FORECASTING INFLATION IN A DATA-RICH ENVIRONMENT: THE BENEFITS OF MACHINE LEARNING METHODS

Marcelo C. Medeiros
Department of Economics
Pontifical Catholic University of Rio de Janeiro
Rua Marquês de São Vicente 225, Gávea
Rio de Janeiro, 22451-900, BRAZIL
E-mail: `mcm@econ.puc-rio.br`

Álvaro Veiga
Department of Electrical Engineering
Pontifical Catholic University of Rio de Janeiro
Rua Marquês de São Vicente 225, Gávea
Rio de Janeiro, 22451-900, BRAZIL
E-mail: `alvf@ele.puc-rio.br`

Gabriel F. R. Vasconcelos
Department of Electrical Engineering
Pontifical Catholic University of Rio de Janeiro
Rua Marquês de São Vicente 225, Gávea
Rio de Janeiro, 22451-900, BRAZIL
E-mail: `gabrielrvsc@yahoo.com.br`

Eduardo Zilberman
Research Department
Central Bank of Chile
Agustinas 1180
Santiago, 867, CHILE
E-mail: `ezilberman@bcentral.cl`

JEL: C22, E31.

Keywords: Big data, inflation forecasting, LASSO, random forests, machine learning.

Acknowledgements: We thank Federico Bandi, Peter R. Hansen, Anders B. Kock and Michael Wolf for helpful comments during the workshop "Trends in Econometrics: Machine Learning, Big Data, and Financial Econometrics", Rio de Janeiro, October 2017. The participants of the "Third International Workshop in Financial Econometrics", Arraial d'Ajuda, October 2017 and at the workshop "Big data, Machine Learning and the Macroeconomy", Oslo, October 2017 are also gratefully acknowledged. Finally, we wish to thank Alberto Cavallo for an insightful discussion during the workshop "Measuring and Analyzing the Economy using Big Data" at the Central Bank of Brazil, November 2017.

Abstract: Inflation forecasting is an important but difficult task. In this paper, we explore the advances in machine learning (ML) methods and the availability of new and rich datasets to forecast US inflation over a long period of out-of-sample observations. Despite the skepticism in the previous literature, we show that ML models with a large number of covariates are systematically more accurate than the benchmarks for several forecasting horizons both in the 1990s and the 2000s. The ML method that deserves more attention is the random forest, which dominated all other models in several cases. The good performance of the random forest method is due not only to its specific method of variable selection but also the potential nonlinearities between past key macroeconomic variables and inflation. The results are robust to inflation measures, different samples, levels of macroeconomic uncertainty, and periods of recession and expansion.

1. Introduction

It is difficult to overemphasize the importance of forecasting inflation in rational economic decision-making. Many contracts concerning employment, sales, tenancy, and debt are set in nominal terms. Therefore, inflation forecasting is of great value to households, businesses and policymakers. In addition, central banks rely on inflation forecasts not only to inform monetary policy but also to anchor inflation expectations and thus enhance policy efficacy. Indeed, as part of an effort to improve economic decision-making, many central banks release their inflation forecasts on a regular basis.

Despite the great benefits of forecasting inflation accurately, improving simple benchmark models has been proven to be a major challenge for both academics and practitioners. As Stock & Watson (2010) emphasize, "it is exceedingly difficult to improve systematically upon simple univariate forecasting models, such as the Atkeson & Ohanian (2001) random walk model [...] or the time-varying unobserved components model in Stock & Watson (2007)." This conclusion is supported by a large literature; see Faust & Wright (2013) for a recent survey. Nonetheless, this literature has so far largely ignored the recent machine learning (ML) and "big data" boom in economics.[1] Moreover, previous works either focused on a restrictive set of variables or were based on a small set of factors computed from a larger number of potential predictors known as "diffusion indexes"; see, for example, Stock & Watson (2002).

"Big data" and ML methods are not passing fads, and investigating whether the combination of these two methods is able to provide more accurate forecasts is of paramount importance. Gu et al. (2018), for example, show recently that machine learning methods coupled with more than 900 potential predictors improve substantially out-of-sample stock return prediction. In a similar spirit, and despite the previous skepticism, we argue that these methods lead to more accurate inflation forecasts. Moreover, this new set of models can also help to uncover the main predictors for future inflation, possibly shedding light on the drivers of price dynamics.

In this paper, we contribute to the literature in a number of ways. First, we robustly show that it is possible to beat the usual univariate benchmarks for inflation forecasting, namely, random walk (RW), autoregressive (AR) and unobserved components stochastic volatility (UCSV) models. We consider several ML models in a data-rich environment with hundreds of variables from the FRED-MD, a monthly database put together by

[1]See Varian (2014) and Mullainathan & Spiess (2017) for discussions of ML methods and big data in economics. In this paper, we call ML models any statistical model that is able to either handle a large set of covariates and/or describe nonlinear mappings nonparametrically. Some of the methods have been around even before the "machines".

McCracken & Ng (2016), to forecast US inflation during more than twenty years of out-of-sample observations and we show that the gains can be as large as 30% in terms of mean squared errors.

Second, we highlight the main set of variables responsible for these forecast improvements. Our results indicate that such set of variables is not sparse, which corroborates the findings of Giannone et al. (2017) warning against the use of sparse predictive models. Indeed, we find that ML models that do not impose sparsity are the best performing ones. In contrast, the high level of aggregation of factor models, which has been one of the most popular models for macroeconomic forecasting, is not adequate.

Finally, we aim to give a guidance for the choice of which class of ML methods should be used for inflation forecasting. Throughout the paper, we pay special attention to a particular ML model, the random forest (RF) of Breiman (2001), which systematically outperforms the benchmarks, factor models and ten additional ML methods covering a wide class of specifications: the least absolute shrinkage and selection operator (LASSO) family, which includes LASSO, adaptive LASSO, elastic net and the adaptive elastic net; ridge regression (RR); Bayesian vector autoregressions (BVAR); and linear ensemble methods such as bagging, boosting, complete subset regressions (CSR) and jackknife model averaging (JMA). RF models are highly nonlinear nonparametric models that have a tradition in statistics but have only recently attracted attention in economics. This late success is partly due to the new theoretical results developed by Scornet et al. (2015) and Wagner & Athey (forthcoming). Notably, Gu et al. (2018) also find that RF, by allowing for nonlinearities, substantially improves stock return predictions.

1.1. **Main takeaways.** First, as mentioned before, and contrary to the previous evidence in Stock & Watson (1999, 2007), Atkeson & Ohanian (2001), and many others, our results show that consistently beating the benchmark specifications is possible. The ML models outperform the univariate alternatives, especially if we consider the 2001–2015 period, when the US inflation was more volatile compared to the 1990s. This is a robust finding for both individual horizons and the accumulated twelve-month forecasts. Second, although there is strong evidence of the existence of a small number of factors that drive the joint dynamics of the potential predictors, factor models deliver inferior forecasts compared to ML alternatives and are inferior to the RW method for the accumulated twelve-month horizon. Furthermore, either replacing standard principal component factors with target factors, as advocated by Bai & Ng (2008), or using boosting to select factors as discussed in Bai & Ng (2009), improves the results only marginally. Third, RR have a superior performance compared to the other linear ML methods, especially for short horizons. However, the RF model delivers the smallest errors for most of the forecasting horizons for both the consumer price index (CPI) and the personal consumption expenditures (PCE) inflation. The gains, in terms of mean squared error reduction, can be, on average, of the order of 30%. This is a robust finding that is independent of the sample considered, the state of the economy or the level of either macroeconomic, financial uncertainty or real uncertainty. The RF model is an ensemble of fully grown regression trees estimated on different bootstrap subsamples of the original data. Therefore, the RF model is a nonsparse, highly nonlinear specification that aims to reduce the high variance of a single regression tree.

To open the black box of ML methods, we compare the variables selected by the adaptive LASSO method, RR, and the RF alternative. Following McCracken & Ng (2016), we classify variables into nine different groups: (i) output and income; (ii) labor market; (iii) housing; (iv) consumption, orders and inventories; (v) money and credit; (vi) interest and exchange rates; (vii) prices; (viii) stock market; (ix) principal component factors computed from the full set of potential predictors. The most important variables for RR and RF models are stable across forecasting horizons but are quite different between the two specifications. While for RR, AR terms, prices and employment are the most important predictors, resembling a sort of backward-looking Phillips curve, RF models give more importance to prices, interest-exchange rates, employment and housing. LASSO selection is quite different across forecasting horizons and is, by construction and in opposition to RF and RR models, sparse. Only AR terms retain their relative importance independent of the horizon and prices gradually lose their relevance until up to six months ahead but partially recover for longer horizons. Output-income are more important for medium-term forecasts. Finally, none of the three classes of models selects either factors or stocks. Not even RR or RF which produce nonsparse variable selection. This result may indicate that the high level of cross-section aggregation of the factors is one possible cause for the poor performance of the factor models.

To disentangle the effects of variable selection from nonlinearity, we also consider alternative models. In the first method, we use the variables selected by the RF model and estimate a linear specification by OLS. In the second method, we estimate an RF specification with only the regressors selected by the adaptive LASSO method. Both models outperform the RF only for one-month-ahead forecasting. For longer horizons, the RF model is still the winner, which provides evidence that both nonlinearity and variable selection play a key role in the superiority of the RF model.

There are many sources of nonlinearities relating the variables selected and inflation that could justify the superiority of the RF model. For instance, the relationship between inflation and employment is nonlinear to the extent that it depends on the degree of slackness in the economy. Another source of nonlinearities is economic uncertainty as this uncertainty increases the option value of economic decision delays if they have an irreversible component (Bloom 2009). For example, if it is expensive to dismiss workers, hiring should be nonlinear on uncertainty. In addition, this real option argument also makes households and businesses less sensitive to changes in economic conditions when uncertainty is high. Hence, the responses of employment and inflation to interest rate decisions are arguably nonlinear on uncertainty. The presence of a zero lower bound on nominal interest rates and the implications of this bound for unconventional monetary policy is another source of nonlinearity among inflation, employment and interest rate variables (Krugman 1998, Eggertsson & Woodford 2003). Finally, to the extent that houses serve as collateral for loans, not only is monetary policy affected (Iacoviello 2005) but also a housing bubble can form, resulting in a deep credit crash (Geanakoplos 2010, Shiller 2014). Needless to say, these events are highly nonlinear and arguably have nonlinear effects on inflation, employment and interest rates.

1.2. **A brief comparison of the recent literature.** The literature on inflation forecasting is vast, and there is substantial evidence that models based on the Phillips curve do not provide good inflation forecasts. Although Stock & Watson (1999) showed that many production-related variables are potential predictors of US inflation, Atkeson & Ohanian (2001) showed that in many cases, the Phillips curve fails to beat even simple naive models. These results inspired researchers to look for different models and variables to improve inflation forecasts. Among the variables used are financial variables (Forni et al. 2003), commodity prices (Chen et al. 2014) and expectation variables (Groen et al. 2013). However, there is no systematic evidence that these models outperform the benchmarks.

More recently, due to the advancements in computational power, theoretical developments in ML, and availability of large datasets, researchers have started to consider the usage of high-dimensional models on top of the well-established (dynamic) factor models of Stock & Watson (2002), Bai & Ng (2003, 2006), and Reichlin et al. (2000, 2004). However, most of these studies have either focused only on a very small subset of ML models or presented a restrictive analysis. For example, Inoue & Kilian (2008) considered bagging, factor models and other linear shrinkage estimators in an exercise to forecast US inflation with a small set of real economic activity indicators. Their study is more limited than ours both in terms of the pool of models considered and richness of the set of predictors. Nevertheless, the authors were among the few voices that suggested that ML techniques can deliver nontrivial gains over univariate benchmarks. Medeiros & Mendes (2016) provided evidence that LASSO-based models outperform both factor and

AR benchmarks in forecasting US CPI. However, the analysis in the paper is restricted to a single ML method for just one-month-ahead inflation forecasting.

Most of the previous papers in the literature have explored only linear ML models but ignored nonlinear alternatives. The reason for this limitation is that most of the papers in the early days considered only univariate nonlinear models that were, in most cases, outperformed by simple benchmarks; see Teräsvirta et al. (2005) for an example. The message of our paper is that the combination of a rich dataset with modern ML tools is responsible for the nontrivial forecasting gains over traditional univariate benchmarks.[2]

Finally, this paper is different from many "horse-races" in the literature, as we not only compare a large number of different models but we also try to clarify the mechanisms why a given class of models is superior than others and not applying ML methods as pure black-boxes specifications.

1.3. **Organization of the paper.** The remainder of this work is organized as follows. Section 2 gives an overview of the dataset used in the paper. Section 3 describes the forecasting methodology. The results are detailed in Section 4. We start by giving a birds-eye view of the full set of results in Section 4.1, whereas in Section 4.2, we analyze the RF performance with respect to the benchmarks. In Section 4.3, we compare all the models. Section 5 concludes. The paper has a number of appendices and supplementary materials that are not for publication. Appendix A documents the dataset used in the paper. Appendix B presents an overview of the different benchmarks and ML methods/models considered in the analysis. Appendix **??** briefly describes the tests used to compare the forecasts from different models. Additional results, including the analysis of other inflation measures, are presented in Appendix C.

2. Data

Our data consist of 122 variables from the FRED-MD database, which is a large monthly macroeconomic dataset designed for empirical analysis in data-rich environments. The dataset is updated in real-time through the FRED database and is available from Michael McCraken's webpage.[3] For further details, we refer to McCracken & Ng (2016).

In this paper, we use the vintage as of January 2016. Our sample goes from January 1960 to December 2015 (672 observations), and only variables with all observations in the defined sample period are used. The out-of-sample window is from January 1990

[2]More recently, Garcia et al. (2017) applied a large number of ML methods, including RFs, to real-time inflation forecasting in Brazil. The results were very promising and indicated a clear superiority of the CSR method of Elliott et al. (2013, 2015). However, an important question is whether this is a particular result for Brazil or if similar findings can be replicated for the US economy.

[3]https://research.stlouisfed.org/econ/mccracken/fred-databases/

to December 2015. All variables are transformed as described in Appendix A. The price indexes are log-differenced only one time. Therefore, π_t is the inflation in month t computed as $\pi_t = \log(\mathsf{P}_t) - \log(\mathsf{P}_{t-1})$, and P_t is a given price index in period t. We consider two different price indexes, namely, the CPI and the PCE. Figure 1 displays the evolution of the CPI inflation rate during the full sample period.

