	5.210***	13.402***	19.229**	21.155***	22.124***	25.167***
(17) Schwartz-Smith mean	5.258***	13.306***	19.235*	21.570**	22.861***	27.819***
(18) Schwartz-Smith median	5.232***	13.166***	18.888***	21.052***	22.151***	26.171***
(19) Mean all	5.492***	12.803**	17.914	20.238	22.023*	30.223
(20) Median all	5.404***	12.920**	18.566	20.936*	22.589*	31.512
(21) Mean selection	5.488***	12.711**	17.768	19.852*	21.165**	27.865**
(22) Median selection	5.302***	12.807**	18.484*	20.735**	22.023**	28.066***
number of observations	*174*	*172*	*169*	*166*	*163*	*151*
best model	*9*	*7*	*15*	*15*	*14*	*14*
R2 oos (%)	*38*	*18*	*17*	*20*	*23*	*40*

quarterly frequency

	h = 1	h = 4	h = 8	h = 12	h = 16	h = 20
(1) RW	11.010	22.959	29.655	32.539	36.055	38.086
(2) RW-drift	11.225	24.664	34.410	42.767	53.545	66.297
(3) RW-drift5	11.648	26.865	37.347	46.878	62.337	86.501
(4) ARIMA	11.360	23.495	31.242	35.526	39.306	42.133
(5) Factor model1	10.562**	21.723	33.511	35.046	40.552	44.298
(6) Factor model2	10.594*	25.321	37.452	42.719	56.459	75.502
(7) Elastic net	10.345***	25.299	30.989	46.638	58.948	82.425
(8) LASSO	11.734	25.703	30.915	46.553	61.238	84.899
(9) Adalasso	11.142	24.642	32.231	59.314	51.161	79.353
(10) Ridge regression	10.829**	24.789	31.916	39.382	50.646	66.150
(11) Random forest	10.278**	23.652	33.200	40.576	49.966	55.411
(12) Quant.reg.forest	10.379**	23.622	33.414	40.837	49.919	55.193
(13) XGBoost	10.383***	23.411	31.895	40.643	49.929	54.485
(14) AF	11.568	21.020*	25.437*	25.902**	28.133**	30.603**
(15) BCAF	11.392	21.746*	26.511**	26.800***	30.671***	37.408
(16) Brent futures	10.130***	21.603***	24.663**	23.615***	26.166**	30.452
(17) Schwartz-Smith mean	9.982***	22.267*	27.315**	29.743**	34.711*	39.978
(18) Schwartz-Smith median	9.924***	21.658***	25.754**	26.500***	29.834**	33.012**
(19) Mean all	10.025**	22.395*	28.989	34.079	40.875	49.150
(20) Median all	10.083***	22.767	29.201	36.080	41.934	50.491
(21) Mean selection	9.892***	21.310**	27.992**	30.643*	35.299	38.829
(22) Median selection	9.969***	21.867***	28.208**	28.557***	33.973**	35.920*
number of observations	*58*	*55*	*51*	*47*	*43*	*39*
best model	*21*	*14*	*16*	*16*	*16*	*16*
R2 oos (%)	*19*	*16*	*30*	*47*	*47*	*36*

Notes: Yellow cells indicate the Top5 best models (lower RMSEs) in each horizon. ***, **, * indicate rejection at 1%, 5% and 10% levels, respectively, using the Clark and West (2007) test. The benchmark is model 1 (random walk without drift). Forecast combinations 19 and 20 are based on models 1-18, whereas combinations 21 and 22 are based on selected models from each class (models 1, 5, 9, 12, 14, 16 and 18).

Next, the classical *bias-variance* trade-off is investigated by decomposing the MSE of each forecasting method into two parts: the forecast variance and the squared forecast bias; see Rapach et al. (2010), Elliott et al. (2013) and Lima and Meng (2017). To do so, one calculates the MSE of any forecast $\widehat{f}_{t+h|t}$ as $\frac{1}{P}\sum_{t=T_1+1}^{T}\left(Y_{t+h}-\widehat{f}_{t+h|t}\right)^2$, and the respective unconditional forecast variance as $\frac{1}{P}\sum_{t=T_1+1}^{T}\left(\widehat{f}_{t+h|t}-\frac{1}{P}\sum_{t=T_1+1}^{T}\widehat{f}_{t+h|t}\right)^2$, where P is the total number of out-of-sample forecasts. The squared forecast bias is computed as the difference between the MSE and the forecast variance.

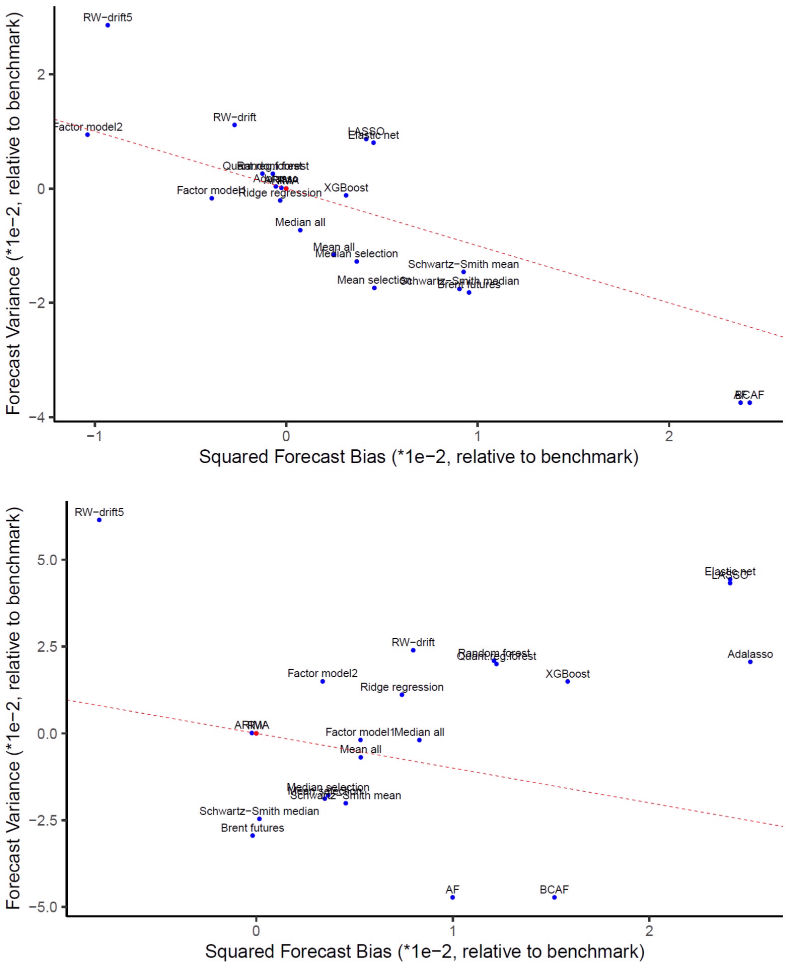
Figure 5 explores the *bias-variance* trade-off, in out-of-sample forecasting, by presenting the relative forecast variance and squared forecast bias of all forecasting methods. The relative forecast variance (squared bias) is calculated as the difference between the forecast variance (squared bias) of the i-th method and the forecast variance (squared bias) of the

[42] According to Alquist et al. (2013), the forecast of real oil price variation can be improved in horizons up to three months, but (in general) cannot be improved for horizons beyond six months. More recently, Duarte et al. (2019) report statistically significant forecast accuracy gains, in respect to the random walk, of optimal forecast combinations (based on a large database of macro and financial variables), exhibiting a R^2oos statistic that reaches 14% for $h = 6$ months.

benchmark approach. This way, the relative forecast variance (and squared bias) for the benchmark is, by construction, equal to zero. Moreover, each point on the red dotted line represents a forecast with the same MSE as the benchmark (red dot). Blue dots to the right and above the red line are forecasts outperformed by the random walk, whereas dots to the left and below it represent forecasts that outperform the benchmark.

Figure 5 - Relative MSE decomposition

($h = 12$ and 24 months, monthly freq.)



Notes: The y-axis and x-axis represent relative forecast variance and squared forecast bias, computed as the difference between the forecast variance (squared bias) of the considered method and the forecast variance (squared bias) of the benchmark (RW). Each point on the red dotted line represents a forecast with the same MSE as the RW (points to the right are forecasts outperformed by the RW and points to the left represent forecasts that outperform the RW).

Note on Figure 5 that, for $h = 12$ months, great part of forecasts beat the random walk. Such performance can be attributed to the ability of those models in substantially reducing the relative forecast variance, while keeping the forecast bias under control. In this sense, for $h = 12$ months, the following models are worth mentioning: factor model 1, AF-BCAF, Brent futures and Schwartz-Smith, besides the mean and median of all (or selected) models. In the same way, for $h = 24$ months, the best models include forecast combination devices or approaches based on Brent futures prices. The good performance of machine learning methods (such as Adalasso, Elastic Net and Random Forest) only applies to short horizons (below six months), in both frequencies; see the Technical Appendixes 8 and 9 for further details.

The previous analysis focused on the MSE decomposition enables a discussion on relative average forecast accuracy. However, such measures alone do not convey any information on how the performance of the competing methods evolves over time. To tackle this issue, Figure 6 shows the Cumulative Squared Prediction Errors (CSPE) of each forecasting method compared to the benchmark, built along the pseudo out-of-sample exercise for $h = 24$ months; see Rapach et al. (2010).

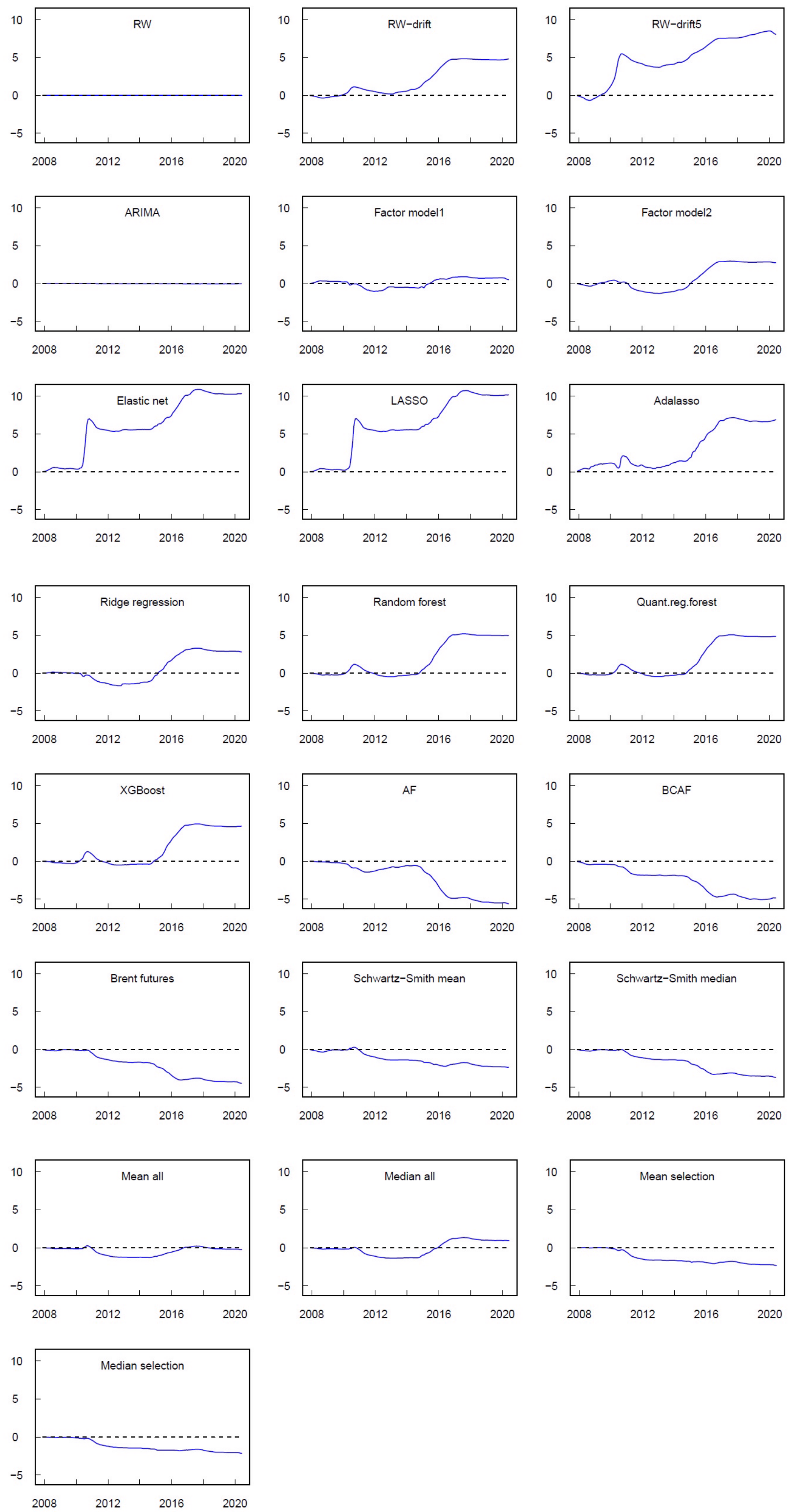
The cumulative performance-analysis depicted in Figure 6 reveals whether a given method *consistently* outperforms the benchmark forecast. For example, the relative good performance of the AF can be attributed to the consistent forecast accuracy gain over the random walk, in particular, obtained between 2014 and 2016 (when occurs a smooth decline of the blue line). On the other hand, for the Elastic Net and LASSO, there is a relevant forecast accuracy loss concentrated in a few months at the beginning of 2011 (when occurs a sharp increase of the blue line). In turn, also note that the mean and median of selected models act here as a *hedge* against high fluctuations on the relative forecast accuracy curve, exhibiting small but consistent gains from 2010 until June 2020 (i.e., smooth decline of the blue line, with no significant fluctuations). The Technical Appendixes 8 and 9 present the CSPE curves for other horizons in both frequencies.

Another interesting analysis is the identification of the most important variables chosen by the machine learning methods to predict the real oil price variation. A first way to investigate such question is to observe the evolution of the number of variables selected (or not) over time, along the pseudo out-of-sample exercise. Figure 7 reveals, among the 630 potential predictors for the Brent real price, which ones were indeed selected (and when), according to the Adalasso and Elastic Net methods, for $h = 1$ or 6 months, in monthly frequency. Note that the overall number of variables selected (blue or red dots), in general, increase with the forecast horizon.

One possible explanation is that the dependent variable (h-period variation of the log of Brent oil real price) tends to be more persistent in longer horizons. In such cases, it can be better explained by the set of covariates, compared to shorter horizons, where the dynamics of the dependent variable approach a white noise pattern.

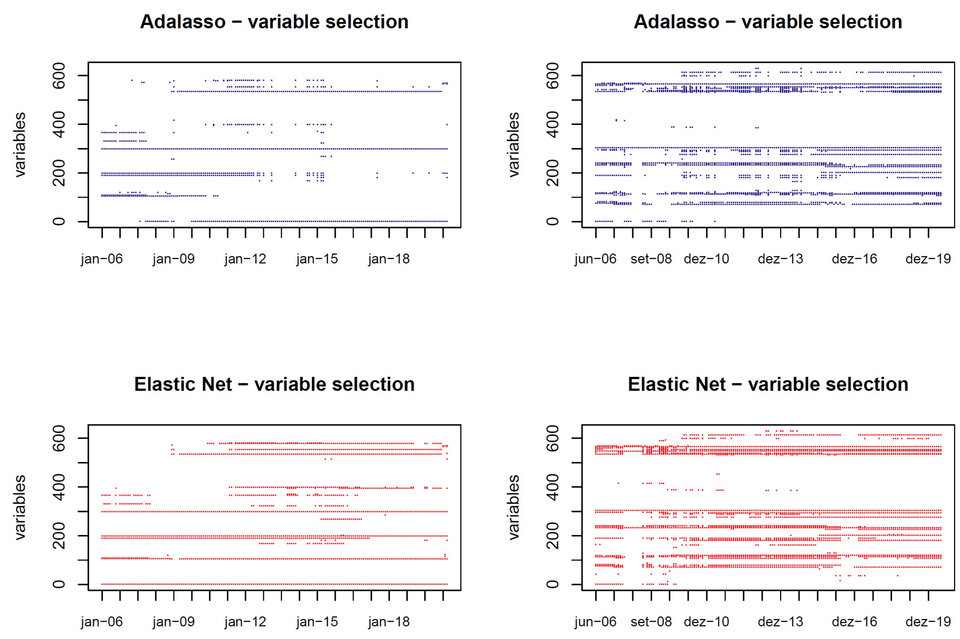
Figure 6 - Cumulative Square Prediction Error (CSPE, divided by $10,000$)

($h = 24$ months, monthly frequency)



Notes: A positively sloped curve in each panel indicates that the conditional model is outperformed by the benchmark, while the opposite holds for a downward sloping curve. Moreover, if the curve is positive (negative) at the end of the period, then the competing method has a higher (lower) MSE than the benchmark over the evaluation period.

Figure 7 - Variable Selection ($h = 1, 6$ months, monthly frequency)



Other interesting analysis that can be done using Figure 7 is checking the existence of structural breaks and the respective change in the set of variables selected as the main drivers of oil price dynamics. In particular, note that, after the global financial crisis in 2007/2008, some variables seem to have lost importance to explain the Brent price variations, whereas other variables started to be selected in a consistent way by the investigated methods. The Technical Appendixes 8 and 9 show similar plots for other horizons in both frequencies.

Our next step is to *qualitatively* investigate the variable selection. In this sense, measures of variable importance in *machine learning* methods generally attribute *scores* to predictors, reflecting the relative importance of each covariate in the overall fit of the model to data; see Hastie et al. (2009, chapter 15). Although this paper does not attempt here to economically (or structurally) interpret the driving-forces behind the machine learning forecasts, further inspecting these models to better understand how they are making forecasts (open the *black-box*) may reveal new statistical relationships in the data, previously overlooked by standard linear models.

Regarding the LASSO family of models, the degree of importance of a given variable $x_{i,t}$ when forecasting $(y_{t+h} - y_t)$ can be computed by $\left|\widehat{\beta}_i\right| * \widehat{\sigma_{x_i}}$, where $\widehat{\beta}_i$ is the estimated coefficient associated with variable $x_{i,t}$, and $\widehat{\sigma_{x_i}}$ is the sample standard deviation of $x_{i,t}$. In the case of standardized variables (zero mean and unit variance), the variable importance is simply $|\beta_i|$.[43]

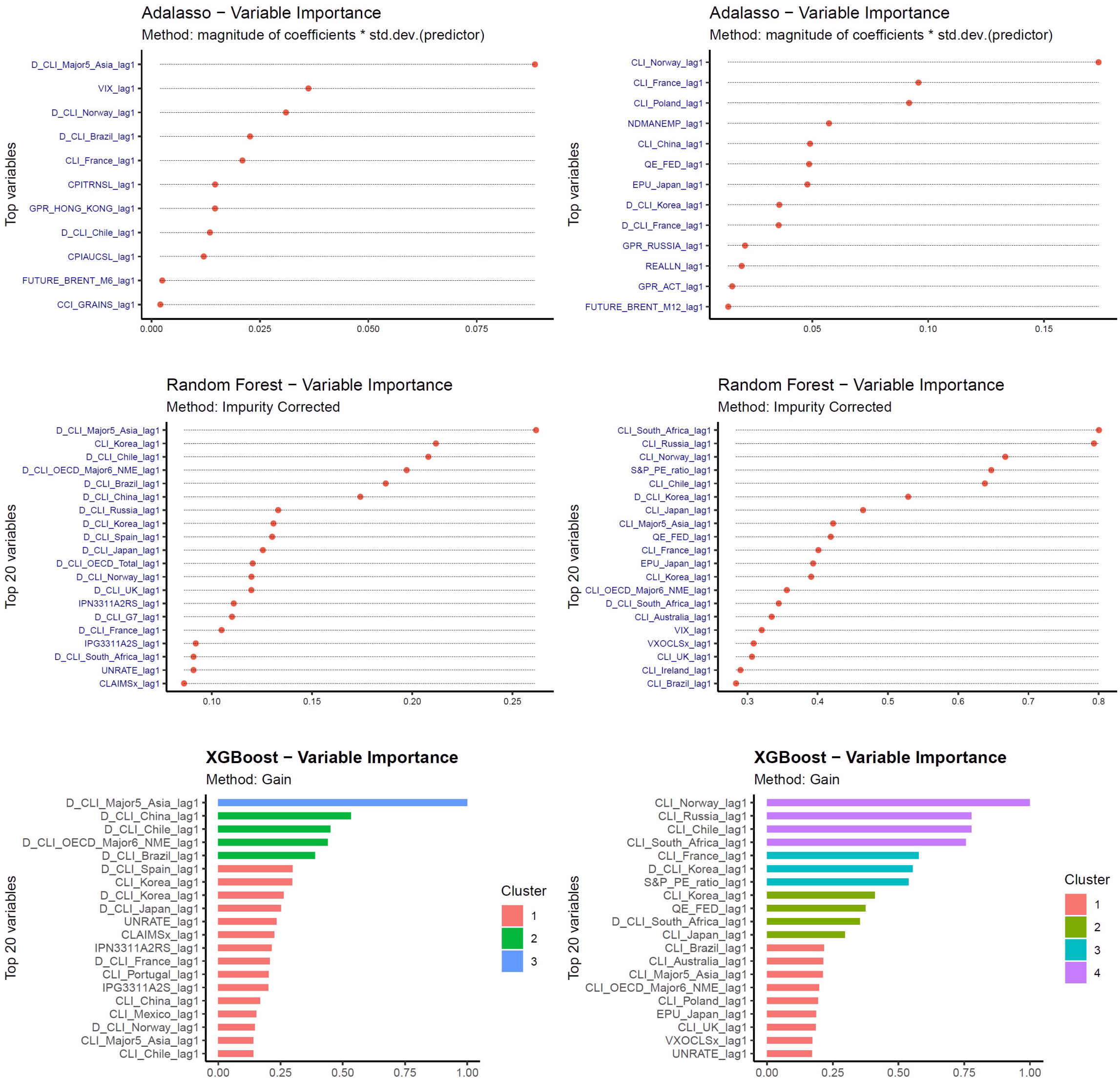
In respect to *random forest* and *quantile regression forest*, variable importance, in general, is computed by using two main methods:[44] (i) "permutation" by Altmann et al. (2010); and

[43] See https://stats.stackexchange.com/questions/14853/variable-importance-from-glmnet

[44] See also Janitza et al. (2018), that proposes for both methods a hypothesis test of no association between the investigated predictor and the dependent variable.

(ii) "impurity-corrected" by Nembrini et al. (2018); see the Technical Appendix 4 for more details on variable Importance in random forest. Figure 8 shows the most important variables in real oil price forecasting, for $h = 6$ and 24 months, based on the full sample, according to models: *adalasso*, *random forest* and *xgboost*; see the Technical Appendixes 8 and 9 for more results.

Figure 8 - Variable importance ($h = 6$ and 24 months, monthly frequency)



Note that the set of most important variables changes according to the investigated horizon. Overall, the *adalasso* is the most parsimonious method, in terms of the number of selected variables, compared to the other methods shown in Figure 8. However, despite the methodological differences, it is worth highlighting the existence of a *common* set of variables selected across the distinct methods.

For instance, considering $h = 6$ months, the most important variable according to the *adalasso*, *random forest* and *xgboost* is the same variable: *D_CLI_Major5_Asia*, that is, the first difference of the leading indicator of economic activity (called CLI), computed by the OECD, for the five biggest countries in Asia. For $h = 24$ months, again one finds a *common* set of most important variables, across the three different methods presented in Figure 8, related to leading indicators of economic activity (*CLI_ Norway* and *CLI_ France*), besides variables associated with economic uncertainty policy in developed countries (*EPU_ Japan*), or even related to the financial asset purchase program implemented by the Federal Reserve in the U.S., also known as *Quantitative Easing* (*QE_ FED*).

The following variables are also worth mentioning as relevant in oil price forecasting (despite some disagreement among the three methods presented in Figure 8): industrial production of durable goods in the U.S. (*IPN3311A2RS, IPG3311A2S*), indicators related to the labour market in the U.S. (*UNRATE, CLAIMSx*) or to the U.S. financial markets (*VIX, VXOCLSx, S&P_ PE_ ratio*). Such results are in line with previous empirical evidence, for instance, suggesting that changes in the nominal price of industrial raw materials, other than crude oil, can be used to improve forecast-accuracy of the oil price in the short run (Barsky and Kilian, 2002).

Variable importance can alternatively be presented by using *word clouds.* Figure 9 presents, for illustrative purpose, the most important variables according to the *xgboost* method, for $h = 24$ months. The variables with the largest font size are the most important ones, whereas variables with similar importance are depicted with the same size and color; see the Technical Appendixes 8 and 9 for further results.

Figure 9 - Word cloud, xgboost ($h = 24$ months, monthly frequency)

3.3 Density Forecast Results

Density forecasts provide much more information about a given variable of interest than a single point forecast such as the expected value or the conditional mean. Indeed, beyond the *location* (or central tendency of the conditional distribution), the density forecast also provides information about the *scale* of such distribution (for instance, related to the second moment of the target variable), besides informing about the existence (or not) of asymmetry, thick tails, among other empirical features of the variable of interest.

This way, density forecasts should be designed to fit the future data well, not only in terms of *location* but also in respect to *scale*. In other words, models that exhibit a poor forecast performance (e.g., in terms of RMSE) will likely produce poor density forecasts too. However, models with superior forecast accuracy (e.g., lower RMSEs) not necessarily generate good density forecasts, since an adequate forecast of the conditional quantiles of the distribution is also crucial. Tables 3, 4 and 5 show a summary of results of the density forecast evaluation, using the three metrics discussed in section 2.2.1; see the Technical Appendix 7 for the full results.

Table 3 presents the empirical coverage results. Ideally, the density forecasts should exhibit an empirical coverage as close as possible to the chosen nominal coverage of 90%. Indeed, in many cases, one finds in Table 3 several figures equal or very close to 90% (green cells). Also note there are many more figures close to the nominal coverage in short horizons than in longer ones. It is worth mentioning the good performance of the density forecasts from the random walk, Brent futures, AF (longer horizons) and the Schwartz-Smith, in monthly frequency, besides the excellent performance of the Schwartz-Smith model in quarterly frequency.

Table 3 - Empirical coverage rate

monthly frequency

	h = 1	h = 3	h = 6	h = 9	h = 12	h = 24
(1) RW	0.85	0.87	0.86	0.88	0.90	0.82
(2) RW-drift	0.85	0.88	0.89	0.89	0.90	0.79
(3) RW-drift5	0.85	0.90	0.90	0.90	0.90	0.77
(4) ARIMA	0.84	0.87	0.87	0.90	0.91	0.83
(5) Factor model1	0.90	0.86	0.84	0.84	0.76	0.73
(6) Factor model2	0.89	0.85	0.87	0.87	0.87	0.73
(7) Elastic net	0.96	0.85	0.83	0.77	0.79	0.75
(8) LASSO	0.95	0.85	0.81	0.79	0.79	0.75
(9) Adalasso	0.97	0.87	0.85	0.76	0.78	0.67
(10) Ridge regression	0.90	0.88	0.86	0.85	0.80	0.71
(11) Random forest	0.92	0.87	0.88	0.89	0.83	0.79
(12) Quant.reg.forest	0.90	0.88	0.88	0.87	0.82	0.79
(13) XGBoost	0.90	0.87	0.88	0.87	0.81	0.78
(14) AF	0.61	0.83	0.88	0.92	0.94	0.94
(15) BCAF	0.62	0.83	0.87	0.92	0.90	0.83
(16) Brent futures	0.93	0.88	0.87	0.90	0.89	0.82
(17) Schwartz-Smith mean	0.92	0.89	0.86	0.83	0.80	0.86
(18) Schwartz-Smith median	0.92	0.89	0.86	0.83	0.80	0.86
(19) Mean all	0.92	0.88	0.88	0.90	0.88	0.80
(20) Median all	0.92	0.88	0.88	0.90	0.91	0.82
(21) Mean selection	0.92	0.88	0.88	0.90	0.89	0.80
(22) Median selection	0.92	0.88	0.89	0.90	0.90	0.83

quarterly frequency

	h = 1	h = 4	h = 8	h = 12	h = 16	h = 20
(1) RW	0.73	0.85	0.76	0.76	0.88	0.90
(2) RW-drift	0.76	0.85	0.69	0.50	0.50	0.27
(3) RW-drift5	0.84	0.87	0.67	0.55	0.35	0.17
(4) ARIMA	0.57	0.85	0.83	0.84	0.79	0.77
(5) Factor model1	0.67	0.74	0.50	0.45	0.47	0.23
(6) Factor model2	0.69	0.83	0.55	0.66	0.44	0.07
(7) Elastic net	0.78	0.74	0.60	0.58	0.53	0.13
(8) LASSO	0.71	0.70	0.57	0.53	0.53	0.13
(9) Adalasso	0.67	0.54	0.55	0.58	0.50	0.17
(10) Ridge regression	0.73	0.61	0.57	0.42	0.44	0.13
(11) Random forest	0.73	0.80	0.71	0.61	0.47	0.23
(12) Quant.reg.forest	0.73	0.83	0.71	0.61	0.47	0.27
(13) XGBoost	0.69	0.74	0.69	0.58	0.44	0.27
(14) AF	0.65	0.80	0.88	0.76	0.79	0.90
(15) BCAF	0.65	0.78	0.81	0.79	0.88	0.93
(16) Brent futures	0.76	0.74	0.74	0.58	0.68	0.80
(17) Schwartz-Smith mean	0.94	0.85	0.81	0.82	0.94	0.93
(18) Schwartz-Smith median	0.94	0.85	0.81	0.82	0.94	0.93
(19) Mean all	0.67	0.85	0.76	0.68	0.50	0.27
(20) Median all	0.69	0.85	0.79	0.71	0.53	0.30
(21) Mean selection	0.76	0.87	0.79	0.66	0.53	0.37
(22) Median selection	0.76	0.85	0.79	0.61	0.59	0.53

Notes: The nominal coverage rate is 90%. The closer the empirical coverage rate is to 90% (green cells) the better is the fit of the density forecast in respect to observed data.

On the other hand, the red or orange cells indicate a poor fit of the density forecast in respect to the observed data. In several cases, such result is due to a poor point forecast (see the respective RMSEs in Table 2), providing an inadequate *location* of the forecasted density and, thus, an empirical coverage far from the nominal one. In this sense, recall that empirical coverages much below the nominal coverage might indicate (besides a bad location) that the variance of the density forecast, for instance, is low compared to the unconditional empirical distribution of the data.

Tables 4 and 5 present the results of the density forecast evaluation in terms of interval score and log predictive density score (LPDS). Overall, these results corroborate the ones previously shown in Table 3. In particular, note that the set of best density forecasts, considering the *interval score*, includes the factor model 1, ridge regression and Schwartz-Smith approach, in monthly frequency, and the Brent futures and Schwartz-Smith, in quarterly frequency; see the Technical Appendix 7 for the full results, in both frequencies.

Table 4 - Interval score

	monthly frequency						quarterly frequency					
	h = 1	h = 3	h = 6	h = 9	h = 12	h = 24	h = 1	h = 4	h = 8	h = 12	h = 16	h = 20
(1) RW	30.2	62.9	151.4	214.1	285.9	208.4	67.4	111.6	148.3	157.9	241.0	310.6
(2) RW-drift	44.9	70.3	125.6	178.4	248.5	346.4	64.0	127.9	208.1	288.6	373.9	496.8
(3) RW-drift5	36.4	65.8	136.9	380.2	541.7	382.0	74.9	145.2	472.4	400.5	600.7	972.8
(4) ARIMA	31.1	65.7	150.1	213.9	286.0	216.1	61.0	106.0	168.6	248.3	346.6	414.0
(5) Factor model1	51.7	75.9	92.1	129.4	135.5	275.4	63.0	104.9	230.2	229.3	269.9	342.5
(6) Factor model2	34.2	89.6	181.1	228.5	257.4	186.7	63.7	119.6	233.6	236.3	375.8	1106.3
(7) Elastic net	58.6	95.8	150.4	220.0	276.8	449.6	56.2	122.5	172.5	331.9	416.0	876.5
(8) LASSO	60.4	86.7	116.7	171.8	186.3	434.8	67.8	127.3	172.9	283.5	422.8	855.4
(9) Adalasso	49.3	83.9	109.2	193.9	147.7	330.8	68.5	162.8	208.3	461.1	365.3	837.1
(10) Ridge regression	27.7	59.8	89.9	122.2	155.5	278.3	68.2	144.0	266.0	392.7	516.8	845.8
(11) Random forest	41.0	88.2	156.7	203.0	249.4	282.3	53.4	115.2	194.0	276.4	441.8	470.2
(12) Quant.reg.forest	33.3	171.3	337.5	406.1	446.6	244.7	52.6	114.9	197.3	280.5	482.7	466.3
(13) XGBoost	34.2	81.0	136.3	185.6	222.9	229.9	55.2	112.9	191.8	270.1	391.0	465.0
(14) AF	57.3	76.9	131.3	172.0	201.0	102.9	66.9	93.6	176.2	296.3	398.5	397.0
(15) BCAF	55.9	76.6	132.8	175.8	207.4	120.6	65.7	112.9	192.0	289.9	394.8	406.8
(16) Brent futures	33.7	75.8	174.6	250.7	319.2	186.7	57.0	122.6	147.7	146.8	111.8	132.7
(17) Schwartz-Smith mean	28.8	67.0	95.4	106.6	110.9	122.4	49.5	106.2	125.0	128.0	146.3	170.9
(18) Schwartz-Smith median	28.8	67.0	95.4	106.6	110.9	122.4	49.5	106.2	125.0	128.0	146.3	170.9
(19) Mean all	87.6	145.2	203.7	229.6	222.2	239.0	53.4	106.9	139.9	172.5	219.5	350.7
(20) Median all	29.4	62.5	128.9	175.4	226.4	247.4	54.4	117.9	193.7	264.6	297.5	322.9
(21) Mean selection	28.2	77.0	242.7	273.9	290.1	195.8	52.7	100.4	136.7	150.6	170.4	208.0
(22) Median selection	23.0	63.0	135.2	186.8	235.9	157.5	53.8	108.0	149.4	142.8	157.0	158.7

Note: A lower score implies a better interval forecast. Yellow cells indicate the Top5 best models in each horizon.