We compare performances not only across models in the out-of-sample window but also in two subsample periods, namely, 1990 to 2000 (132 out-of-sample observations) and 2001 to 2015 (180 out-of-sample observations). In Table 1, we report the mean, standard deviation (Sd), median, maximum, minimum, first-order autocorrelation (AC1), and sum of the first 36 autocorrelations (AC36) for several macroeconomics variables. These variables include CPI monthly inflation (π_t), CPI twelve-month inflation ($\pi_{12,t}$), monthly growth of the industrial production ($\Delta\mathsf{IP}_t$), twelve-month growth of industrial production ($\Delta_{12}\mathsf{IP}_t$) and measures of macroeconomic, financial and real uncertainty computed as in Jurado et al. (2015), and broadly speaking, these measures are the conditional volatility of the unforecastable part of macroeconomic, financial and firm-level variables, respectively. In particular, the authors consider forecasting horizons of one, three and twelve months ahead.[4]

The statistics in Table 1 give an overview of the economic scenario in each subsample. The first sample corresponds to a period of low inflation volatility ($\sigma = 0.17\%$), while in the second sample, inflation is much more volatile ($\sigma = 0.32\%$). However, on average, inflation is higher during 1990-2000 than 2001-2015 and much more persistent as well. Relative to the 1990-2000 period, inflation was more volatile near the recession in the early 1990s. The monthly growth in industrial production is on average higher and less volatile during the first subsample. Finally, uncertainty measures are uniformly higher during 2001-2015, mainly due to the Great Recession.

3. Methodology

Consider the following model:

$$\pi_{t+h} = T_h(\boldsymbol{x}_t) + u_{t+h}, \quad , h = 1, \dots, H, \quad t = 1, \dots, T, \tag{1}$$

where π_{t+h} is the inflation in month $t+h$; $\boldsymbol{x}_t = (x_{1t}, \dots, x_{nt})'$ is a n-vector of covariates, possibly containing lags of π_t and/or common factors as well as a large set of potential predictors; $T_h(\cdot)$ is the mapping between covariates and future inflation; and u_t is a zero-mean random error. The target function $T_h(\boldsymbol{x}_t)$ can be a single model or an ensemble of different specifications. There is a different mapping for each forecasting horizon.

[4] These uncertainty measures are available at Sydney C. Ludvigson's webpage (https://www.sydneyludvigson.com/).

The direct forecasting equation is given by

$$\widehat{\pi}_{t+h|t} = \widehat{T}_{h,t-R_h+1:t}(\boldsymbol{x}_t), \tag{2}$$

where $\widehat{T}_{h,t-R_h+1:t}$ is the estimated target function based on data from time $t - R_h + 1$ up to t and R_h is the window size, which varies according to the forecasting horizon and the number of lagged variables in the model. We consider direct forecasts as we do not make any attempt to predict the covariates. The only exception is the case of the BVAR model, where joint forecasts for all predictors are computed in a straightforward manner following the procedure described in Bańbura et al. (2010).

The forecasts are based on a rolling window framework of fixed length. However, as mentioned before, the actual in-sample number of observations depends on the forecasting horizon. For example, for the 1990–2000 period, the number of observations is $R_h = 360 - h - p - 1$, where p is the number of lags in the model. For 2001–2015, $R_h = 492 - h - p - 1$.

In addition to three benchmark specifications (RW, AR and UCSV models), we consider factor-augmented AR models, sparsity-inducing shrinkage estimators (LASSO, adaptive LASSO, elastic net and adaptive elastic net), other shrinkage methods that do not induce sparsity (RR and BVAR with Minnesota priors), averaging (ensemble) methods (bagging, CSR and JMA)[5] and RF. With respect to the factor-augmented AR models, we consider in addition to the standard factors computed with principal component analysis a set of target factors as advocated by Bai & Ng (2008) and boosted factors as in Bai & Ng (2009). A detailed discussion of the models implemented in this paper can be found in Appendix B. Finally, we also include in the comparison three different model combination schemes, namely, simple average, trimmed average and the median of the forecasts.

We find that the RF, a highly nonlinear method, robustly outperforms other methods. To disentangle the role of variable selection from nonlinearity, we also consider a linear model where the regressors are selected by the RFs (RF/ordinary least squares, OLS) and an RF model with regressors preselected by adaptive LASSO (adaLASSO/RF).

Forecasts for the accumulated inflation over the following twelve months is computed, with the exception of the RW and UCSV models, by aggregating the individual forecasts for each horizon. In the case of the RW and UCSV models, a different specification is used to construct the forecast of the 12-month inflation.

4. Results

In this section, we describe our main results for the CPI. More detailed results, robustness checks and a similar set of results for the PCE and the CPI-core can be all found in the Appendix.

[5]Bagging and CSR can also be viewed as nonsparsity-inducing shrinkage estimators.

The models are compared according to three different statistics, namely, the root mean squared error (RMSE), the mean absolute error (MAE) and the median absolute deviation from the median (MAD), which are defined for each model and forecasting horizon as follows:

$$\mathsf{RMSE}_{m,h} = \frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T} \widehat{e}_{t,m,h}^2,$$

$$\mathsf{MAE}_{m,h} = \frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T} \left| \widehat{e}_{t,m,h} \right|,$$

$$\mathsf{MAD}_{m,h} = \mathsf{median} \left[\left| \widehat{e}_{t,m,h} - \mathsf{median} \left(\widehat{e}_{t,m,h} \right) \right| \right],$$

where $\widehat{e}_{t,m,h} = \pi_t - \widehat{\pi}_{t,m,h}$ and $\widehat{\pi}_{t,m,h}$ is the inflation forecast for month t made by model m with information up to $t-h$. The first two performance measures above are the usual ones in the forecasting literature. MAD, which is less commonly used in empirical papers, is robust to outliers.

To test whether the forecasts from different models are different, we consider a number of tests, namely, the model confidence sets (MCS) as proposed in Hansen et al. (2011), the superior predictive ability (SPA) tests of Hansen (2005), and the multi-horizon uniform SPA test of Quaedvlieg (2017).

4.1. **Overview.** In this section, we report an overview of the main findings of the paper.

Tables 2–4 report a number of statistics for each model across all the forecasting horizons, including the accumulated twelve-month horizon. The first three columns report the average RMSE, the average MAE and the average MAD. Columns (4), (5) and (6) report the number of times (out of thirteen possible horizons)[6] each model achieved the lowest RMSE, MAE, and MAD, respectively. Columns (7)–(10) present, for square and absolute losses, the average p-values of the MCS based either on the range or the $\mathsf{t}_{\mathsf{max}}$ statistics. Columns (11) and (12) show the average p-values of the SPA test for the squared and absolute errors, respectively. Finally, columns (13) and (14) display the p-value of the uniform multi-horizon test for superior predictive ability and the p-value of the MCS based on the multi-horizon comparison of the models, respectively. The uniform SPA test is designed to check for superior performance at each individual horizon. Table 2 shows the results for the full out-of-sample period (1990–2015), whereas Tables 3 and 4 present the results for the subsample periods 1990–2000 and 2001–2015, respectively. The bold figures highlight the best-performing model. The following facts emerge from the tables:

[6]To be precise, monthly inflation from one month up to twelve months ahead and yearly inflation accumulated over the following twelve months.

(1) Machine learning models and the use of a large set of predictors are able to systematically improve the quality of inflation forecasts over traditional benchmarks in the literature. This is a robust and statistically significant result.
(2) The RF model outperforms all the other alternatives in terms of point statistics. The superiority of RF is due both to the variable selection mechanism induced by the method as well as the presence of nonlinearities in the relation between inflation and its predictors. RF has the lowest RMSEs, MAEs, and MADs across the horizons and the highest MCS p-values. The RF model also has the highest p-values in the SPA test, multi-horizon SPA test and multi-horizon MCS. The improvements over the RW in terms of RMSE, MAE and MAD are almost 30% and are more pronounced during the second subsample, where inflation volatility is much higher.
(3) Shrinkage methods also produce more precise forecasts than the benchmarks. Sparsity-inducing methods are slightly worse than nonsparsity-inducing shrinkage methods. Overall, the forecasting performance among shrinkage methods is very similar, and ranking them is difficult.
(4) Factor models are strongly outperformed by other methods. The adoption of boosting and target factors improves the quality of the forecasts produced by factor models only marginally. The poor performance of factor models is more pronounced during the first subsample (low volatility period).
(5) CSR and JMA do not perform well either and are comparable to the factor models.
(6) Forecast ensembles do not bring any significant improvements in any of the performance criteria considered.
(7) In line with Stock & Watson (2007), among the benchmark models, both AR and UCSV outperform the RW alternative. Furthermore, the UCSV model is slightly superior to the AR specification.

4.2. **Results: Random Forests versus Benchmarks.** Tables 5–7 show the results of the comparison between the RF and the benchmark models. Table 5 presents the RMSE, MAE and MAD ratios of the AR, UCSV and RF models with respect to the RW alternative for all the forecasting horizons as well as for the accumulated forecasts over twelve months. The models with the smallest ratios are highlighted in bold. It is clear from the table that the RF model has the smallest ratios for all forecasting horizons.

To check whether this is a robust finding across the full out-of-sample period, we also compute rolling RMSEs, MAEs, and MADs over windows of 48 observations. Table 6 reports the results. The table shows the frequency with which each model achieved the lowest RMSEs, MAEs and MADs as well as the frequency with which each model was the worst-performing alternative among the four competitors. The RF model is the winning specification and is superior to the competitors for the majority of time periods, including the Great Recession. In contrast, the RW model delivers the worst forecasts most of the time. Figures 2, 3, and 4 show the rolling RMSEs, MAEs, and MADs, respectively, over the out-of-sample period. As expected, the performance of the RW deteriorates as the forecasting horizon increases. However, the accomplishments of the RFs seem rather robust.

Finally, Table 7 reports the p-values of the unconditional Giacomini and White (2000) test for superior predictive ability for squared (panel (a)) and absolute errors (panel (b)). Rejections of the null mean that the forecasts are significantly different. It is evident from the table that the RF has forecasts that are significantly different from and superior to the three benchmark models.

4.3. **Results: The Full Picture.** In this section, we compare all models. The main results are shown in Tables 8–10. Table 8 presents the results for the full out-of-sample period, whereas Tables 9 and 10 present the results for the 1990–2000 and 2000-2015 periods, respectively. The tables report the RMSEs and, between parenthesis, the MAEs for all models relative to the RW specification. The error measures were calculated from 132 rolling windows covering the 1990–2000 period and 180 rolling windows covering the 2001–2015 period. Values in bold denote the most accurate model in each horizon. Cells in gray (blue) show the models included in the 50% MCS using the squared error (absolute error) as the loss function. The MCSs were constructed based on the maximum t statistic. The last column in the table reports the number of forecast horizons in which the model was included in the MCS for the square (absolute) loss. The last two rows in the table report the number of models included in the MCS for the square and absolute losses.

Several conclusions come out from the tables and we start by analyzing the full out-of-sample period. Apart from a few short horizons, where either the RF/OLS or the adaLASSO/RF alternatives are the winning models, the RF alternative delivers the smallest ratios in most of the cases. The RF is followed closely by shrinkage models, where RR seems be superior to the other alternatives. RR, RF and the hybrid linear-RF models are the only specifications included in the MCS for all forecasting horizons. Neither RF nor RR impose sparsity, which may corroborate the conclusions of Giannone et al. (2017), who provide evidence against sparsity in several applications. Factor models have very poor results and are almost never included in the MCS. When factors are combined with boosting, there is a small gain, but the results are still greatly inferior to those from RF and shrinkage models. This is particularly curious as there is a correspondence between factor models and RR: RR predictions are weighed combinations of all principal component factors of the set of predictors. Several reasons might explain the difference. (1) lack of clear factor structure in the regressors. This is not the case as shown in Figure 5, where we display the eigenvalues of the covariance matrix of regressors over the forecasting period. As can be seen, there is a small number of dominating factors. (2) There might be factors which explain only a small portion of the total variance of the regressors but have a high predictive power on inflation. Again, we do not think this is the case as target factors as well as boosting are specifically designed to enhance the quality of the predictions but, in this case, do no bring any visible improvement. Furthermore, we allow the ML methods to select factors as well and, as we are going to show latter, they are never selected. (3) Finally, which we believe is the most probable explanation is that although sparsity can be questioned, factor models are a too aggregated representation of the potential predictors. The results of JMA are not encouraging either. Nevertheless, all the competing models outperform the RW for almost all horizons. Finally, forecast combination does not provide any significant gain, which can be explained by the empirical fact that most of the forecasts are positively correlated, as depicted in Figure 6.

Focusing now on the two subsamples, the following conclusions stand out from the tables. The superiority of RF is more pronounced during the 2000–2015 period, when inflation is much more volatile. During this period, RF achieves the smallest RMSE and MAE ratios for almost all horizons. From 1990-2000, the linear shrinkage methods slightly outperform the RF for short horizons. However, RF dominates for long horizons and for the twelve-month forecasts. Among the shrinkage models and during the first period, there is no clear evidence of a single winner. Depending on the horizon, different models perform the best. Another important fact is that there are fewer models included in the MSC during the first subsample.

Finally, we test whether the superiority of the RF model with respect to alternative models depends on the state of the economy. We consider two cases, namely, recessions versus expansions and high versus low macroeconomic uncertainty. The results of the test proposed by Giacomini & White (2006) for squared loss functions are presented in Tables 11 and 12. The tables report the value of the test statistic as well as the respective p-values. As usual, one, two and three asterisks represent rejection of the null hypothesis at 10%, 5%, and 1% significance levels, respectively. In Table 11, the results for expansion periods versus recessions are presented, whereas in Table 12, we consider periods of high macroeconomic uncertainty versus periods of low macroeconomic uncertainty. Periods of high (low) macroeconomic uncertainty are those where uncertainty is higher (lower) than the historical average. For conciseness, we display only the results for the most relevant models.

Inspecting the tables, it is clear that the majority of the statistics are negative, meaning that the RF model is superior than its competitors. For instance, out of 72 entries in each table, the values of the statistics are positive only in four (Table 11) and seven cases (Table 12). However, the differences are not statistically significant during recessions. This result is not surprising as only 34 of the 312 out-of-sample observations are labeled as recessions. Nevertheless, the magnitudes of the differences are much higher during recessions. Turning attention now to periods of high (low) macroeconomic uncertainty, it is evident from Table 12 that the RF model is statistically superior to the benchmark models for both periods, and as in the previous case, the differences are higher in periods of greater uncertainty. As argued above, both the degrees of slackness and uncertainty might be sources of nonlinearities in the economy. The fact that the RF model outperforms competitors in these states of the economy suggests that allowing for nonlinearities is key to improving macroeconomic forecasts.

4.4. **Opening the Black Box: Variable Selection.** In this section, we compare the predictors selected by some of the ML methods, namely, adaLASSO, ridge and RFs. We select these three models for two reasons. First, they are generally the three best-performing models, and second, they have quite different characteristics: while adaLASSO is a true sparsity-inducing method, RR and RF models are only approximately sparse. RR is a linear model, and RF is a highly nonlinear specification.

In principle, this analysis is straightforward with sparsity-inducing shrinkage methods such as the adaLASSO, as the coefficients of potentially irrelevant variables are automatically set to zero.[7] For the other ML methods, the analysis is more complex. To keep the results among models comparable, we adopt the following strategy. For ridge and adaLASSO, the relative importance measure is computed as the average coefficient size (divided by the respective standard deviations of the regressors). To measure the importance of each variable for the RF models, we use out-of-bag (OOB) samples.[8] When the bth tree is grown, the OOB samples are passed down the tree and the prediction accuracy is recorded. Then, the values of the jth variable are randomly permuted in the OOB sample, and the accuracy is again computed. The decrease in accuracy due to the permutation is averaged over all trees and is the measure of the importance of the jth variable in the RF.

[7]Medeiros & Mendes (2016) showed, for example, that under sparsity conditions, the adaLASSO model selection is consistent for high-dimensional time series models in very general settings, i.e., the method correctly selects the "true" set of regressors.

[8]For a given data point $(y_t, \boldsymbol{x}_t')$, the OOB sample is the collection of all bootstrap samples that do not include $(y_t, \boldsymbol{x}_t')$.

Due to space constraints, we cannot show the relative importance for each variable, each lag, each horizon or each estimation window. Therefore, as described in Appendix A, and following McCracken & Ng (2016), we categorize the variables, including lags, into the following eight groups: (i) output and income; (ii) labor market; (iii) housing; (iv) consumption, orders and inventories; (v) money and credit; (vi) interest and exchange rates; (vii) prices; and (viii) stock market. We also consider two additional groups, namely, the principal component factors and autoregressive terms. Furthermore, the results are averaged across all estimation windows.

Figure 7 shows the importance of each variable group for the ridge, adaLASSO, and RF methods for all the twelve forecasting horizons. For all different methods, the values in the plots are re-scaled to sum tone.

The set of the most relevant variables for RR and RF models is quite stable across forecasting horizons but is remarkably different between the two specifications. While for RR, AR terms, prices and employment are the most important predictors, RF models give more importance to prices, interest-exchange rates, employment and housing. LASSO selection is quite different across forecasting horizons, and only AR terms retain their relative importance independent of the horizon. Prices gradually lose their relevance until up to six-months-ahead and partially recover relevance when longer horizons are considered. Output-income are more important for medium-term forecasts. Finally, none of the three classes of models selects either factors or stocks. This result may indicate that the high level of cross-section aggregation of the factors is causing the poor performance.

To compare the degree of sparsity of each model, we report word clouds of the selected variables in Appendix C.1.

5. Conclusions

We show that with the recent advances in ML methods and the availability of new and rich datasets, it is possible to improve inflation forecasts. Models such as LASSO, bagging, RF and others are able to produce more accurate forecasts than the standard benchmarks. These results highlight the benefits of ML methods and rich datasets for macroeconomic forecasting. Although our paper focuses on inflation forecasting in the US, one can easily apply ML methods to forecast other macroeconomic series in a variety of countries. We leave for further research the question as to whether ML methods can systematically outperform standard methods when other macroeconomic series, such as industrial production, and countries are considered.

The RF method deserves special attention as it delivers the smallest errors for most forecasting horizons in the two out-of-sample periods (1990–1999 and 2001–2015). The good performance of the RF is due both to potential nonlinearities in the relationship between inflation and its predictors and the variable selection mechanism of such a model.

The selection of variables for RF models is quite stable across forecasting horizons. These variables are mostly selected from the following four groups of variables: prices, exchange and interest rates, housing and labor market. Although it is difficult to disentangle the precise sources of nonlinearities that the RF method uncovers, this variable selection may shed light on them. In fact, there are many theoretical reasons that nonlinearities may be induced among inflation, interest rate, labor market outcomes and housing. For example, the relationship between inflation and employment depends on the degree of slackness in the economy. In addition, as we argued above, uncertainty might induce nonlinearities among these variables. Finally, part of the out-of-sample window encompasses quarters when the zero lower bound on nominal interest rates is binding, which is another source of nonlinearity. This out-of-sample window also encompasses a period in which a housing bubble led to a credit crunch, which are events with highly nonlinear consequences.

The RF is the winning method not only in the full sample but also in the periods of expansion and recession as well as low uncertainty and high uncertainty. Relative to other methods, the RF performs particularly well in periods of high uncertainty. In addition, the RF also outperforms other methods during and after the Great Recession, when uncertainty skyrocketed and when the zero lower bound was binding. Altogether, these results suggest that the relationships among key macroeconomic variables might be highly nonlinear. If this is the case, the various linear methods widely applied in the profession not only to forecast variables but also to achieve other objectives such as approximate DSGE models might lead to inaccurate results.

Finally, in this paper, we focus on the RF model due to its flexibility and scalability for very large datasets. Nevertheless, alternative nonlinear methods such as deep learning and other semiparametric models should also be considered in future work.

References

Atkeson, A. & Ohanian, L. (2001), 'Are phillips curves useful for forecasting inflation?', *Federal Reserve bank of Minneapolis Quarterly Review* **25**, 2–11.

Bai, J. & Ng, S. (2002), 'Determine the number of factors in approximate factor model', *Econometrica* **70**, 191–221.

Bai, J. & Ng, S. (2003), 'Inferential theory for factor models of large dimensions', *Econometrica* **71**, 135–171.

Bai, J. & Ng, S. (2006), 'Confidence intervals for diffusion index forecasts and inference for factor augmented regressions', *Econometrica* **74**, 1133–1155.

Bai, J. & Ng, S. (2008), 'Forecasting economic time series using targeted predictors', *Journal of Econometrics* **146**, 304–317.

Bai, J. & Ng, S. (2009), 'Boosting diffusion indexes', *Journal of Applied Econometrics* **24**, 607–629.

Bańbura, M., Giannone, D. & Reichlin, L. (2010), 'Large Bayesian vector autoregressions', *Journal of Applied Econometrics* **25**, 71–92.

Berwin, A., Turlach, R. & Weingessel, A. (2013), *quadprog: Functions to solve Quadratic Programming Problems.* R package version 1.5-5.
URL: *https://CRAN.R-project.org/package=quadprog*

Bloom, N. (2009), 'The impact of uncertainty shocks', *Econometrica* **77**(3), 623–685.

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

Breiman, L. (2001), 'Random forests', *Machine Learning* **45**, 5–32.

Chen, Y.-C., Turnovsky, S. & Zivot, E. (2014), 'Forecasting inflation using commodity price aggregates', *Journal of Econometrics* **183**, 117–134.

Eggertsson, G. B. & Woodford, M. (2003), 'Zero bound on interest rates and optimal monetary policy', *Brookings Papers on Economic Activity* (1), 139–233.

Elliott, G., Gargano, A. & Timmermann, A. (2013), 'Complete subset regressions', *Journal of Econometrics* **177**(2), 357–373.

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.

Faust, J. & Wright, J. (2013), Forecasting inflation, *in* G. Elliott & A. Timmermann, eds, 'Handbook of Economic Forecasting', Vol. 2A, Elsevier.

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2003), 'Do financial variables help forecasting inflation and real activity in the euro area?', *Journal of Monetary Economics* **50**, 1243–1255.

Garcia, M. G., Medeiros, M. C. & Vasconcelos, G. F. (2017), 'Real-time inflation forecasting with high-dimensional models: The case of brazil', *International Journal of Forecasting* **33**(3), 679–693.

Geanakoplos, J. (2010), 'The leverage cycle', *NBER Macroeconomics Annual* **24**(1), 1–66.

Giacomini, R. & White, H. (2006), 'Tests of conditional predictive ability', *Econometrica* **74**, 1545–1578.

Giannone, D., Lenza, M. & g. Primiceri (2017), Economic predictions with big data: The illusion of sparsity, Working paper, Northwestern University.

Groen, J., Paap, R. & Ravazzolo, F. (2013), 'Real-time inflation forecasting in a changing world', *Journal of Business and Economic Statistics* **31**, 29–44.

Gu, S., Kelly, B. & Xiu, D. (2018), Empirical asset pricing with machine learning, Working paper, University of Chicago.

Hansen, B. E. & Racine, J. S. (2012), 'Jackknife model averaging', *Journal of Econometrics* **167**(1), 38–46.

Hansen, P. (2005), 'A test for superior predictive ability', *Journal of Business and Economic Statistics* **23**, 365–380.

Hansen, P. R., Lunde, A. & Nason, J. M. (2011), 'The model confidence set', *Econometrica* **79**(2), 453–497.

Hastie, T., Tibshirami, R. & Friedman, J. (2001), *The Elements of Statistical Learning; Data Mining, Inference and Prediction*, Springer.

Hoerl, A. E. & Kennard, R. W. (1970*a*), 'Ridge regression: applications to nonorthogonal problems', *Technometrics* **12**(1), 69–82.

Hoerl, A. E. & Kennard, R. W. (1970*b*), 'Ridge regression: Biased estimation for nonorthogonal problems', *Technometrics* **12**(1), 55–67.

Iacoviello, M. (2005), 'House prices, borrowing constraints, and monetary policy in the business cycle', *American Economic Review* **95**(3), 739–764.

Inoue, A. & Kilian, L. (2008), 'How useful is bagging in forecasting economic time series? a case study of U.S. CPI inflation', *Journal of the American Statistical Association* **103**, 511–522.

Jurado, K., Ludvigson, S. & Ng, S. (2015), 'Measuring uncertainty', *American Economic Review* **105**, 1177–1215.

Krugman, P. (1998), 'It's baaack: Japan's slump and the return of the liquidity trap', *Brookings Papers on Economic Activity* (2), 137–205.

McCracken, M. & Ng, S. (2016), 'FRED-MD: A monthly database for macroeconomic research', *Journal of Business and Economic Statistics* **34**, 574–589.

Medeiros, M. & Mendes, E. (2016), 'ℓ_1-regularization of high-dimensional time-series models with non-gaussian and heteroskedastic errors', *Journal of Econometrics* **191**, 255–271.

Medeiros, M. & Vasconcelos, G. (2016), 'Forecasting macroeconomic variables in data-rich environments', *Economics Letters* **138**, 50–52.

Mullainathan, S. & Spiess, J. (2017), 'Machine learning: An applied econometric approach', *Journal of Economic Perspectives* **31**, 87–106.

Quaedvlieg, R. (2017), Multi-horizon forecast comparison, Working paper, Erasmus School of Economics.

Reichlin, L., Forni, M., Hallin, M. & Lippi, M. (2000), 'The generalised dynamic factor model: Identification and estimation', *Review of Economics and Statistics* **82**, 540–554.

Reichlin, L., Forni, M., Hallin, M. & Lippi, M. (2004), 'The generalised dynamic factor model consistency and rates', *Journal of Econometrics* **119**, 231–255.

Scornet, E., Biau, G. & Vert, J.-P. (2015), 'Consistency of random forests', *Annals of Statistics* **43**, 1716–1741.

Shiller, R. J. (2014), 'Speculative asset prices', *American Economic Review* **104**(6), 1486–1517.

Stock, J. H. & Watson, M. W. (2010), Modeling inflation after the crisis, Technical report, National Bureau of Economic Research.

Stock, J. & Watson, M. (1999), 'Forecasting inflation', *Journal of Monetary Economics* **44**, 293–335.

Stock, J. & Watson, M. (2002), 'Macroeconomic forecasting with diffusion indexes', *Journal of Business and Economic Statistics* **20**, 147–162.

Stock, J. & Watson, M. (2007), 'Why has US inflation become harder to forecast?', *Journal of Money, Credit and Banking* **39**, 3–33.

Teräsvirta, T., van Dijk, D. & Medeiros, M. (2005), 'Linear models, smooth transition autoregressions and neural networks for forecasting macroeconomic time series: A re-examination (with discussion)', *International Journal of Forecasting* **21**, 755–774.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the LASSO', *Journal of the Royal Statistical Society. Series B (Methodological)* **58**, 267–288.

Varian, H. (2014), 'Big data: New tricks for econometrics', *Journal of Economic Perspectives* **28**, 3–28.

Wagner, I. & Athey, S. (forthcoming), 'Estimation and inference of heterogeneous treatment effects using random forests', *Journal of the American Statistical Association* .

Zhang, X., Wan, A. T. & Zou, G. (2013), 'Model averaging by jackknife criterion in models with dependent data', *Journal of Econometrics* **174**(2), 82–94.

Zhao, P. & Yu, B. (2006), 'On model selection consistency of lasso', *Journal of Machine learning research* **7**(Nov), 2541–2563.

Zou, H. (2006), 'The adaptive lasso and its oracle properties', *Journal of the American statistical association* **101**(476), 1418–1429.

Zou, H. & Hastie, T. (2005), 'Regularization and variable selection via the elastic net', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(2), 301–320.

TABLES AND FIGURES

Figure 1. Inflation rate (CPI, PCE and CPI core) from 1960 to 2015.

The figure shows the time evolution of the consumer price index (CPI), the personal consumption expenditures (PCE) and the core CPI inflation measures from January 1960 to December 2015 (672 observations). Inflation is computed as $\pi_t = \log(\mathsf{p}_t) - \log(\mathsf{p}_{t-1})$, where p_t represents each one of the price measures considered in this paper. Shaded areas represent recession periods.