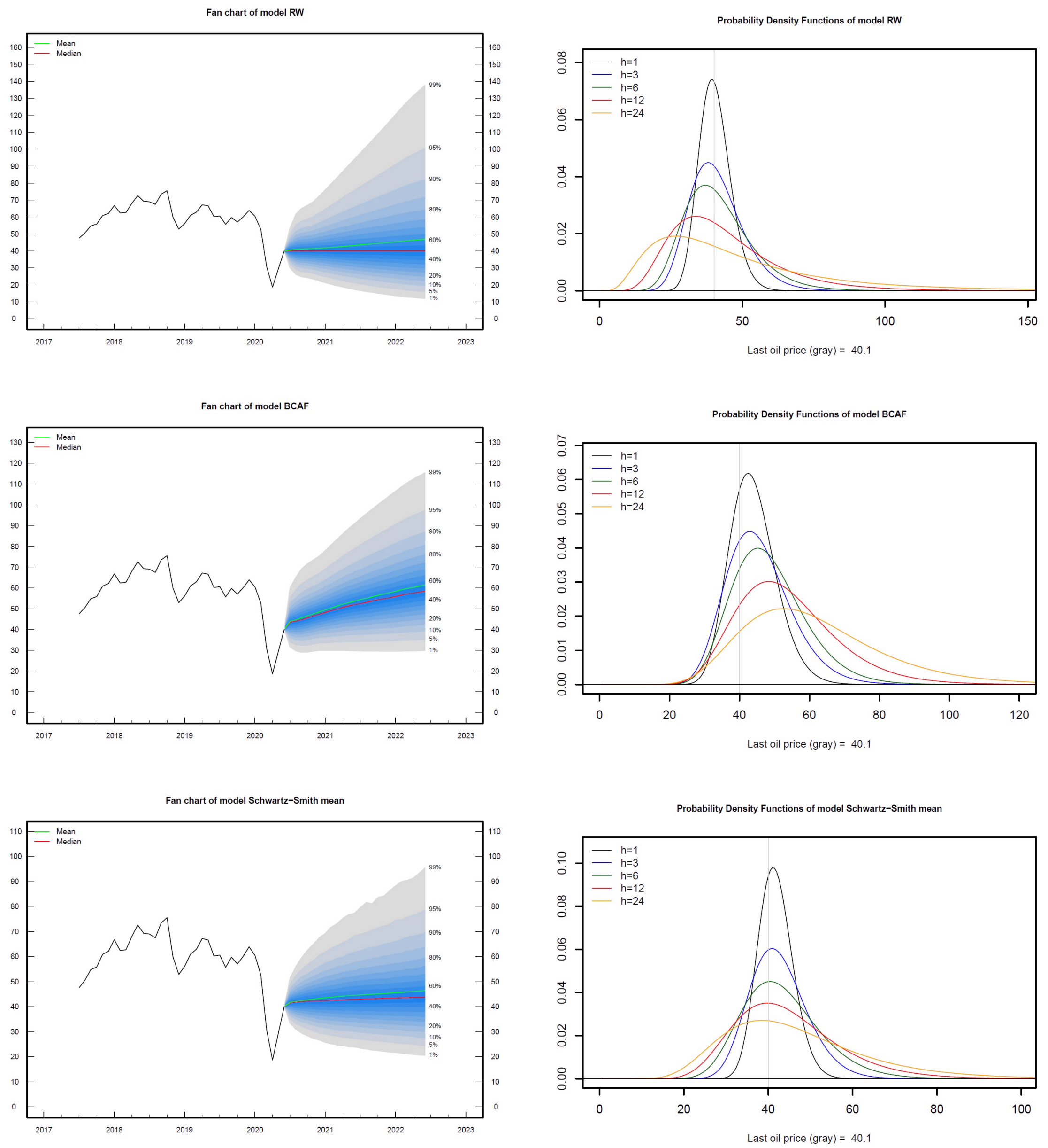
Table 5 - Log predictive density score (LPDS)

	monthly frequency						quarterly frequency					
	h = 1	h = 3	h = 6	h = 9	h = 12	h = 24	h = 1	h = 4	h = 8	h = 12	h = 16	h = 20
(1) RW	-3.42	-4.21	-4.57	-4.56	-4.71	-4.83	-4.14	-4.87	-5.09	-4.87	-4.95	-4.94
(2) RW-drift	-3.46	-4.23	-4.59	-4.58	-4.76	-5.08	-4.20	-4.97	-5.67	-5.98	-6.17	-6.86
(3) RW-drift5	-3.47	-4.21	-4.43	-4.61	-4.79	-5.31	-4.00	-5.17	-6.81	-8.75	-10.88	-13.34
(4) ARIMA	-3.41	-4.22	-4.56	-4.56	-4.73	-4.86	-4.32	-4.80	-4.91	-4.87	-5.08	-5.43
(5) Factor model1	-3.30	-4.19	-4.54	-4.65	-4.98	-5.77	-4.58	-4.65	-5.75	-5.62	-6.38	-7.51
(6) Factor model2	-3.18	-4.28	-4.45	-4.55	-4.75	-5.12	-4.45	-4.89	-6.26	-5.61	-8.15	-43.21
(7) Elastic net	-3.15	-4.13	-4.79	-5.50	-5.22	-5.62	-4.58	-5.10	-5.29	-6.11	-6.69	-14.15
(8) LASSO	-3.16	-4.14	-4.78	-5.43	-5.18	-5.73	-4.75	-5.19	-5.29	-5.81	-6.39	-12.80
(9) Adalasso	-3.21	-4.15	-4.75	-5.68	-5.53	-7.01	-4.83	-5.86	-5.58	-9.51	-6.51	-14.85
(10) Ridge regression	-3.20	-4.17	-4.53	-4.72	-5.05	-6.64	-4.59	-5.94	-9.68	-11.87	-14.63	-22.53
(11) Random forest	-3.23	-4.20	-4.42	-4.51	-4.73	-4.98	-4.03	-4.84	-5.62	-6.04	-6.86	-7.48
(12) Quant.reg.forest	-3.22	-4.21	-4.42	-4.52	-4.74	-4.98	-3.99	-4.84	-5.64	-6.09	-6.86	-7.50
(13) XGBoost	-3.21	-4.21	-4.40	-4.52	-4.79	-5.01	-4.11	-4.95	-5.68	-5.95	-6.91	-7.68
(14) AF	-4.24	-4.29	-4.45	-4.44	-4.55	-4.50	-4.64	-4.65	-4.73	-4.70	-4.79	-4.82
(15) BCAF	-4.21	-4.29	-4.46	-4.46	-4.57	-4.56	-4.64	-4.90	-4.82	-4.62	-4.69	-4.92
(16) Brent futures	-3.03	-4.10	-4.38	-4.50	-4.61	-4.66	-4.06	-5.11	-5.94	-5.03	-4.77	-4.81
(17) Schwartz-Smith mean	-3.06	-4.11	-4.39	-4.49	-4.66	-4.79	-4.10	-5.30	-6.95	-5.96	-5.80	-5.93
(18) Schwartz-Smith median	-3.05	-4.09	-4.36	-4.47	-4.63	-4.70	-4.07	-5.06	-6.42	-5.42	-5.28	-5.17
(19) Mean all	-3.17	-4.15	-4.38	-4.44	-4.58	-4.79	-4.05	-4.75	-5.00	-4.94	-5.33	-6.43
(20) Median all	-3.14	-4.13	-4.45	-4.48	-4.64	-4.79	-4.10	-4.83	-4.92	-4.90	-5.21	-6.21
(21) Mean selection	-3.13	-4.13	-4.36	-4.40	-4.53	-4.66	-4.00	-4.63	-5.00	-4.89	-5.14	-5.61
(22) Median selection	-3.08	-4.10	-4.39	-4.47	-4.62	-4.69	-4.04	-4.77	-5.18	-4.87	-5.06	-5.20

Note: A higher score implies a better density forecast. Yellow cells indicate the Top5 best models in each horizon.

In turn, Figure 10 shows the *truly* out-of-sample density forecasts of the real oil prices (at constant prices of June 2020) built at monthly frequency. To do so, the so-called *fan charts* are built to illustrate the evolution of oil price point forecasts, along the term-structure of horizons, plotted together with the uncertainty (blue shades) associated with each forecasting method and considered horizon. Figure 10 also presents the *probability density functions* (PDFs) for selected horizons. Note the asymmetry of the estimated densities, consequence of the log-normality assumption discussed in section 2.2.

Figure 10 - Fan charts and probability density functions (PDFs)



Finally, *point* forecasts selected[45] at monthly frequency indicate the Brent oil real price, for June 2022, ranging from US$ 40 to 58. The same models, estimated at quarterly frequency, predict the price of oil for June 2025 to be between US$ 30 and 50.

In respect to risk management of oil prices, the *density* forecasts provide a 90% probability interval forecast of the Brent oil real price. For example, according to the AF monthly model, with 90% of probability, the oil price for June 2022 will be in the range of US$ 28 and 90 (and for June 2025, between US$ 8 and 126, according to the quarterly frequency estimation). The same interval forecast from the Schwartz-Smith model, for June 2022, ranges from US$ 24 to 79 (and for June 2025, from US$ 19 to 101). Such intervals can be useful, for instance, when hedging against extreme oil price fluctuations.

[45] Based on the RMSEs shown in the Technical Appendix 6, we select here the following models: random walk without drift, random forest, quantile regression forest, xgboost, AF, BCAF, Brent futures and Schwartz-Smith (mean and median).

4 Conclusions

The price of oil is considered a key variable in macroeconomic because its dynamics affect the global economy. The linkages between the oil price fluctuations and several macroeconomic aggregates have been extensively investigated in the literature. Nonetheless, the current context of *big data*, coupled with novel machine learning tools, allows one to further investigate potential oil price nonlinearities, so far hidden or not considered by traditional statistical models.

In this sense, this paper studies the forecast accuracy of 22 competing methods, which are used to build point forecasts of the Brent oil price variation. The selected suite of methods includes recent machine learning techniques based on regression trees, more traditional *machine learning* approaches using regularization procedures, standard econometric models and forecast combinations, besides the structural factor model of Schwartz and Smith (2000). In order to evaluate the predictive power of each method, an extensive pseudo out-of-sample forecasting exercise is conducted, in both monthly and quarterly frequencies, where each method produces point and density forecasts for horizons from one month up to five years.

According to Alquist et al. (2013) the no-change forecast of the real price of oil can typically be improved upon horizons up to three months, but generally not at horizons beyond half a year. This paper provides evidence that reduced-form models based on machine learning algorithms can indeed reduce the out-of-sample MSE in the short run compared to the no-change forecast. Our main findings are the following:

(i) in respect to point forecasts, the Adaptive LASSO model presents the lowest RMSE in our shortest (one-month) horizon. In this sense, the *machine learning* methods (e.g., Adalasso, Elastic Net), together with the BCAF forecasts, exhibit a good performance in the short run, providing forecasts statistically superior to the ones from a random walk without drift, in horizons from one to three months;

(ii) the forecasting exercise in monthly frequency also revealed that other models from the LASSO family, the Brent future prices and the median of the Schwartz-Smith density forecast provide the best forecasts in horizons up to six months. For longer horizons, considering the same frequency, the forecast combination techniques AF and BCAF, and the mean (or median) of models gain importance, together with the Brent future prices and the forecasts from the Schwartz-Smith model. In quarterly frequency, the best forecasts come from the approaches AF and BCAF, Brent future prices and Schwartz-Smith;

(iii) in both frequencies, and several cases, the forecast accuracy gains in respect to the benchmark model (random walk without drift) are statistically significant and reach two-digit figures, in percentage terms: the R^2 out-of-sample statistics, for the best model in each horizon, range from 14% and 40% in monthly frequency, and between 9% and 49% in quarterly frequency, which represents an improvement in respect to the previous literature;

(iv) regarding density forecasts, it is worth mentioning the relatively good performance of the density forecasts, using our proposed approach, built from the random walk, Brent future prices, forecast combination AF (longer horizons) and the Schwartz-Smith model.

In sum, the forecasting methods applied here to solve an important economic forecasting problem (including some fresh machine learning nonlinear algorithms as well as traditional econometric approaches) can be useful to help improving the set of tools currently used by academics and market agents to build oil price forecasts, thereby offering a valuable contribution to the field of macroeconomic forecasting.

References

[1] Aastveit, K.A., H.C. Bjørnland, and L.A. Thorsud, 2015, What drives Oil Prices? Emerging Versus Developing Economies? *Journal of Applied Econometrics* 30(1), 1013-1028.

[2] Adolfson, M., Linde, J., Villani, M., 2005. Forecasting Performance of an Open Economy Dynamic Stochastic General Equilibrium Model. Sveriges Riksbank Working Paper n.190.

[3] Alquist, R., Kilian, L., Vigfusson, R.J., 2013. Forecasting the Price of Oil. Handbook of Economic Forecasting, Vol. 2A, Chapter 8. Elsevier.

[4] Altmann, A., Tolosi, L., Sander, O., Lengauer, T., 2010. Permutation importance: a corrected feature importance measure. *Bioinformatics* 26, 1340-1347.

[5] Apostol, T.M., 1967. *Calculus, Vol. 1: One-Variable Calculus with an Introduction to Linear Algebra* (2nd Ed.), New York: John Wiley & Sons.

[6] Araujo, G.S., Gaglianone, W.P., 2020. Machine learning methods for inflation forecasting in Brazil: New contenders versus classical models. Mimeo.

[7] Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. *Econometrica* 70, 191-221.

[8] Bai, J., Ng, S., 2008. Forecasting economic time series using targeted predictors. *Journal of Econometrics* 146, 304-317.

[9] Baker, S.R., Bloom, N., Davis, S.J., 2015. Measuring Economic Policy Uncertainty. NBER Working Paper 21633, National Bureau of Economic Research.

[10] Bańbura, M., Giannone, D., Modugno, M., Reichlin, L., 2013. Now-casting and the real-time data flow. Working Paper Series n.1564, European Central Bank.

[11] Barsky, R.B., Kilian,L., 2002. Do we really know that oil caused the great stagflation? A monetary alternative. In: Bernanke, B.S., Rogoff, K. (Eds.), NBER Macroeconomics Annual 2001. MIT Press, Cambridge, 137-183.

[12] Batchelor, R., 2007. Bias in macroeconomic forecasts. *International Journal of Forecasting* 23(2), 189-203.

[13] Baumeister, C., Kilian, L., 2012. Real-time forecasts of the real price of oil. *Journal of Business and Economic Statistics* 30, 326-336.

[14] Baumeister, C., Kilian, L., 2015. Forecasting the Real Price of Oil in a Changing World: A Forecast Combination Approach. *Journal of Business and Economic Statistics* 33(3), 338-351.

[15] Baumeister, C., Kilian, L., 2016. Forty Years of Oil Price Fluctuations: Why the Price of Oil May Still Surprise Us. *Journal of Economic Perspectives* 30(1), 139-160.

[16] Bjørnland, H.C., Larsen, V.H., Maih, J., 2018. Oil and Macroeconomic (In)stability. *American Economic Journal: Macroeconomics* 10(4), 128-151.

[17] Bjørnland, H.C., Zhulanova, J., 2018. The Shale Oil Boom and the U.S. Economy: Spillovers and Time-Varying Effects. CAMP Working paper 8/2018.

[18] Breiman, L., 2001. Random forests. *Machine Learning* 45, 5-32.

[19] Caldara, D., Iacoviello, M., 2018. Measuring Geopolitical Risk. FRB International Finance Discussion Paper n. 1222, Board of Governors - Fed.

[20] Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. The 22nd SIGKDD Conference on Knowledge Discovery and Data Mining. Mimeo.

[21] Cheng, K., Huang, N., Shi, Z., 2019. Survey-Based Forecasting: To Average or Not To Average. Mimeo.

[22] Clark, T.E., 2011. Real-Time Density Forecasts from Bayesian Vector Autoregressions with Stochastic Volatility. *Journal of Business and Economic Statistics* 29, 327-341.

[23] Clark, T.E., West, K.D., 2007. Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. *Journal of Econometrics* 138, 291-311.

[24] Cologni, A., Manera, M., 2008. Oil prices, inflation and interest rates in a structural cointegrated VAR model for the G-7 countries. *Energy Economics* 38, 856-888.

[25] Cortazar, G., Naranjo, L., 2006. An N-factor Gaussian model of oil futures prices. *Journal of Futures Markets: Futures, Options, and Other Derivative Products* 26(3), 243-268.

[26] Cortazar, G., Kovacevic, I., Schartz, E., 2015. Expected commodity returns and pricing models. *Energy Economics* 49, 60-71.

[27] Duarte, A.M., Gaglianone, W.P., Guillén, O.T.C., Issler, J.V., 2019. Commodity Prices and Global Economic Activity: A Derived-Demand Approach. Mimeo.

[28] Elliott, G., Gargano, A., Timmermann, A., 2013. Complete subset regressions. *Journal of Econometrics* 177(2), 357-373.

[29] Elliott, G., Gargano, A., Timmermann, A., 2015. Complete subset regressions with large-dimensional sets of predictors. *Journal of Economic Dynamics and Control* 54, 86-110.

[30] Forni, M., Hallim, M., Lippi, M., Reichlin, L., 2000. The generalized dynamic factor model: Identification and estimation. *Review of Economics and Statistics* 82, 540-554.

[31] Gaglianone, W.P., Issler, J.V., 2019. Microfounded Forecasting. Ensaios Econômicos EPGE n.813, Getulio Vargas Foundation.

[32] Garcia, M.G.P., Medeiros, M.C., Vasconcelos, G.F.R., 2017. Real-time inflation forecasting with high-dimensional models: The case of Brazil. *International Journal of Forecasting* 33, 679-693.

[33] Gargano, A., Timmermann, A., 2014. Forecasting commodity price indexes using macroeconomic and financial predictors *International Journal of Forecasting* 30, 825-843.

[34] Gibson, R., Schwartz, E.S., 1990. Stochastic convenience yield and the pricing of oil contingent claims. *Journal of Finance* 45, 959-976.

[35] Gneiting, T., 2011. Making and Evaluating Point Forecasts. *Journal of the American Statistical Association* 106(494), 746-762.

[36] Gneiting, T., Raftery, A.E., 2007. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association* 102(477), 359-378.

[37] Gogolin, F., Kearney, F., Lucey, B.M., Peat, M., Vigne, S.A., 2018. Uncovering long term relationships between oil prices and the economy: A time-varying cointegration analysis. *Energy Economics* 76, 584-593.

[38] Goyal, A.,Welch, I., 2008. A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies* 21(4), 1455-1508.

[39] Granger, C.W.J., Ramanathan, R., 1984. Improved methods of combining forecasting. *Journal of Forecasting* 3, 197-204.

[40] Hall, A.S., 2018. Machine Learning Approaches to Macroeconomic Forecasting. *Federal Reserve Bank of Kansas City Economic Review*, 4th quarter of 2018, 63-81.

[41] Hamilton, J.D., 2003. What is an oil shock? *Journal of Econometrics* 113, 363-398.

[42] Hamilton, J.D., Herrera, A., 2004. Oil Shocks and Aggregate Macroeconomic Behavior: The Role of Monetary Policy. *Journal of Money, Credit, and Banking* 36(2), 265-286.

[43] Hansen, B.E., 2019. Econometrics. Current manuscript, https://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf

[44] Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd edition. Springer-Verlag, New York.

[45] Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12(1), 55-67.

[46] Hong, H., Yogo, M., 2012. What does futures market interest tell us about the macroeconomy and asset prices? *Journal of Financial Economics* 105, 473-490.

[47] Isserlis, L., 1938. Tramp shipping cargoes and freights. *Journal of the Royal Statistical Society* 101 (1), 53-134.

[48] Issler, J.V., Lima, L.R., 2009. A Panel Data Approach to Economic Forecasting: The Bias-corrected Average Forecast. *Journal of Econometrics* 152 (2), 153-164.

[49] Janitza, S., Celik, E., Boulesteix, A.-L., 2018. A computationally fast variable importance test for random forests for high-dimensional data. *Advances in Data Analysis and Classification* 12(4), 885-915.

[50] Judge, G.G., Hill, R.C., Griffiths, W.E., Lütkepohl, H., Lee, T.-C., 1988. *Introduction to the Theory and Practice of Econometrics*. New York, Wiley.

[51] Jung, J.K., Patnam, M., Ter-Martirosyan, A., 2018. An Algorithmic Crystal Ball: Forecasts-based on Machine Learning. IMF Working Paper WP/18/230.

[52] Kilian, L., Murphy, D., 2014. The Role of Inventories and Speculative Trading in the Global Market for Crude Oil. *Journal of Applied Econometrics* 29 (3), 454-478.

[53] Kilian, L., Vigfusson, R.J., 2013. Do Oil Prices Help Forecast U.S. Real GDP? The Role of Nonlinearities and Asymmetries. J*ournal of Business and Economic Statistics* 31(1), 78-93.

[54] Kilian, L., Vigfusson, R.J., 2017. The Role of Oil Price Shocks in Causing U.S. Recessions. *Journal of Money, Credit, and Banking* 49(8), 1747-1776.

[55] Koenker, R., 2005. *Quantile Regression*. Cambridge University Press.

[56] Laster, D., Bennett, P., Geoum, I., 1999. Rational bias in macroeconomic forecasts. *The Quarterly Journal of Economics* 114(1), 293-318.

[57] Lima, L.R., Meng, F., 2017. Out-of-sample return predictability: a quantile combination approach. *Journal of Applied Econometrics* 32(4), 877-895.

[58] Marcellino, M., Stock, J., Watson, M., 2006. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics* 135, 499-526.

[59] McCracken, M.W., Ng, S., 2015. FRED-MD: A Monthly Database for Macroeconomic Research. Working Paper 2015-012B, Federal Reserve Bank of St. Louis.

[60] Medeiros, M., Mendes, E., 2016. L1-regularization of high-dimensional time-series models with flexible innovations. *Journal of Econometrics* 191, 255-271.

[61] Medeiros, M., Vasconcelos, G.F.R., de Freitas, E.H., 2016. Forecasting Brazilian Inflation with High Dimensional Models. *Brazilian Review of Econometrics* 36(2), 223-254.

[62] Meinshausen, N., 2006. Quantile Regression Forests. *Journal of Machine Learning Research* 7, 983-999.

[63] Miller, J.I., Ni, S., 2011. Long-Term Oil Price Forecasts: A New Perspective on Oil and the Macroeconomy. *Macroeconomic Dynamics* 15(S3), 396-415.

[64] Miller, J.I., Ratti, R., 2009. Crude Oil and Stock Markets: Stability, Instability, and Bubbles. *Energy Economics* 31(4), 559-568.

[65] Mohaddes, K., Pesaran, M.H., 2016. Oil Prices and the Global Economy: Is It Different This Time Around? IMF Working Paper WP/16/210.

[66] Morales-Arias, L., Moura, G.V., 2013. Adaptive forecasting of exchange rates with panel data. *International Journal of Forecasting* 29, 493-509.

[67] Morde, V., Setty, V.A., 2019. XGBoost Algorithm: Long May She Reign! Mimeo.

[68] Neal, B., Mittal, S., Baratin, A., Tantia, V., Scicluna, M., Lacoste-Julien, S., Mitliagkas, I., 2018. A Modern Take on the Bias-Variance Tradeoff in Neural Networks. Mimeo, available at https://arxiv.org/abs/1810.08591

[69] Nembrini, S., Koenig, I.R., Wright, M.N., 2018. The revival of the Gini Importance? *Bioinformatics* 34(21), 3711-3718.

[70] Phillips, P.C.B., Moon, H.R., 1999. Linear regression limit theory for nonstationary panel data. *Econometrica* 67 (5), 1057-1111.

[71] Rapach, D. E., Strauss, J. K., & Zhou, G., 2010. Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *The Review of Financial Studies* 23(2), 821-862.

[72] Ravazzolo F., Rothman, P., 2012. Oil and U.S. GDP: A Real-Time Out-of-Sample Examination. *Journal of Money, Credit and Banking* 45(2-3), 449-463.

[73] Schwartz, E., Smith, J.E., 2000. Short-Term Variations and Long-Term Dynamics in Commodity Prices. *Management Science* 46 (7), 893-911.

[74] Stock, J., Watson, M., 2002. Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association* 97(460), 1167-1179.

[75] Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society* 58(1), 267-288.

[76] U.S. Energy Information Administration, 2020. What drives crude oil prices? Mimeo.

[77] Varian, H.R., 2014. Big data: New tricks for econometrics. *Journal of Economic Perspectives* 28(2), 3-28.

[78] Yu, L., Zhao, Y., Tang, L., Yang, Z., 2019. Online big data-driven oil consumption forecasting with Google trends. *International Journal of Forecasting* 35, 213-223.

[79] Zagaglia, P., 2010. Macroeconomic factors and oil futures prices: a data-rich model. *Energy Economics* 32, 409-417.

[80] Zou, H., 2006. The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association* 101 (476), 1418-1429.

[81] Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society* 67(2), 301-320.

[82] Zou, H., Hastie, T., Tibshirani, R., 2007. On the degrees of freedom of the lasso. *The Annals of Statistics* 35, 2173-2192.

– Technical Appendix –

"Machine Learning and Oil Price Point and Density Forecasting"

Contents:

Appendix 1. Forecast combination and bias-correction

In this section, we discuss econometric techniques used to optimally forecast the oil prices under a quadratic *risk function*. These tools are appropriate to forecast a weakly stationary and ergodic univariate process $\{y_t\}$ from a vast number of individual forecasts that are combined to generate an optimal forecast. Such individual forecasts are the outcome of different econometric models that must be estimated before the forecast combination. We label the forecasts of y_t (h-period change of the logarithm of the real price of oil), computed using the information set lagged h periods, $f_{i,t}^h$, $i = 1, 2, \ldots, N$. This way, $f_{i,t}^h$ are h periods ahead forecasts and N is the number of estimated models used to predict y_t.

Issler and Lima (2009) consider three consecutive distinct time sub-periods. The first sub-period is labeled the "estimation sample", where models are usually fitted to forecast y_t subsequently. The next sub-period is labeled the post-model-estimation or "training sample", where realizations of y_t are usually confronted with forecasts produced in the estimation sample, and weights and bias-correction terms are estimated.[1] The final sub-period is where genuine out-of-sample forecast is entertained. In this setup, the individual forecasts $f_{i,t}^h$ are considered approximations of the optimal forecast ($\mathbb{E}_{t-h}(y_t)$), as follows:

$$f_{i,t}^h = \mathbb{E}_{t-h}(y_t) + k_i^h + \varepsilon_{i,t}^h, \tag{1}$$

where k_i^h is the time invariant bias and $\varepsilon_{i,t}^h$ is the error term of model i, such that $\mathbb{E}(\varepsilon_{i,t}^h) = 0$ for all i, t, and h. Here, the optimal forecast is the *common feature* of all individual forecasts and k_i^h and $\varepsilon_{i,t}^h$ arise due to model misspecification. The term k_i^h is assumed to be identically distributed (but not independently), i.e., $k_i^h \sim i.d.\left(B^h, \sigma_{k^h}^2\right)$.

Issler and Lima proposed the feasible Bias-Corrected Average Forecast (BCAF) $\frac{1}{N}\sum_{i=1}^N f_{i,t}^h - \widehat{B^h}$, where $\widehat{B^h}$ is a consistent estimate of B^h, that obeys:

$$\underset{(T,N\to\infty)_{seq}}{\text{plim}} \left(\frac{1}{N} \sum_{i=1}^{N} f_{i,t}^h - \widehat{B^h} \right) = \mathbb{E}_{t-h}(y_t), \tag{2}$$

where $\underset{(T,N\to\infty)_{seq}}{\text{plim}}$ is the probability limit using the asymptotic structure proposed in Phillips and Moon (1999). Therefore, the BCAF is an optimal forecast device. They also proposed a test for zero bias, that is, $H_0 : B^h = 0$, using the approach of Conley (1999). Note that if H_0 is not rejected, there is no need to use the bias correction. In this case, the optimal forecast will be the simple cross-section average forecast:

$$\frac{1}{N} \sum_{i=1}^{N} f_{i,t}^h.$$

[1] See Laster et al. (1999) e Batchelor (2007).

Gaglianone and Issler (2019) proposed an extended setup, now including two sources of bias: the intercept bias k_i^h and the slope bias β_i^h, as follows:

$$f_{i,t}^h = k_i^h + \beta_i^h \mathbb{E}_{t-h}(y_t) + \varepsilon_{i,t}^h. \tag{3}$$

By comparing (1) with (3), it becomes clear that the first setup is a special case of the second framework, where $\beta_i = 1$ for all i. Gaglianone and Issler proposed the use of GMM to estimate the model parameters $\theta = \left(\overline{B^h}, \overline{\beta^h}\right)'$, where $\overline{B^h} = \frac{1}{N}\sum_{i=1}^{N} k_i^h$ and $\overline{\beta^h} = \frac{1}{N}\sum_{i=1}^{N} \beta_i^h$, are cross-section averages for each h. Starting with (3), and using the law of iterated expectations with valid observable instruments z_{t-s}, where $s \geq h$, it follows that:

$$\mathbb{E}\left[\left(f_{i,t}^h - k_i^h - \beta_i^h y_t\right) \otimes z_{t-s}\right] = 0, \tag{4}$$

which is valid for all $i = 1, ..., N$, $t = 1, ..., T$, and $h = 1, ..., H$. The system of equations (4) has $2NH$ parameters and (at least) $2NH$ moment conditions, provided that $dim(z_{t-s}) > 2$, which is necessary for overidentification. Despite that, a problem remains: as long as $N \to \infty$, the number of parameters in (4) diverges, which works against consistency. To overcome this *curse of dimensionality*, one can use the cross-section averages of $f_{i,t}^h$, k_i^h and $\beta_i^h y_t$, resulting in the following moment restrictions:

$$\mathbb{E}\left[\left(\overline{f_{\cdot,t}^h} - \overline{B^h} - \overline{\beta^h} y_t\right) \otimes z_{t-s}\right] = 0, \tag{5}$$

for $t = 1, ..., T$, and $h = 1, ..., H$, where $\overline{f_{\cdot,t}^h} = \frac{1}{N}\sum_{i=1}^{N} f_{i,t}^h$, $\overline{B^h} = \frac{1}{N}\sum_{i=1}^{N} k_i^h$ and $\overline{\beta^h} = \frac{1}{N}\sum_{i=1}^{N} \beta_i^h$, represent cross-section averages for each h. Finally, Gaglianone and Issler show how to obtain consistent estimates of the model parameters $\theta = \left(\overline{B^h}, \overline{\beta^h}\right)'$ using GMM and the previous cross-section averages within different asymptotic setups.

Appendix 2. The factor model of Schwartz and Smith (2000) and others

Schwartz and Smith (2000) assume the logarithm of the spot oil price, $\ln(S_t)$, defined in continuous time, can be decomposed as follows:

$$\ln(S_t) = \chi_t + \xi_t,$$

where χ_t and ξ_t are, respectively, the short-run price deviation and the equilibrium price level. The authors also assume that the first term is zero mean-reverting, following an Ornstein-Uhlenbeck process, whereas the second term follows a Brownian motion, as follows:

$$\begin{aligned} d\chi_t &= -\kappa\chi_t dt + \sigma_\chi dz_\chi, \\ d\xi_t &= \mu_\xi dt + \sigma_\xi dz_\xi, \end{aligned}$$

where dz_χ and dz_ξ are correlated terms from a standard Brownian motion, such that $dz_\chi dz_\xi = \rho_{\chi\xi}dt$. The first and second centered moments of $(\chi_t, \xi_t)'$ are, respectively:

$$\begin{aligned} \mathbb{E}\left[(\chi_t, \xi_t)\right] &= \left[e^{-\kappa t}\chi_0, \xi_0 + \mu_\xi t\right], \text{ and} \\ \text{COV}\left[(\chi_t, \xi_t)\right] &= \begin{bmatrix} \left(1-e^{-2\kappa t}\right)\frac{\sigma_\chi^2}{2\kappa} & \left(1-e^{-\kappa t}\right)\frac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa} \\ \left(1-e^{-\kappa t}\right)\frac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa} & \sigma_\xi^2 t \end{bmatrix}, \end{aligned}$$

where χ_0 and ξ_0 are initial conditions. Given these initial conditions, and the assumption of log-normality for $\ln(S_t)$, the future spot prices have mean and variance, respectively, given as follows:

$$\begin{aligned} \mathbb{E}\left[\ln(S_t)\right] &= e^{-\kappa t}\chi_0 + \xi_0 + \mu_\xi t, \text{ and,} \qquad (6) \\ \text{VAR}\left[\ln(S_t)\right] &= \left(1-e^{-2\kappa t}\right)\frac{\sigma_\chi^2}{2\kappa} + \sigma_\xi^2 t + 2\left(1-e^{-\kappa t}\right)\frac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa}, \end{aligned}$$

which implies the following expected future spot prices:

$$\mathbb{E}\left[S_t\right] = \exp\left\{\mathbb{E}\left[\ln(S_t)\right] + \frac{1}{2}\text{VAR}\left[\ln(S_t)\right]\right\},$$

and, therefore,

$$\begin{aligned} \ln\left(\mathbb{E}\left[S_t\right]\right) &= \mathbb{E}\left[\ln(S_t)\right] + \frac{1}{2}\text{VAR}\left[\ln(S_t)\right] \\ &= e^{-\kappa t}\chi_0 + \xi_0 + \mu_\xi t + \frac{1}{2}\left[\left(1-e^{-2\kappa t}\right)\frac{\sigma_\chi^2}{2\kappa} + \sigma_\xi^2 t + 2\left(1-e^{-\kappa t}\right)\frac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa}\right]. \end{aligned}$$

By considering $t \to \infty$ in the last expression, the log of expected prices can be calculated as long as the horizon increases, i.e.,

$$\lim_{t\to\infty} \ln\left(\mathbb{E}\left[S_t\right]\right) = \left(\xi_0 + \frac{\sigma_\chi^2}{4\kappa} + \frac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa}\right) + \left(\mu_\xi + \frac{1}{2}\sigma_\xi^2\right) t.$$

This way, in the long-run, the expected spot price behave as if it has started with an "effective long-run price" equal to $\exp\left(\xi_0 + \frac{\sigma_\chi^2}{4\kappa} + \frac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa}\right)$ and further increased at the rate $\left(\mu_\xi + \frac{1}{2}\sigma_\xi^2\right)$. Note this effective long-run price is slightly different from the equilibrium price $(\exp(\xi_0))$, where the difference reflects the contribution of the short-run volatility related to the expected spot prices.

Schwartz and Smith discuss a risk-neutral process and the respective cash-flow evaluation. The authors argue that, under risk-neutral probabilities, it follows that:

$$\begin{aligned} d\chi_t &= \left(-\kappa\chi_t + \lambda_\chi\right) dt + \sigma_\chi dz_\chi^*, \\ d\xi_t &= \underbrace{\left(\mu_\xi - \lambda_\xi\right)}_{\mu_\xi^*} dt + \sigma_\xi dz_\xi^*, \end{aligned}$$

where $dz_\chi^* dz_\xi^* = \rho_{\chi\xi} dt$, and show that:

$$\begin{aligned} \mathbb{E}^*\left[(\chi_t, \xi_t)\right] &= \left[e^{-\kappa t}\chi_0 - \left(1 - e^{-\kappa t}\right)\frac{\lambda_\chi}{\kappa}, \xi_0 + \mu_\xi^* t\right], \text{ and} \\ \text{COV}^*\left[(\chi_t, \xi_t)\right] &= COV\left[(\chi_t, \xi_t)\right], \end{aligned}$$

where $\mathbb{E}^*$ denotes the respective variable under risk-neutral probabilities, instead of physical probabilities, and λ_χ is the average adjustment needed in the Ornstein-Uhlenbeck process under risk-neutral probabilities. Thus, it follows that:

$$\begin{aligned} \mathbb{E}^*\left[\ln(S_t)\right] &= e^{-\kappa t}\chi_0 + \xi_0 - \left(1 - e^{-\kappa t}\right)\frac{\lambda_\chi}{\kappa} + \mu_\xi^* t, \text{ and,} \\ \text{VAR}^*\left[\ln(S_t)\right] &= \text{VAR}\left[\ln(S_t)\right]. \end{aligned} \tag{7}$$

This way, by comparing (6) with (7), note the risk premium decreases the log of the expected spot price by:

$$\left(1 - e^{-\kappa t}\right)\frac{\lambda_\chi}{\kappa} + \lambda_\xi t.$$

After discussing the risk-neutral approach based on future contracts, with maturities $T_1, T_2, \cdots, T_n$, Schwartz and Smith show that the state-space form that represents $\ln(S_t)$ is the following:

$$\mathbf{X}_t = \mathbf{c} + \mathbf{G}\mathbf{X}_{t-1} + \boldsymbol{\omega}_t, \tag{8}$$

$$\mathbf{y}_t = \mathbf{d}_t + \mathbf{F}_t'\mathbf{X}_t + \mathbf{v}_t, \tag{9}$$

where $\mathbf{X}_t = (\chi_t, \xi_t)'$, $\mathbf{c} = \left(0, \mu_\xi \Delta t\right)'$, and Δt represents the duration of time periods, $\boldsymbol{\omega}_t$ is a 2×1 vector of disturbances with covariance matrix COV$(\boldsymbol{\omega}_t)$ =COV$[(\chi_t, \xi_t)]$, $\mathbf{y}_t = \left[\ln(F_{T_1}), \ln(F_{T_2}), \cdots, \ln(F_{T_n})\right]'$, where $F_{T_1}, F_{T_2}, \cdots, F_{T_N}$

are future prices, respectively, with maturities $T_1, T_2, \cdots, T_n$, with constant terms $\mathbf{d}_t = [A(T_1), A(T_2), \cdots, A(T_n)]'$ associated to them, $\mathbf{F}_t = \left[e^{-\kappa T_1}\mathbf{1}, e^{-\kappa T_2}\mathbf{1}, \cdots, e^{-\kappa T_n}\mathbf{1}\right]'$, with $\mathbf{1} = (1,1)'$, and $\mathbf{v}_t$ is a $n \times 1$ vector of normal disturbances without serial correlation, zero mean, and $\text{COV}[\mathbf{v}_t] = \mathbf{V}$. The total number of periods is n_T, i.e., $t = 1, 2, \cdots, n_T$, and:

$$\mathbf{G} = \begin{bmatrix} e^{-\kappa \Delta t} & 0 \\ 0 & 1 \end{bmatrix}.$$

The equation (8) is often called the *transition equation,* whereas the equation (9) is known as the *measurement equation.* Under a joint normality hypothesis for $(\boldsymbol{\omega}_t, \mathbf{v}_t)'$ it is easy to estimate the parameters of interest associated to equations (8) and (9) using the Kalman filter (maximum likelihood estimates). These parameters are set in vector $\left(\kappa, \sigma_\chi, \mu_\xi, \sigma_\xi, \rho_{\chi\xi}, \lambda_\chi, \mu_\xi^*\right)'$. Implicitly, future prices with different maturities are being used to identify the short- and long-run price components. Schwartz and Smith employ a Bayesian approach in the model estimation, using a multivariate Gaussian distribution as *a priori* distribution. Besides, they also use Kalman filter techniques in the steady state, where $\text{COV}[(\chi_t, \xi_t)]$ becomes independent of the initial conditions assumed in the filter.