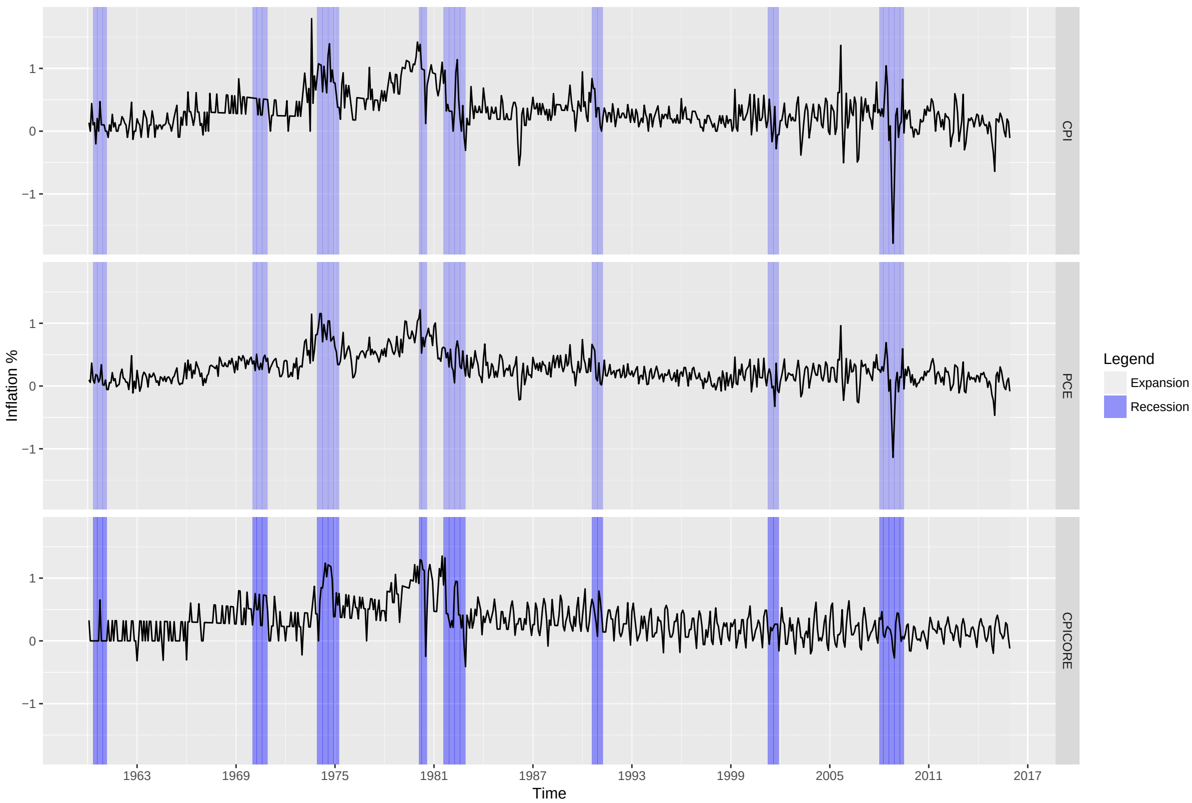


Table 1. Descriptive Statistics

The table reports the mean, standard deviation (Sd), median, maximum, minimum, first-order autocorrelation (AC1), and sum of the first 36 autocorrelations (AC36) for several macroeconomics variables, namely: CPI monthly inflation (pi_t), CPI twelve-month inflation ($\pi_{12,t}$), monthly growth of industrial production ($\Delta\mathsf{IP}_t$), twelve-month growth of industrial production ($\Delta_{12}\mathsf{IP}_t$), and measures of macroeconomic, financial and real uncertainty for one, three and twelve months.

Statistic	π_t	$\pi_{12,t}$	$\Delta\mathsf{IP}_t$	$\Delta_{12}\mathsf{IP}_t$	$\mathcal{U}_{\mathsf{fin}}(1)$	$\mathcal{U}_{\mathsf{fin}}(3)$	$\mathcal{U}_{\mathsf{fin}}(12)$	$\mathcal{U}_{\mathsf{macro}}(1)$	$\mathcal{U}_{\mathsf{macro}}(3)$	$\mathcal{U}_{\mathsf{macro}}(12)$	$\mathcal{U}_{\mathsf{real}}(1)$	$\mathcal{U}_{\mathsf{real}}(3)$	$\mathcal{U}_{\mathsf{real}}(12)$
						Full Sample: 1990–2015							
Mean	0.203	2.395	0.154	1.929	0.891	0.934	0.979	0.640	0.779	0.905	0.613	0.724	0.862
Sd	0.268	1.199	0.642	4.181	0.186	0.149	0.055	0.081	0.083	0.047	0.046	0.041	0.022
Median	0.201	2.569	0.187	2.661	0.827	0.882	0.964	0.623	0.763	0.897	0.601	0.715	0.858
Max	1.367	6.067	2.026	8.298	1.549	1.432	1.137	1.022	1.174	1.121	0.826	0.935	0.971
Min	-1.786	-1.978	-4.398	-16.721	0.641	0.729	0.903	0.551	0.686	0.847	0.539	0.661	0.826
AC1	0.415	0.926	0.245	0.973	0.987	0.988	0.992	0.987	0.989	0.992	0.969	0.979	0.985
AC36	0.812	5.079	1.727	6.980	14.358	14.743	15.879	15.460	15.535	14.906	11.495	11.640	11.960
						First Sample: 1990–2000							
Mean	0.245	2.789	0.302	3.898	0.897	0.941	0.985	0.609	0.748	0.891	0.597	0.711	0.857
Sd	0.168	0.815	0.515	2.432	0.184	0.150	0.058	0.052	0.055	0.031	0.029	0.026	0.016
Median	0.214	2.782	0.335	4.163	0.869	0.922	0.979	0.587	0.723	0.879	0.594	0.707	0.859
Max	0.946	6.067	2.026	8.298	1.277	1.232	1.081	0.777	0.902	0.955	0.673	0.771	0.889
Min	-0.058	1.367	-1.173	-3.552	0.659	0.745	0.909	0.551	0.686	0.853	0.539	0.661	0.826
AC1	0.229	0.893	0.075	0.939	0.983	0.985	0.990	0.981	0.987	0.991	0.926	0.956	0.976
AC36	1.678	5.482	1.311	9.489	21.454	21.680	22.257	15.273	16.836	16.476	6.560	8.632	10.321
						Second Sample: 2001–2015							
Mean	0.172	2.130	0.045	0.605	0.886	0.928	0.975	0.663	0.802	0.915	0.624	0.734	0.866
Sd	0.320	1.337	0.703	4.577	0.188	0.149	0.053	0.090	0.092	0.053	0.053	0.047	0.024
Median	0.181	2.074	0.120	2.066	0.812	0.869	0.952	0.643	0.779	0.900	0.608	0.720	0.858
Max	1.367	5.352	1.545	8.148	1.549	1.432	1.137	1.022	1.174	1.121	0.826	0.935	0.971
Min	-1.786	-1.978	-4.398	-16.721	0.641	0.729	0.903	0.554	0.691	0.847	0.555	0.681	0.840
AC1	0.441	0.928	0.264	0.972	0.977	0.978	0.981	0.986	0.988	0.991	0.974	0.981	0.985
AC36	0.339	4.195	0.237	1.933	8.074	8.195	8.573	11.493	11.287	11.995	9.321	9.460	10.816

Table 2. Forecasting Results: Summary statistics for the out-of-sample period from 1990–2015

The table reports for each model a number of different summary statistics across all the forecasting horizons, including the accumulated twelve-month horizon. Columns (1), (2) and (3) report the average root mean square error (RMSE), the average mean absolute error (MAE) and the average median absolute deviation (MAD). Columns (4), (5) and (6) report the number of times (across horizons) each model achieved the lowest RMSE, MAE, and MAD, respectively. Columns (7)–(10) present for square and absolute losses, the average p values based either on the range or the t_max statistics as described in Hansen et al. (2011). Columns (11) and (12) show the average p values of the superior predictive ability (SPA) test proposed by Hansen (2005). Finally, columns (13) and (14) display the p-value of the multi-horizon test for superior predictive ability proposed by Quaedvlieg (2017).

	Forecasting Precision						Model Confidence Set				Sup. Pred. Ability		Multi-Horizon SPA	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Model	ave. RMSE	ave MAE	ave. MAD	# min. RMSE	# min. MAE	# min. MAD	ave. p.v. range sq	ave. p.v. range abs	ave. p.v. Tmax sq	ave p.v. Tmax abs	ave. p.v. sq	ave. p.v. abs	p.v. model	p.v. test
RW	1.000	1.000	1.000	0	0	0	0.111	0.017	0.116	0.053	0.010	0.000	0.000	0.000
AR	0.834	0.838	0.749	0	0	0	0.276	0.047	0.541	0.235	0.056	0.008	0.000	0.000
UCSV	0.815	0.815	0.836	0	0	0	0.202	0.143	0.532	0.346	0.047	0.029	0.000	0.000
LASSO	0.780	0.786	0.724	0	0	0	0.310	0.349	0.741	0.540	0.160	0.192	0.729	0.729
adaLASSO	0.783	0.775	0.746	0	0	0	0.332	0.449	0.701	0.637	0.171	0.355	0.184	0.184
ElNet	0.780	0.791	0.720	0	0	2	0.320	0.320	0.734	0.530	0.178	0.174	0.889	0.889
adaElnet	0.781	0.775	0.738	0	0	2	0.367	0.477	0.722	0.650	0.204	0.386	0.048	0.048
Ridge	0.762	0.765	0.774	0	0	0	0.589	0.572	0.781	0.676	0.496	0.505	0.398	0.398
BVAR	0.799	0.812	0.780	0	0	0	0.278	0.355	0.713	0.502	0.135	0.145	0.036	0.036
Bagging	0.795	0.830	0.875	0	0	0	0.322	0.153	0.621	0.313	0.111	0.019	0.037	0.037
CSR	0.818	0.813	0.779	0	0	0	0.353	0.187	0.629	0.392	0.140	0.063	0.000	0.000
JMA	0.870	0.915	0.969	0	0	0	0.100	0.018	0.219	0.065	0.018	0.000	0.000	0.000
Factor	0.835	0.857	0.833	0	0	0	0.179	0.054	0.382	0.137	0.049	0.011	0.000	0.000
T. Factor	0.829	0.862	0.862	0	0	0	0.082	0.022	0.313	0.100	0.011	0.001	0.000	0.000
Boosting	0.828	0.878	0.940	0	0	0	0.234	0.102	0.635	0.278	0.027	0.004	0.000	0.000
RF	**0.737**	**0.731**	**0.692**	**9**	**11**	**6**	**0.712**	**0.749**	**0.943**	**0.963**	**0.933**	**0.942**	**1**	**1**
Mean	0.769	0.765	0.752	0	0	0	0.526	0.532	0.766	0.607	0.451	0.461	0.830	0.889
T.Mean	0.768	0.763	0.730	0	0	0	0.431	0.517	0.758	0.685	0.414	0.551	0.826	0.826
Median	0.769	0.764	0.731	0	0	0	0.415	0.502	0.751	0.674	0.367	0.509	0.804	0.889
RF/OLS	0.767	0.770	0.786	1	1	0	0.529	0.542	0.782	0.684	0.504	0.516	0.963	0.963
adaLASSO/RF	0.751	0.746	0.713	3	1	3	0.621	0.714	0.837	0.776	0.589	0.651	0.807	0.889

Table 3. Forecasting Results: Summary statistics for the out-of-sample period from 1990–2000

The table reports for each model a number of different summary statistics across all the forecasting horizons, including the accumulated twelve-month horizon. Columns (1), (2) and (3) report the average root mean square error (RMSE), the average mean absolute error (MAE) and the average median absolute deviation (MAD). Columns (4), (5) and (6) report the number of times (across horizons) each model achieved the lowest RMSE, MAE, and MAD, respectively. Columns (7)–(10) present for square and absolute losses, the average p values based either on the range or the t_max statistics as described in Hansen et al. (2011). Columns (11) and (12) show the average p values of the superior predictive ability (SPA) test proposed by Hansen (2005). Finally, columns (13) and (14) display the p-value of the multi-horizon test for superior predictive ability proposed by Quaedvlieg (2017).

	Forecasting Precision						Model Confidence Set				Sup. Pred. Ability		Multi-Horizon SPA	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Model	ave. RMSE	ave MAE	ave. MAD	# min. RMSE	# min. MAE	# min. MAD	ave. p.v. range sq	ave. p.v. range abs	ave. p.v. Tmax sq	ave p.v. Tmax abs	ave. p.v. sq	ave. p.v. abs	p.v. model	p.v. test
RW	1.000	1.000	1.000	0	0	0	0.176	0.137	0.373	0.419	0.014	0.019	0.000	0.000
AR	0.871	0.924	0.709	0	0	2	0.436	0.211	0.701	0.519	0.236	0.097	0.011	0.011
UCSV	0.879	0.889	0.844	0	0	0	0.408	0.486	0.667	0.731	0.184	0.278	0.036	0.036
LASSO	0.879	0.947	0.740	0	0	0	0.187	0.122	0.617	0.489	0.041	0.021	0.000	0.000
adaLASSO	0.838	0.864	0.772	1	1	1	0.615	0.625	0.850	0.896	0.486	0.534	0.773	0.775
ElNet	0.890	0.971	0.725	0	0	1	0.162	0.107	0.538	0.355	0.073	0.028	0.007	0.007
adaElnet	0.843	0.874	0.753	0	2	0	0.605	0.624	0.833	0.883	0.479	0.532	0.732	0.794
Ridge	0.826	0.872	0.805	2	0	0	0.539	0.576	0.809	0.770	0.482	0.463	0.594	0.794
BVAR	0.939	1.017	0.795	0	1	1	0.397	0.311	0.562	0.368	0.130	0.105	0.183	0.183
Bagging	0.907	0.943	0.949	0	0	0	0.287	0.415	0.501	0.614	0.077	0.162	0.005	0.011
CSR	0.849	0.900	0.808	2	0	0	0.396	0.333	0.716	0.605	0.306	0.218	0.794	0.794
JMA	1.055	1.092	1.105	0	0	0	0.093	0.077	0.202	0.184	0.003	0.002	0.000	0.000
Factor	0.970	1.057	0.902	0	0	0	0.075	0.087	0.284	0.185	0.009	0.016	0.000	0.000
T. Factor	1.007	1.121	0.873	0	0	0	0.024	0.009	0.163	0.077	0.001	0.000	0.000	0.000
Boosting	1.004	1.111	1.164	0	0	0	0.328	0.230	0.469	0.363	0.035	0.020	0.000	0.000
RF	**0.793**	**0.821**	**0.655**	**5**	**6**	**5**	**0.915**	**0.857**	**0.976**	**0.927**	**0.981**	**0.927**	**1**	**1**
Mean	0.820	0.862	0.771	0	0	0	0.699	0.625	0.873	0.844	0.694	0.579	0.433	0.479
T.Mean	0.825	0.870	0.745	1	0	0	0.613	0.445	0.868	0.846	0.579	0.401	0.775	0.775
Median	0.827	0.873	0.743	0	0	0	0.566	0.447	0.855	0.828	0.479	0.401	0.741	0.794
RF/OLS	0.822	0.867	0.835	0	1	0	0.691	0.623	0.880	0.857	0.653	0.510	0.479	0.479
adaLASSO/RF	0.812	0.846	0.703	2	2	3	0.780	0.726	0.878	0.876	0.725	0.646	0.127	0.794

Table 4. Forecasting Results: Summary statistics for the out-of-sample period from 2001–2015

The table reports for each model a number of different summary statistics across all the forecasting horizons, including the accumulated twelve-month horizon. Columns (1), (2) and (3) report the average root mean square error (RMSE), the average mean absolute error (MAE) and the average median absolute deviation (MAD). Columns (4), (5) and (6) report the number of times (across horizons) each model achieved the lowest RMSE, MAE, and MAD, respectively. Columns (7)–(10) present for square and absolute losses, the average p values based either on the range or the t_max statistics as described in Hansen et al. (2011). Columns (11) and (12) show the average p values of the superior predictive ability (SPA) test proposed by Hansen (2005). Finally, columns (13) and (14) display the p-value of the multi-horizon test for superior predictive ability proposed by Quaedvlieg (2017).

	Forecasting Precision						Model Confidence Set				Sup. Pred. Ability		Multi-Horizon SPA	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Model	ave. RMSE	ave MAE	ave. MAD	# min. RMSE	# min. MAE	# min. MAD	ave. p.v. range sq	ave. p.v. range abs	ave. p.v. Tmax sq	ave p.v. Tmax abs	ave. p.v. sq	ave. p.v. abs	p.v. model	p.v. test
RW	1.000	1.000	1.000	0	0	0	0.137	0.007	0.137	0.037	0.014	0.000	0.000	0.000
AR	0.828	0.807	0.794	0	0	0	0.308	0.107	0.514	0.228	0.067	0.022	0.000	0.000
UCSV	0.805	0.788	0.799	0	0	0	0.272	0.226	0.596	0.395	0.087	0.063	0.000	0.000
LASSO	0.762	0.728	0.670	0	0	2	0.483	0.702	0.876	0.798	0.347	0.630	0.972	0.992
adaLASSO	0.773	0.743	0.704	0	0	0	0.424	0.474	0.784	0.669	0.205	0.300	0.181	0.186
ElNet	0.760	0.726	0.663	0	0	**3**	0.513	0.726	0.875	0.813	0.423	0.711	0.974	0.992
adaElnet	0.771	0.739	0.699	0	0	0	0.460	0.534	0.838	0.716	0.264	0.355	0.132	0.132
Ridge	0.751	0.727	0.682	0	0	1	0.672	0.755	0.844	0.803	0.626	0.694	0.206	0.206
BVAR	0.773	0.737	0.688	0	0	1	0.475	0.690	0.861	0.822	0.290	0.488	0.766	0.766
Bagging	0.775	0.789	0.785	0	0	0	0.560	0.247	0.760	0.315	0.234	0.051	0.186	0.186
CSR	0.813	0.781	0.714	0	0	0	0.418	0.306	0.624	0.468	0.164	0.127	0.000	0.000
JMA	0.837	0.852	0.861	0	0	0	0.291	0.077	0.446	0.130	0.077	0.011	0.002	0.002
Factor	0.810	0.783	0.742	0	0	0	0.377	0.268	0.597	0.440	0.131	0.087	0.001	0.001
T. Factor	0.795	0.767	0.734	0	0	2	0.381	0.315	0.630	0.489	0.175	0.152	0.012	0.012
Boosting	0.795	0.792	0.763	0	0	1	0.467	0.399	0.821	0.582	0.147	0.087	0.002	0.002
RF	**0.728**	**0.698**	**0.696**	**9**	**8**	**3**	**0.837**	**0.862**	**0.952**	**0.962**	**0.922**	**0.909**	**1**	**1**
Mean	0.761	0.729	0.694	0	1	0	0.560	0.569	0.860	0.795	0.506	0.578	0.976	0.986
T.Mean	0.759	0.724	0.680	0	0	0	0.548	0.724	0.874	0.833	0.515	0.810	0.983	0.983
Median	0.759	0.725	0.683	0	0	0	0.539	0.650	0.870	0.824	0.474	0.719	0.986	0.986
RF/OLS	0.758	0.735	0.684	1	1	0	0.640	0.745	0.894	0.789	0.535	0.623	0.971	0.992
adaLASSO/RF	0.742	0.711	0.688	3	3	0	0.684	0.728	0.913	0.888	0.614	0.655	0.992	0.992

Table 5. Forecasting Results: RMSE, MAE and MAD Ratios (1990–2015)

The table reports the root mean squared error (RMSE), mean absolute error (MAE) and median absolute deviation from the median (MAD) ratios with respect to the random walk model for the full out-of-sample period (1990–2015). The statistics for the best-performing model are highlighted in bold.

Panel (a): RMSE Ratio

Model	Forecasting Horizon 1	2	3	4	5	6	7	8	9	10	11	12	Acc.
AR	0.902	0.809	0.790	0.805	0.786	0.791	0.783	0.764	0.779	0.824	0.837	0.753	1.218
UCSV	0.954	0.816	0.797	0.813	0.783	0.777	0.784	0.776	0.770	0.804	0.832	0.781	0.908
RF	**0.844**	**0.731**	**0.706**	**0.738**	**0.711**	**0.715**	**0.718**	**0.712**	**0.722**	**0.763**	**0.773**	**0.685**	**0.766**

Panel (b): MAE Ratio

Model	Forecasting Horizon 1	2	3	4	5	6	7	8	9	10	11	12	Acc.
AR	0.874	0.791	0.782	0.805	0.802	0.806	0.777	0.760	0.807	0.847	0.861	0.764	1.220
UCSV	0.911	0.817	0.786	0.803	0.801	0.795	0.796	0.787	0.784	0.799	0.851	0.777	0.894
RF	**0.811**	**0.721**	**0.711**	**0.749**	**0.727**	**0.728**	**0.699**	**0.681**	**0.717**	**0.747**	**0.767**	**0.668**	**0.774**

Panel (c): MAD Ratio

Model	Forecasting Horizon 1	2	3	4	5	6	7	8	9	10	11	12	Acc.
AR	0.738	0.703	0.815	0.822	0.828	0.755	0.664	0.685	0.767	0.697	0.769	0.600	0.889
UCSV	0.876	0.770	0.832	0.906	0.878	0.790	0.761	0.835	0.857	0.829	0.884	0.777	0.876
RF	**0.698**	**0.633**	**0.772**	**0.841**	**0.750**	**0.728**	**0.653**	**0.639**	**0.728**	**0.685**	**0.706**	**0.575**	**0.587**

Table 6. Forecasting Results: Ranking of Models (1990–2015)

The table reports the frequency with which each model achieved the best (worst) performance statistics over a rolling window period of four years (48 observations).

Panel (a): Lowest Rolling RMSE

	Forecasting Horizon												
Model	1	2	3	4	5	6	7	8	9	10	11	12	Acc.
RW	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.038	0.000	0.083
AR	0.083	0.049	0.000	0.158	0.011	0.011	0.098	0.177	0.117	0.128	0.113	0.000	0.000
UCSV	0.023	0.049	0.211	0.098	0.181	0.109	0.004	0.030	0.192	0.109	0.094	0.000	0.236
RF	**0.894**	**0.902**	**0.789**	**0.743**	**0.808**	**0.879**	**0.898**	**0.792**	**0.691**	**0.762**	**0.755**	**1.000**	**0.681**

Panel (b): Lowest Rolling MAE

	Forecasting Horizon												
Model	1	2	3	4	5	6	7	8	9	10	11	12	Acc.
RW	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.094	0.000	0.039
AR	0.166	0.034	0.000	0.049	0.000	0.023	0.098	0.136	0.132	0.064	0.049	0.023	0.000
UCSV	0.151	0.177	0.257	0.226	0.151	0.155	0.000	0.072	0.242	0.226	0.094	0.023	0.201
RF	**0.683**	**0.789**	**0.743**	**0.725**	**0.849**	**0.823**	**0.902**	**0.792**	**0.626**	**0.709**	**0.762**	**0.955**	**0.760**

Panel (c): Lowest Rolling MAD

	Forecasting Horizon												
Model	1	2	3	4	5	6	7	8	9	10	11	12	Acc.
RW	0.136	0.000	0.034	0.000	0.091	0.045	0.004	0.000	0.034	0.053	0.026	0.045	0.020
AR	0.234	0.162	0.234	0.230	0.121	0.234	0.264	0.321	0.109	0.147	0.423	0.234	0.039
UCSV	0.038	0.192	0.268	0.328	0.094	0.117	0.034	0.008	0.219	0.087	0.034	0.019	0.051
RF	**0.592**	**0.645**	**0.464**	**0.442**	**0.694**	**0.604**	**0.698**	**0.672**	**0.638**	**0.713**	**0.517**	**0.702**	**0.890**

Panel (d): Highest Rolling RMSE

	Forecasting Horizon												
Model	1	2	3	4	5	6	7	8	9	10	11	12	Acc.
RW	**0.823**	**0.985**	**0.936**	**0.992**	**0.977**	**0.996**	**1.000**	**1.000**	**0.864**	**0.804**	**0.713**	**0.849**	**0.000**
AR	0.000	0.004	0.057	0.008	0.023	0.004	0.000	0.000	0.136	0.189	0.287	0.151	0.969
UCSV	0.177	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.031
RF	0.000	0.011	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.008	0.000	0.000	0.000

Panel (e): Highest Rolling MAE

	Forecasting Horizon												
Model	1	2	3	4	5	6	7	8	9	10	11	12	Acc.
RW	**0.898**	**0.943**	**0.864**	**1.000**	**0.992**	**0.966**	**1.000**	**0.943**	**0.819**	**0.728**	**0.687**	**0.770**	**0.031**
AR	0.083	0.034	0.128	0.000	0.008	0.034	0.000	0.057	0.181	0.272	0.283	0.230	0.862
UCSV	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.106
RF	0.000	0.023	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000

Panel (f): Highest Rolling MAD

	Forecasting Horizon												
Model	1	2	3	4	5	6	7	8	9	10	11	12	Acc.
RW	**0.683**	**0.940**	**0.853**	**0.804**	**0.796**	**0.657**	**0.811**	**0.943**	**0.774**	**0.804**	**0.789**	**0.921**	**0.315**
AR	0.053	0.026	0.117	0.098	0.102	0.068	0.042	0.019	0.098	0.034	0.019	0.008	0.512
UCSV	0.215	0.034	0.026	0.098	0.042	0.260	0.140	0.038	0.106	0.143	0.192	0.072	0.173
RF	0.049	0.000	0.004	0.000	0.060	0.015	0.008	0.000	0.023	0.019	0.000	0.000	0.000

Table 7. Forecasting Results: Superior Predictive Ability Test (1990–2015)

The table reports the p-values of the unconditional Giacomini-White test for superior predictive ability between the random forest models and each of the benchmark models. The test is based on the full out-of-sample period. Panel (a) presents the results for squared errors, while panel (b) shows the results for absolute errors.

Panel (a): Giacomini-White Test (Sq. Errors)

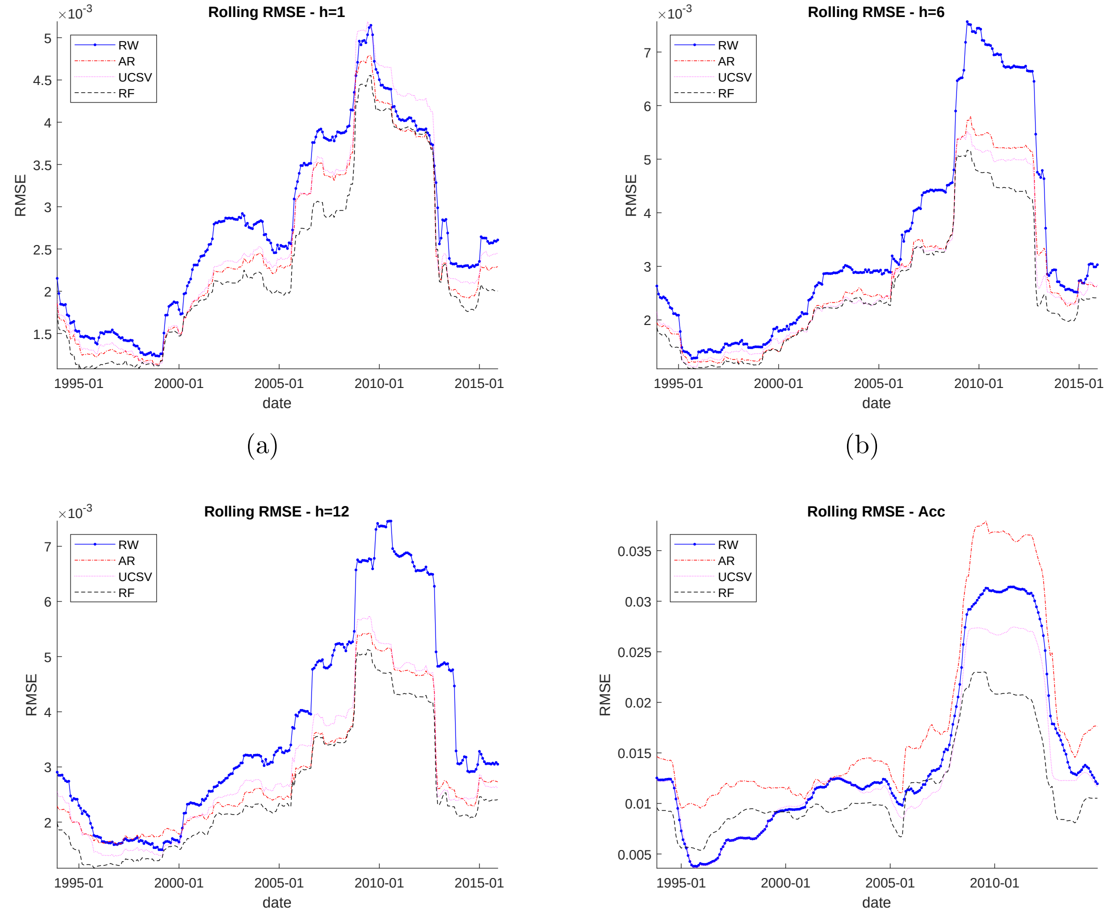
	Forecasting Horizon												
Model	1	2	3	4	5	6	7	8	9	10	11	12	Acc.
RW	0.003	0.000	0.000	0.001	0.006	0.012	0.010	0.003	0.003	0.027	0.024	0.001	0.049
AR	0.002	0.010	0.023	0.045	0.024	0.024	0.056	0.075	0.047	0.062	0.008	0.000	0.021
UCSV	0.003	0.003	0.013	0.055	0.055	0.024	0.001	0.000	0.003	0.038	0.002	0.000	0.072

Panel (b): Giacomini-White Test (Abs. Errors)

	Forecasting Horizon												
Model	1	2	3	4	5	6	7	8	9	10	11	12	Acc.
RW	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.023
AR	0.000	0.000	0.001	0.012	0.002	0.005	0.009	0.017	0.007	0.004	0.000	0.000	0.000
UCSV	0.000	0.000	0.010	0.029	0.003	0.007	0.000	0.000	0.007	0.054	0.008	0.000	0.078

FIGURE 2. Rolling RMSE.