Based on the framework above, Cortazar and Naranjo (2006) proposed a Gaussian model with $N-$factors to explain the stochastic behavior of future oil prices, which is estimated using all price available information, in contrast to the traditional approaches that aggregate data for a set of maturities. The model is calibrated using a Kalman filter procedure that allows for a number of time-dependent daily observations. The model shows a relatively good performance, requiring at least three factors to explain the term structure of future prices, but four factors to properly adjust the volatility term structure.

Cortazar et al. (2015) argue that stochastic models of commodity prices have considerably evolved in respect to both structure and state variable interpretation. However, it is not well emphasized in the literature that those models, besides providing a risk-neutral distribution for future prices, also give their physical distribution. Although the parameters of the risk-neutral distribution can be more precisely estimated (and, in general, are statistically significant), some parameters of the physical distribution are typically measured with huge confidence bands, and are not statistically significant. This way, in order to improve the model performance, some parameters – in particular, the risk premium parameters – must be obtained from other sources. In this sense, to reduce the uncertainty related to the *future* risk premium estimates, a model restriction can be made using the CAPM (Capital Asset Pricing Model) approach, that is, the authors set the term structure of risk premium based on a satellite CAPM model. Using such restriction, Cortazar et alli argue that the estimate of the physical distribution becomes stable and reliable.

Appendix 3. Random Forest and Quantile Regression Forest

Random Forest

In this section, we first discuss how to properly grow a single regression tree and (automatically) decide on both the splitting variables and split points. Hastie et al. (2009) propose the following algorithm, in the context of CART (classification and regression tree) models:

(i) consider a splitting variable j and split point s, and define the pair of half-planes:

$$R_1(j,s) = \{X \mid X_j \leq s\} \quad \text{and} \quad R_2(j,s) = \{X \mid X_j > s\}, \tag{10}$$

(ii) find the splitting variable j and split point s that solve the minimization problem:

$$\min_{j,s}\left[\min_{c_1}\sum_{x_i\in R_1(j,s)}(y_i-c_1)^2+\min_{c_2}\sum_{x_i\in R_2(j,s)}(y_i-c_2)^2\right], \tag{11}$$

where the previous inner minimizations, for any choice j and s, can be solved by:

$$\widehat{c_1} = E(y_i \mid x_i \in R_1(j,s)) \quad \text{and} \quad \widehat{c_2} = E(y_i \mid x_i \in R_2(j,s)). \tag{12}$$

Note that for a given splitting variable, the computation of the optimal split point s can be easily done. Thus, by searching through all covariates, the determination of the best pair (j,s) is feasible. Then, based on the best split one divides the data into the two resulting regions R_1 and R_2 and repeat the splitting process on each of the two regions. This process is repeated on all of the resulting regions until some stopping rule is applied. Finally, the forecast of Y conditioned on the covariate space X, which is partitioned into L regions $R_l(j,s)$, $l = 1, ..., L$, according to the regression tree approach, is the following:

$$E_{\text{regression tree}}(Y \mid X) = \sum_{l=1}^{L} c_l \mathbf{1}_{\{X\in R_l(j,s)\}}, \tag{13}$$

where $c_l = E(Y \mid X \in R_l(j,s))$. To sum it up, the *regression tree* can be estimated by repeating the three steps below, for each terminal node of the tree, until the minimum number of observations at each node is achieved:[2]

(1) randomly select m out of p covariates as possible split variables;[3]

[2] The size of a tree is a tuning parameter governing the model's complexity, and the optimal size should be adaptively chosen from the data. The preferred strategy is to stop the splitting process when some minimum node size is reached. Typically, for regression problems with p predictors, the literature recommends to use $m = p/3$ (rounded down) in each split, with a minimum node size of 5 as the default; see Hastie et al. (2009, chapter 15.3) for more details.

[3] The reduction of the tuning parameter m will, in general, reduce the correlation between any pair of trees.

(2) select the best variable/split point among the m candidates;
(3) split the node into two child nodes.

Next, we represent mathematically the *Random Forest* (RF) model, following the discussion in Meinshausen (2006). Consider n independent observations (Y_i, X_i), for $i = 1, ..., n$, and let θ be the random parameter vector that determines how a tree $T(\theta)$ is grown, that is, characterizes the tree in terms of split variables, cut-points at each node, and terminal-node values. Also, let $\Im$ be the space in which X lives, that is $X : \Omega \rightarrow \Im$, where $\Im \subseteq \mathbb{R}^p$ and $p \in \mathbb{N}_+$ is the dimensionality of the set of covariates X.

Every leaf of a tree (terminal node) $l = 1, ..., L$ corresponds to a subspace of $\Im$, that is $R_l \subseteq \Im$. For every $x \in \Im$, there is one (and only one) leaf l such that $x \in R_l$ (corresponding to the leaf that is obtained when dropping x down the tree). Denote this leaf by $l(x, \theta)$ for tree $T(\theta)$. The prediction of a single tree $T(\theta)$ conditioned on $X = x$ is obtained by averaging over the observed values in leaf $l(x, \theta)$. Let the weight vector $w_i(x, \theta)$ be given by a positive constant if observation X_i is part of leaf $l(x, \theta)$ and 0 if it is not. The weights sum to one, such that:

$$w_i(x, \theta) = \frac{\mathbf{1}_{\{X_i \in R_{l(x,\theta)}\}}}{\sum_{j=1}^{n} \mathbf{1}_{\{X_j \in R_{l(x,\theta)}\}}}. \tag{14}$$

The forecasting model based on a single *regression tree*, conditioned on a covariate $X = x$, is then the weighted average of the original observations Y_i, for all $i = 1, ..., n$, that is:

$$E_{\text{regression tree}}(Y \mid X = x) = \sum_{i=1}^{n} w_i(x, \theta) Y_i. \tag{15}$$

Note that conditional on the knowledge of the subregions R_l, for $l = 1, ..., L$, the relationship between inflation Y and the set of covariates X is approximated here by a piecewise constant model, where each leaf represents a distinct regime; see Garcia et al. (2017). Now, using *random forest*, the conditional mean above is approximated by the averaged prediction of K single trees, each constructed with a parameter vector θ_k, $k = 1, ..., K$. Let $w_i(x)$ be the average of $w_i(x, \theta_k)$ over this collection of trees, as follows:

$$w_i(x) = \frac{1}{K} \sum_{k=1}^{K} w_i(x, \theta_k). \tag{16}$$

The RF forecast is the averaged response of all trees, as follows:[4]

[4] According to Hastie et al. (2009), tree learning is invariant under scaling and various other transformations (and it is robust to inclusion of irrelevant covariates), however it is seldom accurate. In particular, large trees tend to learn highly irregular patterns and overfit their training sets, thus producing low bias but very high prediction variance. In order to reduce such high variance, random forests average multiple decision trees, trained on different parts of the training set. This often comes at the expense of a small increase in the bias, but usually improves the overall performance of the model.

$$E_{\text{random forest}}\left(Y \mid X=x\right)=\sum_{i=1}^{n} w_i(x)Y_i. \tag{17}$$

Note that the approximation of the conditional mean of Y given $X = x$ is given by a weighted sum over all observations. The weights vary with the covariate and tend to be large for those observations $i \in \{1, ..., n\}$ where the conditional distribution of Y, given $X = X_i$, is similar to the conditional distribution of Y given $X = x$.

Quantile Regression Forest (QRF)

The *quantile regression forest* algorithm proposed by Meinshausen (2006) to compute the estimate of the conditional distribution function can be summarized as follows:

(a) grow trees $T(\theta_k)$, for $k = 1, ..., K$, as in random forests. However, for every leaf (on each tree) consider all observations in the leaf, not just their average.

(b) for a given $X = x$, drop x down all trees. Compute the weight $w_i(x, \theta_k)$ of observation $i \in \{1, ..., n\}$ for every tree as in (14). Compute weight $w_i(x)$ for every observation $i \in \{1, ..., n\}$ as an average over $w_i(x, \theta_k)$, for all $k = 1, ..., K$, as in (16).

(c) compute the estimate of the distribution function as in (20) for all $y \in \mathbb{R}$, using the weights from the previous step (b).

This way, conditional quantiles can be inferred with QRF as a generalization of random forests. The idea is to provide a non-parametric way of estimating conditional quantiles for a high-dimensional set of predictor variables. According to Meinshausen (2006), the QRF algorithm is shown to be consistent and competitive in terms of predictive power. First, recall that the conditional distribution function (CDF) of Y, given $X = x$, is given by:

$$F\left(y \mid X=x\right)=\Pr\left(Y \leq y \mid X=x\right)=E\left(I_{\{Y\leq y\}} \mid X=x\right). \tag{18}$$

Also, recall that the conditional quantile of Y, given $X = x$, at quantile level τ, is given by:

$$Q_\tau\left(Y \mid X=x\right)=\inf\{y : F(y \mid X=x) \geq \tau\}. \tag{19}$$

In other words, for a continuous distribution function of Y, conditional on $X = x$, the probability of Y being smaller than $Q_\tau(\cdot)$ is equal to τ. Now, similarly to the random forest approximation of the conditional mean, define an approximation to $E\left(I_{\{Y\leq y\}} \mid X=x\right)$ by the weighted mean over the observations of $I_{\{Y\leq y\}}$, as follows:

$$\widehat{F}\left(y \mid X=x\right)=\sum_{i=1}^{n} w_i(x) I_{\{Y_i\leq y\}}, \tag{20}$$

using the same weights $w_i(x)$ for random forests, as defined above. Estimates $\widehat{Q}_\tau(\cdot)$ of the conditional quantiles $Q_\tau(\cdot)$ can, thus, be obtained by simply plugging $\widehat{F}\left(y \mid X=x\right)$, instead of $F\left(y \mid X=x\right)$, into (19).

On the other hand, the conditional mean of Y can be approximated by a combination of conditional quantiles. It is not a novel approach in the literature. Indeed, it has a long tradition in statistics (see Judge et al., 1988) and has been previously applied in the forecasting literature; see Lima and Meng (2017). We follow here the approach of Araujo and Gaglianone (2020), that proposed a quantile combination approach using the QRF algorithm to build conditional mean forecasts of Y. This could be accomplished by integrating the conditional quantile function of Y over the entire domain $\tau \in [0,1]$, as follows (see Koenker, 2005, p.302):

$$E(Y \mid X = x) = \int_0^1 Q_\tau (Y \mid X = x)\, d\tau. \tag{21}$$

The conditional mean of Y, based on the QRF approach, can thus be approximated[5] by a sum of estimated conditional quantiles, as follows:[6]

$$\int_0^1 Q_\tau (Y \mid X = x)\, d\tau = \lim_{P \to \infty} \left(\sum_{p=1}^{P} \widehat{Q}_{\tau_p} (Y \mid X = x)\, \Delta\tau_p \right). \tag{22}$$

Therefore, one can build conditional mean forecasts of Y through equations (19), (20), (21) and (22).

Appendix 4. Variable Importance in Random Forest

Random forests are among the most popular machine learning methods due to their relatively good forecasting accuracy, robustness and ease of use. In contrast to parametric methods, random forests are fully non-parametric and can deal with nonlinear effects, thus offering a great model flexibility in practical applications. Furthermore, RF can even be applied in the statistically challenging setting in which the number of variables is higher than the number of observations. This makes random forests especially attractive for complex high-dimensional data applications; see Janitza et al. (2018).

Nonetheless, a suitable understanding of the *black box* mechanism behind the random forest method is of greatest importance. Nowadays, machine-learning models are often deployed to production without a proper understanding of why exactly the algorithms make the decisions they do. As these new tools become more relevant in everyday life, model interpretability becomes one of the most important problems in machine learning these days. In particular, regarding the use of RF as a forecasting

[5] By applying the second fundamental theorem of calculus (or the Newton-Leibniz axiom) on the sum of quantiles, the Riemann integral is obtained in the limit $P \to \infty$ (see Apostol, 1967) and the partitions $\Delta\tau_p = \frac{1}{P+1}$ get finer (i.e., $\Delta\tau_p \to 0$ as long as $P \to \infty$).

[6] We rely on the fact that the conditional quantiles are consistenly estimated using the QRF approach.

device, it is critical to comprehend the key variable interactions that are providing the predictive accuracy.

One attempt to tackle this issue is to compute the so-called "variable importance measures", by attributing scores to the variables, which reflect their relative importance in the overall model accuracy. Such measures can be used to identify relevant features, perform variable selection and quantify the prediction strength of each variable, allowing one to rank the variables according to their predictive abilities. See Hastie et al. (2009, chapter 15) for further details.[7]

A global insight into the random forest's behavior can be obtained by computing the two main variable importance measures, based on the "permutation" approach of Altmann et al. (2010) and on the "impurity-corrected" method of Nembrini et al. (2018). Moreover, one can carry out the Janitza et al. (2018) hypothesis test of no association between the predictor and the dependent variable for both measures.

The permutation method, also known as the mean decrease in accuracy, is one of the most common variable importance measures, and it is computed from the change in prediction accuracy when removing any association between the dependent variable (response) and a given regressor (i.e., feature or predictor), with large changes indicating that the predictor is important.[8] One disadvantage of the permutation approach is to produce biased outcomes when predictors are highly correlated. In addition, adding a correlated variable to the RF model can decrease the importance of another variable. Furthermore, the permutation importance is very computationally intensive in the case of high dimensional data.

Alternative importance measures based on impurity (i.e., how well the regression trees split the variables) are popular because they are simple, fast to compute and can be more robust to data perturbations compared with those based on permutation.[9] However, the impurity importance is known to be biased towards variables with more categories or more possible split points. Also, when the dataset has two (or more) correlated variables, any of them can be selected as predictor. Nevertheless, once

[7] There are many other ways on the lookout for opening the ML black box. Just to mention a few examples: (i) Partial Dependence Plots (PDP), which show the marginal effect of a given predictor on the outcome of a ML model; and (ii) Surrogate Models (SM), which are auxiliary interpretable models (e.g., linear regression), built to approximate the predictions of a ML model in order to understand the black box outcomes by analyzing (and interpreting) the surrogate model's responses.

[8] According to Nembrini et al. (2018): *"To calculate the permutation importance of the variable xi, its original association with the response y is broken by randomly permuting the values of all individuals for xi. With this permuted data, the tree-wise out-of-bag (OOB) estimate of the prediction error is computed. The difference between this estimate and the OOB error without permutation, averaged over all trees, is the permutation importance of the variable xi. This procedure is repeated for all variables of interest x1,. . . ,xp. The larger the permutation importance of a variable, the more relevant the variable is for the overall prediction accuracy."*

[9] Recall that random forest consists of a number of decision trees. Every node in the trees is a condition on a given variable, and it is designed to optimally split the dataset into two parts so that overall model accuracy can be improved. The measure based on which the (locally) optimal condition is chosen is called impurity (or variance, in the case of the regression trees). This way, one can compute how much each variable reduces the weighted impurity in a tree. For a forest, the impurity reduction from each variable can be averaged and a ranking of variables can be constructed according to this importance measure.

one of these (correlated) variables is used as predictor, the importance of others is significantly reduced, since the impurity these other variables can decrease is already reduced by the first selected variable.[10] In this sense, Nembrini et al.(2018) propose the "corrected impurity" importance measure, which is unbiased in terms of the number of categories and category frequencies and is computationally efficient (i.e., almost as fast as the standard impurity importance and much faster than the permutation importance).

Besides building a ranking of importance, it is also crucial to statistically check whether a given predictor is important (or not) in respect to the dependent variable of the RF model. According to Janitza et al. (2018), the variable importance depends on many different factors, including aspects related to the data (e.g., correlations, signal-to-noise ratio or the total number of variables) as well as on the random forest specific factors (such as the choice of the number of randomly drawn candidate predictor variables for each split node). Therefore, there is no universally applicable threshold that can be used to statistically discriminate between important and non-important variables. Nonetheless, several hypothesis-testing approaches have been developed. The permutation-based tests entail the repeated computation of random forests. While for low-dimensional settings those approaches might be computationally tractable, for high-dimensional models (e.g., including thousands of predictors), computing time might become enormous. In this sense, Janitza et al. (2018) propose a variable importance test that is appropriate for high-dimensional data where many variables do not carry any information related to the dependent variable. According to the authors, the testing approach, based on cross-validation procedures, shows at least comparable power at a substantially smaller computation time.

[10] This is not an issue in respect to model forecasting, but regarding model interpretation, it can lead to the incorrect conclusion that one of the variables is a strong predictor while the others (correlated variables) are not important, while, in reality, they are all close in respect to their statistical relationship with the dependent variable. This effect can be attenuated by using random variable selection at each node (instead of using all possible variables) when growing a tree within the random forest setup.

Appendix 5. Database

Table 5.1 - List of macroeconomic and financial variables

	Category	Name	Source	Nickname
1	Brent oil real price	Europe Brent Spot FOB U$/BBL Daily, deflated by PPI all commodities	FRED, Thomson Reuters	BRENT_REAL
2	Output and Income	Real Personal Income	FRED-MD	RPI
3	Output and Income	Real personal income ex transfer receipts	FRED-MD	W875RX1
4	Output and Income	IP Index	FRED-MD	INDPRO
5	Output and Income	IP: Final Products and Nonindustrial Supplies	FRED-MD	IPFPNSS
6	Output and Income	IP: Final Products (Market Group)	FRED-MD	IPFINAL
7	Output and Income	IP: Consumer Goods	FRED-MD	IPCONGD
8	Output and Income	IP: Durable Consumer Goods	FRED-MD	IPDCONGD
9	Output and Income	IP: Nondurable Consumer Goods	FRED-MD	IPNCONGD
10	Output and Income	IP: Business Equipment	FRED-MD	IPBUSEQ
11	Output and Income	IP: Materials	FRED-MD	IPMAT
12	Output and Income	IP: Durable Materials	FRED-MD	IPDMAT
13	Output and Income	IP: Nondurable Materials	FRED-MD	IPNMAT
14	Output and Income	IP: Manufacturing (SIC)	FRED-MD	IPMANSICS
15	Output and Income	IP: Residential Utilities	FRED-MD	IPB51222s
16	Output and Income	IP: Fuels	FRED-MD	IPFUELS
17	Output and Income	Capacity Utilization: Manufacturing	FRED-MD	CUMFNS
18	Labor market	Help-Wanted Index for United States	FRED-MD	HWI
19	Labor market	Ratio of Help Wanted/No. Unemployed	FRED-MD	HWIURATIO
20	Labor market	Civilian Labor Force	FRED-MD	CLF16OV
21	Labor market	Civilian Employment	FRED-MD	CE16OV
22	Labor market	Civilian Unemployment Rate	FRED-MD	UNRATE
23	Labor market	Average Duration of Unemployment (Weeks)	FRED-MD	UEMPMEAN
24	Labor market	Civilians Unemployed - Less Than 5 Weeks	FRED-MD	UEMPLT5
25	Labor market	Civilians Unemployed for 5-14 Weeks	FRED-MD	UEMP5TO14
26	Labor market	Civilians Unemployed - 15 Weeks & Over	FRED-MD	UEMP15OV
27	Labor market	Civilians Unemployed for 15-26 Weeks	FRED-MD	UEMP15T26
28	Labor market	Civilians Unemployed for 27 Weeks and Over	FRED-MD	UEMP27OV
29	Labor market	Initial Claims	FRED-MD	CLAIMSx
30	Labor market	All Employees: Total nonfarm	FRED-MD	PAYEMS
31	Labor market	All Employees: Goods-Producing Industries	FRED-MD	USGOOD
32	Labor market	All Employees: Mining and Logging: Mining	FRED-MD	CES1021000001
33	Labor market	All Employees: Construction	FRED-MD	USCONS
34	Labor market	All Employees: Manufacturing	FRED-MD	MANEMP
35	Labor market	All Employees: Durable goods	FRED-MD	DMANEMP
36	Labor market	All Employees: Nondurable goods	FRED-MD	NDMANEMP
37	Labor market	All Employees: Service-Providing Industries	FRED-MD	SRVPRD
38	Labor market	All Employees: Trade, Transportation & Utilities	FRED-MD	USTPU
39	Labor market	All Employees: Wholesale Trade	FRED-MD	USWTRADE
40	Labor market	All Employees: Retail Trade	FRED-MD	USTRADE
41	Labor market	All Employees: Financial Activities	FRED-MD	USFIRE
42	Labor market	All Employees: Government	FRED-MD	USGOVT
43	Labor market	Avg Weekly Hours : Goods-Producing	FRED-MD	CES0600000007
44	Labor market	Avg Weekly Overtime Hours : Manufacturing	FRED-MD	AWOTMAN
45	Labor market	Avg Weekly Hours : Manufacturing	FRED-MD	AWHMAN
46	Labor market	Avg Hourly Earnings : Goods-Producing	FRED-MD	CES0600000008
47	Labor market	Avg Hourly Earnings : Construction	FRED-MD	CES2000000008
48	Labor market	Avg Hourly Earnings : Manufacturing	FRED-MD	CES3000000008
49	Housing	Housing Starts: Total New Privately Owned	FRED-MD	HOUST
50	Housing	Housing Starts, Northeast	FRED-MD	HOUSTNE
51	Housing	Housing Starts, Midwest	FRED-MD	HOUSTMW
52	Housing	Housing Starts, South	FRED-MD	HOUSTS
53	Housing	Housing Starts, West	FRED-MD	HOUSTW
54	Housing	New Private Housing Permits (SAAR)	FRED-MD	PERMIT
55	Housing	New Private Housing Permits, Northeast (SAAR)	FRED-MD	PERMITNE
56	Housing	New Private Housing Permits, Midwest (SAAR)	FRED-MD	PERMITMW
57	Housing	New Private Housing Permits, South (SAAR)	FRED-MD	PERMITS
58	Housing	New Private Housing Permits, West (SAAR)	FRED-MD	PERMITW
59	Consumption, orders, and inventorie	Real personal consumption expenditures	FRED-MD	DPCERA3M086SBEA
60	Consumption, orders, and inventorie	Real Manu. and Trade Industries Sales	FRED-MD	CMRMTSPLx
61	Consumption, orders, and inventorie	Retail and Food Services Sales	FRED-MD	RETAILx
62	Consumption, orders, and inventorie	New Orders for Consumer Goods	FRED-MD	ACOGNO
63	Consumption, orders, and inventorie	New Orders for Durable Goods	FRED-MD	AMDMNOx
64	Consumption, orders, and inventorie	New Orders for Nondefense Capital Goods	FRED-MD	ANDENOx
65	Consumption, orders, and inventorie	Unfilled Orders for Durable Goods	FRED-MD	AMDMUOx
66	Consumption, orders, and inventorie	Total Business Inventories	FRED-MD	BUSINVx
67	Consumption, orders, and inventorie	Total Business: Inventories to Sales Ratio	FRED-MD	ISRATIOx
68	Consumption, orders, and inventorie	Consumer Sentiment Index	FRED-MD	UMCSENTx

Table 5.1 - List of macroeconomic and financial variables (cont.)

69	Money and credit	M1 Money Stock	FRED-MD	M1SL
70	Money and credit	M2 Money Stock	FRED-MD	M2SL
71	Money and credit	Real M2 Money Stock	FRED-MD	M2REAL
72	Money and credit	Monetary Base	FRED-MD	BOGMBASE
73	Money and credit	Total Reserves of Depository Institutions	FRED-MD	TOTRESNS
74	Money and credit	Reserves Of Depository Institutions	FRED-MD	NONBORRES
75	Money and credit	Commercial and Industrial Loans	FRED-MD	BUSLOANS
76	Money and credit	Real Estate Loans at All Commercial Banks	FRED-MD	REALLN
77	Money and credit	Total Nonrevolving Credit	FRED-MD	NONREVSL
78	Money and credit	Nonrevolving consumer credit to Personal Income	FRED-MD	CONSPI
79	Money and credit	MZM Money Stock	FRED-MD	MZMSL
80	Money and credit	Consumer Motor Vehicle Loans Outstanding	FRED-MD	DTCOLNVHFNM
81	Money and credit	Total Consumer Loans and Leases Outstanding	FRED-MD	DTCTHFNM
82	Money and credit	Securities in Bank Credit at All Commercial Banks	FRED-MD	INVEST
83	Interest and exchange rates	Effective Federal Funds Rate	FRED-MD	FEDFUNDS
84	Interest and exchange rates	3-Month AA Financial Commercial Paper Rate	FRED-MD	CP3Mx
85	Interest and exchange rates	3-Month Treasury Bill	FRED-MD	TB3MS
86	Interest and exchange rates	6-Month Treasury Bill	FRED-MD	TB6MS
87	Interest and exchange rates	1-Year Treasury Rate	FRED-MD	GS1
88	Interest and exchange rates	5-Year Treasury Rate	FRED-MD	GS5
89	Interest and exchange rates	10-Year Treasury Rate	FRED-MD	GS10
90	Interest and exchange rates	Moody's Seasoned Aaa Corporate Bond Yield	FRED-MD	AAA
91	Interest and exchange rates	Moody's Seasoned Baa Corporate Bond Yield	FRED-MD	BAA
92	Interest and exchange rates	3-Month Commercial Paper Minus FEDFUNDS	FRED-MD	COMPAPFFx
93	Interest and exchange rates	3-Month Treasury C Minus FEDFUNDS	FRED-MD	TB3SMFFM
94	Interest and exchange rates	6-Month Treasury C Minus FEDFUNDS	FRED-MD	TB6SMFFM
95	Interest and exchange rates	1-Year Treasury C Minus FEDFUNDS	FRED-MD	T1YFFM
96	Interest and exchange rates	5-Year Treasury C Minus FEDFUNDS	FRED-MD	T5YFFM
97	Interest and exchange rates	10-Year Treasury C Minus FEDFUNDS	FRED-MD	T10YFFM
98	Interest and exchange rates	Moody's Aaa Corporate Bond Minus FEDFUNDS	FRED-MD	AAAFFM
99	Interest and exchange rates	Moody's Baa Corporate Bond Minus FEDFUNDS	FRED-MD	BAAFFM
100	Interest and exchange rates	Trade Weighted U.S. Dollar Index	FRED-MD	TWEXAFEGSMTHx
101	Interest and exchange rates	Switzerland / U.S. Foreign Exchange Rate	FRED-MD	EXSZUSx
102	Interest and exchange rates	Japan / U.S. Foreign Exchange Rate	FRED-MD	EXJPUSx
103	Interest and exchange rates	U.S. / U.K. Foreign Exchange Rate	FRED-MD	EXUSUKx
104	Interest and exchange rates	Canada / U.S. Foreign Exchange Rate	FRED-MD	EXCAUSx
105	Prices	PPI: Finished Goods	FRED-MD	WPSFD49207
106	Prices	PPI: Finished Consumer Goods	FRED-MD	WPSFD49502
107	Prices	PPI: Intermediate Materials	FRED-MD	WPSID61
108	Prices	PPI: Crude Materials	FRED-MD	WPSID62
109	Prices	Crude Oil, spliced WTI and Cushing	FRED-MD	OILPRICEx
110	Prices	PPI: Metals and metal products	FRED-MD	PPICMM
111	Prices	CPI: All Items	FRED-MD	CPIAUCSL
112	Prices	CPI: Apparel	FRED-MD	CPIAPPSL
113	Prices	CPI: Transportation	FRED-MD	CPITRNSL
114	Prices	CPI: Medical Care	FRED-MD	CPIMEDSL
115	Prices	CPI: Commodities	FRED-MD	CUSR0000SAC
116	Prices	CPI: Durables	FRED-MD	CUSR0000SAD
117	Prices	CPI: Services	FRED-MD	CUSR0000SAS
118	Prices	CPI: All Items Less Food	FRED-MD	CPIULFSL
119	Prices	CPI: All items less shelter	FRED-MD	CUSR0000SA0L2
120	Prices	CPI: All items less medical care	FRED-MD	CUSR0000SA0L5
121	Prices	Personal Cons. Expend.: Chain Index	FRED-MD	PCEPI
122	Prices	Personal Cons. Exp: Durable goods	FRED-MD	DDURRG3M086SBEA
123	Prices	Personal Cons. Exp: Nondurable goods	FRED-MD	DNDGRG3M086SBEA
124	Prices	Personal Cons. Exp: Services	FRED-MD	DSERRG3M086SBEA
125	Stock market	S&P's Common Stock Price Index: Composite	FRED-MD	S&P500
126	Stock market	S&P's Common Stock Price Index: Industrials	FRED-MD	S&P_indust
127	Stock market	S&P's Composite Common Stock: Dividend Yield	FRED-MD	S&P_div_yield
128	Stock market	S&P's Composite Common Stock: Price-Earnings Ratio	FRED-MD	S&P_PE_ratio
129	Stock market	CBOE S&P 100 Volatility Index: VXO	FRED-MD	VXOCLSx
130	Industrial Production	Production of Total Industry in Austria	FRED	AUTPROINDMISMEI
131	Industrial Production	Production of Total Industry in Belgium	FRED	BELPROINDMISMEI
132	Industrial Production	Production of Total Industry in Brazil	FRED	BRAPROINDMISMEI
133	Industrial Production	Production of Total Industry in Canada	FRED	CANPROINDMISMEI
134	Industrial Production	Production of Total Industry in Czech Republic	FRED	CZEPROINDMISMEI
135	Industrial Production	Production of Total Industry in Germany	FRED	DEUPROINDMISMEI
136	Industrial Production	Production of Total Industry in Denmark	FRED	DNKPROINDMISMEI
137	Industrial Production	Production of Total Industry in Spain	FRED	ESPPROINDMISMEI
138	Industrial Production	Production of Total Industry in Finland	FRED	FINPROINDMISMEI
139	Industrial Production	Production of Total Industry in France	FRED	FRAPROINDMISMEI
140	Industrial Production	Production of Total Industry in the United Kingdom	FRED	GBRPROINDMISMEI

Table 5.1 - List of macroeconomic and financial variables (cont.)

141	Industrial Production	Production of Total Industry in Greece	FRED	GRCPROINDMISMEI
142	Industrial Production	Production of Total Industry in Hungary	FRED	HUNPROINDMISMEI
143	Industrial Production	Production of Total Industry in Ireland	FRED	IRLPROINDMISMEI
144	Industrial Production	Production of Total Industry in Israel	FRED	ISRPROINDMISMEI
145	Industrial Production	Production of Total Industry in Italy	FRED	ITAPROINDMISMEI
146	Industrial Production	Production of Total Industry in Japan	FRED	JPNPROINDMISMEI
147	Industrial Production	Production of Total Industry in Korea	FRED	KORPROINDMISMEI
148	Industrial Production	Production of Total Industry in Netherlands	FRED	NLDPROINDMISMEI
149	Industrial Production	Production of Total Industry in Norway	FRED	NORPROINDMISMEI
150	Industrial Production	Production of Total Industry in Poland	FRED	POLPROINDMISMEI
151	Industrial Production	Production of Total Industry in Portugal	FRED	PRTPROINDMISMEI
152	Industrial Production	Production of Total Industry in Slovak Republic	FRED	SVKPROINDMISMEI
153	Industrial Production	Production of Total Industry in Sweden	FRED	SWEPROINDMISMEI
154	Industrial Production	Production of Total Industry in Turkey	FRED	TURPROINDMISMEI
155	Industrial Production in the U.S.	Industrial Production: Durable Goods: Iron and steel products	FRED	IPG3311A2S
156	Industrial Production in the U.S.	Industrial Production: Durable Goods: Alumina and aluminum production and processin	FRED	IPG3313S
157	Industrial Production in the U.S.	Industrial Production: Durable Goods: Raw steel	FRED	IPN3311A2RS
158	Industrial Production in the U.S.	Industrial Production: Durable Goods: Automotive products	FRED	IPB51110S
159	Industrial Production in the U.S.	Industrial Production: Durable Goods: Cement and concrete product	FRED	IPG3273S
160	Industrial Production in the U.S.	Industrial Production: Durable manufacturing: Primary metal	FRED	IPG331S
161	Industrial Production in the U.S.	Industrial Production: Durable manufacturing: Machinery	FRED	IPG333S
162	Industrial Production in the U.S.	Industrial Production: Durable manufacturing: Aerospace and miscellaneous transporta	FRED	IPG3364T9S
163	Industrial Production in the U.S.	Industrial Production: Nondurable manufacturing: Petroleum and coal products	FRED	IPG324S
164	Industrial Production in the U.S.	Industrial Production: Nondurable manufacturing: Chemical	FRED	IPG325S
165	Industrial Production in the U.S.	Industrial Production: Nondurable manufacturing: Plastics and rubber products	FRED	IPG326S
166	Industrial Production in the U.S.	Industrial Production: Nondurable Goods: Petroleum refineries	FRED	IPG32411S
167	Industrial Production in the U.S.	Industrial Production: Nondurable Goods: Pharmaceutical and medicine	FRED	IPG3254S
168	Industrial Production in the U.S.	Industrial Production: Nondurable Goods: Plastics material and resin	FRED	IPN325211S
169	Industrial Production in the U.S.	Industrial Production: Nondurable Goods: Chemical products	FRED	IPB51213S
170	Industrial Production in the U.S.	Industrial Production: Construction supplies	FRED	IPB54100S
171	Industrial Production in the U.S.	Industrial Production: Non-energy, total	FRED	IPX5001ES
172	Industrial Production in the U.S.	Industrial Production: Energy Materials: Energy, total	FRED	IPB50089S
173	Industrial Production in the U.S.	Industrial Production: Electric power generation, transmission, and distribution	FRED	IPG2211S
174	Industrial Production in the U.S.	Industrial Production: Mining: Crude oil	FRED	IPG211111CS
175	Industrial Production in the U.S.	Industrial Production: Mining: Crude petroleum and natural gas extraction	FRED	IPG211111S
176	Industrial Production in the U.S.	Industrial Production: Mining: Oil and gas extraction	FRED	IPG211S
177	Industrial Production in the U.S.	Industrial Production: Mining: Copper, nickel, lead, and zinc mining	FRED	IPG21223S
178	Industrial Production in the U.S.	Industrial Production: Mining: Natural gas	FRED	IPN211111GS
179	Industrial Production in the U.S.	Industrial Production: Mining: Coal mining	FRED	IPN2121S
180	Industrial Production in the U.S.	Industrial Production: Mining: Iron ore mining	FRED	IPN21221S
181	Industrial Production in the U.S.	Industrial Production: Mining: Drilling oil and gas wells	FRED	IPN213111S
182	Economic indicators for the U.S.	University of Michigan: Consumer Sentiment	FRED	UMCSENT
183	Economic indicators for the U.S.	Leading Index for the United States	FRED	USSLIND
184	Economic indicators for the U.S.	NBER based Recession Indicators for the United States	FRED	USREC
185	Economic uncertainty	Policy-related economic uncertainty index	Economic Policy Uncertainty	EPU_Brazil
186	Economic uncertainty	Policy-related economic uncertainty index	Economic Policy Uncertainty	EPU_Canada
187	Economic uncertainty	Policy-related economic uncertainty index	Economic Policy Uncertainty	EPU_France
188	Economic uncertainty	Policy-related economic uncertainty index	Economic Policy Uncertainty	EPU_Ireland
189	Economic uncertainty	Policy-related economic uncertainty index	Economic Policy Uncertainty	EPU_Japan
190	Economic uncertainty	Policy-related economic uncertainty index	Economic Policy Uncertainty	EPU_Korea
191	Economic uncertainty	Policy-related economic uncertainty index	Economic Policy Uncertainty	EPU_US
192	Economic uncertainty	Policy-related economic uncertainty index	Economic Policy Uncertainty	EPU_Sweden
193	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_ARGENTINA
194	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_BRAZIL
195	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_CHINA
196	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_COLOMBIA
197	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_HONG_KONG
198	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_INDIA
199	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_INDONESIA
200	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_ISRAEL
201	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_KOREA
202	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_MALAYSIA
203	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_MEXICO
204	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_PHILIPPINES
205	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_RUSSIA
206	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_SAUDI_ARABIA
207	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_SOUTH_AFRICA
208	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_THAILAND
209	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_TURKEY

Table 5.1 - List of macroeconomic and financial variables (cont.)