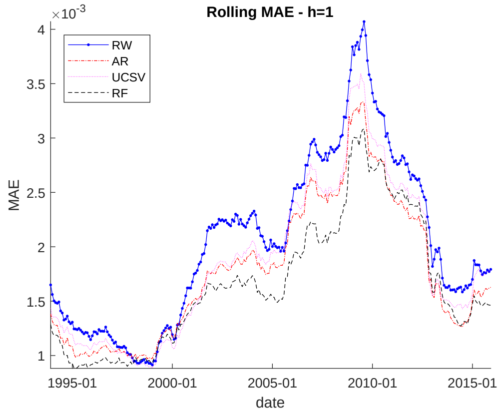
The figure displays the root mean squared errors (RMSE) computed over rolling windows of 48 observations. Panel (a) displays the results for one-month-ahead forecasts ($h = 1$), panel (b) displays the results for six-months-ahead forecasts ($h = 6$), panel (c) displays the results for twelve-months-ahead forecasts ($h = 12$), and finally, Panel (d) displays the results for the accumulated twelve month forecasts.



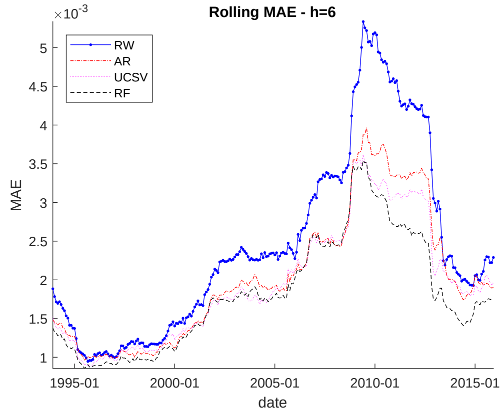
(a) (b) (c) (d)

Figure 3. Rolling MAE.

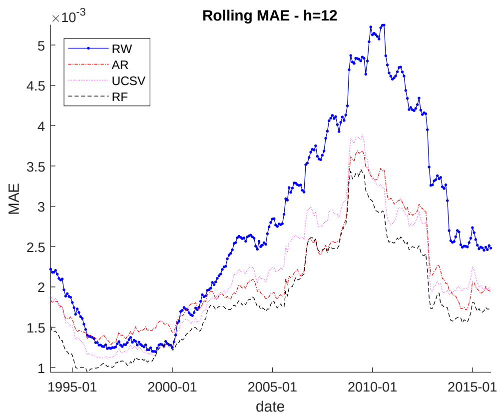
The figure displays the mean absolute errors (MAE) computed over rolling windows of 48 observations. Panel (a) displays the results for one-month-ahead forecasts ($h = 1$), panel (b) displays the results for six-months-ahead forecasts ($h = 6$), panel (c) displays the results for twelve-months-ahead forecasts ($h = 12$), and finally, panel (d) displays the results for the accumulated twelve month forecasts.



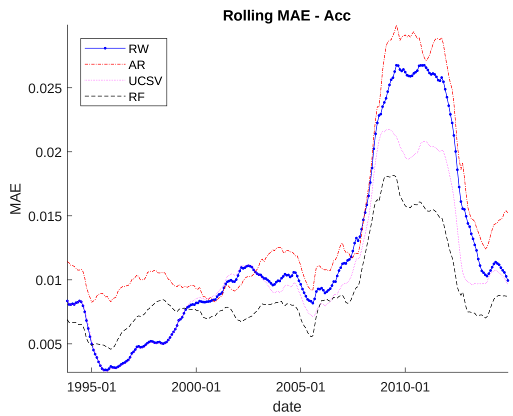
(a)



(b)



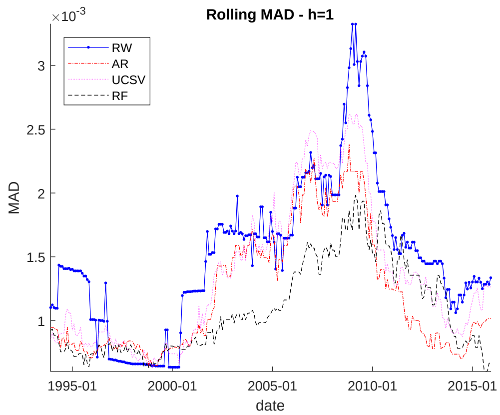
(c)



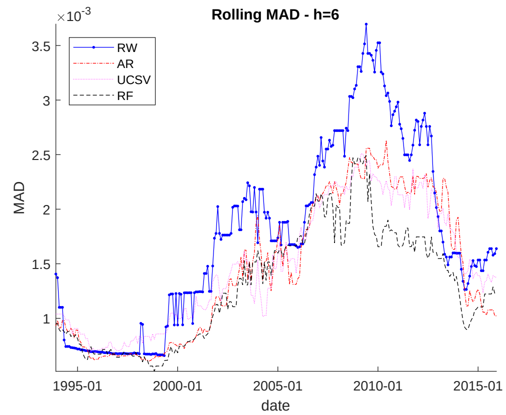
(d)

Figure 4. Rolling MAD.

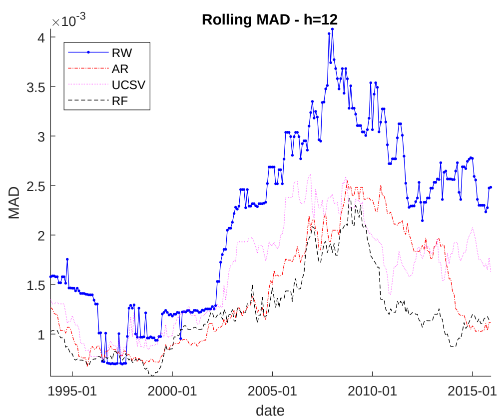
The figure displays the mean absolute deviation from the median (MAD) computed over rolling windows of 48 observations. Panel (a) displays the results for one-month-ahead forecasts ($h = 1$), panel (b) displays the results for six-months-ahead forecasts ($h = 6$), Panel (c) displays the results for twelve-months-ahead forecasts ($h = 12$), and finally, panel (d) displays the results for the accumulated twelve month forecasts.



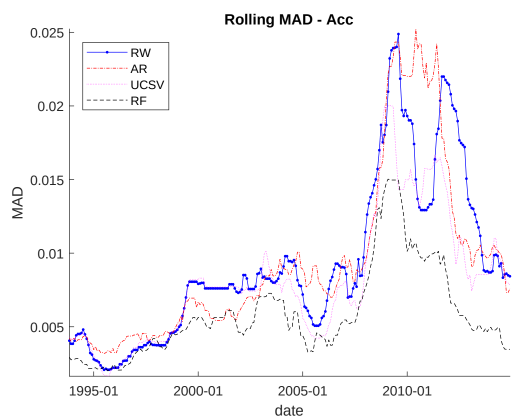
(a)



(b)



(c)



(d)

Table 8. Forecasting Errors for the CPI from 1990 to 2015

The table shows the root mean squared error (RMSE) and, between parenthesis, the mean absolute errors (MAE) for all models relative to the random walk (RW). The error measures were calculated from 132 rolling windows covering the 1990-2000 period and 180 rolling windows covering the 2001-2015 period. Values in bold show the most accurate model in each horizon. Cells in gray (blue) show the models included in the 50% model confidence set (MCS) using the squared error (absolute error) as loss function. The MCSs were constructed based on the maximum t-statistic. The last column in the table reports in how many horizons the row model was included in the MCS for square (absolute) loss. The last two rows in the table report how many models were included in the MCS for square and absolute losses.

	Consumer Price Index 1990–2015													
	Forecasting Horizon													
RMSE/(MAE)	1	2	3	4	5	6	7	8	9	10	11	12	Acc.	RMSE count (MAE count)
RW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1)
AR	0.90	0.81	0.79	0.81	0.79	0.79	0.78	0.76	0.78	0.82	0.84	0.75	1.22	9
	(0.87)	(0.79)	(0.78)	(0.81)	(0.80)	(0.81)	(0.78)	(0.76)	(0.81)	(0.85)	(0.86)	(0.76)	(1.22)	(0)
UCSV	0.95	0.82	0.80	0.81	0.78	0.78	0.78	0.78	0.77	0.80	0.83	0.78	0.91	8
	(0.91)	(0.82)	(0.79)	(0.80)	(0.80)	(0.79)	(0.80)	(0.79)	(0.78)	(0.80)	(0.85)	(0.78)	(0.89)	(3)
LASSO	0.83	0.75	0.73	0.76	0.74	0.75	0.75	0.73	0.75	0.80	0.82	0.73	0.98	11
	(0.82)	(0.74)	(0.73)	(0.78)	(0.77)	(0.75)	(0.74)	(0.71)	(0.76)	(0.81)	(0.84)	(0.74)	(1.04)	(9)
adaLASSO	0.84	0.76	0.74	0.77	0.75	0.75	0.76	0.75	0.76	0.80	0.83	0.72	0.96	11
	(0.81)	(0.75)	(0.72)	(0.77)	(0.75)	(0.74)	(0.73)	(0.71)	(0.75)	(0.79)	(0.84)	(0.73)	(0.96)	(11)
ElNet	0.83	0.75	0.73	0.76	0.75	0.74	0.75	0.74	0.76	0.81	0.82	0.73	0.98	11
	(0.82)	(0.74)	(0.73)	(0.78)	(0.78)	(0.76)	(0.75)	(0.71)	(0.77)	(0.81)	(0.85)	(0.75)	(1.05)	(9)
adaElnet	0.84	0.75	0.73	0.77	0.75	0.75	0.75	0.74	0.76	0.80	0.81	0.73	0.96	11
	(0.82)	(0.74)	(0.72)	(0.76)	(0.75)	(0.74)	(0.73)	(0.71)	(0.75)	(0.79)	(0.83)	(0.75)	(0.97)	(11)
Ridge	0.85	0.73	0.72	0.75	0.74	0.75	0.75	0.73	0.74	0.77	0.78	0.70	0.89	13
	(0.83)	(0.72)	(0.72)	(0.77)	(0.76)	(0.76)	(0.73)	(0.71)	(0.74)	(0.77)	(0.79)	(0.71)	(0.93)	(13)
BVAR	0.86	0.76	0.75	0.77	0.74	0.76	0.77	0.76	0.77	0.82	0.83	0.74	1.07	11
	(0.87)	(0.73)	(0.75)	(0.79)	(0.78)	(0.78)	(0.76)	(0.76)	(0.81)	(0.83)	(0.85)	(0.76)	(1.09)	(8)
Bagging	0.83	0.76	0.76	0.80	0.78	0.79	0.83	0.81	0.78	0.82	0.83	0.74	0.82	11
	(0.84)	(0.78)	(0.79)	(0.87)	(0.86)	(0.85)	(0.83)	(0.80)	(0.80)	(0.84)	(0.86)	(0.78)	(0.88)	(4)
CSR	0.85	0.77	0.76	0.79	0.77	0.79	0.79	0.77	0.79	0.83	0.84	0.76	1.13	11
	(0.84)	(0.76)	(0.75)	(0.79)	(0.79)	(0.79)	(0.76)	(0.74)	(0.79)	(0.83)	(0.84)	(0.77)	(1.11)	(3)
JMA	0.99	0.82	0.84	0.85	0.84	0.81	0.91	0.86	0.84	0.95	0.92	0.80	0.88	2
	(0.99)	(0.85)	(0.89)	(0.94)	(0.96)	(0.90)	(0.91)	(0.87)	(0.93)	(0.96)	(0.96)	(0.83)	(0.91)	(1)
Factor	0.87	0.78	0.78	0.79	0.78	0.78	0.80	0.81	0.82	0.84	0.84	0.78	1.17	4
	(0.88)	(0.80)	(0.80)	(0.82)	(0.82)	(0.80)	(0.78)	(0.80)	(0.87)	(0.87)	(0.87)	(0.82)	(1.21)	(0)
T. Factor	0.88	0.79	0.78	0.80	0.77	0.79	0.79	0.80	0.80	0.82	0.83	0.78	1.17	3
	(0.87)	(0.82)	(0.81)	(0.84)	(0.83)	(0.84)	(0.80)	(0.80)	(0.84)	(0.87)	(0.86)	(0.80)	(1.23)	(0)
Boosting	0.95	0.77	0.76	0.78	0.77	0.79	0.79	0.78	0.79	0.83	0.84	0.74	1.17	10
	(0.96)	(0.80)	(0.81)	(0.85)	(0.84)	(0.86)	(0.84)	(0.82)	(0.85)	(0.86)	(0.86)	(0.75)	(1.32)	(1)
RF	0.84	**0.73**	**0.71**	0.74	**0.71**	**0.72**	**0.72**	0.71	0.72	**0.76**	**0.77**	**0.68**	**0.77**	13
	(0.81)	**(0.72)**	**(0.71)**	(0.75)	**(0.73)**	**(0.73)**	**(0.70)**	**(0.68)**	**(0.72)**	**(0.75)**	**(0.77)**	**(0.67)**	**(0.77)**	(13)
Mean	0.83	0.75	0.73	0.76	0.74	0.74	0.75	0.74	0.75	0.77	0.78	0.71	0.95	12
	(0.81)	(0.74)	(0.73)	(0.76)	(0.76)	(0.75)	(0.73)	(0.71)	(0.75)	(0.76)	(0.78)	(0.70)	(0.97)	(12)
T.Mean	0.84	0.74	0.73	0.75	0.74	0.74	0.75	0.73	0.74	0.78	0.79	0.71	0.95	12
	(0.81)	(0.74)	(0.72)	(0.76)	(0.75)	(0.74)	(0.72)	(0.70)	(0.74)	(0.77)	(0.79)	(0.70)	(0.96)	(12)
Median	0.84	0.75	0.72	0.76	0.74	0.74	0.75	0.73	0.74	0.78	0.79	0.71	0.94	12
	(0.81)	(0.74)	(0.72)	(0.76)	(0.76)	(0.74)	(0.73)	(0.70)	(0.74)	(0.77)	(0.79)	(0.71)	(0.97)	(12)
RF/OLS	**0.81**	0.73	0.72	0.75	0.74	0.75	0.75	0.74	0.74	0.78	0.79	0.71	0.94	13
	(0.79)	(0.73)	(0.72)	(0.76)	(0.76)	(0.76)	(0.73)	(0.72)	(0.75)	(0.78)	(0.81)	(0.72)	(0.97)	(13)
adaLASSO/RF	0.85	0.76	0.72	**0.73**	0.73	0.72	0.72	**0.71**	**0.72**	0.79	0.82	0.70	0.80	13
	(0.82)	(0.73)	(0.72)	**(0.74)**	(0.74)	(0.73)	(0.71)	(0.68)	(0.72)	(0.79)	(0.82)	(0.68)	(0.82)	(13)
RMSE count	14	15	16	17	19	19	17	15	17	18	19	8	8	
MAE count	(12)	(11)	(12)	(11)	(10)	(12)	(13)	(11)	(10)	(15)	(16)	(7)	(9)	

FIGURE 5. Eigenvalues of the matrix of contemporaneous regressor.

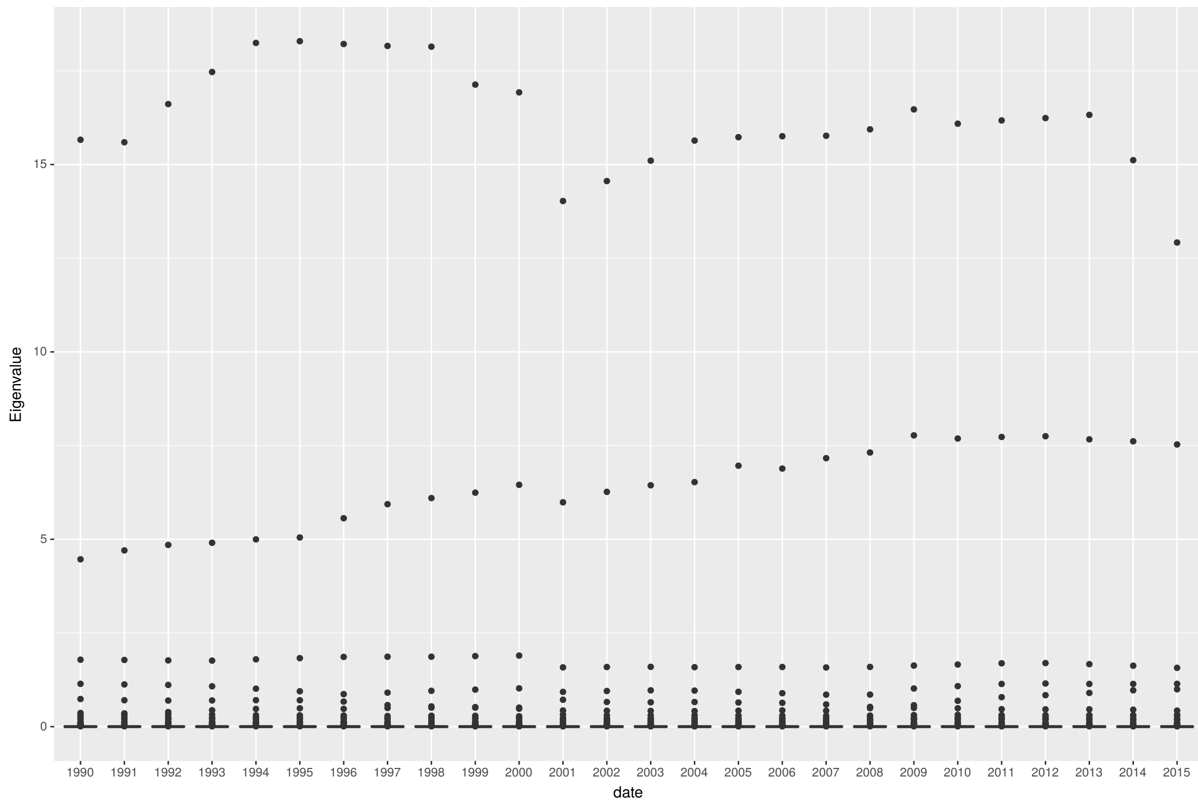


Figure 6. Correlation of the Forecasts for the CPI from 1990 to 2015

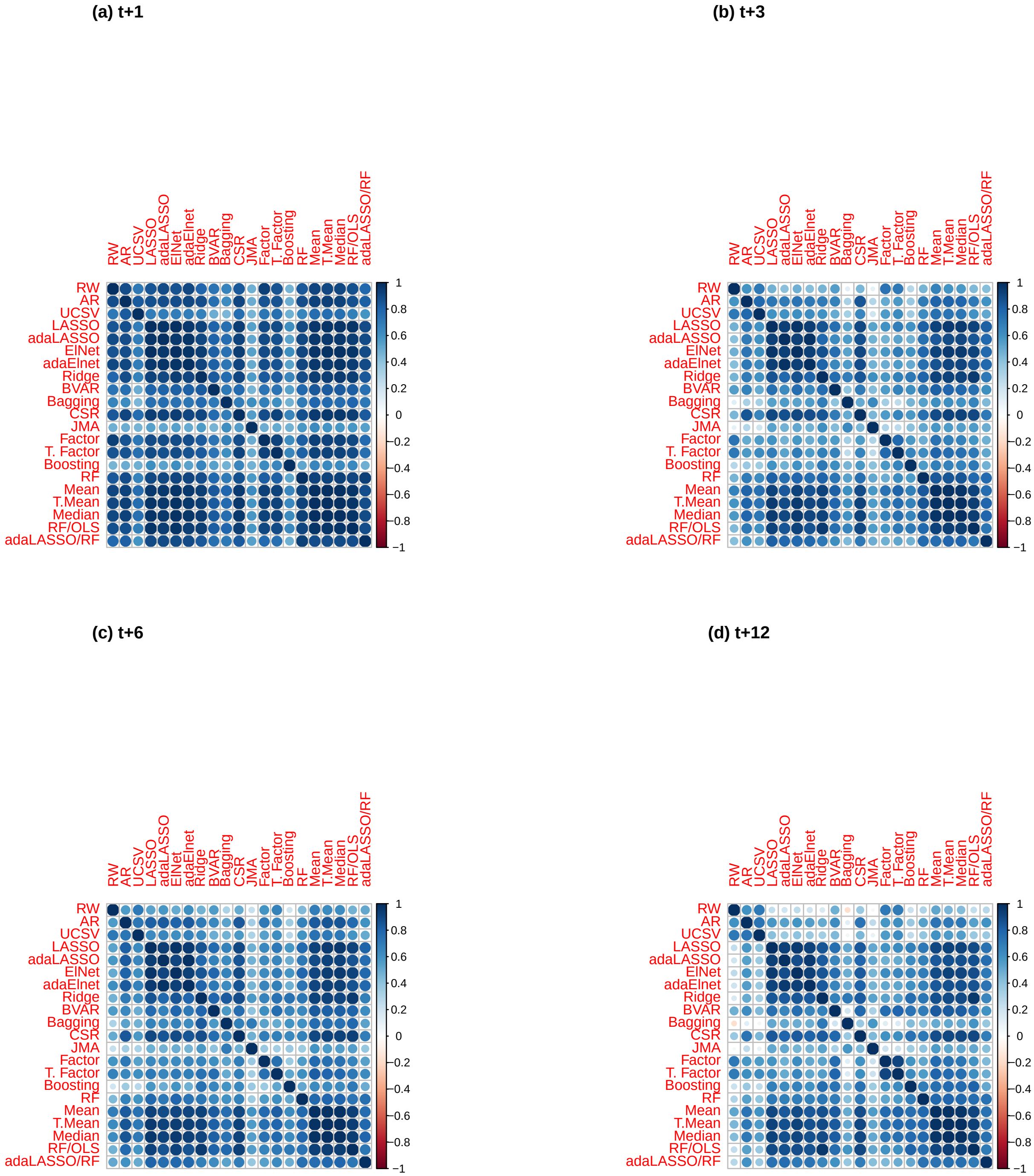


Table 9. Forecasting Errors for the CPI from 1990 to 2000

The table shows the root mean squared error (RMSE), and between parenthesis, the mean absolute errors (MAE) for all models relative to the random walk (RW). The error measures were calculated from 132 rolling windows covering the 1990-2000 period and 180 rolling windows covering the 2001-2015 period. Values in bold show the most accurate model in each horizon. Cells in gray (blue) show the models included in the 50% model confidence set (MCS) using the squared error (absolute error) as loss functions. The MCSs were constructed based on the maximum t-statistic. The last column in the table reports in how many horizons the row model was included in the MCS for square (absolute) loss. The last two rows in the table report how many models were included in the MCS for square and absolute losses.

Consumer Price Index 1990–2000														
	Forecasting Horizon													
RMSE/(MAE)	1	2	3	4	5	6	7	8	9	10	11	12	Acc.	RMSE count (MAE count)
RW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	3
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(4)
AR	0.84	0.82	0.88	0.82	0.78	0.79	0.79	0.80	0.87	0.89	0.95	0.85	1.24	10
	(0.88)	(0.83)	(0.92)	(0.83)	(0.81)	(0.84)	(0.84)	(0.80)	(0.94)	(0.98)	(1.04)	(0.94)	(1.38)	(6)
UCSV	0.86	0.84	0.87	0.87	0.85	0.85	0.86	0.85	0.86	0.89	0.94	0.88	1.00	8
	(0.88)	(0.85)	(0.88)	(0.87)	(0.86)	(0.86)	(0.87)	(0.84)	(0.88)	(0.91)	(0.96)	(0.89)	(1.02)	(11)
LASSO	0.83	0.82	0.88	0.83	0.79	0.78	0.80	0.81	0.88	0.92	0.97	0.85	1.24	9
	(0.88)	(0.84)	(0.92)	(0.84)	(0.83)	(0.84)	(0.88)	(0.83)	(0.96)	(1.02)	(1.08)	(0.96)	(1.41)	(5)
adaLASSO	0.81	0.82	0.87	0.83	0.75	**0.75**	0.77	0.77	0.85	0.87	0.92	0.82	1.03	13
	(0.84)	(0.82)	(0.86)	(0.80)	(0.73)	**(0.77)**	(0.81)	(0.77)	(0.90)	(0.92)	(1.00)	(0.89)	(1.08)	(13)
ElNet	0.81	0.81	0.88	0.83	0.80	0.79	0.82	0.81	0.92	0.92	1.00	0.89	1.26	7
	(0.86)	(0.84)	(0.92)	(0.86)	(0.86)	(0.85)	(0.92)	(0.83)	(1.02)	(1.02)	(1.14)	(1.02)	(1.47)	(4)
adaElnet	0.81	0.82	0.86	0.80	0.74	0.75	0.77	0.78	0.87	0.87	0.92	0.87	1.06	12
	(0.85)	(0.83)	**(0.86)**	(0.77)	**(0.73)**	(0.78)	(0.81)	(0.78)	(0.92)	(0.93)	(1.00)	(0.95)	(1.13)	(12)
Ridge	**0.79**	**0.77**	0.86	0.80	0.76	0.80	0.80	0.80	0.86	0.85	0.88	0.76	0.99	12
	(0.83)	(0.78)	(0.90)	(0.81)	(0.78)	(0.84)	(0.85)	(0.79)	(0.90)	(0.92)	(0.96)	(0.82)	(1.15)	(12)
BVAR	0.97	0.80	0.92	0.83	0.77	0.84	0.87	0.90	1.00	0.98	1.02	0.88	1.43	6
	(1.00)	**(0.77)**	(0.96)	(0.88)	(0.84)	(0.93)	(0.98)	(0.95)	(1.12)	(1.10)	(1.16)	(1.01)	(1.56)	(1)
Bagging	0.85	0.86	1.02	0.92	0.90	0.91	0.90	0.86	0.91	0.91	0.93	0.79	1.02	8
	(0.86)	(0.87)	(1.04)	(0.95)	(0.93)	(0.95)	(0.92)	(0.82)	(0.94)	(0.95)	(0.99)	(0.87)	(1.15)	(8)
CSR	0.83	0.85	0.89	0.81	0.77	0.76	**0.76**	**0.76**	0.85	0.88	0.91	0.81	1.11	10
	(0.89)	(0.89)	(0.92)	(0.82)	(0.79)	(0.81)	(0.82)	(0.76)	(0.91)	(0.95)	(0.97)	(0.89)	(1.25)	(8)
JMA	0.94	1.01	1.17	0.99	1.03	1.01	1.06	1.03	1.21	1.13	1.13	0.93	1.00	1
	(1.00)	(1.02)	(1.19)	(1.01)	(1.07)	(1.05)	(1.06)	(1.01)	(1.29)	(1.19)	(1.20)	(0.98)	(1.08)	(2)
Factor	0.87	0.85	0.98	0.90	0.89	0.86	0.84	0.90	1.02	0.97	1.04	0.98	1.51	1
	(0.96)	(0.92)	(1.05)	(0.97)	(0.92)	(0.90)	(0.88)	(0.91)	(1.14)	(1.09)	(1.15)	(1.14)	(1.72)	(1)
T. Factor	0.87	0.91	1.01	0.98	0.92	0.94	0.86	0.91	1.04	1.02	1.02	0.95	1.62	0
	(0.93)	(0.98)	(1.13)	(1.07)	(1.02)	(1.05)	(0.94)	(0.93)	(1.16)	(1.18)	(1.15)	(1.10)	(1.91)	(0)
Boosting	0.96	0.90	1.05	0.91	0.88	0.95	0.95	0.97	1.02	0.96	0.97	0.81	1.66	5
	(1.09)	(0.98)	(1.16)	(0.98)	(0.97)	(1.06)	(1.06)	(1.03)	(1.12)	(1.06)	(1.07)	(0.89)	(1.92)	(3)
RF	0.79	0.78	0.85	**0.77**	0.73	0.76	0.76	0.77	0.82	**0.82**	**0.85**	**0.72**	**0.87**	13
	(0.82)	(0.78)	(0.88)	(0.77)	(0.76)	(0.79)	**(0.78)**	**(0.75)**	**(0.86)**	**(0.86)**	**(0.89)**	(0.76)	**(0.94)**	(12)
Mean	0.80	0.79	0.85	0.79	0.76	0.77	0.77	0.77	0.84	0.84	0.87	0.78	1.02	13
	(0.83)	(0.81)	(0.87)	(0.80)	(0.79)	(0.81)	(0.81)	(0.76)	(0.90)	(0.91)	(0.94)	(0.85)	(1.11)	(12)
T.Mean	0.80	0.80	**0.85**	0.79	0.75	0.76	0.77	0.77	0.85	0.84	0.89	0.79	1.04	13
	(0.84)	(0.82)	(0.87)	(0.79)	(0.77)	(0.80)	(0.81)	(0.78)	(0.91)	(0.91)	(0.97)	(0.87)	(1.15)	(12)
Median	0.80	0.80	0.85	0.79	0.75	0.76	0.77	0.77	0.85	0.85	0.89	0.79	1.05	13
	(0.84)	(0.83)	(0.88)	(0.79)	(0.78)	(0.80)	(0.82)	(0.77)	(0.91)	(0.91)	(0.97)	(0.87)	(1.16)	(12)
RF/OLS	0.80	0.80	0.86	0.78	0.74	0.77	0.77	0.78	0.85	0.85	0.88	0.76	1.01	13
	(0.82)	(0.82)	(0.89)	(0.79)	(0.76)	(0.81)	(0.82)	(0.78)	(0.90)	(0.92)	(0.96)	(0.83)	(1.14)	(12)
adaLASSO/RF	0.79	0.81	0.91	0.77	**0.72**	0.77	0.77	0.77	**0.82**	0.89	0.90	0.72	0.89	12
	(0.84)	(0.81)	(0.94)	**(0.77)**	(0.73)	(0.81)	(0.81)	(0.77)	(0.86)	(0.94)	(0.99)	**(0.76)**	(0.95)	(12)
RMSE count	12	18	14	16	11	14	10	15	14	18	16	11	13	
MAE count	(12)	(15)	(9)	(10)	(3)	(14)	(13)	(15)	(15)	(16)	(14)	(13)	(13)	

Table 10. Forecasting Errors for the CPI from 2001 to 2015

The table shows the root mean squared error (RMSE), and between parenthesis, the mean absolute errors (MAE) for all models relative to the random walk (RW). The error measures were calculated from 132 rolling windows covering the 1990-2000 period and 180 rolling windows covering the 2001-2015 period. Values in bold show the most accurate model in each horizon. Cells in gray (blue) show the models included in the 50% model confidence set (MCS) using the squared error (absolute error) as loss functions. The MCSs were constructed based on the maximum t-statistic. The last column in the table reports in how many horizons the row model was included in the MCS for square (absolute) loss. The last two rows in the table report how many models were included in the MCS for square and absolute losses.