210	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_UKRAINE
211	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_VENEZUELA
212	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR
213	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_THREAT
214	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_ACT
215	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_BROAD
216	Economic uncertainty	Geopolitical Risk Index of Caldara and Iacoviello	Geopolitical Risk	GPR_NARROW
217	Leading Indicator	OECD Composite Leading Indicator (CLI) for Australia	OECD	CLI_Australia
218	Leading Indicator	OECD Composite Leading Indicator (CLI) for Austria	OECD	CLI_Austria
219	Leading Indicator	OECD Composite Leading Indicator (CLI) for Belgium	OECD	CLI_Belgium
220	Leading Indicator	OECD Composite Leading Indicator (CLI) for Brazil	OECD	CLI_Brazil
221	Leading Indicator	OECD Composite Leading Indicator (CLI) for Canada	OECD	CLI_Canada
222	Leading Indicator	OECD Composite Leading Indicator (CLI) for Chile	OECD	CLI_Chile
223	Leading Indicator	OECD Composite Leading Indicator (CLI) for China	OECD	CLI_China
224	Leading Indicator	OECD Composite Leading Indicator (CLI) for Denmark	OECD	CLI_Denmark
225	Leading Indicator	OECD Composite Leading Indicator (CLI) for Finland	OECD	CLI_Finland
226	Leading Indicator	OECD Composite Leading Indicator (CLI) for France	OECD	CLI_France
227	Leading Indicator	OECD Composite Leading Indicator (CLI) for Germany	OECD	CLI_Germany
228	Leading Indicator	OECD Composite Leading Indicator (CLI) for Greece	OECD	CLI_Greece
229	Leading Indicator	OECD Composite Leading Indicator (CLI) for Hungary	OECD	CLI_Hungary
230	Leading Indicator	OECD Composite Leading Indicator (CLI) for Ireland	OECD	CLI_Ireland
231	Leading Indicator	OECD Composite Leading Indicator (CLI) for Italy	OECD	CLI_Italy
232	Leading Indicator	OECD Composite Leading Indicator (CLI) for Japan	OECD	CLI_Japan
233	Leading Indicator	OECD Composite Leading Indicator (CLI) for Korea	OECD	CLI_Korea
234	Leading Indicator	OECD Composite Leading Indicator (CLI) for Mexico	OECD	CLI_Mexico
235	Leading Indicator	OECD Composite Leading Indicator (CLI) for Netherlands	OECD	CLI_Netherlands
236	Leading Indicator	OECD Composite Leading Indicator (CLI) for Norway	OECD	CLI_Norway
237	Leading Indicator	OECD Composite Leading Indicator (CLI) for Poland	OECD	CLI_Poland
238	Leading Indicator	OECD Composite Leading Indicator (CLI) for Portugal	OECD	CLI_Portugal
239	Leading Indicator	OECD Composite Leading Indicator (CLI) for Russia	OECD	CLI_Russia
240	Leading Indicator	OECD Composite Leading Indicator (CLI) for South_Africa	OECD	CLI_South_Africa
241	Leading Indicator	OECD Composite Leading Indicator (CLI) for Spain	OECD	CLI_Spain
242	Leading Indicator	OECD Composite Leading Indicator (CLI) for Sweden	OECD	CLI_Sweden
243	Leading Indicator	OECD Composite Leading Indicator (CLI) for Switzerland	OECD	CLI_Switzerland
244	Leading Indicator	OECD Composite Leading Indicator (CLI) for Turkey	OECD	CLI_Turkey
245	Leading Indicator	OECD Composite Leading Indicator (CLI) for United Kingdom	OECD	CLI_UK
246	Leading Indicator	OECD Composite Leading Indicator (CLI) for United States of America	OECD	CLI_USA
247	Leading Indicator	OECD Composite Leading Indicator (CLI) for Euro area (19 countries)	OECD	CLI_Euro_area
248	Leading Indicator	OECD Composite Leading Indicator (CLI) for Big four European	OECD	CLI_Big4_European
249	Leading Indicator	OECD Composite Leading Indicator (CLI) for G7	OECD	CLI_G7
250	Leading Indicator	OECD Composite Leading Indicator (CLI) for NAFTA	OECD	CLI_NAFTA
251	Leading Indicator	OECD Composite Leading Indicator (CLI) for Major five Asia	OECD	CLI_Major5_Asia
252	Leading Indicator	OECD Composite Leading Indicator (CLI) for OECD Europe	OECD	CLI_OECD_Europe
253	Leading Indicator	OECD Composite Leading Indicator (CLI) for OECD Total	OECD	CLI_OECD_Total
254	Leading Indicator	OECD Composite Leading Indicator (CLI) for OECD Major six NME	OECD	CLI_OECD_Major6_NME
255	Real business conditions in the U.S.	Aruoba-Diebold-Scotti Business Conditions Index	Federal Reserve Bank of Philadelphia	ADS_index
256	Quantitative Easing	Total Assets (US$ trillions), Federal Reserve	Federal Reserve Bank of St. Louis	QE_FED
257	Quantitative Easing	Total Assets (US$ trillions), Federal Reserve + European Central Bank + Bank of Japan	Federal Reserve Bank of St. Louis	QE_FED_ECB_BOJ
258	Energy Outlook	Liquid Fuels Consumption, World (million barrels per day)	Short-Term Energy Outlook, U.S. EIA	STEO.PATC_WORLD.M
259	Energy Outlook	Liquid Fuels Consumption, OECD (million barrels per day)	Short-Term Energy Outlook, U.S. EIA	STEO.PATC_OECD.M
260	Energy Outlook	Liquid Fuels Consumption, non-OECD (million barrels per day)	Short-Term Energy Outlook, U.S. EIA	STEO.PATC_NON_OECD.M
261	Energy Outlook	Crude Oil Production Capacity, OPEC (million barrels per day)	Short-Term Energy Outlook, U.S. EIA	STEO.COPC_OPEC.M
262	Energy Outlook	Petroleum Product Supply, Total (million barrels per day)	Short-Term Energy Outlook, U.S. EIA	STEO.PASUPPLY.M
263	Energy Outlook	Crude Oil Production, U.S. (million barrels per day)	Short-Term Energy Outlook, U.S. EIA	STEO.COPRPUS.M
264	Energy Outlook	Crude Oil and Other Liquids Inventory, U.S. (million barrels)	Short-Term Energy Outlook, U.S. EIA	STEO.PASC_US.M
265	Energy Outlook	Petroleum Net Imports, U.S. (million barrels per day)	Short-Term Energy Outlook, U.S. EIA	STEO.PAIMPORT.M
266	Energy Outlook	Net Inventory Withdrawals, Crude Oil and Other Liquids, U.S. (million barrels per day)	Short-Term Energy Outlook, U.S. EIA	STEO.T3_STCHANGE_US.M
267	Energy Outlook	Natural Gas Henry Hub Spot Price, U.S. (dollars per thousand cubic feet)	Short-Term Energy Outlook, U.S. EIA	STEO.NGHHMCF.M
268	Energy Outlook	Cost of Coal Delivered to Electric Generating Plants, U.S. (dollars per million Btu)	Short-Term Energy Outlook, U.S. EIA	STEO.CLEUDUS.M
269	Energy Outlook	Coal Production, U.S. (million short tons)	Short-Term Energy Outlook, U.S. EIA	STEO.CLPRPUS_TON.M
270	Energy Outlook	Coal Consumption, U.S. (million short tons)	Short-Term Energy Outlook, U.S. EIA	STEO.CLTCPUS_TON.M
271	Energy Outlook	Consumption of Electricity, U.S. (billion kilowatthours)	Short-Term Energy Outlook, U.S. EIA	STEO.ELCOTWH.M
272	Energy Outlook	Raw Steel Production, U.S. (million short tons per day)	Short-Term Energy Outlook, U.S. EIA	STEO.RSPRPUS.M
273	Energy Outlook	Aircraft Utilization, U.S. (revenue ton-miles/day thousands)	Short-Term Energy Outlook, U.S. EIA	STEO.RMZZPUS.M
274	Energy Outlook	Vehicle Miles Traveled, U.S. (million miles/day)	Short-Term Energy Outlook, U.S. EIA	STEO.MVVMPUS.M
275	Financial markets	Baltic Exchange Dry Index (BDI)	Thomson Reuters	BALTIC_DRY
276	Financial markets	CBOE SPX VOLATILITY VIX	Thomson Reuters	VIX
277	Financial markets	US Dollar index DXY	Thomson Reuters	US_DOLLAR_INDEX
278	Financial markets	MSCI Emerging Markets U$	Thomson Reuters	MSCI_EM
279	Financial markets	MSCI World U$	Thomson Reuters	MSCI_WORLD
280	Financial markets	EURO STOXX 50	Thomson Reuters	EURO_STOXX50

Table 5.1 - List of macroeconomic and financial variables (cont.)

281	Financial markets	S&P 500 ES ENERGY	Thomson Reuters	SP500_ENERGY
282	Financial markets	S&P GSCI Energy Total Return - RETURN IND. (OFCL)	Thomson Reuters	SP_GSCI_ENERGY
283	Financial markets	CRB BLS Spot Index (1967=100)	Thomson Reuters	CRB
284	Financial markets	CRB BLS Spot Index Raw Industrials	Thomson Reuters	CRB_RAW_IND
285	Financial markets	CRB BLS Spot Index Metals	Thomson Reuters	CRB_METALS
286	Financial markets	CRB BLS Spot Index Foodstuffs	Thomson Reuters	CRB_FOOD
287	Financial markets	CRB BLS Spot Index Fats & Oils	Thomson Reuters	CRB_FATS
288	Financial markets	CRB BLS Spot Index Livestock	Thomson Reuters	CRB_LIVESTOCK
289	Financial markets	CRB BLS Spot Index Textiles	Thomson Reuters	CRB_TEXTI
290	Financial markets	Thomson Reuters Equal Weight Continuous Commodity Index (CCI) Energy 1967 = 100	Thomson Reuters	CCI_ENERGY67
291	Financial markets	Thomson Reuters Equal Weight Continuous Commodity Index (CCI) Energy 1977=100	Thomson Reuters	CCI_ENERGY77
292	Financial markets	Thomson Reuters Equal Weight Continuous Commodity Index (CCI) Industrials	Thomson Reuters	CCI_IND
293	Financial markets	Thomson Reuters Equal Weight Continuous Commodity Index (CCI) Precious Metals	Thomson Reuters	CCI_PREC_METALS
294	Financial markets	Thomson Reuters Equal Weight Continuous Commodity Index (CCI) Grains & Oilseed	Thomson Reuters	CCI_GRAINS
295	Financial markets	Thomson Reuters Equal Weight Continuous Commodity Index (CCI) Interest Rates	Thomson Reuters	CCI_INTEREST
296	Financial markets	Thomson Reuters Equal Weight Continuous Commodity Index (CCI) Livestock Index	Thomson Reuters	CCI_LIVESTOCK
297	Financial markets	Thomson Reuters Equal Weight Continuous Commodity Index (CCI) Softs Index	Thomson Reuters	CCI_SOFT
298	Financial markets	Refinitiv Equal Weight CCI	Thomson Reuters	CCI_REFINITIV
299	Financial markets	Futures Brent crude oil, Intercontinental Exchange (ICE), 1 month	Thomson Reuters	FUTURE_BRENT_M1
300	Financial markets	Futures Brent crude oil, Intercontinental Exchange (ICE), 2 months	Thomson Reuters	FUTURE_BRENT_M2
301	Financial markets	Futures Brent crude oil, Intercontinental Exchange (ICE), 3 months	Thomson Reuters	FUTURE_BRENT_M3
302	Financial markets	Futures Brent crude oil, Intercontinental Exchange (ICE), 4 months	Thomson Reuters	FUTURE_BRENT_M4
303	Financial markets	Futures Brent crude oil, Intercontinental Exchange (ICE), 5 months	Thomson Reuters	FUTURE_BRENT_M5
304	Financial markets	Futures Brent crude oil, Intercontinental Exchange (ICE), 6 months	Thomson Reuters	FUTURE_BRENT_M6
305	Financial markets	Futures Brent crude oil, Intercontinental Exchange (ICE), 7 months	Thomson Reuters	FUTURE_BRENT_M7
306	Financial markets	Futures Brent crude oil, Intercontinental Exchange (ICE), 8 months	Thomson Reuters	FUTURE_BRENT_M8
307	Financial markets	Futures Brent crude oil, Intercontinental Exchange (ICE), 9 months	Thomson Reuters	FUTURE_BRENT_M9
308	Financial markets	Futures Brent crude oil, Intercontinental Exchange (ICE), 10 months	Thomson Reuters	FUTURE_BRENT_M10
309	Financial markets	Futures Brent crude oil, Intercontinental Exchange (ICE), 11 months	Thomson Reuters	FUTURE_BRENT_M11
310	Financial markets	Futures Brent crude oil, Intercontinental Exchange (ICE), 12 months	Thomson Reuters	FUTURE_BRENT_M12
311	Financial markets	Futures Brent crude oil, Intercontinental Exchange (ICE), 24 months	Thomson Reuters	FUTURE_BRENT_M24
312	Financial markets	Futures Brent crude oil, Intercontinental Exchange (ICE), 36 months	Thomson Reuters	FUTURE_BRENT_M36
313	Financial markets	Futures Brent crude oil, Intercontinental Exchange (ICE), 48 months	Thomson Reuters	FUTURE_BRENT_M48
314	Financial markets	Futures Brent crude oil, Intercontinental Exchange (ICE), 60 months	Thomson Reuters	FUTURE_BRENT_M60
315	Financial markets	Futures Brent crude oil, Intercontinental Exchange (ICE), 72 months	Thomson Reuters	FUTURE_BRENT_M72

Appendix 6. Point forecast results

Table 6.1 - Monthly Frequency - *Root Mean Squared Error (RMSE)*

	h = 1	h = 2	h = 3	h = 4	h = 5	h = 6	h = 7	h = 8	h = 9	h = 10	h = 11	h = 12	h = 13	h = 14	h = 15	h = 16	h = 17	h = 18	h = 19	h = 20	h = 21	h = 22	h = 23	h = 24
(1) RW	6.574	10.637	13.723	16.088	17.968	19.432	20.546	21.448	22.184	22.771	23.332	23.995	24.633	25.137	25.470	25.921	26.486	27.103	27.628	28.141	28.594	29.177	29.888	30.485
(2) RW-drift	6.618	10.763	13.972	16.479	18.516	20.146	21.425	22.495	23.394	24.125	24.839	25.689	26.532	27.250	27.821	28.510	29.311	30.157	30.942	31.738	32.465	33.339	34.372	35.333
(3) RW-drift5	6.696	11.040	14.474	17.232	19.544	21.423	22.954	24.219	25.240	26.054	26.813	27.716	28.622	29.408	30.053	30.803	31.639	32.511	33.330	34.201	35.048	36.045	37.167	38.260
(4) ARIMA	6.550	10.639	13.764	16.141	18.001	19.466	20.582	21.473	22.204	22.761	23.328	23.969	24.615	25.134	25.466	25.892	26.439*	27.064	27.586*	28.100*	28.539*	29.117*	29.846	30.466
(5) Factor model1	6.006***	10.119**	13.439**	16.034	17.951	19.119	20.516	20.894	21.540	21.912	22.229	22.800*	23.577	23.893*	25.459	24.966	25.340*	27.286	27.731	28.419	29.088	29.558	30.067	31.038
(6) Factor model2	5.625**	10.259*	13.788	15.775	17.437	19.124	20.070	21.332	21.643	22.472	23.292	23.794	24.920	25.654	26.634	27.251	28.161	28.674	29.010	29.320	29.919	30.979	31.833	33.357
(7) Elastic net	5.192***	10.019***	12.419*	14.868	16.051	18.035	18.936	20.240	22.400	25.046	27.273	26.486	25.484	27.490	27.532	27.887	29.507	29.892	31.717	33.419	36.368	39.392	40.308	40.171
(8) LASSO	5.221***	9.827***	12.447*	14.905	16.089	17.929	18.934	20.226	22.455	25.021	27.251	26.535	24.830	27.424	27.652	28.143	29.562	30.027	31.884	33.832	36.548	39.567	39.702	40.037
(9) Adalasso	5.174***	9.859**	12.454*	14.885	16.854	18.184	18.951	20.348	23.751	24.315	26.098	23.958	25.811	26.405	27.891	30.103	31.924	33.159	31.488	31.572	33.664	34.845	35.304	37.235
(10) Ridge regression	5.754***	10.041***	12.935**	15.555	16.922	17.970	18.795	19.982	21.024	22.146	23.255	23.489	24.340	25.389	25.881	26.400	26.905	27.552	28.385	29.278	30.472	31.374	32.370	33.386
(11) Random forest	5.710***	10.107**	13.258**	15.677	17.639	19.117	20.310	21.140	21.991	22.727	23.446	24.385	24.940	25.720	26.382	27.137	28.080	29.000	30.089	31.202	32.237	33.322	34.383	35.487
(12) Quant.reg.forest	5.742***	10.181**	13.375**	15.702	17.688	19.078	20.307	21.140	21.948	22.611	23.356	24.278	24.947	25.640	26.232	26.856	27.923	28.764	29.975	31.017	32.173	33.208	34.270	35.373
(13) XGBoost	5.741***	10.343***	13.593***	16.064**	17.437**	19.085**	20.313*	21.201*	21.899	22.760	23.503	24.394	25.092	25.540	26.247	27.136	28.355	29.180	30.092	31.193	32.578	33.359	34.258	35.174
(14) AF	9.489	11.896	13.897	15.576	16.824	17.814	18.593	19.274	19.828	20.253	20.584	20.946*	21.246*	21.408*	21.465*	21.699*	22.041*	22.392*	22.620*	22.800*	22.934	23.130	23.391*	23.605*
(15) BCAF	9.376	11.737	13.723	15.412	16.665	17.665	18.458	19.183	19.800	20.284*	20.663*	21.060**	21.406**	21.588***	21.687***	21.970**	22.381**	22.826***	23.156***	23.469***	23.690***	23.968**	24.341**	24.678**
(16) Brent futures	5.210***	10.035***	13.402***	15.945**	17.826**	19.229**	20.107***	20.747***	21.155***	21.561***	21.841***	22.124***	22.490***	22.716***	22.778***	22.847***	23.107***	23.419***	23.704***	24.000***	24.308***	24.541***	24.864***	25.167***
(17) Schwartz-Smith mean	5.258***	10.003***	13.306***	15.834***	17.737**	19.235*	20.188**	20.962**	21.570**	22.107**	22.504**	22.861***	23.336***	23.668***	23.835***	24.037***	24.418***	24.907***	25.381***	25.871***	26.339***	26.760***	27.275***	27.819***
(18) Schwartz-Smith median	5.232***	9.930***	13.166***	15.626***	17.458***	18.888***	19.784***	20.485***	21.052***	21.542***	21.858***	22.151***	22.559***	22.815***	22.906***	23.043***	23.345***	23.778***	24.140***	24.536***	24.944***	25.269***	25.708***	26.171***
(19) Mean all	5.492***	9.860***	12.803**	15.084	16.694	17.914	18.912	19.689	20.238	20.915	21.642	22.023*	22.454*	23.057*	23.457*	23.921*	24.400*	25.087*	25.765*	26.662	27.969	28.942	29.572	30.223
(20) Median all	5.404***	9.972***	12.920**	15.318*	17.106*	18.566	19.648*	20.429*	20.936*	21.613	22.116	22.589*	23.245*	23.797*	24.286	24.695*	25.389*	26.183	26.899	27.831	28.989	29.811	30.741	31.512
(21) Mean selection	5.488***	9.747***	12.711**	14.934	16.630*	17.768	18.639	19.264	19.852*	20.249	20.872	21.165**	21.505**	22.049**	22.402**	22.537**	22.668***	23.432**	23.890***	24.638**	25.876**	26.776**	27.361***	27.865**
(22) Median selection	5.302***	9.821***	12.807**	15.109**	17.066**	18.484*	19.496**	20.235**	20.735**	21.507*	21.834**	22.023**	22.626***	22.943***	23.225***	23.370***	23.828***	24.665***	25.032***	25.575***	26.128***	26.876***	27.377***	28.066***
number of observations	*174*	*173*	*172*	*171*	*170*	*169*	*168*	*167*	*166*	*165*	*164*	*163*	*162*	*161*	*160*	*159*	*158*	*157*	*156*	*155*	*154*	*153*	*152*	*151*
best model	*9*	*21*	*7*	*7*	*7*	*15*	*15*	*15*	*15*	*21*	*14*	*14*	*14*	*14*	*14*	*14*	*14*	*14*	*14*	*14*	*14*	*14*	*14*	*14*
R2 oos (%)	*38*	*16*	*18*	*14*	*20*	*17*	*19*	*20*	*20*	*20*	*22*	*23*	*25*	*27*	*28*	*29*	*30*	*31*	*32*	*34*	*35*	*37*	*38*	*40*

Notes: Yellow cells indicate the Top5 best models (lower RMSEs) in each horizon. ***, **, * indicate rejection at 1%, 5% and 10% levels, respectively, using the Clark and West (2007) test. The benchmark is model 1 (random walk without drift). Forecast combinations 19 and 20 are based on models 1-18, whereas combinations 21 and 22 are based on selected models from each class (models 1, 5, 9, 12, 14, 16 and 18).

Table 6.2 - Quarterly Frequency - *Root Mean Squared Error (RMSE)*

	h = 1	h = 2	h = 3	h = 4	h = 5	h = 6	h = 7	h = 8	h = 9	h = 10	h = 11	h = 12	h = 13	h = 14	h = 15	h = 16	h = 17	h = 18	h = 19	h = 20
(1) RW	11.010	17.626	20.970	22.959	24.546	26.027	27.727	29.655	31.077	31.929	31.918	32.539	33.496	34.375	35.305	36.055	36.314	36.860	37.527	38.086
(2) RW-drift	11.225	18.295	22.159	24.664	26.893	29.058	31.553	34.410	36.947	39.009	40.513	42.767	45.346	47.875	50.645	53.545	56.187	59.244	62.787	66.297
(3) RW-drift5	11.648	19.562	24.101	26.865	29.273	31.531	34.191	37.347	40.144	42.393	44.261	46.878	49.744	53.160	57.397	62.337	67.467	73.440	80.003	86.501
(4) ARIMA	11.360	17.922	21.355	23.495	25.595	27.318	29.025	31.242	32.967	34.521	34.478	35.526	36.586	37.367	38.471	39.306	39.606	40.393	41.282	42.133
(5) Factor model1	10.562**	17.870	19.857	21.723	26.353	26.572	32.661	33.511	36.023	36.069	35.151	35.046	34.856	34.936	36.805	40.552	40.320	40.245	36.755*	44.298
(6) Factor model2	10.594*	16.789*	21.734	25.321	27.362	30.297	31.559	37.452	33.634	38.329	41.611	42.719	40.945	45.119	48.836	56.459	55.433	58.385	68.132	75.502
(7) Elastic net	10.345***	18.800	25.730	25.299	29.420	33.463	34.778	30.989	31.186	35.253	40.536	46.638	61.136	55.895	55.571	58.948	56.756	53.273	63.344	82.425
(8) LASSO	11.734	18.542	25.831	25.703	29.238	37.729	33.494	30.915	31.458	34.625	40.713	46.553	60.870	55.079	57.347	61.238	57.654	52.730	62.415	84.899
(9) Adalasso	11.142	17.468**	24.904	24.642	30.573	35.128	51.256	32.231	30.938	39.133	52.071	59.314	92.263	57.752	56.692	51.161	51.321	41.185	60.583	79.353
(10) Ridge regression	10.829**	18.559	21.860	24.789	32.330	42.588	31.844	31.916	35.854	41.270	36.270	39.382	50.096	48.837	49.630	50.646	50.751	58.641	57.971	66.150
(11) Random forest	10.278**	18.267	21.902	23.652	25.952	27.650	30.776	33.200	35.534	37.697	38.906	40.576	42.299	44.527	47.685	49.966	51.880	52.932	53.702	55.411
(12) Quant.reg.forest	10.379**	18.198	21.732	23.622	25.699	27.420	30.566	33.414	35.898	37.853	39.474	40.837	42.491	44.848	47.696	49.919	51.609	52.676	53.096	55.193
(13) XGBoost	10.383***	19.043	23.204	23.411	25.894	27.005	30.041	31.895	34.872	36.706	37.797	40.643	41.553	43.271	48.129	49.929	49.859	50.842	49.185	54.485
(14) AF	11.568	16.957	19.594*	21.020*	22.125**	23.103*	24.353*	25.437*	26.071*	26.204**	25.638**	25.902**	26.475**	27.072**	27.641**	28.133**	28.331**	29.125*	29.822**	30.603**
(15) BCAF	11.392	16.972*	19.952*	21.746*	23.113**	24.220**	25.535**	26.511**	27.141**	27.022**	26.441**	26.800***	27.416**	28.177***	29.423***	30.671***	31.717**	33.484*	35.731	37.408
(16) Brent futures	10.130***	18.126	20.612	21.603***	22.314***	22.732***	23.934***	24.663**	25.219**	25.097**	24.105**	23.615***	23.913**	24.458**	25.186**	26.166**	26.672**	27.733**	28.971*	30.452
(17) Schwartz-Smith mean	9.982***	18.096	20.940	22.267*	23.408**	24.259**	26.021**	27.315**	28.467**	29.106**	29.040**	29.743**	30.771**	32.093**	33.241**	34.711*	35.539	36.900	38.308	39.978
(18) Schwartz-Smith median	9.924***	17.783	20.457*	21.658***	22.557***	23.206***	24.715***	25.754**	26.642**	26.879**	26.327**	26.500***	27.122***	28.065**	28.779**	29.834**	30.224**	31.012**	31.820**	33.012**
(19) Mean all	10.025**	17.282**	20.373*	22.395*	24.435	25.758	27.884	28.989	30.145	32.115	32.936	34.079	35.477	37.725	39.342	40.875	41.461	41.568	44.799	49.150
(20) Median all	10.083***	17.544**	20.597**	22.767	24.926	26.243	28.473	29.201	30.148	33.519	34.339	36.080	36.842	39.514	40.604	41.934	43.642	43.610	46.498	50.491
(21) Mean selection	9.892***	17.007**	19.671*	21.310**	23.434**	24.451**	26.715*	27.992**	28.977**	30.491***	30.578**	30.643*	31.415***	33.151*	34.624	35.299	35.168*	33.972**	36.136	38.829
(22) Median selection	9.969***	16.903**	20.263***	21.867***	23.669***	24.615**	26.647**	28.208**	28.885***	30.122***	29.507***	28.557***	29.012***	30.329***	31.832**	33.973**	33.617**	34.514*	34.320**	35.920*
number of observations	*58*	*57*	*56*	*55*	*54*	*53*	*52*	*51*	*50*	*49*	*48*	*47*	*46*	*45*	*44*	*43*	*42*	*41*	*40*	*39*
best model	*21*	*6*	*14*	*14*	*14*	*16*	*16*	*16*	*16*	*16*	*16*	*16*	*16*	*16*	*16*	*16*	*16*	*16*	*16*	*16*
R2 oos (%)	*19*	*9*	*12*	*16*	*18*	*23*	*25*	*30*	*34*	*38*	*42*	*47*	*49*	*49*	*49*	*47*	*46*	*43*	*40*	*36*

Notes: Yellow cells indicate the Top5 best models (lower RMSEs) in each horizon. ***, **, * indicate rejection at 1%, 5% and 10% levels, respectively, using the Clark and West (2007) test. The benchmark is model 1 (random walk without drift). Forecast combinations 19 and 20 are based on models 1-18, whereas combinations 21 and 22 are based on selected models from each class (models 1, 5, 9, 12, 14, 16 and 18).

Appendix 7. Density forecast results

Table 7.1 - Monthly Frequency - *Empirical Coverage Rate*

	h = 1	h = 2	h = 3	h = 4	h = 5	h = 6	h = 7	h = 8	h = 9	h = 10	h = 11	h = 12	h = 13	h = 14	h = 15	h = 16	h = 17	h = 18	h = 19	h = 20	h = 21	h = 22	h = 23	h = 24
(1) RW	0.85	0.87	0.87	0.86	0.88	0.86	0.87	0.85	0.88	0.90	0.90	0.90	0.90	0.88	0.88	0.85	0.85	0.86	0.86	0.83	0.83	0.83	0.82	0.82
(2) RW-drift	0.85	0.86	0.88	0.88	0.89	0.89	0.89	0.88	0.89	0.89	0.90	0.90	0.88	0.87	0.87	0.86	0.85	0.85	0.86	0.82	0.81	0.81	0.80	0.79
(3) RW-drift5	0.85	0.88	0.90	0.91	0.90	0.90	0.90	0.89	0.90	0.90	0.91	0.90	0.90	0.89	0.89	0.87	0.86	0.86	0.86	0.83	0.83	0.83	0.83	0.77
(4) ARIMA	0.84	0.85	0.87	0.88	0.88	0.87	0.88	0.86	0.90	0.89	0.90	0.91	0.90	0.86	0.87	0.87	0.87	0.86	0.85	0.84	0.83	0.81	0.82	0.83
(5) Factor model1	0.90	0.87	0.86	0.87	0.84	0.84	0.88	0.79	0.84	0.79	0.78	0.76	0.76	0.76	0.75	0.75	0.76	0.73	0.72	0.73	0.72	0.74	0.72	0.73
(6) Factor model2	0.89	0.88	0.85	0.86	0.85	0.87	0.91	0.91	0.87	0.86	0.88	0.87	0.81	0.83	0.83	0.81	0.81	0.80	0.80	0.76	0.77	0.76	0.76	0.73
(7) Elastic net	0.96	0.87	0.85	0.82	0.80	0.83	0.83	0.78	0.77	0.74	0.75	0.79	0.76	0.76	0.81	0.79	0.83	0.82	0.82	0.79	0.77	0.76	0.73	0.75
(8) LASSO	0.95	0.87	0.85	0.82	0.80	0.81	0.83	0.81	0.79	0.76	0.77	0.79	0.76	0.74	0.80	0.79	0.83	0.83	0.81	0.79	0.78	0.75	0.73	0.75
(9) Adalasso	0.97	0.89	0.87	0.82	0.82	0.85	0.84	0.83	0.76	0.73	0.74	0.78	0.73	0.78	0.76	0.75	0.73	0.73	0.74	0.77	0.76	0.75	0.73	0.67
(10) Ridge regression	0.90	0.87	0.88	0.85	0.84	0.86	0.87	0.84	0.85	0.86	0.81	0.80	0.80	0.78	0.78	0.75	0.74	0.72	0.73	0.71	0.71	0.72	0.69	0.71
(11) Random forest	0.92	0.87	0.87	0.87	0.89	0.88	0.92	0.87	0.89	0.87	0.89	0.83	0.82	0.84	0.86	0.83	0.83	0.84	0.82	0.79	0.79	0.81	0.80	0.79
(12) Quant.reg.forest	0.90	0.86	0.88	0.86	0.88	0.88	0.91	0.87	0.87	0.87	0.89	0.82	0.82	0.84	0.86	0.83	0.83	0.84	0.82	0.80	0.80	0.80	0.80	0.79
(13) XGBoost	0.90	0.87	0.87	0.85	0.89	0.88	0.91	0.85	0.87	0.87	0.85	0.81	0.82	0.82	0.84	0.82	0.83	0.81	0.82	0.79	0.79	0.78	0.78	0.78
(14) AF	0.61	0.76	0.83	0.85	0.86	0.88	0.92	0.89	0.92	0.92	0.94	0.94	0.90	0.92	0.92	0.88	0.90	0.92	0.93	0.90	0.90	0.93	0.93	0.94
(15) BCAF	0.62	0.78	0.83	0.85	0.86	0.87	0.92	0.89	0.92	0.90	0.88	0.90	0.90	0.89	0.89	0.87	0.87	0.86	0.87	0.86	0.83	0.84	0.85	0.83
(16) Brent futures	0.93	0.90	0.88	0.88	0.88	0.87	0.91	0.92	0.90	0.90	0.88	0.89	0.91	0.88	0.87	0.87	0.87	0.85	0.84	0.83	0.83	0.83	0.82	0.82
(17) Schwartz-Smith mean	0.92	0.90	0.89	0.88	0.88	0.86	0.87	0.86	0.83	0.82	0.81	0.80	0.80	0.80	0.83	0.83	0.83	0.85	0.84	0.86	0.83	0.86	0.85	0.86
(18) Schwartz-Smith median	0.92	0.90	0.89	0.88	0.88	0.86	0.87	0.86	0.83	0.82	0.81	0.80	0.80	0.80	0.83	0.83	0.83	0.85	0.84	0.86	0.83	0.86	0.85	0.86
(19) Mean all	0.92	0.88	0.88	0.87	0.87	0.88	0.91	0.89	0.90	0.87	0.89	0.88	0.88	0.85	0.86	0.85	0.84	0.85	0.84	0.82	0.81	0.81	0.80	0.80
(20) Median all	0.92	0.87	0.88	0.88	0.88	0.88	0.92	0.92	0.90	0.90	0.89	0.91	0.88	0.87	0.87	0.85	0.85	0.85	0.84	0.82	0.81	0.81	0.80	0.82
(21) Mean selection	0.92	0.88	0.88	0.87	0.87	0.88	0.91	0.91	0.90	0.89	0.90	0.89	0.88	0.88	0.86	0.86	0.85	0.84	0.85	0.82	0.81	0.81	0.82	0.80
(22) Median selection	0.92	0.88	0.88	0.88	0.88	0.89	0.92	0.92	0.90	0.90	0.89	0.90	0.88	0.86	0.89	0.87	0.87	0.85	0.84	0.83	0.81	0.81	0.81	0.83

Notes: The nominal coverage rate is 90%. The closer the empirical coverage rate is to 90% (green cells) the better is the fit of the density forecast in respect to observed data.

Table 7.2 - Monthly Frequency - *Interval Score*

	h = 1	h = 2	h = 3	h = 4	h = 5	h = 6	h = 7	h = 8	h = 9	h = 10	h = 11	h = 12	h = 13	h = 14	h = 15	h = 16	h = 17	h = 18	h = 19	h = 20	h = 21	h = 22	h = 23	h = 24
(1) RW	30.2	48.3	62.9	82.3	106.4	151.4	174.7	196.5	214.1	236.4	319.8	285.9	309.8	326.7	346.0	347.7	343.2	344.2	343.6	313.5	287.7	255.8	216.7	208.4
(2) RW-drift	44.9	61.7	70.3	80.5	100.1	125.6	144.0	158.4	178.4	205.9	267.5	248.5	266.1	282.8	301.8	320.7	338.5	364.8	361.7	363.6	365.7	353.6	351.2	346.4
(3) RW-drift5	36.4	54.2	65.8	79.6	107.1	136.9	278.9	338.4	380.2	436.0	528.4	541.7	584.6	623.7	665.1	664.6	669.8	676.5	681.6	658.2	629.1	577.3	481.1	382.0
(4) ARIMA	31.1	49.8	65.7	93.6	119.9	150.1	169.8	189.5	213.9	242.0	322.3	286.0	300.2	309.1	321.8	314.4	316.4	339.2	303.9	291.3	263.5	241.3	225.3	216.1
(5) Factor model1	51.7	67.1	75.9	78.6	86.4	92.1	97.3	101.0	129.4	120.3	124.9	135.5	142.8	148.8	161.0	173.1	200.5	394.7	247.1	250.5	259.1	266.8	277.0	275.4
(6) Factor model2	34.2	60.6	89.6	130.5	154.8	181.1	202.7	215.0	228.5	242.2	398.6	257.4	267.0	266.8	274.6	259.2	261.9	261.4	242.6	231.8	211.1	196.2	207.6	186.7
(7) Elastic net	58.6	81.3	95.8	109.9	129.8	150.4	176.2	196.6	220.0	231.9	290.4	276.8	396.3	306.8	323.7	344.0	363.6	404.2	417.0	435.8	453.2	441.3	446.2	449.6
(8) LASSO	60.4	79.2	86.7	90.2	97.4	116.7	145.4	142.3	171.8	185.6	210.3	186.3	214.8	198.8	209.1	245.8	294.1	304.6	330.3	357.6	389.6	414.1	420.8	434.8
(9) Adalasso	49.3	71.4	83.9	95.3	100.0	109.2	121.8	132.5	193.9	168.4	168.8	147.7	166.7	166.6	168.4	171.4	182.0	190.4	193.8	221.2	258.5	280.8	298.8	330.8
(10) Ridge regression	27.7	46.9	59.8	75.5	81.4	89.9	111.4	112.4	122.2	192.3	234.3	155.5	166.3	191.1	185.5	304.4	203.8	208.3	220.6	240.4	248.3	253.0	271.0	278.3
(11) Random forest	41.0	59.4	88.2	112.5	133.8	156.7	173.0	188.4	203.0	223.9	251.4	249.4	259.9	269.2	286.1	284.0	288.3	301.4	325.8	301.4	291.6	295.1	281.5	282.3
(12) Quant.reg.forest	33.3	53.4	171.3	227.9	278.8	337.5	370.3	387.1	406.1	423.4	452.1	446.6	455.3	446.7	446.9	426.2	422.4	410.6	397.7	307.8	298.6	336.7	244.8	244.7
(13) XGBoost	34.2	54.9	81.0	104.4	113.1	136.3	184.6	169.6	185.6	205.9	251.3	222.9	229.4	235.0	303.8	248.1	254.3	264.2	309.3	248.9	241.1	241.6	230.4	229.9
(14) AF	57.3	56.1	76.9	97.0	111.6	131.3	140.7	158.6	172.0	188.1	215.9	201.0	200.8	200.5	197.3	194.7	186.9	180.4	168.1	148.2	132.8	122.0	110.8	102.9
(15) BCAF	55.9	55.2	76.6	97.9	113.2	132.8	144.6	160.0	175.8	194.0	221.9	207.4	207.2	209.7	207.2	204.0	195.3	190.1	178.7	158.0	144.5	135.8	125.9	120.6
(16) Brent futures	33.7	56.2	75.8	101.4	139.9	174.6	195.9	226.9	250.7	275.9	313.4	319.2	336.5	337.1	384.1	318.1	317.1	309.1	320.6	284.7	259.4	222.9	195.4	186.7
(17) Schwartz-Smith mean	28.8	51.5	67.0	77.5	86.5	95.4	99.2	103.0	106.6	109.6	110.8	110.9	111.4	111.5	111.0	112.4	113.0	115.6	116.9	119.4	120.1	120.8	121.5	122.4
(18) Schwartz-Smith median	28.8	51.5	67.0	77.5	86.5	95.4	99.2	103.0	106.6	109.6	110.8	110.9	111.4	111.5	111.0	112.4	113.0	115.6	116.9	119.4	120.1	120.8	121.5	122.4
(19) Mean all	87.6	121.9	145.2	158.4	183.7	203.7	215.7	222.9	229.6	236.8	258.5	222.2	226.1	231.4	235.2	236.7	237.7	240.1	238.9	239.7	243.7	234.2	237.4	239.0
(20) Median all	29.4	50.2	62.5	83.3	103.9	128.9	141.0	158.7	175.4	230.6	241.7	226.4	238.9	258.7	268.0	270.8	272.5	278.5	291.4	281.4	274.5	259.9	256.3	247.4
(21) Mean selection	28.2	50.5	77.0	167.0	210.9	242.7	259.7	267.4	273.9	296.6	424.7	290.1	290.1	293.7	279.3	272.5	282.4	270.0	263.3	249.1	222.8	187.3	201.9	195.8
(22) Median selection	23.0	46.4	63.0	84.0	111.1	135.2	150.6	168.3	186.8	222.2	232.9	235.9	247.0	253.5	254.8	266.9	251.4	253.4	285.7	228.4	212.4	188.9	169.8	157.5

Note: A lower score implies a better interval forecast. Yellow cells indicate the Top5 best models in each horizon.

Table 7.3 - Monthly Frequency - *Log Predictive Density Score (LPDS)*

	h = 1	h = 2	h = 3	h = 4	h = 5	h = 6	h = 7	h = 8	h = 9	h = 10	h = 11	h = 12	h = 13	h = 14	h = 15	h = 16	h = 17	h = 18	h = 19	h = 20	h = 21	h = 22	h = 23	h = 24
(1) RW	-3.42	-3.92	-4.21	-4.40	-4.44	-4.57	-4.59	-4.59	-4.56	-4.61	-4.70	-4.71	-4.74	-4.75	-4.75	-4.78	-4.82	-4.88	-4.90	-4.92	-4.89	-4.88	-4.86	-4.83
(2) RW-drift	-3.46	-3.96	-4.23	-4.40	-4.45	-4.59	-4.61	-4.53	-4.58	-4.65	-4.76	-4.76	-4.80	-4.82	-4.84	-4.87	-4.93	-5.00	-5.03	-5.06	-5.06	-5.06	-5.06	-5.08
(3) RW-drift5	-3.47	-3.95	-4.21	-4.33	-4.36	-4.43	-4.54	-4.61	-4.61	-4.67	-4.77	-4.79	-4.83	-4.87	-4.91	-4.95	-5.01	-5.09	-5.11	-5.15	-5.17	-5.20	-5.24	-5.31
(4) ARIMA	-3.41	-3.92	-4.22	-4.38	-4.46	-4.56	-4.57	-4.54	-4.56	-4.63	-4.73	-4.73	-4.76	-4.77	-4.77	-4.79	-4.83	-4.90	-4.92	-4.94	-4.90	-4.89	-4.88	-4.86
(5) Factor model1	-3.30	-3.87	-4.19	-4.45	-4.50	-4.54	-4.56	-4.62	-4.65	-4.81	-4.84	-4.98	-5.07	-5.11	-5.19	-5.14	-5.24	-5.45	-5.57	-5.60	-5.64	-5.66	-5.72	-5.77
(6) Factor model2	-3.18	-3.87	-4.28	-4.42	-4.46	-4.45	-4.46	-4.50	-4.55	-4.63	-4.74	-4.75	-4.86	-4.86	-4.88	-4.90	-4.94	-4.98	-4.99	-5.02	-5.02	-5.01	-5.06	-5.12
(7) Elastic net	-3.15	-3.79	-4.13	-4.47	-4.57	-4.79	-4.93	-5.25	-5.50	-5.87	-5.87	-5.22	-5.08	-5.06	-4.98	-5.02	-5.03	-5.04	-5.10	-5.23	-5.34	-5.42	-5.54	-5.62
(8) LASSO	-3.16	-3.80	-4.14	-4.46	-4.55	-4.78	-4.86	-5.04	-5.43	-5.73	-5.66	-5.18	-5.09	-5.05	-5.00	-5.04	-5.05	-5.03	-5.11	-5.21	-5.34	-5.45	-5.53	-5.73
(9) Adalasso	-3.21	-3.83	-4.15	-4.44	-4.60	-4.75	-4.79	-4.98	-5.68	-5.82	-5.84	-5.53	-6.01	-5.94	-5.65	-5.60	-5.61	-5.44	-5.37	-5.58	-6.06	-6.35	-6.51	-7.01
(10) Ridge regression	-3.20	-3.77	-4.17	-4.47	-4.50	-4.53	-4.62	-4.65	-4.72	-4.84	-5.00	-5.05	-5.12	-5.23	-5.28	-5.35	-5.42	-5.56	-5.67	-5.80	-5.95	-6.08	-6.35	-6.64
(11) Random forest	-3.23	-3.85	-4.20	-4.35	-4.37	-4.42	-4.42	-4.47	-4.51	-4.59	-4.67	-4.73	-4.76	-4.78	-4.78	-4.81	-4.86	-4.91	-4.95	-4.98	-4.99	-4.98	-4.97	-4.98
(12) Quant.reg.forest	-3.22	-3.86	-4.21	-4.35	-4.38	-4.42	-4.43	-4.49	-4.52	-4.60	-4.68	-4.74	-4.77	-4.79	-4.78	-4.81	-4.86	-4.91	-4.95	-4.98	-4.99	-4.99	-4.97	-4.98
(13) XGBoost	-3.21	-3.83	-4.21	-4.34	-4.32	-4.40	-4.47	-4.48	-4.52	-4.62	-4.75	-4.79	-4.81	-4.82	-4.82	-4.86	-4.90	-4.95	-4.96	-5.01	-5.01	-5.02	-5.02	-5.01
(14) AF	-4.24	-4.14	-4.29	-4.40	-4.43	-4.45	-4.41	-4.42	-4.44	-4.48	-4.55	-4.55	-4.57	-4.55	-4.54	-4.56	-4.59	-4.62	-4.62	-4.62	-4.58	-4.56	-4.53	-4.50
(15) BCAF	-4.21	-4.12	-4.29	-4.41	-4.44	-4.46	-4.43	-4.43	-4.46	-4.50	-4.58	-4.57	-4.59	-4.57	-4.56	-4.57	-4.60	-4.64	-4.65	-4.65	-4.62	-4.59	-4.57	-4.56
(16) Brent futures	-3.03	-3.77	-4.10	-4.28	-4.34	-4.38	-4.38	-4.48	-4.50	-4.51	-4.57	-4.61	-4.64	-4.66	-4.66	-4.65	-4.67	-4.72	-4.76	-4.77	-4.74	-4.72	-4.69	-4.66
(17) Schwartz-Smith mean	-3.06	-3.79	-4.11	-4.31	-4.34	-4.39	-4.40	-4.43	-4.49	-4.55	-4.61	-4.66	-4.68	-4.69	-4.70	-4.69	-4.73	-4.79	-4.84	-4.86	-4.84	-4.83	-4.81	-4.79
(18) Schwartz-Smith median	-3.05	-3.77	-4.09	-4.28	-4.32	-4.36	-4.37	-4.41	-4.47	-4.53	-4.59	-4.63	-4.65	-4.66	-4.67	-4.66	-4.68	-4.74	-4.78	-4.79	-4.76	-4.75	-4.72	-4.70
(19) Mean all	-3.17	-3.82	-4.15	-4.31	-4.35	-4.38	-4.37	-4.41	-4.44	-4.49	-4.59	-4.58	-4.60	-4.61	-4.61	-4.61	-4.66	-4.72	-4.75	-4.78	-4.81	-4.78	-4.79	-4.79
(20) Median all	-3.14	-3.82	-4.13	-4.32	-4.37	-4.45	-4.39	-4.43	-4.48	-4.55	-4.63	-4.64	-4.67	-4.69	-4.69	-4.70	-4.73	-4.78	-4.81	-4.84	-4.85	-4.80	-4.81	-4.79
(21) Mean selection	-3.13	-3.80	-4.13	-4.31	-4.35	-4.36	-4.34	-4.37	-4.40	-4.46	-4.54	-4.53	-4.54	-4.54	-4.55	-4.53	-4.57	-4.61	-4.65	-4.68	-4.70	-4.66	-4.66	-4.66
(22) Median selection	-3.08	-3.79	-4.10	-4.30	-4.36	-4.39	-4.37	-4.42	-4.47	-4.55	-4.61	-4.62	-4.66	-4.65	-4.64	-4.66	-4.68	-4.73	-4.76	-4.78	-4.77	-4.72	-4.70	-4.69

Note: A higher score implies a better density forecast. Yellow cells indicate the Top5 best models in each horizon.

Table 7.4 - Quarterly Frequency - *Empirical Coverage Rate*

	h = 1	h = 2	h = 3	h = 4	h = 5	h = 6	h = 7	h = 8	h = 9	h = 10	h = 11	h = 12	h = 13	h = 14	h = 15	h = 16	h = 17	h = 18	h = 19	h = 20
(1) RW	0.73	0.81	0.79	0.85	0.82	0.82	0.79	0.76	0.71	0.58	0.64	0.76	0.65	0.81	0.83	0.88	0.91	0.94	0.87	0.90
(2) RW-drift	0.76	0.77	0.81	0.85	0.82	0.77	0.72	0.69	0.61	0.55	0.51	0.50	0.51	0.47	0.54	0.50	0.48	0.44	0.29	0.27
(3) RW-drift5	0.84	0.83	0.81	0.87	0.82	0.75	0.72	0.67	0.63	0.58	0.54	0.55	0.51	0.47	0.46	0.35	0.30	0.31	0.26	0.17
(4) ARIMA	0.57	0.81	0.79	0.85	0.89	0.86	0.88	0.83	0.83	0.63	0.77	0.84	0.73	0.75	0.74	0.79	0.85	0.84	0.81	0.77
(5) Factor model1	0.67	0.73	0.83	0.74	0.62	0.64	0.60	0.50	0.51	0.43	0.49	0.45	0.41	0.47	0.49	0.47	0.36	0.38	0.42	0.23
(6) Factor model2	0.69	0.81	0.85	0.83	0.82	0.82	0.74	0.55	0.76	0.60	0.69	0.66	0.65	0.53	0.49	0.44	0.45	0.28	0.19	0.07
(7) Elastic net	0.78	0.73	0.70	0.74	0.69	0.80	0.79	0.60	0.61	0.68	0.59	0.58	0.46	0.50	0.51	0.53	0.45	0.34	0.23	0.13
(8) LASSO	0.71	0.73	0.66	0.70	0.69	0.80	0.81	0.57	0.59	0.63	0.59	0.53	0.43	0.50	0.46	0.53	0.42	0.31	0.23	0.13
(9) Adalasso	0.67	0.71	0.57	0.54	0.62	0.70	0.65	0.55	0.61	0.70	0.56	0.58	0.46	0.50	0.49	0.50	0.33	0.53	0.39	0.17
(10) Ridge regression	0.73	0.73	0.70	0.61	0.53	0.66	0.51	0.57	0.49	0.45	0.54	0.42	0.30	0.39	0.40	0.44	0.39	0.31	0.13	0.13
(11) Random forest	0.73	0.81	0.81	0.80	0.80	0.80	0.74	0.71	0.61	0.60	0.62	0.61	0.51	0.47	0.51	0.47	0.39	0.41	0.29	0.23
(12) Quant.reg.forest	0.73	0.79	0.81	0.83	0.82	0.80	0.77	0.71	0.61	0.60	0.62	0.61	0.51	0.47	0.51	0.47	0.42	0.41	0.32	0.27
(13) XGBoost	0.69	0.77	0.74	0.74	0.78	0.77	0.74	0.69	0.61	0.58	0.59	0.58	0.51	0.53	0.49	0.44	0.45	0.34	0.29	0.27
(14) AF	0.65	0.77	0.77	0.80	0.80	0.80	0.81	0.88	0.68	0.68	0.74	0.76	0.76	0.72	0.80	0.79	0.91	0.94	0.94	0.90
(15) BCAF	0.65	0.73	0.79	0.78	0.78	0.80	0.79	0.81	0.78	0.65	0.74	0.79	0.76	0.81	0.86	0.88	0.94	0.94	0.90	0.93
(16) Brent futures	0.76	0.77	0.83	0.74	0.80	0.80	0.74	0.74	0.73	0.63	0.59	0.58	0.57	0.56	0.63	0.68	0.79	0.84	0.84	0.80
(17) Schwartz-Smith mean	0.94	0.85	0.85	0.85	0.87	0.86	0.84	0.81	0.80	0.83	0.85	0.82	0.92	0.89	0.94	0.94	0.94	0.91	0.94	0.93
(18) Schwartz-Smith median	0.94	0.85	0.85	0.85	0.87	0.86	0.84	0.81	0.80	0.83	0.85	0.82	0.92	0.89	0.94	0.94	0.94	0.91	0.94	0.93
(19) Mean all	0.67	0.81	0.81	0.85	0.80	0.82	0.81	0.76	0.66	0.60	0.67	0.68	0.59	0.58	0.54	0.50	0.45	0.44	0.32	0.27
(20) Median all	0.69	0.81	0.81	0.85	0.84	0.84	0.81	0.79	0.73	0.63	0.69	0.71	0.57	0.61	0.60	0.53	0.58	0.53	0.35	0.30
(21) Mean selection	0.76	0.81	0.81	0.87	0.82	0.84	0.79	0.79	0.66	0.60	0.64	0.66	0.54	0.53	0.51	0.53	0.61	0.56	0.48	0.37
(22) Median selection	0.76	0.79	0.79	0.85	0.84	0.82	0.81	0.79	0.68	0.55	0.62	0.61	0.62	0.58	0.63	0.59	0.70	0.63	0.61	0.53

Notes: The nominal coverage rate is 90%. The closer the empirical coverage rate is to 90% (green cells) the better is the fit of the density forecast in respect to observed data.

Table 7.5 - Quarterly Frequency - *Interval Score*

	h = 1	h = 2	h = 3	h = 4	h = 5	h = 6	h = 7	h = 8	h = 9	h = 10	h = 11	h = 12	h = 13	h = 14	h = 15	h = 16	h = 17	h = 18	h = 19	h = 20
(1) RW	67.4	91.0	125.2	111.6	121.4	125.6	135.2	148.3	151.8	164.0	159.3	157.9	200.0	215.0	234.6	241.0	250.1	293.5	338.3	310.6
(2) RW-drift	64.0	92.0	159.8	127.9	147.2	160.0	183.5	208.1	230.0	261.4	280.2	288.6	318.3	345.2	340.4	373.9	381.3	403.7	440.2	496.8
(3) RW-drift5	74.9	94.5	151.9	145.2	164.2	169.1	220.3	472.4	278.9	343.9	387.7	400.5	470.7	525.8	545.3	600.7	665.7	789.3	875.2	972.8
(4) ARIMA	61.0	96.7	108.4	106.0	111.1	125.4	146.8	168.6	196.1	214.2	253.7	248.3	302.3	321.8	339.6	346.6	345.9	371.6	409.8	414.0
(5) Factor model1	63.0	95.6	75.3	104.9	156.0	143.5	190.0	230.2	225.7	238.1	239.8	229.3	231.3	221.8	223.1	269.9	266.1	317.8	277.4	342.5
(6) Factor model2	63.7	82.7	102.2	119.6	134.7	146.5	164.3	233.6	173.7	217.6	223.1	236.3	237.2	269.1	339.3	375.8	435.8	570.2	837.4	1106.3
(7) Elastic net	56.2	108.5	144.2	122.5	152.0	199.6	168.6	172.5	192.7	198.8	263.5	331.9	526.4	410.6	437.5	416.0	443.0	419.9	576.2	876.5
(8) LASSO	67.8	106.6	150.7	127.3	148.4	384.0	176.3	172.9	191.0	185.2	243.2	283.5	507.4	386.4	439.3	422.8	408.6	367.4	508.1	855.4
(9) Adalasso	68.5	105.4	157.8	162.8	172.4	167.4	240.2	208.3	180.0	229.0	329.2	461.1	703.1	427.5	465.0	365.3	452.6	312.6	589.5	837.1
(10) Ridge regression	68.2	109.6	115.2	144.0	245.5	261.8	244.5	266.0	335.4	405.6	332.7	392.7	582.8	552.6	499.0	516.8	522.0	640.2	728.4	845.8
(11) Random forest	53.4	95.6	1045.7	115.2	140.9	153.1	183.1	194.0	216.8	249.5	265.4	276.4	317.6	330.5	367.9	441.8	415.5	442.5	423.0	470.2
(12) Quant.reg.forest	52.6	95.4	138.6	114.9	135.1	149.0	179.6	197.3	221.9	252.8	273.5	280.5	322.4	335.7	368.3	482.7	414.5	440.7	415.0	466.3
(13) XGBoost	55.2	99.9	200.9	112.9	149.6	155.0	198.4	191.8	211.8	233.7	247.1	270.1	291.9	302.1	366.6	391.0	387.1	453.7	368.9	465.0
(14) AF	66.9	90.7	99.2	93.6	103.8	124.2	155.8	176.2	210.9	249.5	272.9	296.3	336.2	359.2	380.7	398.5	400.6	403.4	405.4	397.0
(15) BCAF	65.7	97.2	113.3	112.9	122.9	143.6	175.5	192.0	218.9	247.3	267.1	289.9	330.9	353.4	374.6	394.8	397.9	406.3	412.8	406.8
(16) Brent futures	57.0	107.0	108.5	122.6	123.8	126.7	140.8	147.7	155.6	169.7	159.8	146.8	133.1	122.3	115.3	111.8	138.1	122.5	110.8	132.7
(17) Schwartz-Smith mean	49.5	95.3	105.2	106.2	107.3	110.1	123.1	125.0	123.7	124.5	125.4	128.0	132.0	136.0	140.8	146.3	148.8	155.7	165.7	170.9
(18) Schwartz-Smith median	49.5	95.3	105.2	106.2	107.3	110.1	123.1	125.0	123.7	124.5	125.4	128.0	132.0	136.0	140.8	146.3	148.8	155.7	165.7	170.9
(19) Mean all	53.4	95.1	110.2	106.9	119.8	126.0	140.7	139.9	149.9	171.4	178.8	172.5	196.9	212.7	237.8	219.5	229.6	268.5	270.3	350.7
(20) Median all	54.4	95.4	119.4	117.9	151.1	162.8	188.8	193.7	207.3	249.1	259.7	264.6	284.5	306.8	330.2	297.5	297.7	280.0	294.7	322.9
(21) Mean selection	52.7	93.7	96.9	100.4	115.0	120.7	136.7	136.7	146.7	163.6	168.3	150.6	167.2	171.6	187.7	170.4	179.1	152.0	195.5	208.0
(22) Median selection	53.8	93.4	109.5	108.0	123.3	127.9	146.4	149.4	150.5	172.3	163.9	142.8	144.6	145.4	161.1	157.0	202.8	163.1	138.4	158.7

Note: A lower score implies a better interval forecast. Yellow cells indicate the Top5 best models in each horizon.

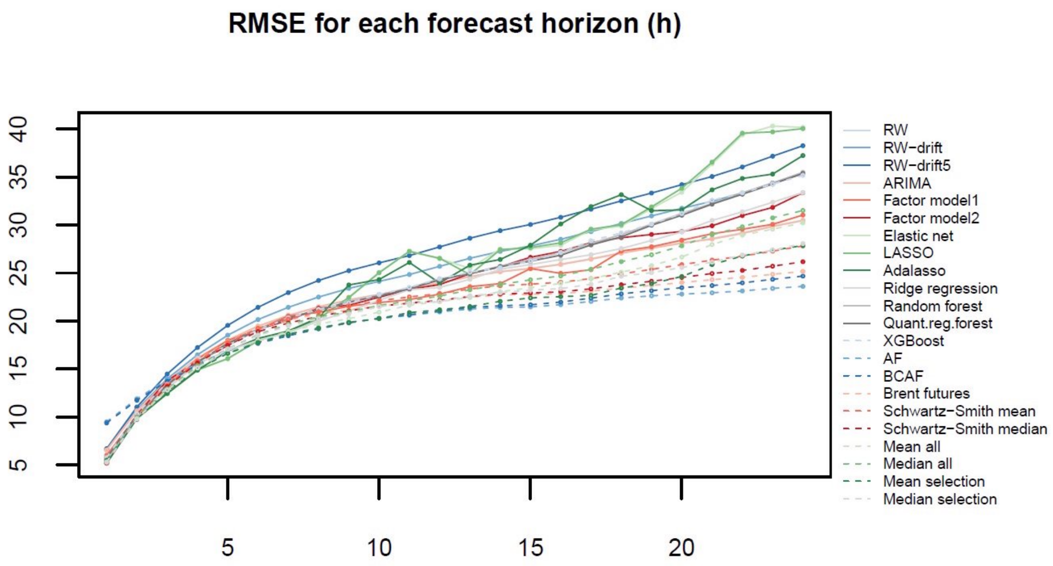
Table 7.6 - Quarterly Frequency - *Log Predictive Density Score (LPDS)*

	h = 1	h = 2	h = 3	h = 4	h = 5	h = 6	h = 7	h = 8	h = 9	h = 10	h = 11	h = 12	h = 13	h = 14	h = 15	h = 16	h = 17	h = 18	h = 19	h = 20
(1) RW	-4.14	-4.82	-4.89	-4.87	-4.84	-4.84	-4.92	-5.09	-5.12	-5.24	-5.00	-4.87	-5.12	-5.09	-5.04	-4.95	-4.84	-4.85	-4.94	-4.94
(2) RW-drift	-4.20	-4.69	-4.86	-4.97	-5.17	-5.37	-5.56	-5.67	-5.78	-6.09	-6.25	-5.98	-6.20	-6.48	-6.21	-6.17	-6.11	-6.18	-6.44	-6.86
(3) RW-drift5	-4.00	-4.52	-5.03	-5.17	-5.54	-5.84	-6.28	-6.81	-7.34	-8.34	-9.55	-8.75	-9.58	-10.70	-10.35	-10.88	-11.10	-11.85	-12.82	-13.34
(4) ARIMA	-4.32	-4.99	-4.80	-4.80	-4.66	-4.58	-4.78	-4.91	-4.89	-4.90	-4.99	-4.87	-5.17	-5.15	-5.10	-5.08	-4.96	-5.10	-5.30	-5.43
(5) Factor model1	-4.58	-5.16	-4.30	-4.65	-5.21	-5.04	-5.48	-5.75	-5.73	-5.77	-5.71	-5.62	-5.59	-5.54	-5.69	-6.38	-6.27	-6.72	-6.44	-7.51
(6) Factor model2	-4.45	-4.40	-4.62	-4.89	-4.92	-4.99	-5.21	-6.26	-5.16	-5.93	-5.55	-5.61	-5.69	-6.10	-7.95	-8.15	-8.52	-12.44	-25.79	-43.21
(7) Elastic net	-4.58	-5.89	-5.32	-5.10	-5.20	-5.04	-5.25	-5.29	-5.34	-5.31	-5.94	-6.11	-12.15	-6.72	-7.09	-6.69	-6.75	-6.84	-8.59	-14.15
(8) LASSO	-4.75	-5.59	-5.51	-5.19	-5.25	-5.20	-5.29	-5.29	-5.34	-5.19	-5.72	-5.81	-11.46	-6.38	-7.00	-6.39	-6.21	-6.57	-7.27	-12.80
(9) Adalasso	-4.83	-5.67	-5.73	-5.86	-5.68	-5.37	-5.58	-5.58	-5.25	-5.61	-6.27	-9.51	-9.80	-7.62	-7.40	-6.51	-7.73	-6.61	-10.72	-14.85
(10) Ridge regression	-4.59	-5.91	-5.30	-5.94	-9.87	-9.20	-9.22	-9.68	-11.25	-15.74	-10.48	-11.87	-20.73	-18.40	-15.58	-14.63	-13.42	-16.16	-18.13	-22.53
(11) Random forest	-4.03	-4.78	-4.70	-4.84	-5.14	-5.20	-5.53	-5.62	-5.71	-6.01	-6.06	-6.04	-6.40	-6.42	-6.81	-6.86	-6.96	-7.08	-7.08	-7.48
(12) Quant.reg.forest	-3.99	-4.76	-4.69	-4.84	-5.01	-5.13	-5.47	-5.64	-5.76	-6.07	-6.16	-6.09	-6.49	-6.51	-6.84	-6.86	-6.98	-7.10	-7.03	-7.50
(13) XGBoost	-4.11	-4.85	-4.92	-4.95	-5.32	-5.35	-5.68	-5.68	-5.75	-5.82	-5.85	-5.95	-6.05	-6.10	-6.66	-6.91	-6.71	-6.92	-6.64	-7.68
(14) AF	-4.64	-4.97	-4.84	-4.65	-4.60	-4.74	-4.78	-4.73	-4.76	-4.87	-4.77	-4.70	-4.82	-4.80	-4.78	-4.79	-4.70	-4.71	-4.77	-4.82
(15) BCAF	-4.64	-5.14	-5.14	-4.90	-4.82	-5.02	-4.97	-4.82	-4.80	-4.84	-4.70	-4.62	-4.75	-4.72	-4.68	-4.69	-4.61	-4.67	-4.79	-4.92
(16) Brent futures	-4.06	-5.33	-4.92	-5.11	-5.62	-5.66	-5.87	-5.94	-5.79	-5.87	-5.38	-5.03	-4.90	-4.84	-4.79	-4.77	-4.66	-4.62	-4.65	-4.81
(17) Schwartz-Smith mean	-4.10	-5.32	-5.12	-5.30	-5.74	-6.52	-7.23	-6.95	-6.50	-6.38	-6.15	-5.96	-5.96	-5.96	-5.90	-5.80	-5.54	-5.45	-5.60	-5.93
(18) Schwartz-Smith median	-4.07	-5.20	-4.95	-5.06	-5.32	-5.23	-6.60	-6.42	-6.21	-5.91	-5.59	-5.42	-5.40	-5.42	-5.35	-5.28	-5.08	-4.95	-4.97	-5.17
(19) Mean all	-4.05	-4.92	-4.71	-4.75	-4.85	-4.80	-4.97	-5.00	-5.03	-5.18	-5.08	-4.94	-5.26	-5.35	-5.48	-5.33	-5.32	-5.42	-5.82	-6.43
(20) Median all	-4.10	-4.84	-4.78	-4.83	-4.93	-4.81	-4.96	-4.92	-4.90	-5.17	-5.05	-4.90	-5.12	-5.26	-5.31	-5.21	-5.16	-5.26	-5.59	-6.21
(21) Mean selection	-4.00	-4.92	-4.58	-4.63	-4.77	-4.75	-4.98	-5.00	-5.01	-5.13	-5.05	-4.89	-5.16	-5.14	-5.30	-5.14	-5.11	-4.94	-5.23	-5.61
(22) Median selection	-4.04	-4.84	-4.70	-4.77	-4.99	-4.97	-5.17	-5.18	-5.12	-5.27	-5.08	-4.87	-4.96	-5.00	-5.11	-5.06	-4.97	-4.82	-4.87	-5.20

Note: A higher score implies a better density forecast. Yellow cells indicate the Top5 best models in each horizon.

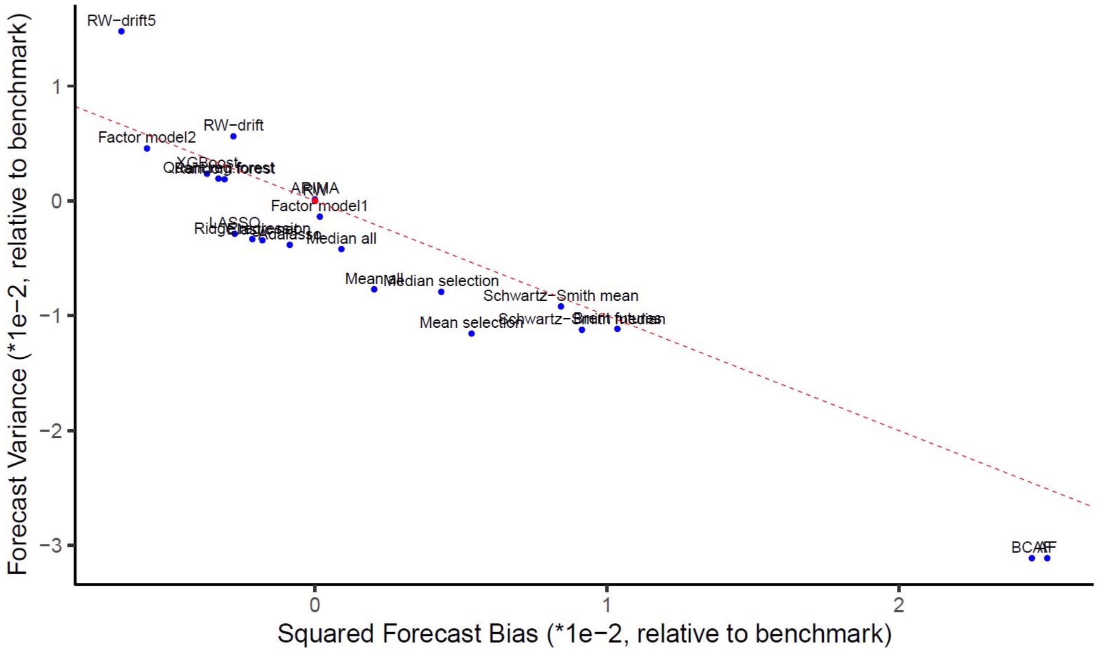
Appendix 8. Other results - monthly frequency

Figure 8.1 - Root Mean Squared Error (RMSE)



Note: RMSE (vertical axis) computed along the pseudo out-of-sample exercise for each forecast horizon (horizontal axis).

Figure 8.2 - MSE Decomposition $(h = 6, 12)$



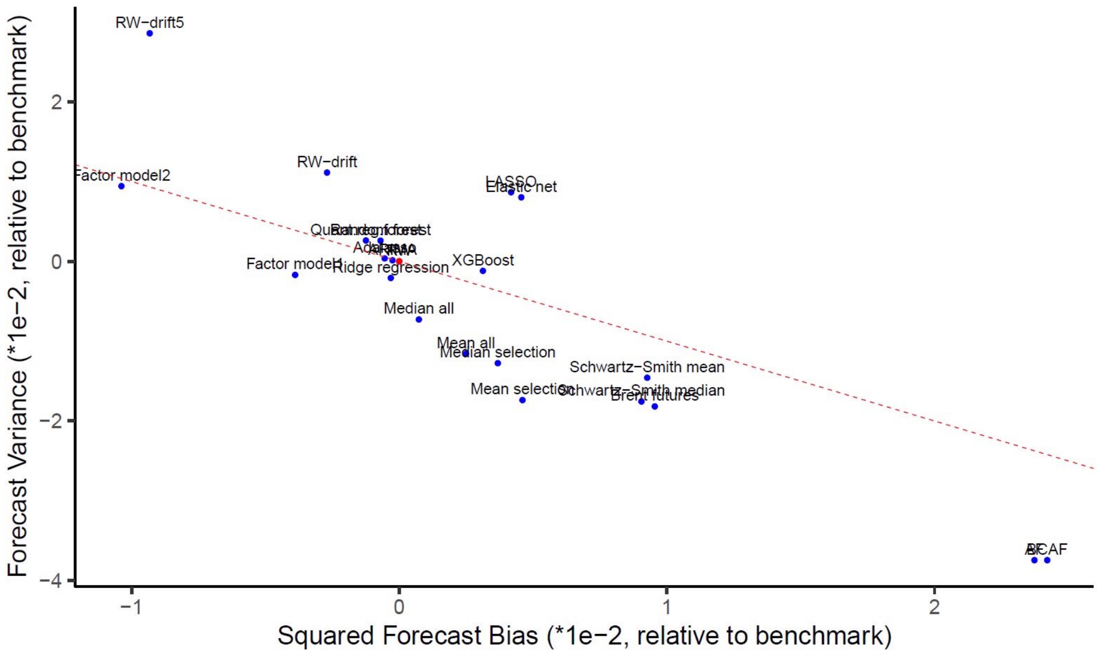
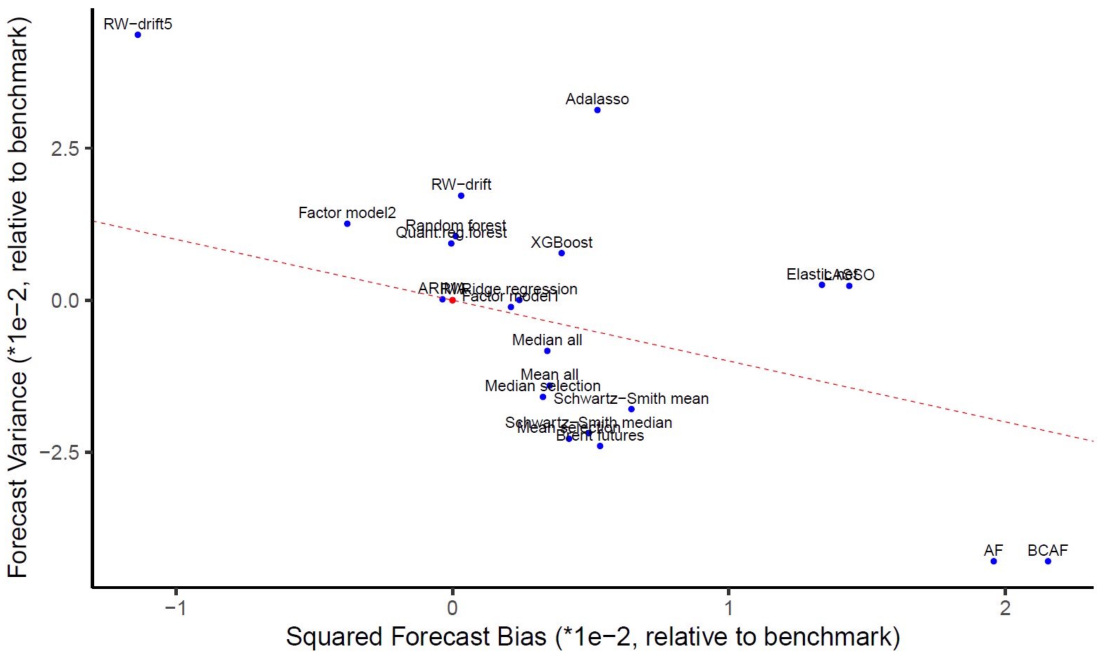


Figure 8.3 - MSE Decomposition ($h = 18, 24$)



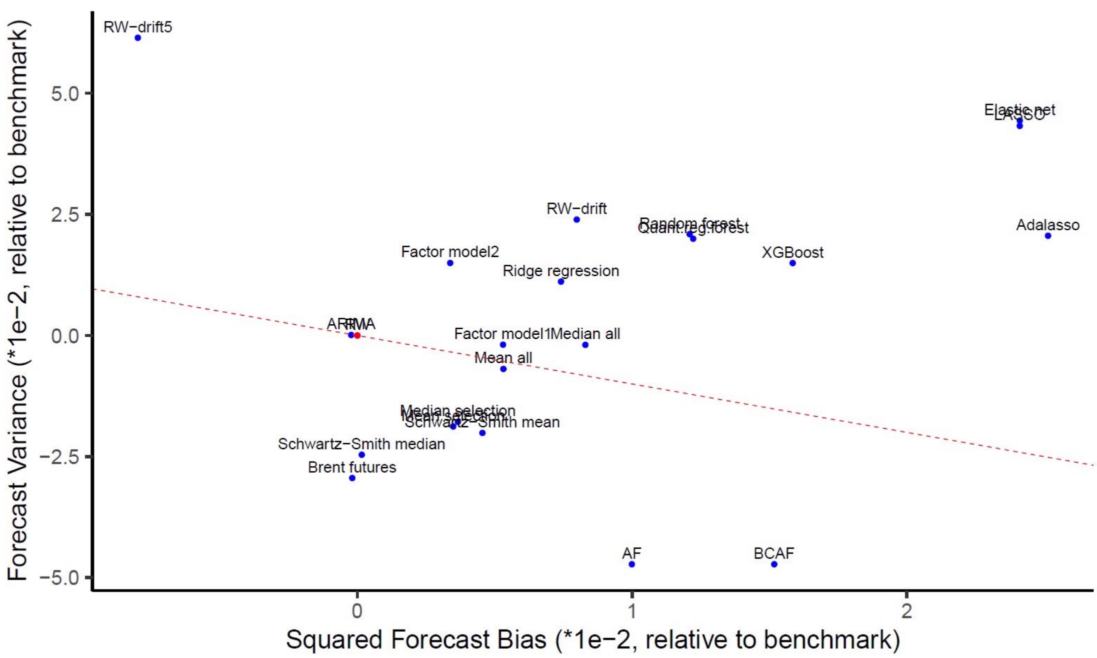


Figure 8.4 - Oil price change and forescasts ($h = 1, 6, 12, 24$)

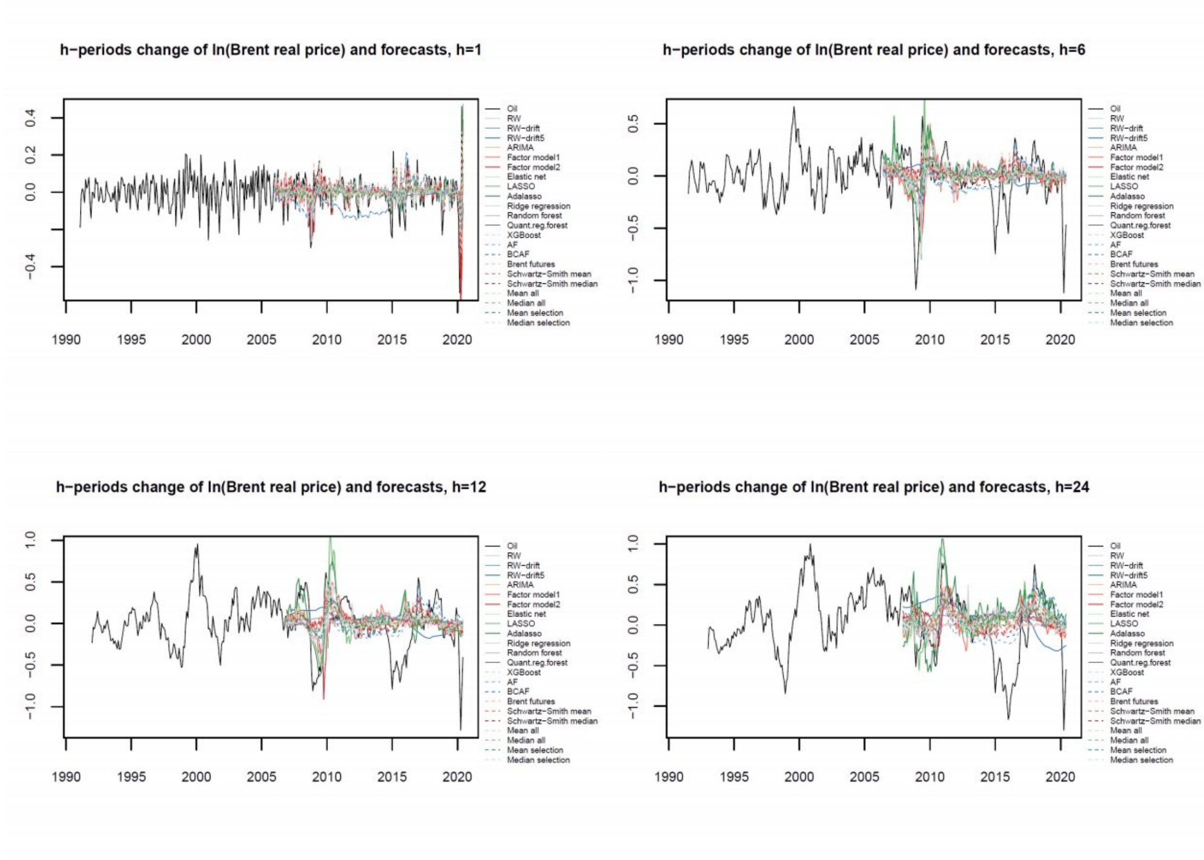


Figure 8.5 - Pseudo out-of-sample forecasts ($h = 1, ..., 24$)

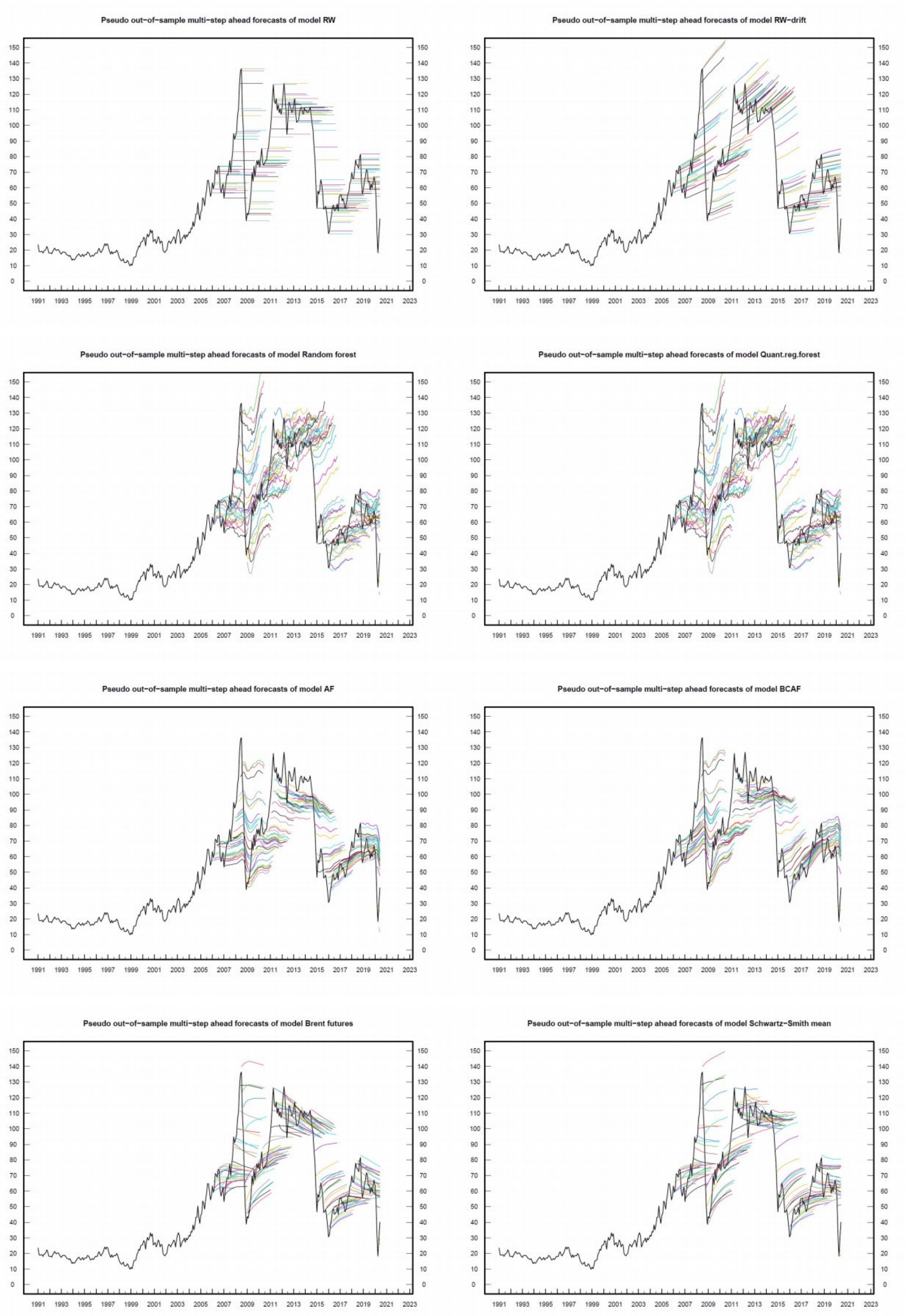
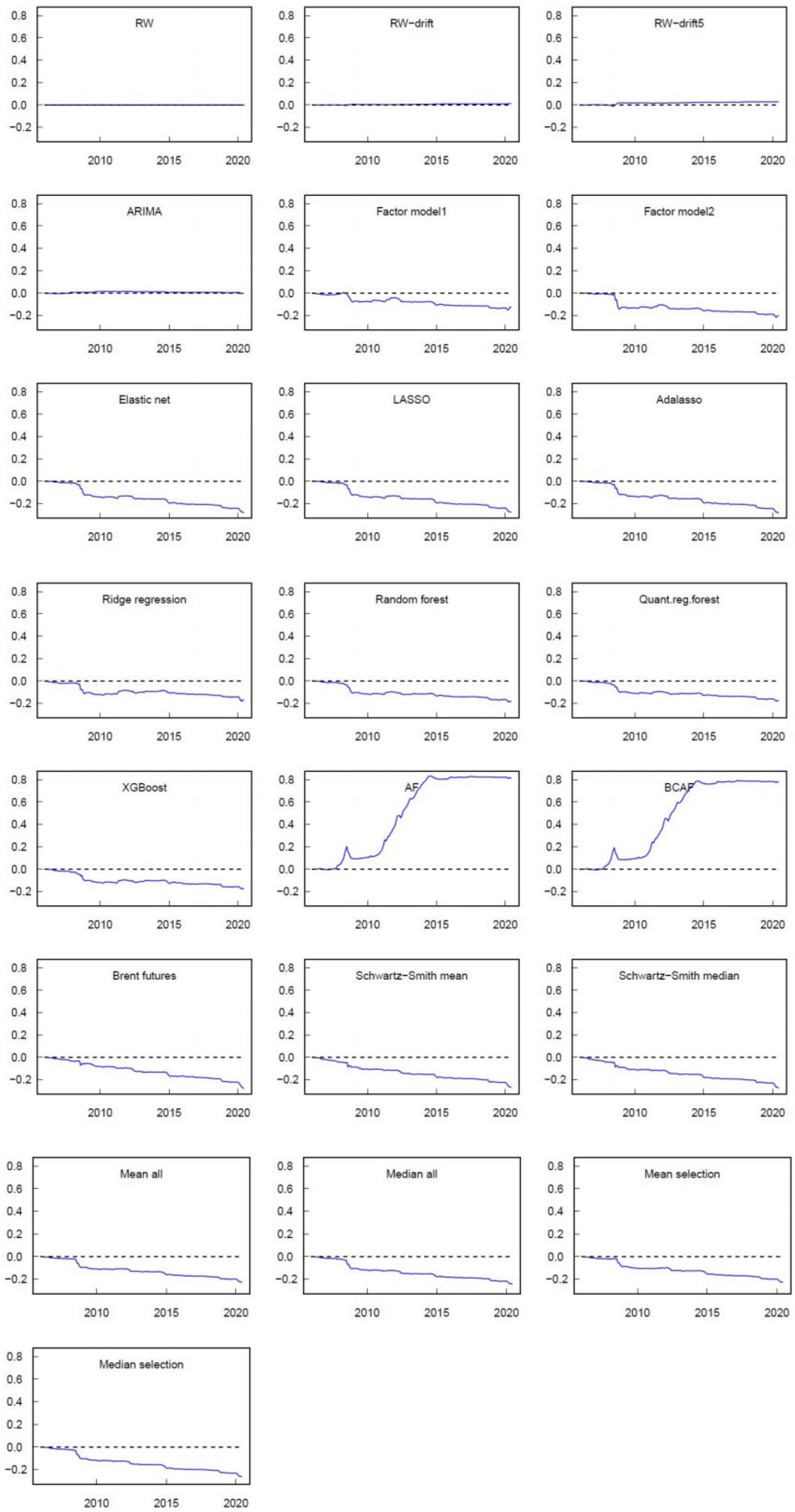
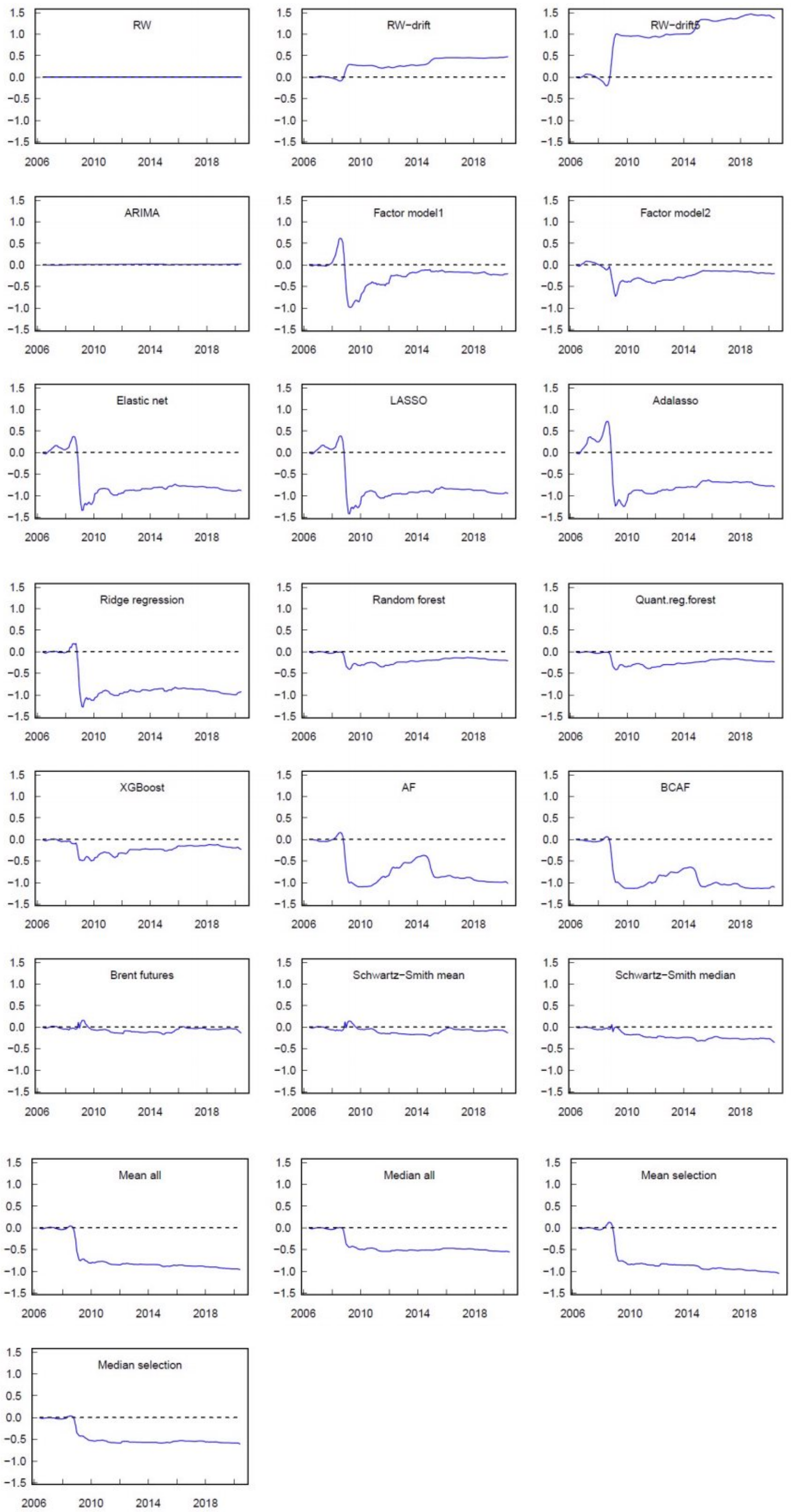


Figure 8.6 - Cumulative Square Prediction Error ($h = 1$)



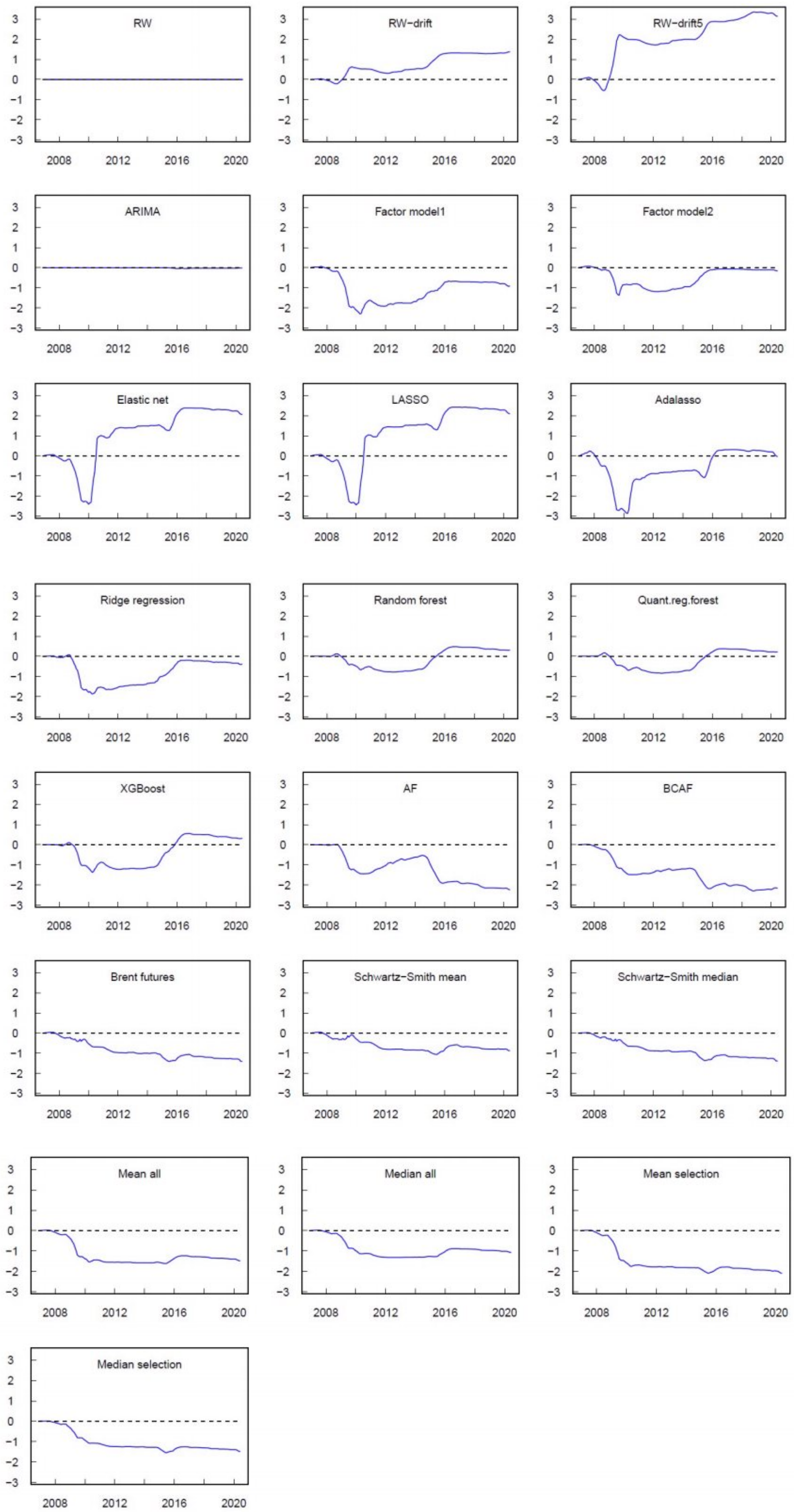
Notes: Graphs show time series plots of the differences (over time) between the Cumulative Squared Prediction Error (CSPE, divided by 10,000) of a given model and the CSPE of the benchmark model (RW). Figures above (below) zero indicate that the benchmark is better (worse).

Figure 8.7 - Cumulative Square Prediction Error ($h = 6$)



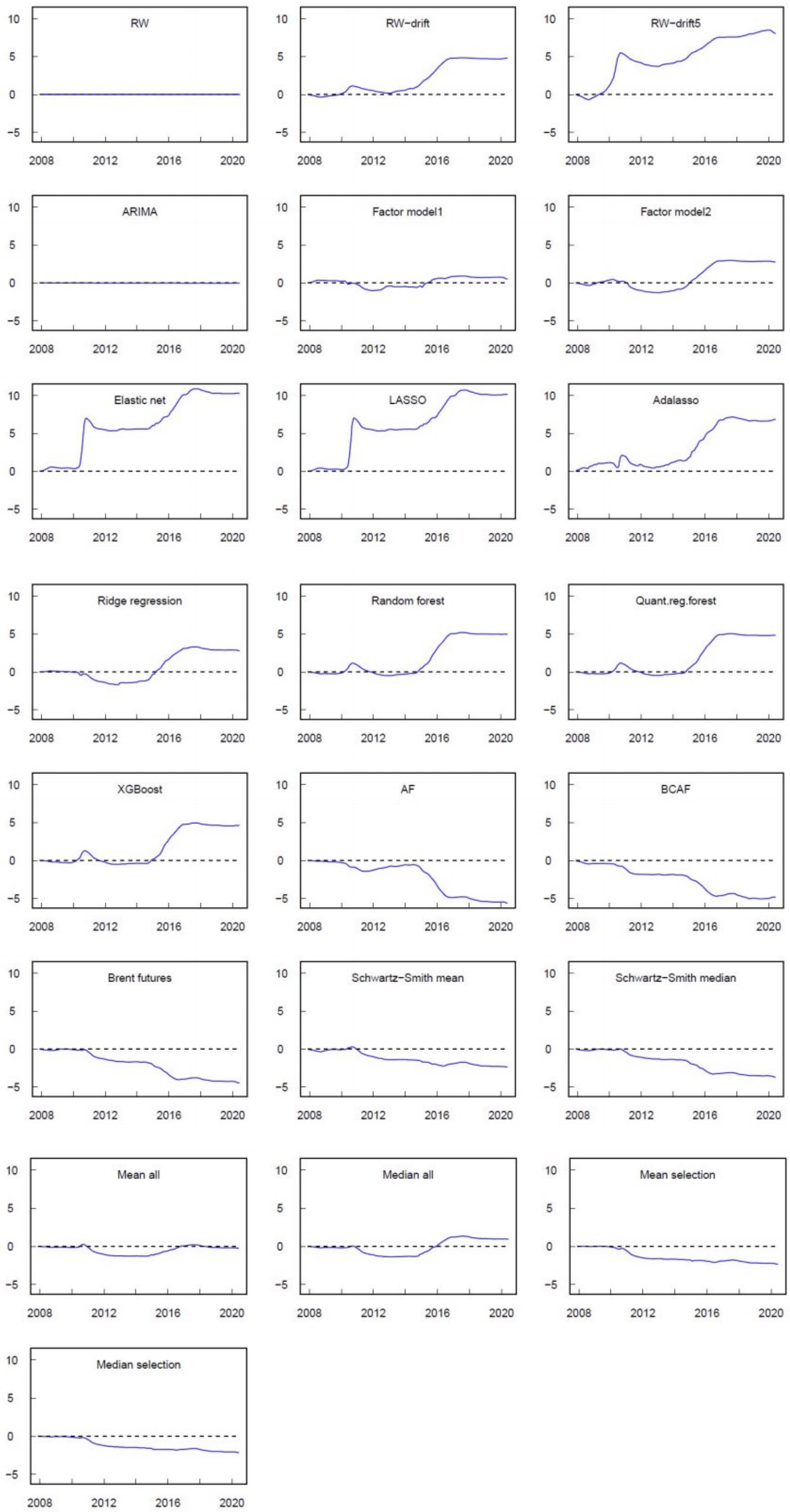
Notes: Graphs show time series plots of the differences (over time) between the Cumulative Squared Prediction Error (CSPE, divided by 10,000) of a given model and the CSPE of the benchmark model (RW). Figures above (below) zero indicate that the benchmark is better (worse).

Figure 8.8 - Cumulative Square Prediction Error ($h = 12$)



Notes: Graphs show time series plots of the differences (over time) between the Cumulative Squared Prediction Error (CSPE, divided by 10,000) of a given model and the CSPE of the benchmark model (RW). Figures above (below) zero indicate that the benchmark is better (worse).

Figure 8.9 - Cumulative Square Prediction Error ($h = 24$)



Notes: Graphs show time series plots of the differences (over time) between the Cumulative Squared Prediction Error (CSPE, divided by 10,000) of a given model and the CSPE of the benchmark model (RW). Figures above (below) zero indicate that the benchmark is better (worse).

Figure 8.10 - Number of selected variables (lasso, adalasso, elastic net)

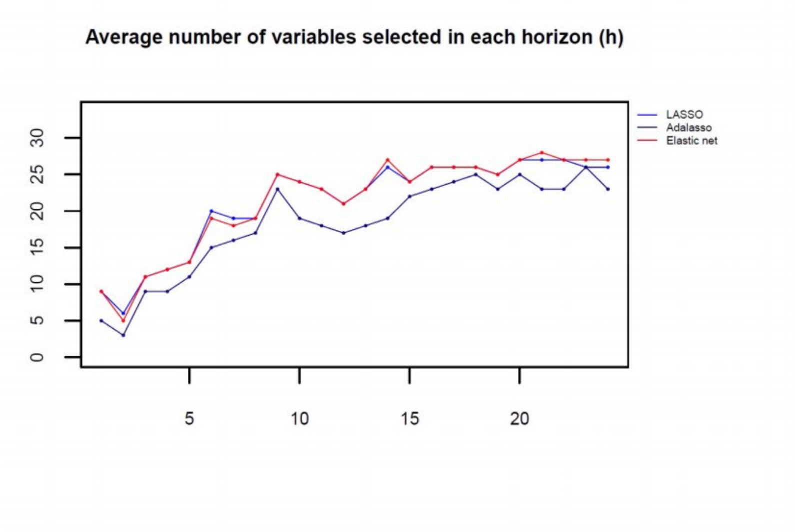


Figure 8.11 - Variable selection over time ($h = 1, 6$)

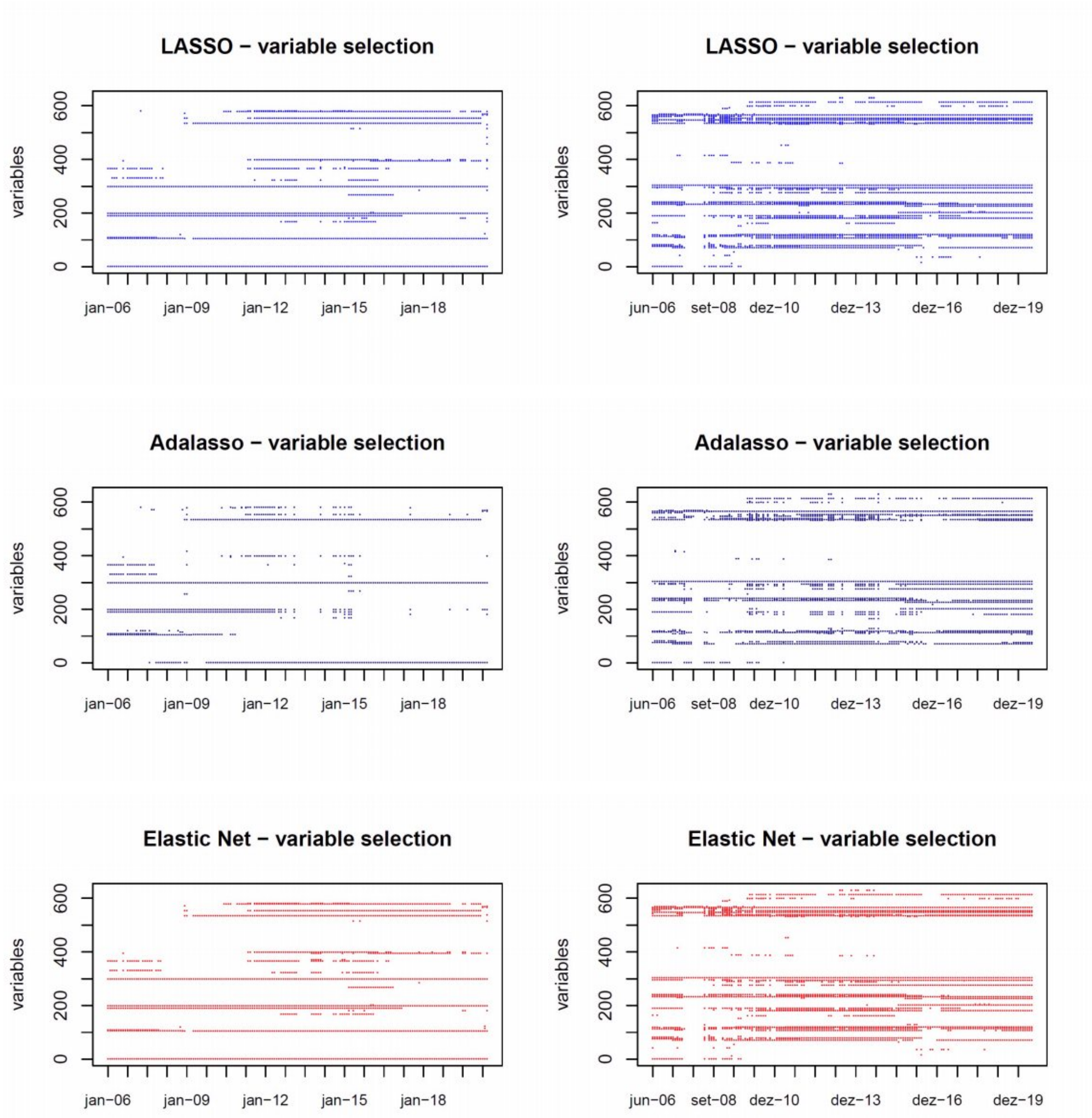


Figure 8.12 - Variable selection over time ($h = 12, 24$)

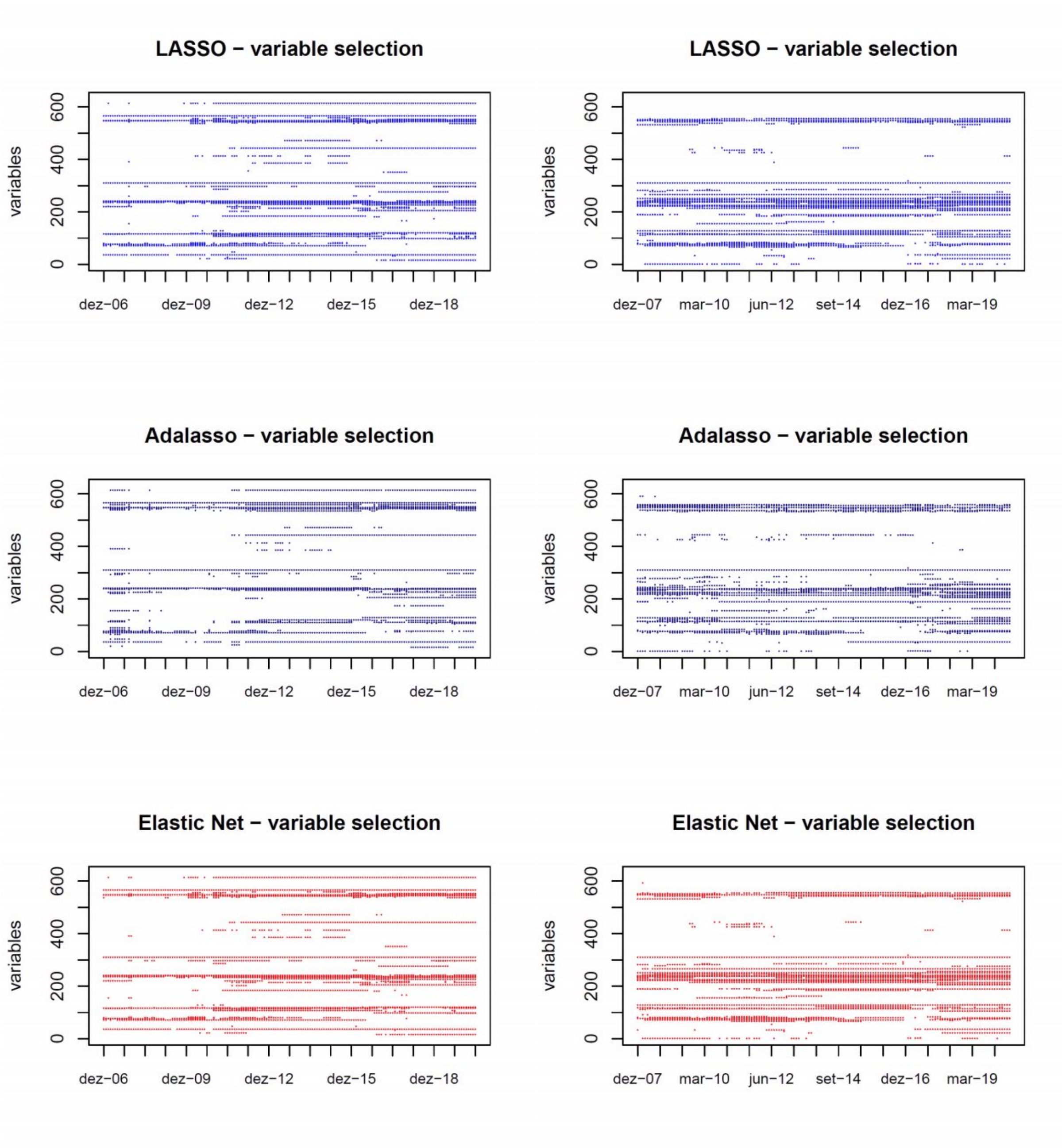


Figure 8.13 - Variable importance ($h = 1, 6$)

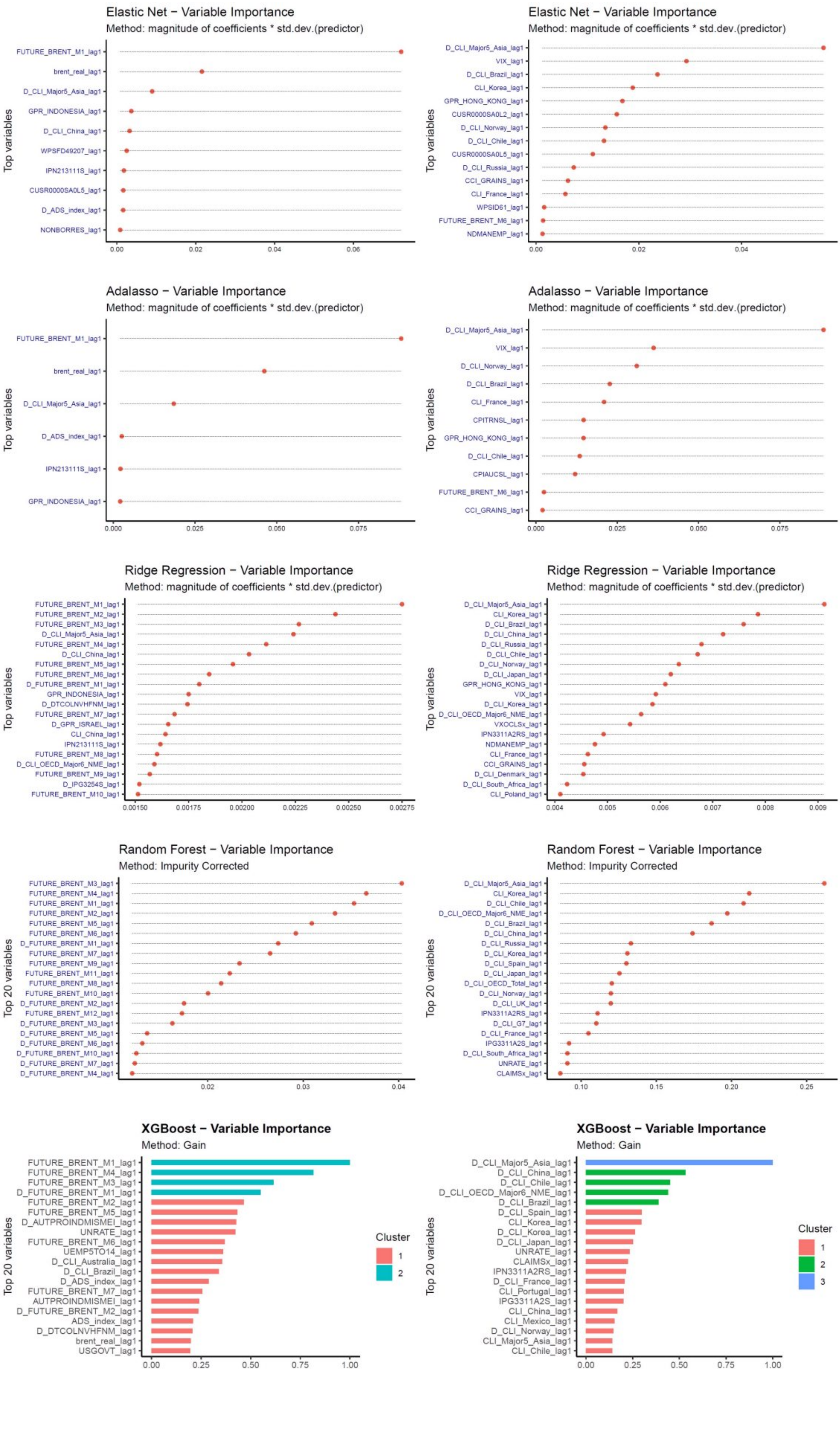


Figure 8.14 - Variable importance ($h = 12, 24$)

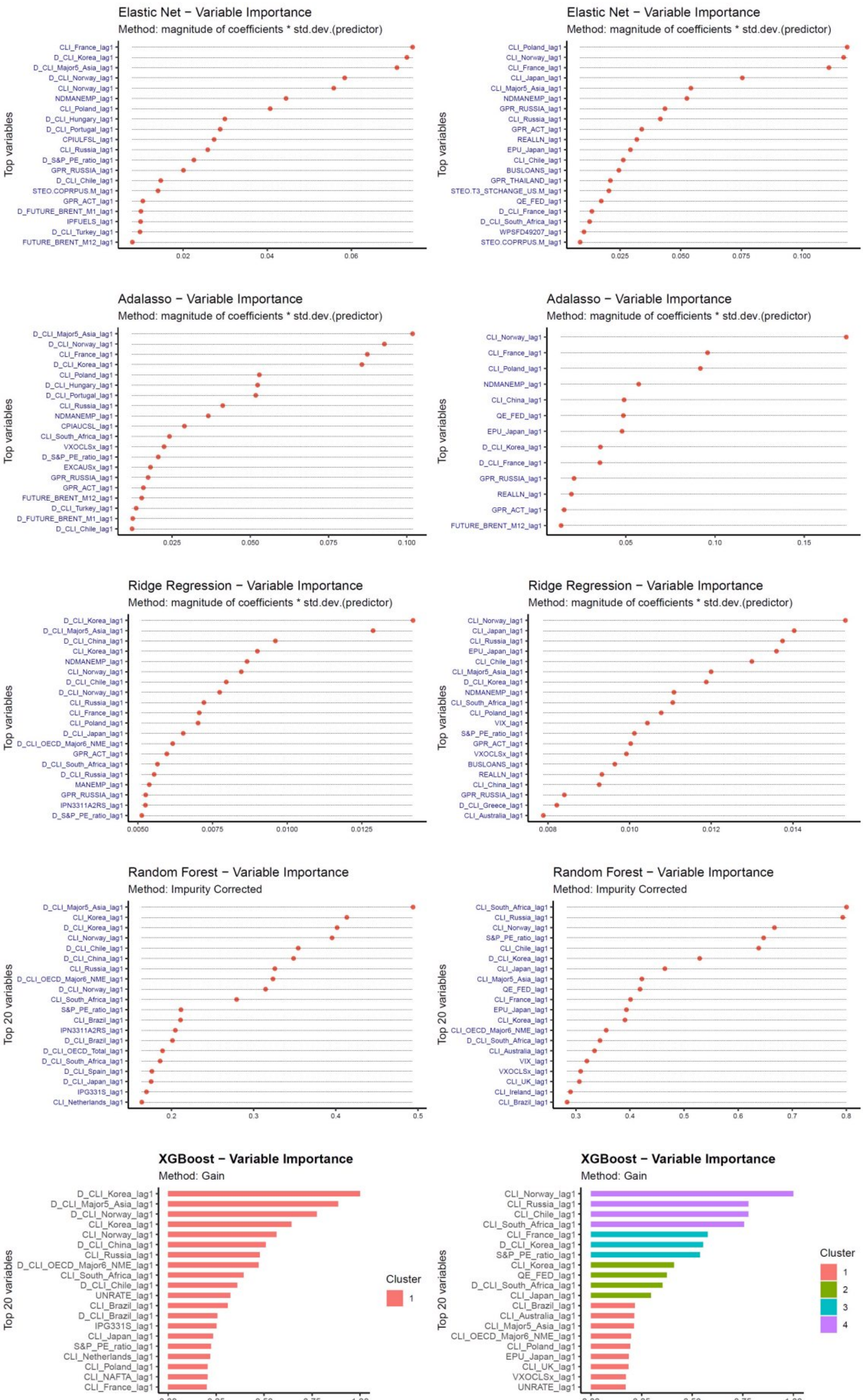
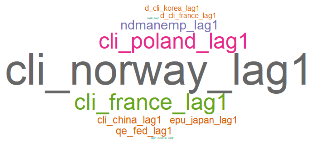
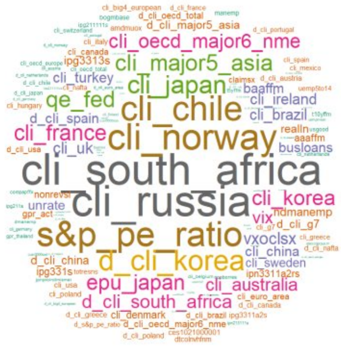


Figure 8.15 - Word clouds ($h = 24$)

Panel (a): elastic net (left) and adalasso (right)



Panel (b): ridge regression (left) and random forest (right)



Panel (c): xgboost

Figure 8.16 - Fan charts and probability density functions (PDFs)

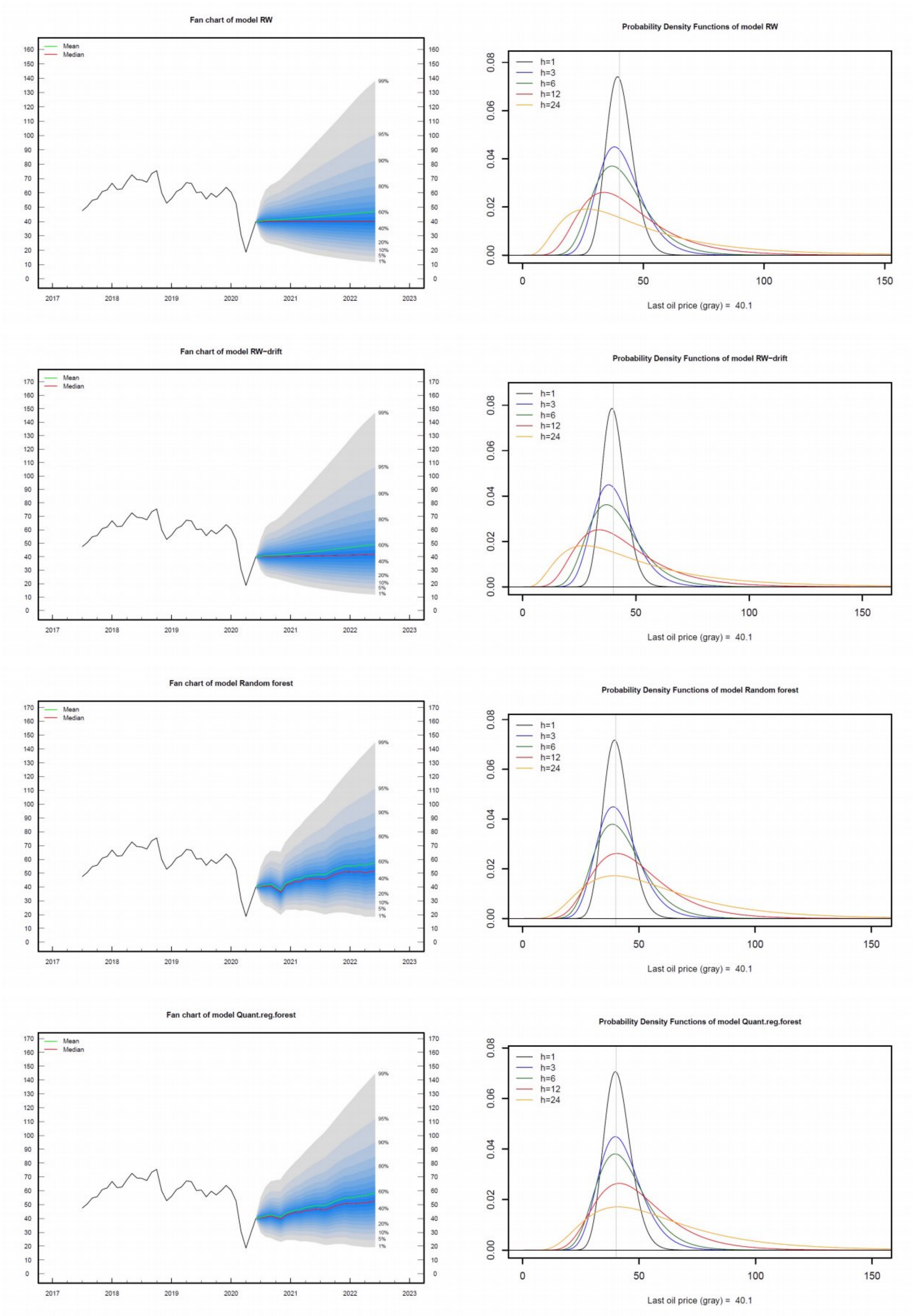
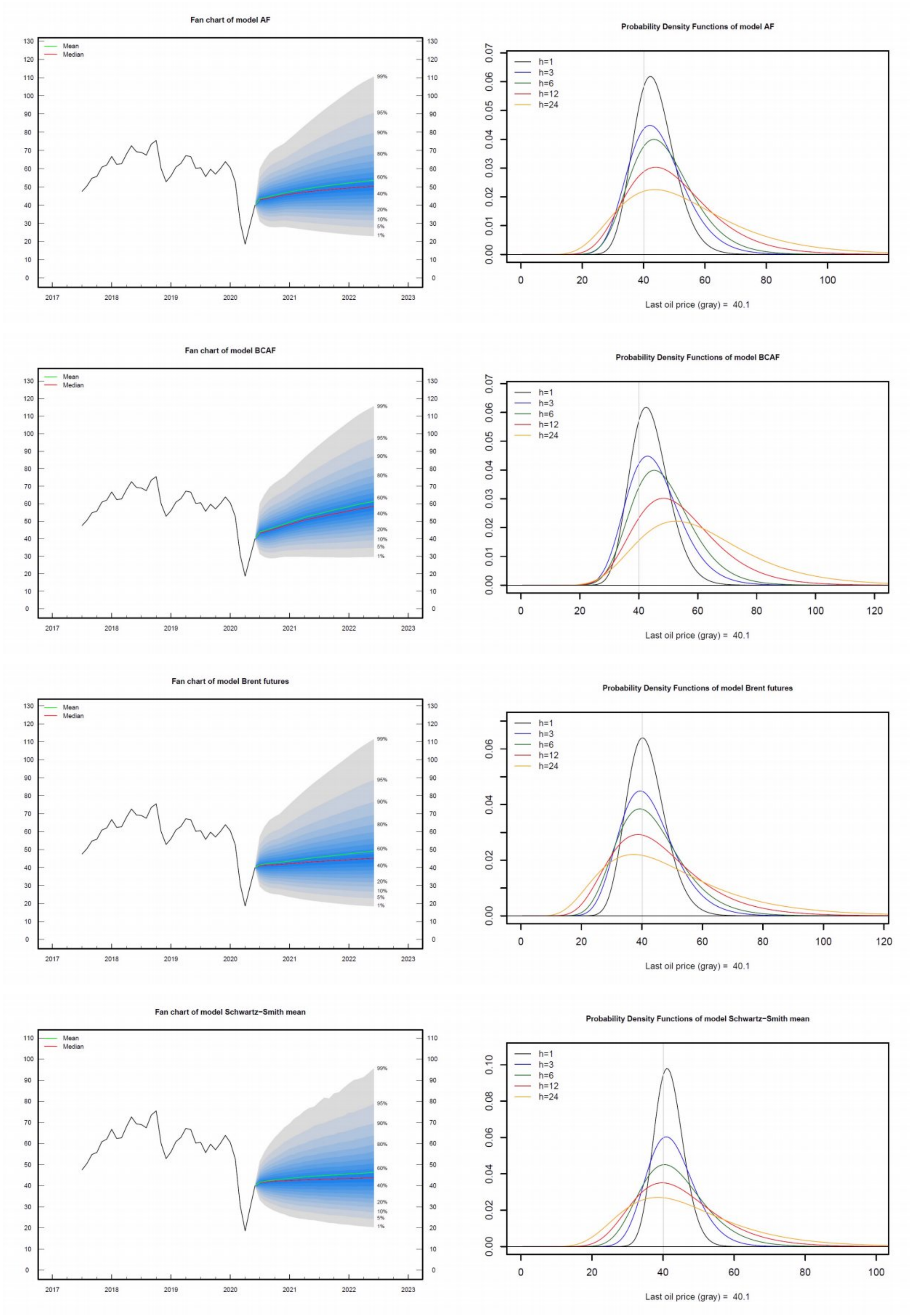
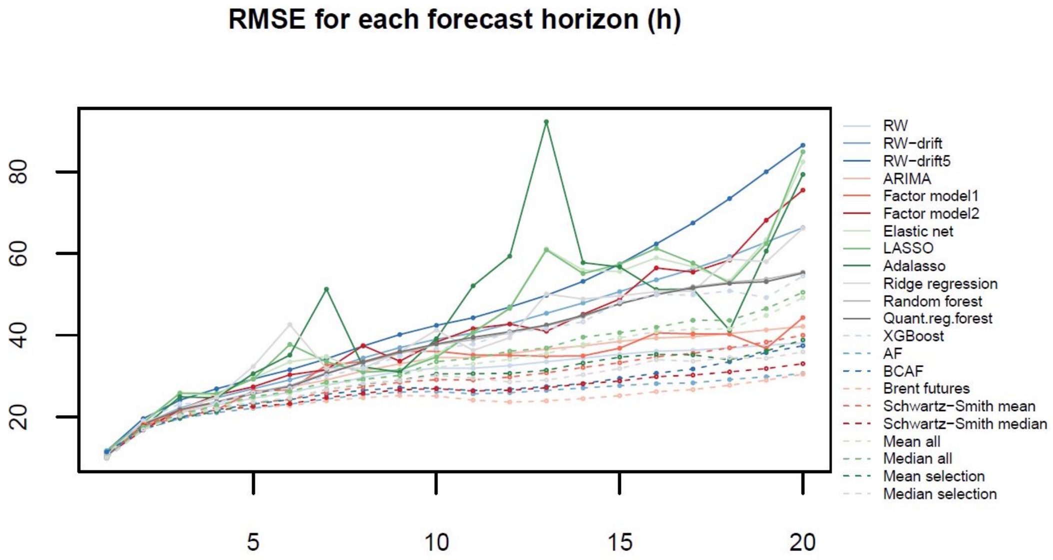


Figure 8.17 - Fan charts and probability density functions (PDFs)



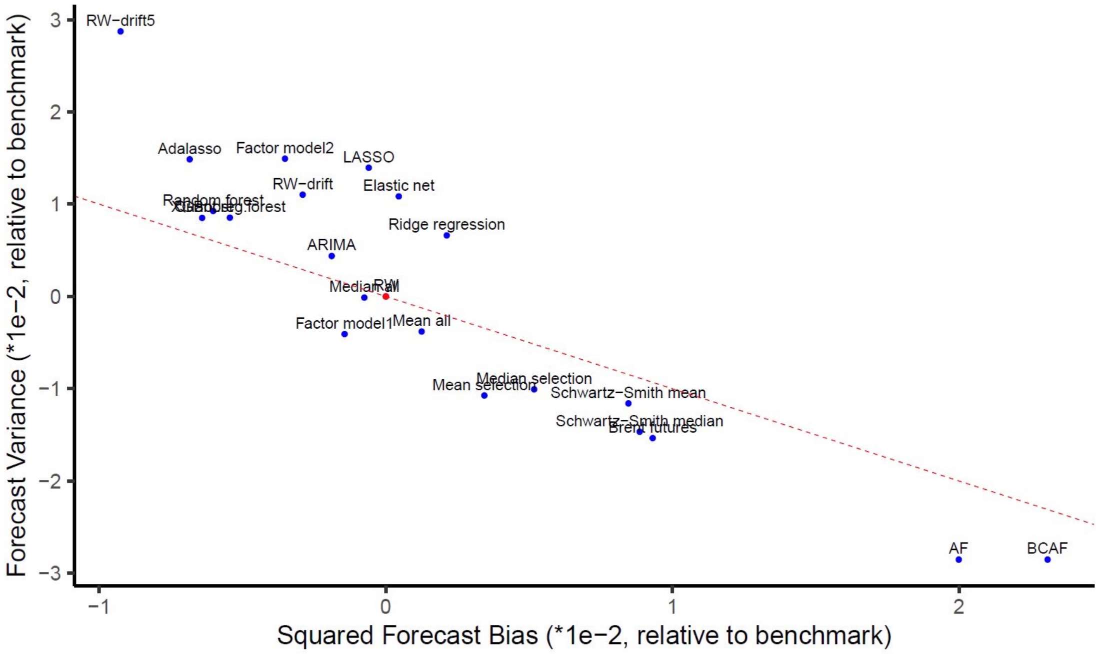
Appendix 9. Other results - quarterly frequency

Figure 9.1 - Root Mean Squared Error (RMSE)



Note: RMSE (vertical axis) computed along the pseudo out-of-sample exercise for each forecast horizon (horizontal axis).

Figure 9.2 - MSE Decomposition $(h = 4, 8)$



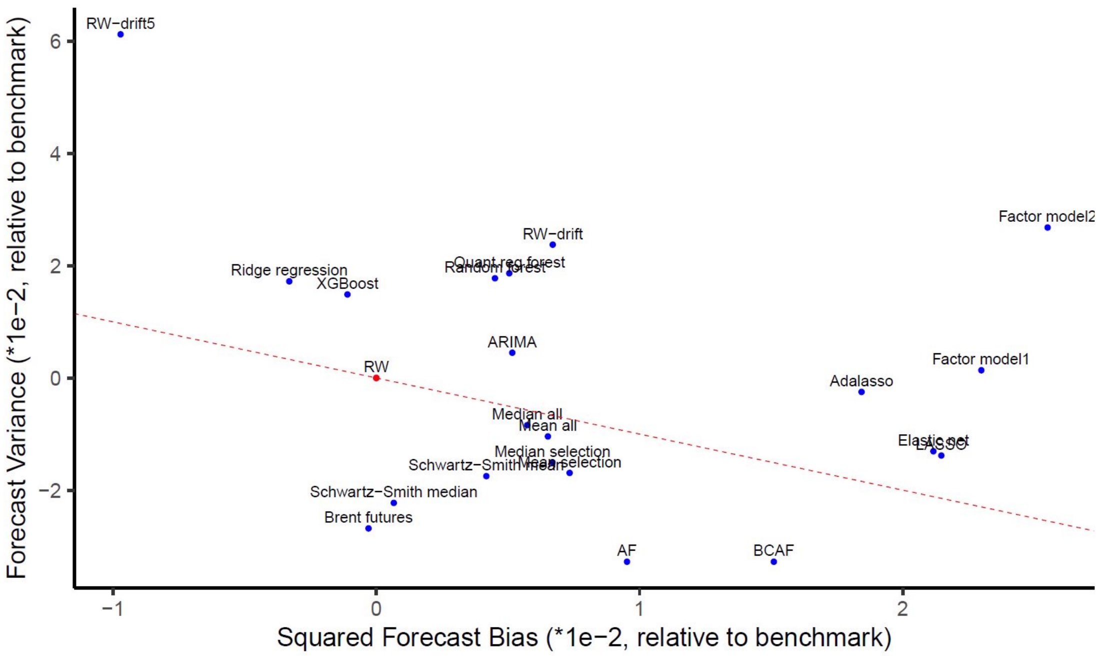


Figure 9.3 - MSE Decomposition $(h = 14, 20)$

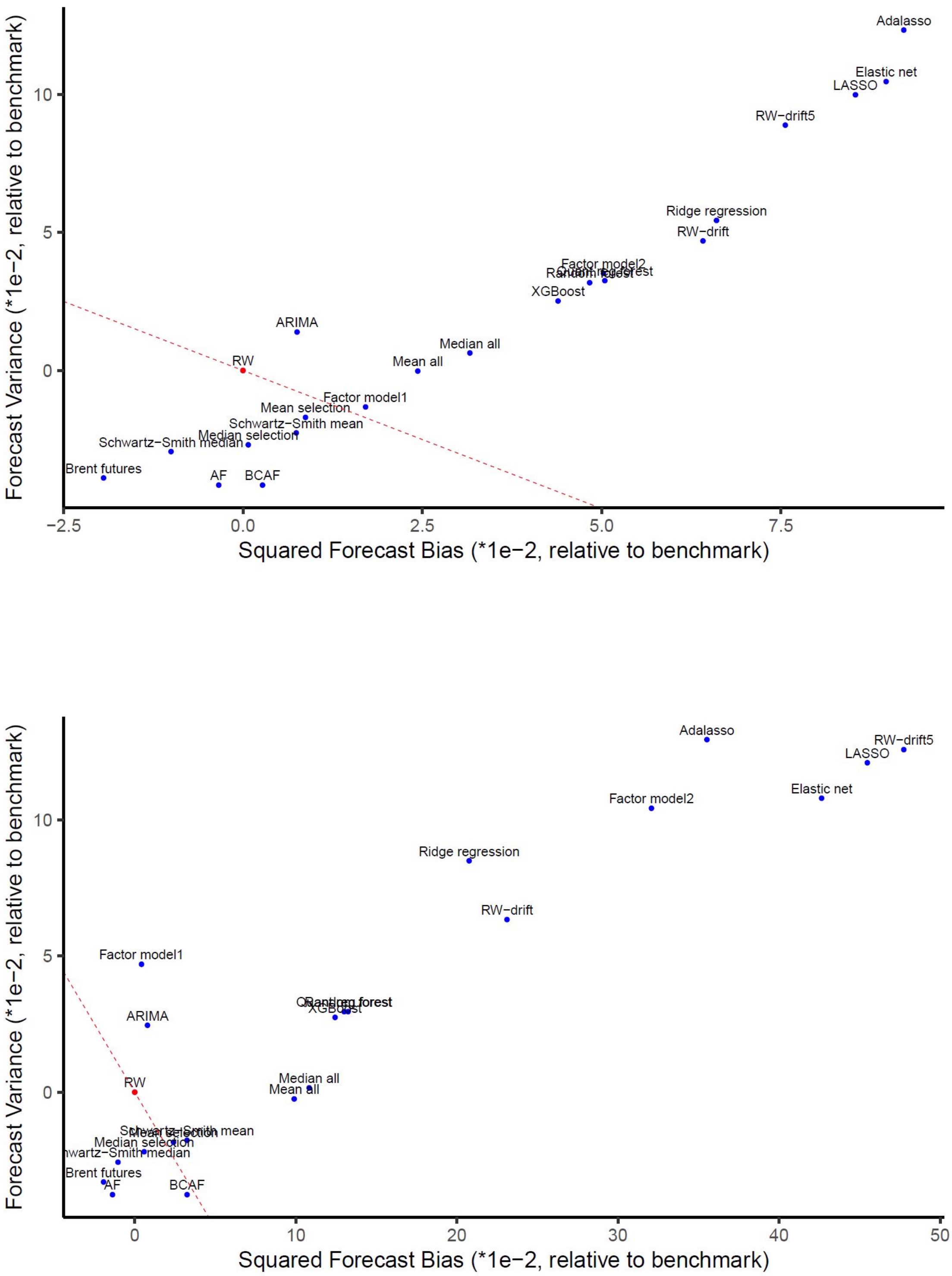


Figure 9.4 - Oil price change and forecasts ($h = 1, 4, 8, 20$)

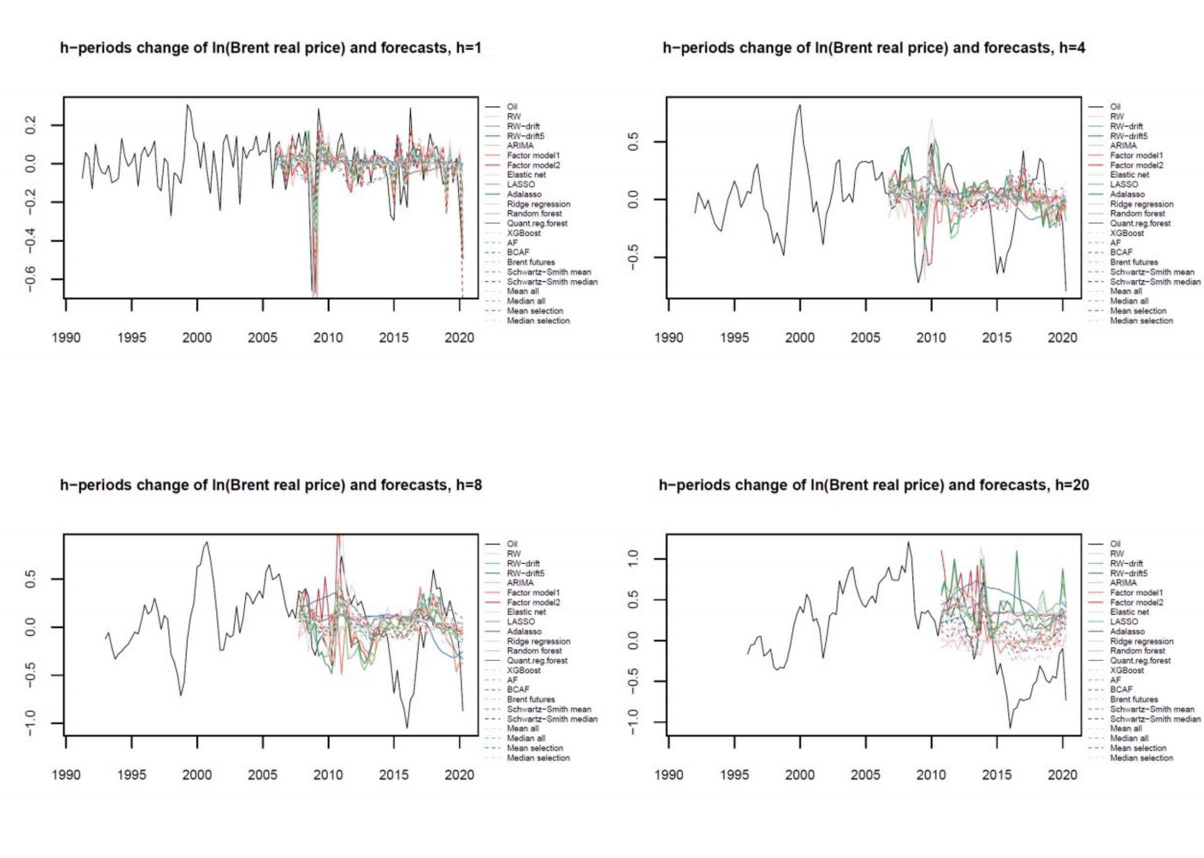


Figure 9.5 - Pseudo out-of-sample forecasts ($h = 1, ..., 20$)

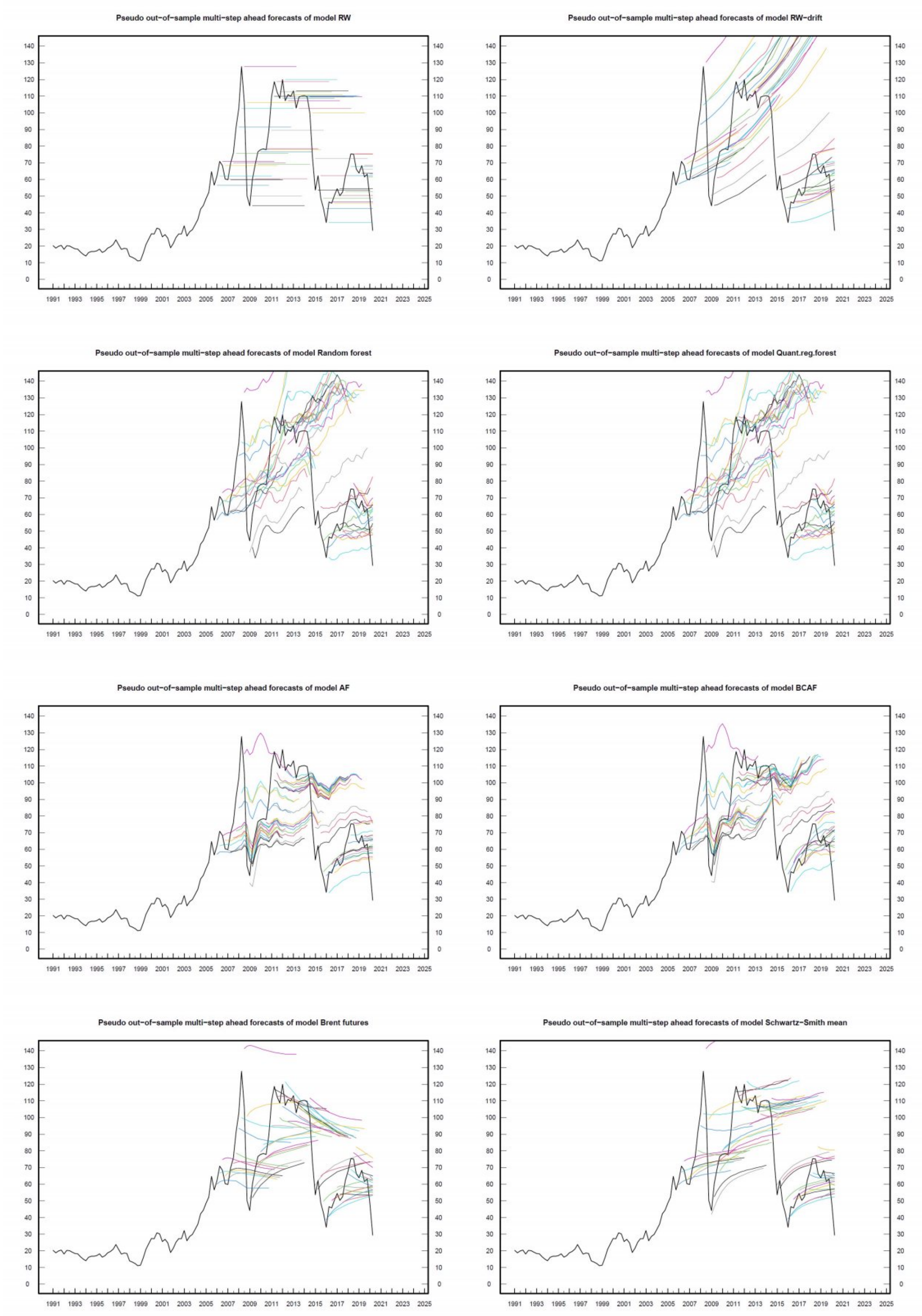
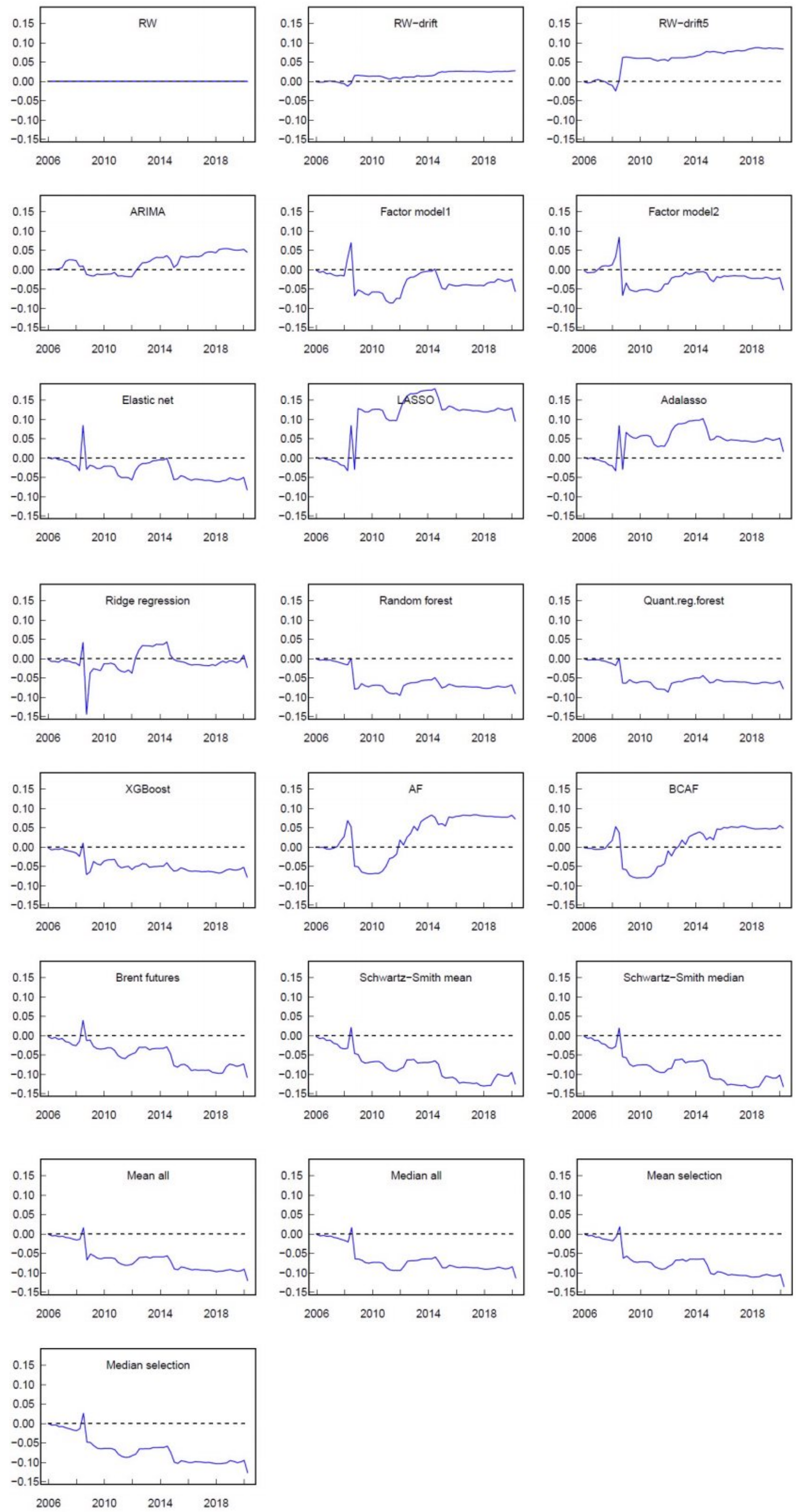
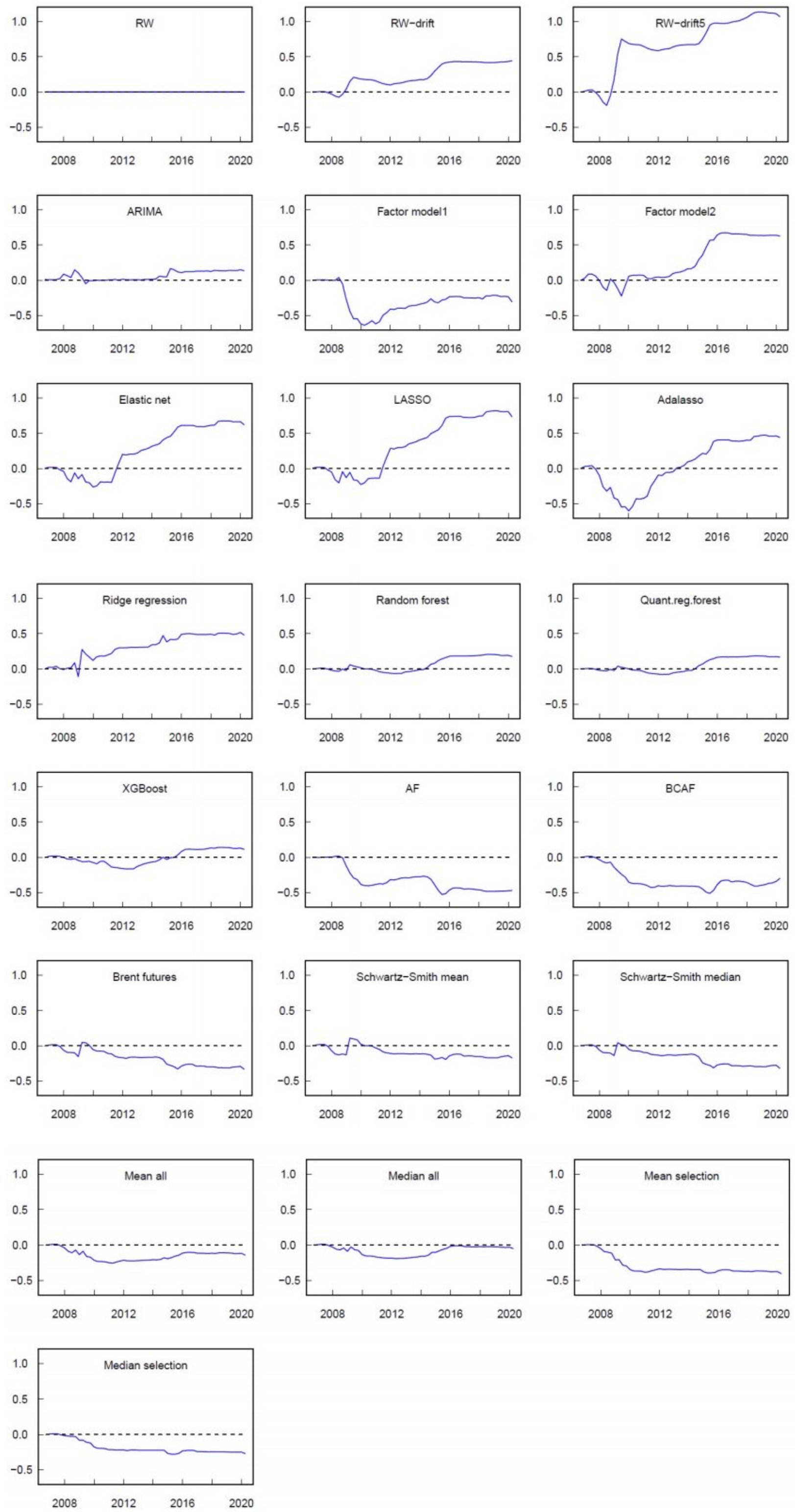


Figure 9.6 - Cumulative Square Prediction Error ($h = 1$)



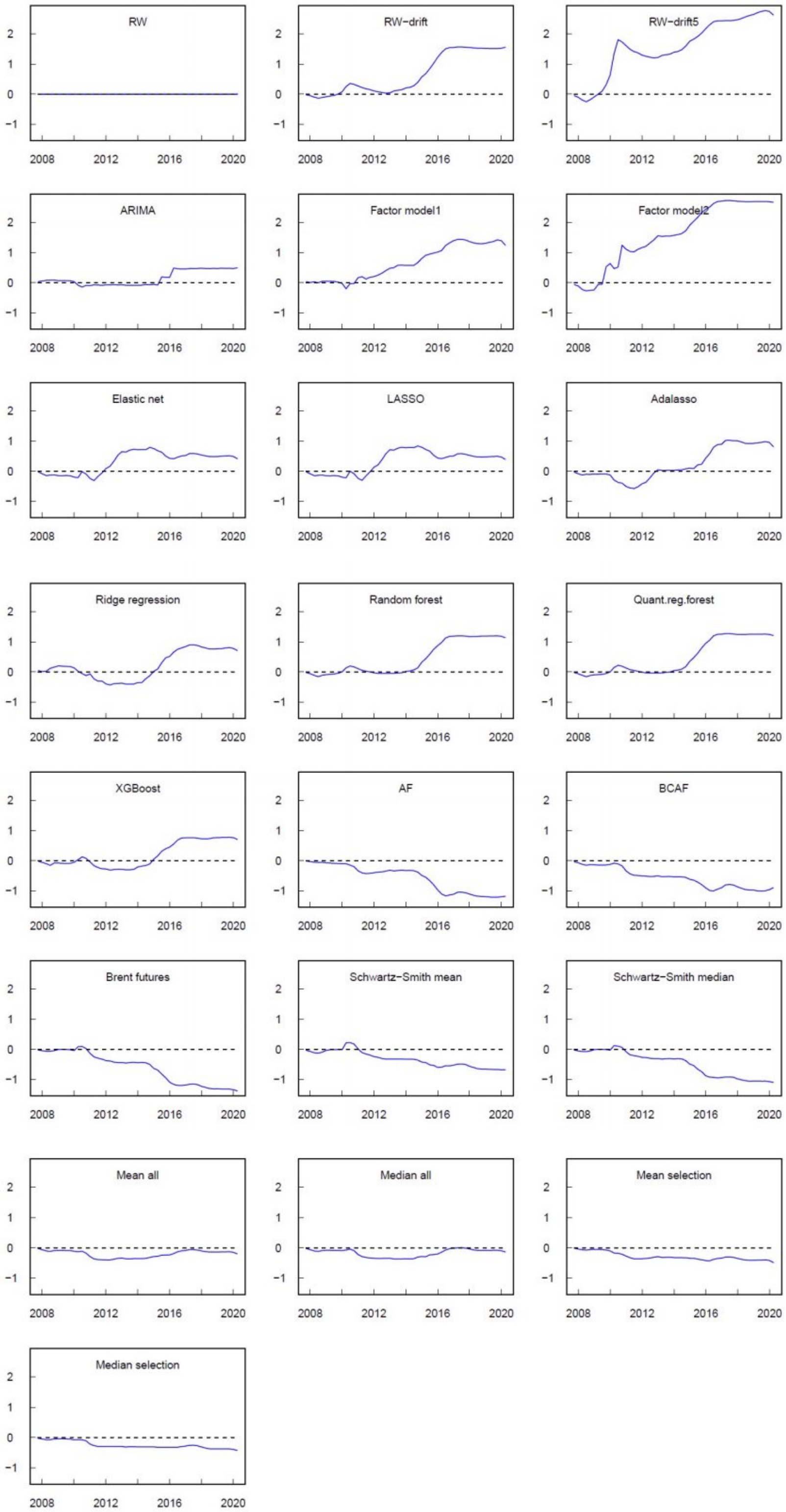
Notes: Graphs show time series plots of the differences (over time) between the Cumulative Squared Prediction Error (CSPE, divided by 10,000) of a given model and the CSPE of the benchmark model (RW). Figures above (below) zero indicate that the benchmark is better (worse).

Figure 9.7 - Cumulative Square Prediction Error ($h = 4$)



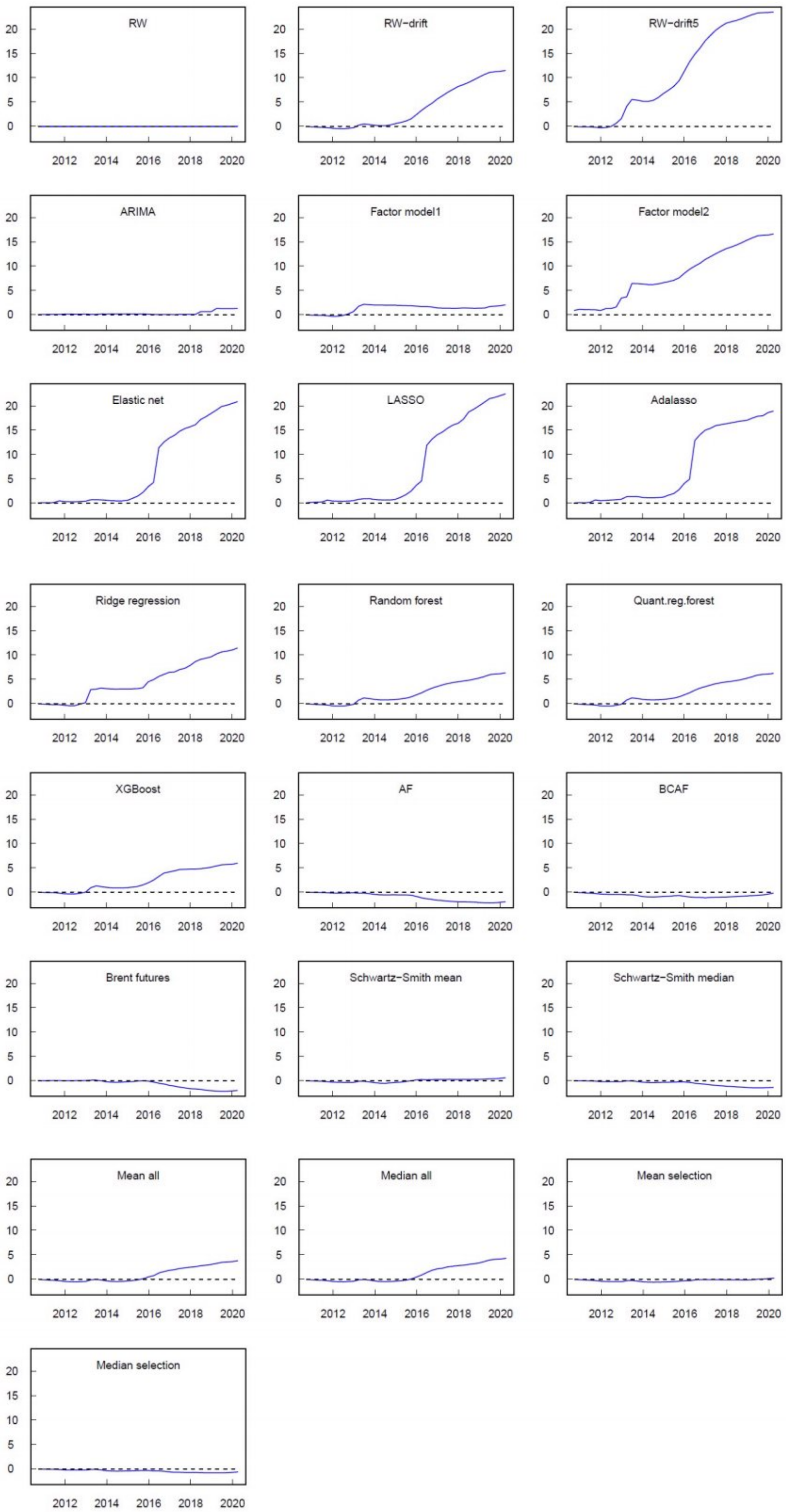
Notes: Graphs show time series plots of the differences (over time) between the Cumulative Squared Prediction Error (CSPE, divided by 10,000) of a given model and the CSPE of the benchmark model (RW). Figures above (below) zero indicate that the benchmark is better (worse).

Figure 9.8 - Cumulative Square Prediction Error ($h = 8$)



Notes: Graphs show time series plots of the differences (over time) between the Cumulative Squared Prediction Error (CSPE, divided by 10,000) of a given model and the CSPE of the benchmark model (RW). Figures above (below) zero indicate that the benchmark is better (worse).

Figure 9.9 - Cumulative Square Prediction Error ($h = 20$)



Notes: Graphs show time series plots of the differences (over time) between the Cumulative Squared Prediction Error (CSPE, divided by 10,000) of a given model and the CSPE of the benchmark model (RW). Figures above (below) zero indicate that the benchmark is better (worse).

Figure 9.10 - Number of selected variables (lasso, adalasso, elastic net)

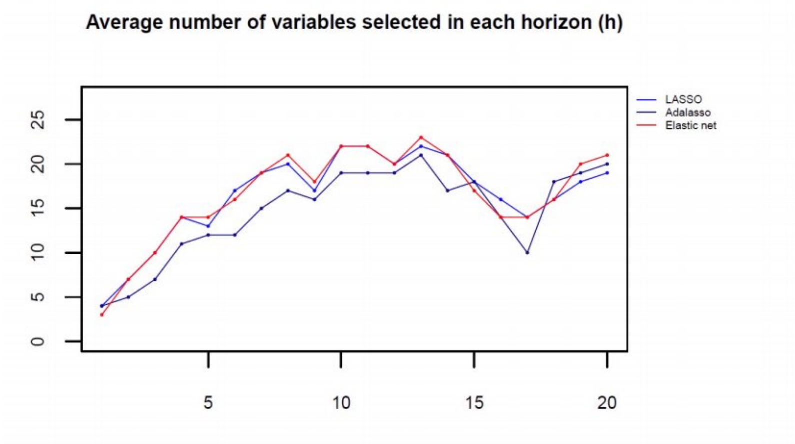


Figure 9.11 - Variable selection over time ($h = 1, 4$)

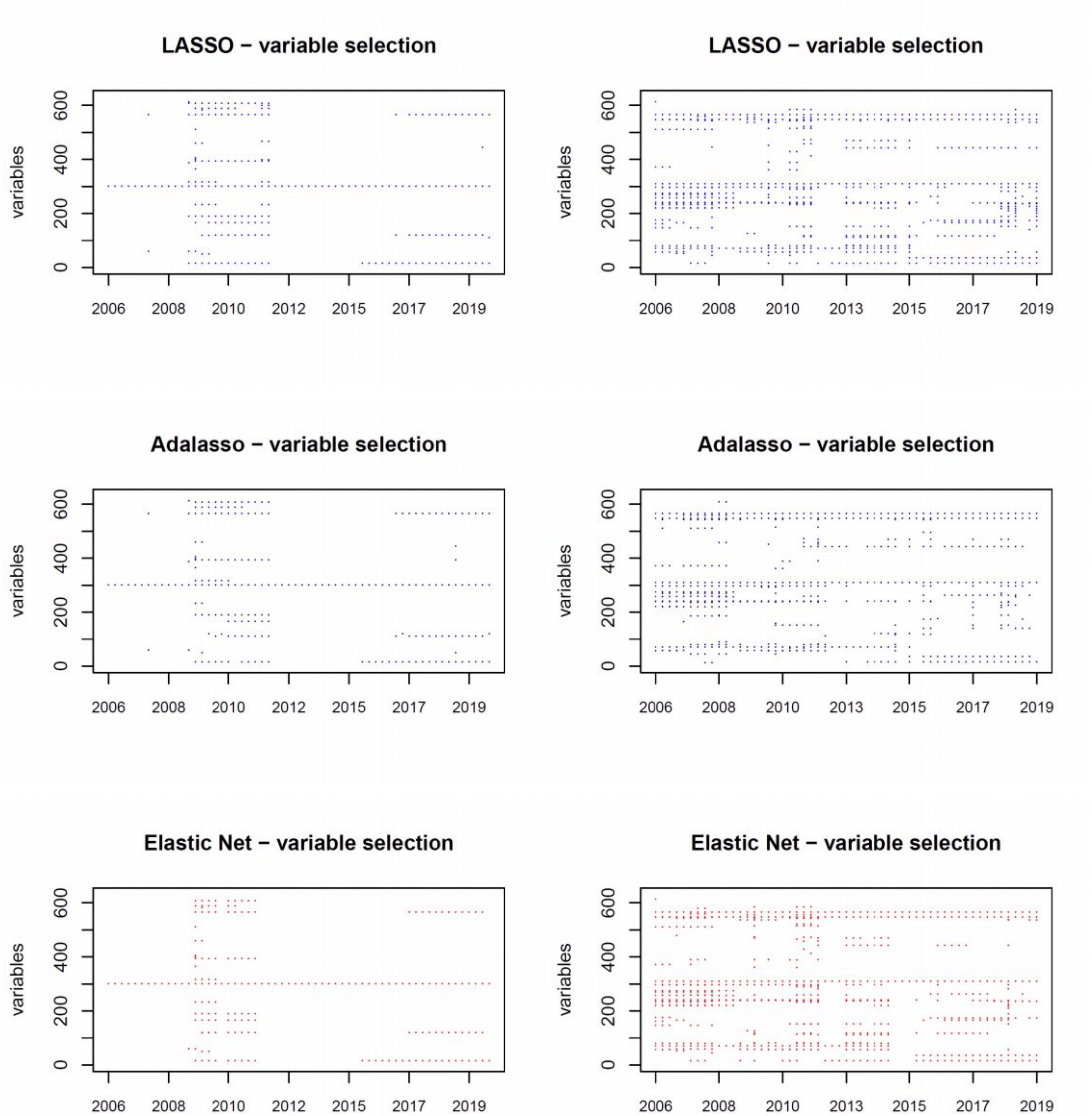


Figure 9.12 - Variable selection over time ($h = 8, 20$)

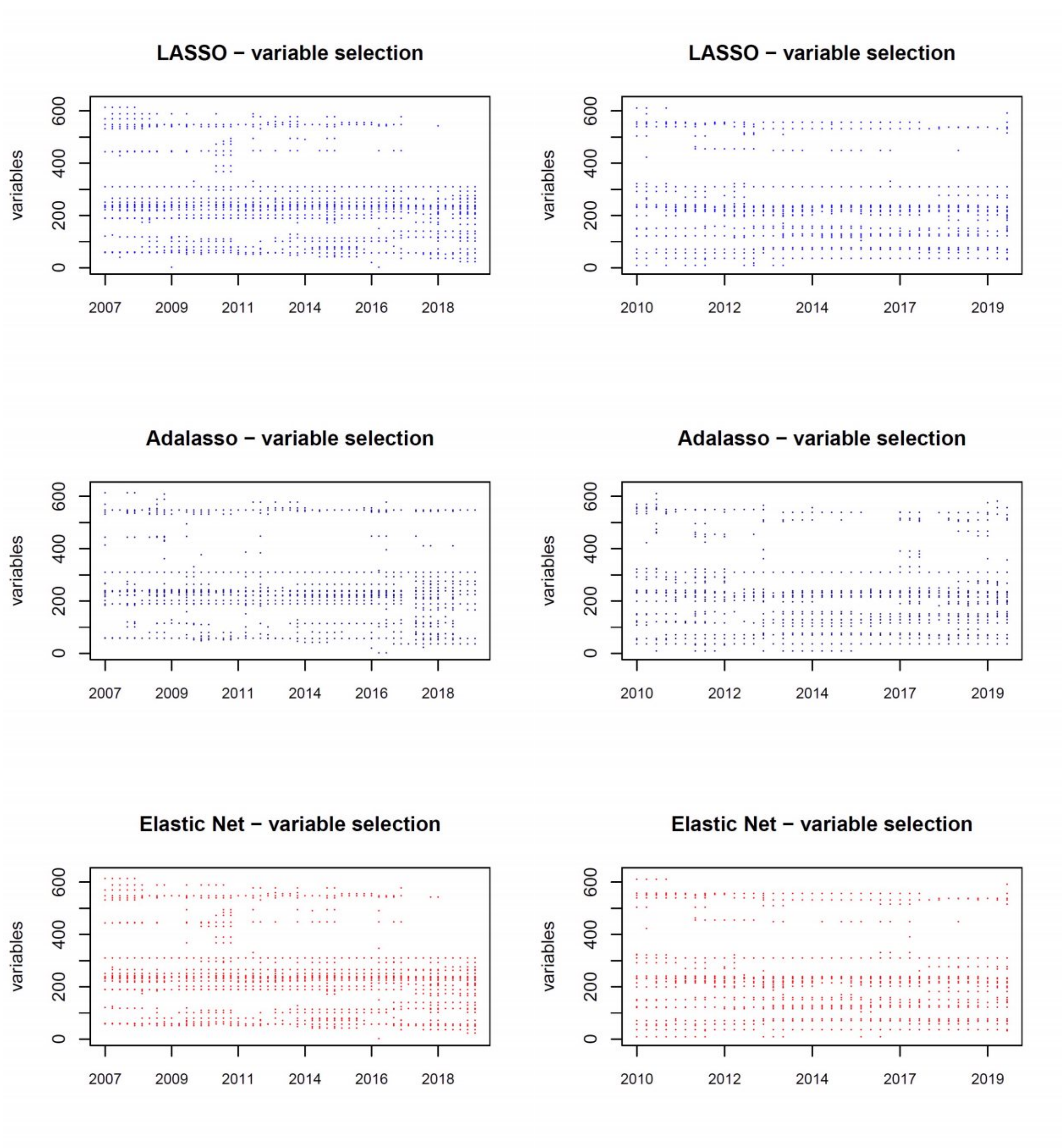


Figure 9.13 - Variable importance ($h = 1, 4$)

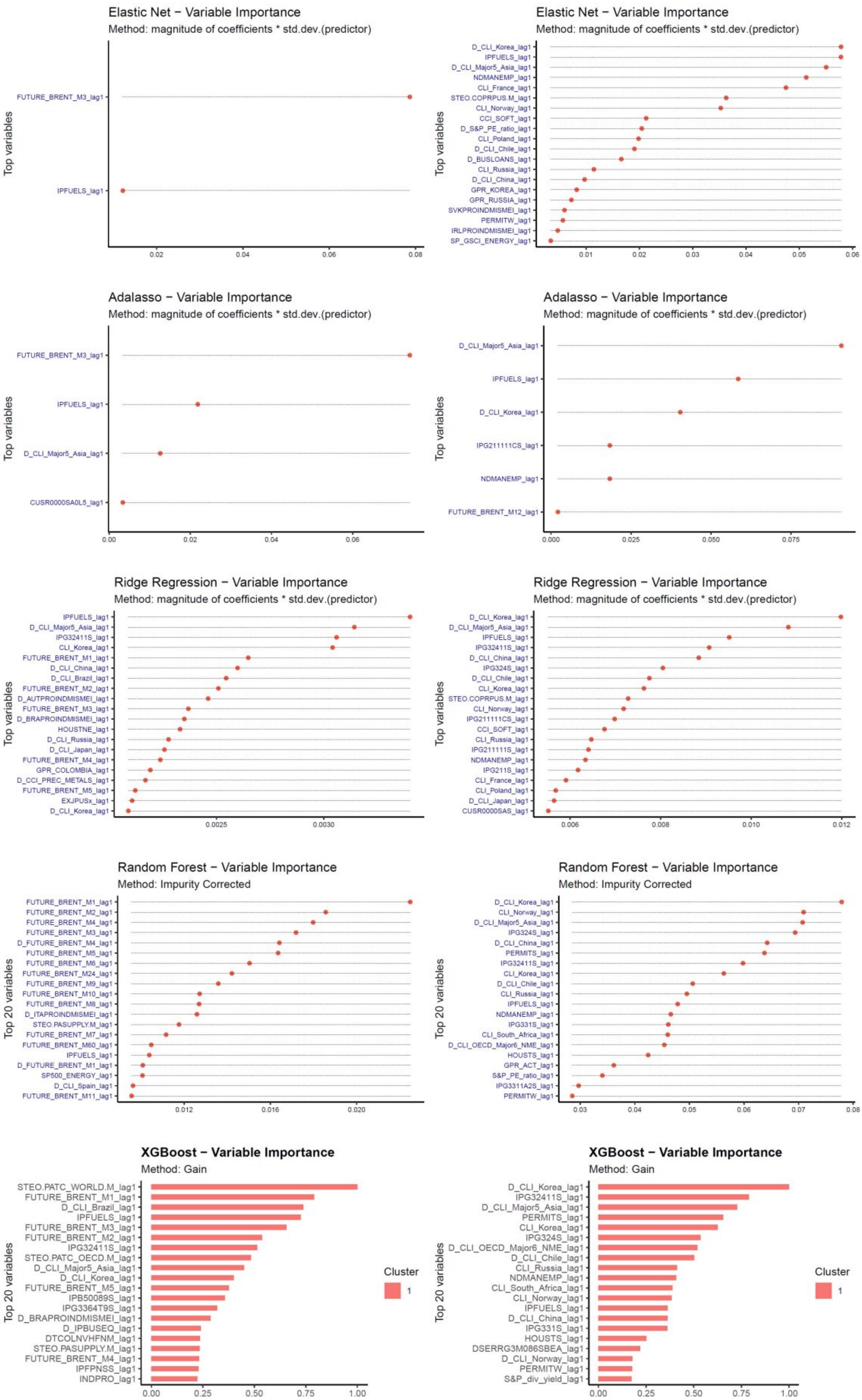


Figure 9.14 - Variable importance ($h = 8, 20$)

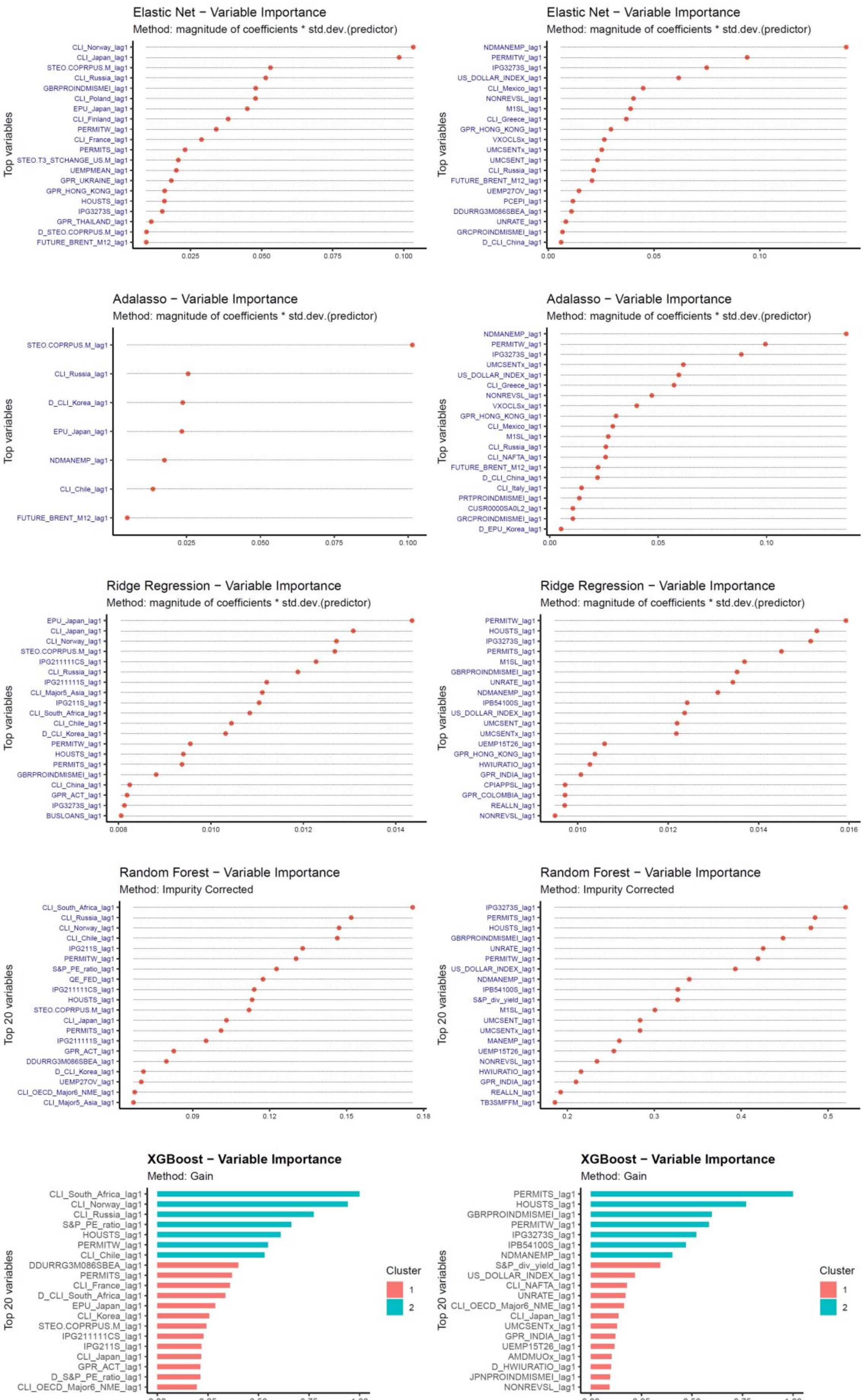
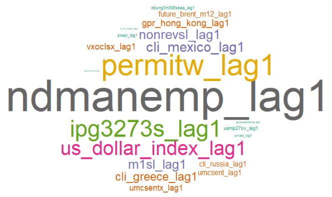
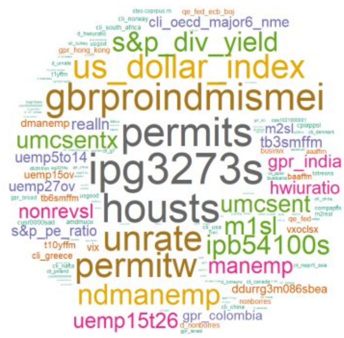


Figure 9.15 - Word clouds ($h = 20$)

Panel (a): elastic net (left) and adalasso (right)



Panel (b): ridge regression (left) and random forest (right)



Panel (c): xgboost

Figure 9.16 - Fan charts and probability density functions (PDFs)

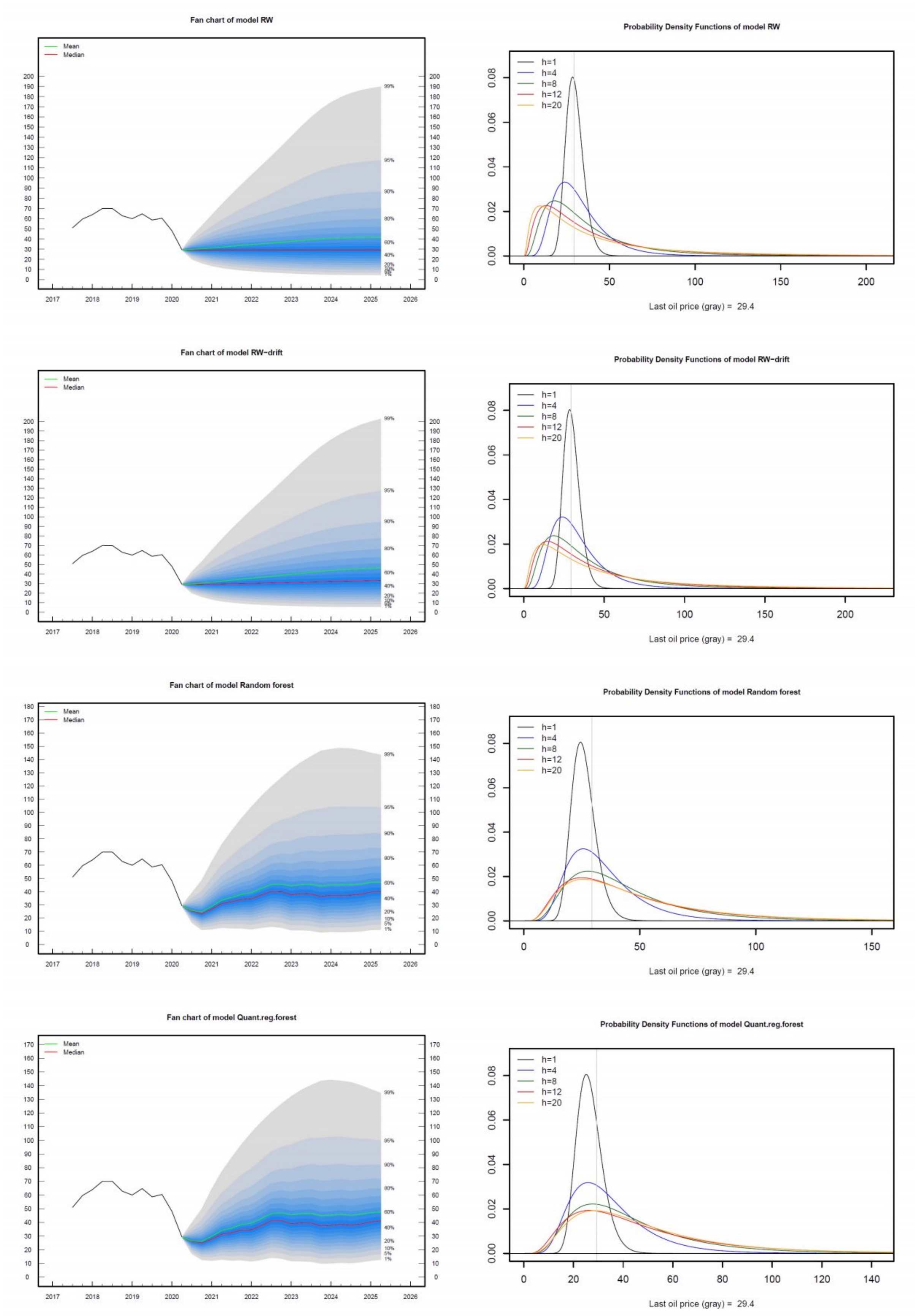


Figure 9.17 - Fan charts and probability density functions (PDFs)