Consumer Price Index 2001–2015														
	Forecasting Horizon													
RMSE/(MAE)	1	2	3	4	5	6	7	8	9	10	11	12	Acc.	RMSE count (MAE count)
RW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(0)
AR	0.92	0.81	0.78	0.80	0.79	0.79	0.78	0.76	0.77	0.81	0.82	0.73	1.21	7
	(0.87)	(0.78)	(0.74)	(0.79)	(0.80)	(0.80)	(0.75)	(0.75)	(0.76)	(0.80)	(0.79)	(0.70)	(1.17)	(0)
UCSV	0.98	0.81	0.79	0.80	0.77	0.77	0.77	0.76	0.76	0.79	0.81	0.76	0.89	9
	(0.93)	(0.81)	(0.76)	(0.77)	(0.78)	(0.77)	(0.77)	(0.77)	(0.75)	(0.76)	(0.81)	(0.73)	(0.85)	(5)
LASSO	0.84	0.74	0.71	0.75	0.74	0.74	0.75	0.72	0.74	0.78	0.79	0.70	0.91	13
	(0.79)	(0.71)	(0.67)	(0.75)	(0.74)	(0.72)	(0.69)	(0.67)	(0.69)	(0.73)	(0.75)	(0.65)	(0.91)	(12)
adaLASSO	0.84	0.75	0.72	0.76	0.75	0.75	0.76	0.74	0.75	0.79	0.81	0.70	0.93	13
	(0.80)	(0.72)	(0.68)	(0.76)	(0.76)	(0.73)	(0.71)	(0.69)	(0.71)	(0.75)	(0.78)	(0.67)	(0.92)	(11)
ElNet	0.84	0.74	0.71	0.74	0.73	0.74	0.74	0.73	0.73	0.79	0.79	0.70	0.91	13
	(0.80)	(0.70)	(0.67)	(0.74)	(0.74)	(0.72)	(0.69)	(0.67)	(0.69)	(0.73)	(0.74)	(0.64)	(0.92)	(12)
adaElnet	0.85	0.74	0.72	0.76	0.75	0.75	0.75	0.74	0.74	0.79	0.80	0.70	0.93	12
	(0.81)	(0.71)	(0.68)	(0.76)	(0.76)	(0.73)	(0.70)	(0.68)	(0.70)	(0.74)	(0.76)	(0.67)	(0.92)	(11)
Ridge	0.86	0.72	0.70	0.75	0.73	0.74	0.74	0.72	0.72	0.76	0.77	0.69	0.86	12
	(0.83)	(0.70)	(0.67)	(0.75)	(0.75)	(0.74)	(0.69)	(0.68)	(0.69)	(0.71)	(0.73)	(0.67)	(0.86)	(12)
BVAR	0.83	0.75	0.72	0.75	0.74	0.74	0.75	0.74	0.74	0.79	0.79	0.72	0.99	13
	(0.81)	(0.72)	(0.68)	(0.75)	(0.75)	(0.73)	(0.69)	(0.69)	(0.70)	(0.73)	(0.74)	(0.67)	(0.93)	(12)
Bagging	0.82	0.74	0.72	0.78	0.76	0.77	0.81	0.80	0.76	0.80	0.81	0.73	0.77	11
	(0.84)	(0.74)	(0.71)	(0.83)	(0.84)	(0.82)	(0.80)	(0.79)	(0.76)	(0.80)	(0.82)	(0.74)	(0.80)	(4)
CSR	0.86	0.75	0.74	0.78	0.78	0.79	0.80	0.77	0.78	0.82	0.83	0.75	1.12	10
	(0.82)	(0.71)	(0.69)	(0.78)	(0.79)	(0.78)	(0.74)	(0.73)	(0.75)	(0.78)	(0.80)	(0.73)	(1.07)	(5)
JMA	1.00	0.78	0.79	0.83	0.80	0.77	0.89	0.83	0.79	0.91	0.88	0.77	0.84	5
	(0.99)	(0.78)	(0.79)	(0.92)	(0.91)	(0.84)	(0.85)	(0.82)	(0.81)	(0.88)	(0.87)	(0.78)	(0.85)	(1)
Factor	0.87	0.77	0.75	0.77	0.76	0.77	0.79	0.80	0.79	0.81	0.81	0.74	1.10	9
	(0.84)	(0.76)	(0.72)	(0.77)	(0.78)	(0.76)	(0.74)	(0.76)	(0.78)	(0.79)	(0.77)	(0.69)	(1.04)	(5)
T. Factor	0.88	0.76	0.74	0.76	0.74	0.76	0.78	0.78	0.76	0.78	0.80	0.74	1.05	9
	(0.85)	(0.75)	(0.71)	(0.74)	(0.75)	(0.76)	(0.75)	(0.75)	(0.74)	(0.76)	(0.76)	(0.69)	(1.00)	(6)
Boosting	0.95	0.75	0.72	0.76	0.74	0.76	0.77	0.75	0.76	0.81	0.81	0.73	1.03	12
	(0.91)	(0.72)	(0.70)	(0.79)	(0.78)	(0.79)	(0.76)	(0.75)	(0.76)	(0.79)	(0.79)	(0.69)	(1.13)	(8)
RF	0.86	**0.72**	**0.69**	0.73	**0.71**	**0.71**	**0.71**	0.70	0.71	**0.75**	**0.76**	**0.68**	**0.74**	13
	(0.81)	**(0.70)**	(0.66)	(0.74)	**(0.71)**	**(0.70)**	**(0.67)**	(0.66)	**(0.67)**	**(0.70)**	(0.72)	**(0.63)**	**(0.72)**	(13)
Mean	0.84	0.74	0.72	0.75	0.74	0.74	0.75	0.74	0.73	0.76	0.77	0.69	0.93	13
	(0.80)	(0.71)	(0.69)	(0.74)	(0.75)	(0.73)	(0.70)	(0.70)	(0.70)	(0.71)	**(0.72)**	(0.65)	(0.92)	(11)
T.Mean	0.85	0.73	0.71	0.75	0.73	0.74	0.74	0.73	0.73	0.77	0.78	0.70	0.92	13
	(0.80)	(0.71)	(0.67)	(0.74)	(0.74)	(0.72)	(0.69)	(0.68)	(0.69)	(0.72)	(0.72)	(0.64)	(0.90)	(12)
Median	0.85	0.73	0.71	0.75	0.73	0.74	0.74	0.73	0.73	0.77	0.78	0.70	0.92	13
	(0.80)	(0.70)	(0.67)	(0.74)	(0.75)	(0.72)	(0.69)	(0.68)	(0.69)	(0.72)	(0.73)	(0.65)	(0.90)	(12)
RF/OLS	**0.81**	0.72	0.71	0.75	0.74	0.75	0.75	0.73	0.73	0.77	0.78	0.70	0.92	13
	(0.78)	(0.70)	(0.67)	(0.75)	(0.76)	(0.74)	(0.70)	(0.69)	(0.70)	(0.73)	(0.76)	(0.68)	(0.91)	(12)
adaLASSO/RF	0.87	0.75	0.69	**0.72**	0.74	0.71	0.72	**0.70**	**0.71**	0.77	0.80	0.70	0.77	13
	(0.81)	(0.70)	**(0.66)**	**(0.73)**	(0.74)	(0.71)	(0.68)	**(0.65)**	(0.67)	(0.73)	(0.75)	(0.66)	(0.77)	(13)
RMSE count	11	17	15	19	18	18	18	17	19	19	20	19	17	
MAE count	(13)	(15)	(16)	(16)	(16)	(16)	(15)	(11)	(12)	(16)	(13)	(12)	(6)	

TABLE 11. Conditional Giacomini-White Test for Superior Predictive Ability: Recession and Expansions

The table reports the p-value of the Giacomini & White's (2005) test of superior predictive ability of the Random Forest model with respect to each one of the alternative models considered in the paper. The test is for squared loss functions and is conducted for expansion versus recessions periods as labeled by the NBER.

	RW	AR	UCSV	adaLASSO	adaElnet	Ridge	Boosting	RF/OLS	adaLASSO/RF
					1-month-ahead				
Expansion	$-2.56 \times 10^{-6***}$ (5.32×10^{-7})	$-8.48 \times 10^{-7***}$ (2.51×10^{-7})	$-1.04 \times 10^{-6***}$ (3.24×10^{-7})	-1.45×10^{-7} (1.09×10^{-7})	-1.61×10^{-7} (1.08×10^{-7})	-1.72×10^{-7} (1.53×10^{-7})	$-1.16 \times 10^{-6***}$ (3.65×10^{-7})	-9.68×10^{-9} (1.33×10^{-7})	-6.07×10^{-8} (1.58×10^{-7})
Recession	-1.06×10^{-6} (4.02×10^{-6})	-7.66×10^{-7} (1.58×10^{-6})	-6.63×10^{-6} (5.11×10^{-6})	2.21×10^{-6} (1.86×10^{-6})	1.25×10^{-6} (1.28×10^{-6})	9.32×10^{-7} (1.40×10^{-6})	-5.71×10^{-6} (3.56×10^{-6})	4.94×10^{-6} (3.50×10^{-6})	-5.77×10^{-7} (8.46×10^{-7})
					3-month-ahead				
Expansion	$-5.25 \times 10^{-6***}$ (1.36×10^{-6})	$-7.33 \times 10^{-7***}$ (2.49×10^{-7})	$-1.30 \times 10^{-6***}$ (4.44×10^{-7})	-3.01×10^{-7} (1.75×10^{-7})	-1.72×10^{-7} (1.71×10^{-7})	-1.69×10^{-7} (1.66×10^{-7})	$-1.24 \times 10^{-6***}$ (3.78×10^{-7})	-1.54×10^{-7} (1.58×10^{-7})	-7.05×10^{-8} (1.69×10^{-7})
Recession	$-2.39 \times 10^{-5*}$ (1.41×10^{-5})	-1.07×10^{-5} (6.80×10^{-6})	-7.48×10^{-6} (5.00×10^{-6})	-3.56×10^{-6} (2.23×10^{-6})	-3.76×10^{-6} (2.38×10^{-6})	-1.34×10^{-6} (1.90×10^{-6})	-1.35×10^{-6} (1.03×10^{-6})	-2.16×10^{-6} (1.77×10^{-6})	-1.30×10^{-6} (1.80×10^{-6})
					12-month-ahead				
Expansion	$-8.01 \times 10^{-6***}$ (2.45×10^{-6})	$-1.32 \times 10^{-6***}$ (2.60×10^{-7})	$-1.78 \times 10^{-6***}$ (4.38×10^{-7})	$-7.36 \times 10^{-7**}$ (3.15×10^{-7})	$-8.92 \times 10^{-7***}$ (2.90×10^{-7})	-2.00×10^{-7} (2.68×10^{-7})	$-1.03 \times 10^{-6***}$ (3.36×10^{-7})	-4.90×10^{-7} (3.14×10^{-7})	-2.68×10^{-7} (2.06×10^{-7})
Recession	-1.17×10^{-5} (8.29×10^{-6})	-3.36×10^{-6} (3.79×10^{-6})	-5.98×10^{-6} (4.88×10^{-6})	$-2.01 \times 10^{-6**}$ (1.01×10^{-6})	$-2.07 \times 10^{-6*}$ (1.22×10^{-6})	$-1.49 \times 10^{-6*}$ (8.28×10^{-7})	$-3.22 \times 10^{-6*}$ (1.97×10^{-6})	$-1.39 \times 10^{-6***}$ (4.93×10^{-7})	-1.72×10^{-6} (1.66×10^{-6})
					accumulated				
Expansion	$-4.34 \times 10^{-5*}$ (2.57×10^{-5})	$-1.18 \times 10^{-4} ***$ (1.90×10^{-5})	-1.50×10^{-5} (2.11×10^{-5})	$-4.90 \times 10^{-5***}$ (1.10×10^{-5})	$-5.05 \times 10^{-5***}$ (1.12×10^{-5})	$-4.68 \times 10^{-5***}$ (1.80×10^{-5})	$-2.27 \times 10^{-4***}$ (7.11×10^{-5})	$-4.92 \times 10^{-5***}$ (1.79×10^{-5})	$-1.53 \times 10^{-5***}$ (4.64×10^{-6})
Recession	$-6.30 \times 10^{-4***}$ (2.52×10^{-4})	$-11.73 \times 10^{-4**}$ (5.25×10^{-4})	$-4.43 \times 10^{-4**}$ (2.00×10^{-4})	$-3.91 \times 10^{-4**}$ (1.96×10^{-4})	$-3.85 \times 10^{-4**}$ (1.90×10^{-4})	$-9.59 \times 10^{-5} ***$ (1.45×10^{-4})	$-4.98 \times 10^{-5} ***$ (2.08×10^{-4})	-3.06×10^{-4} (2.30×10^{-4})	4.90×10^{-6} (5.53×10^{-5})

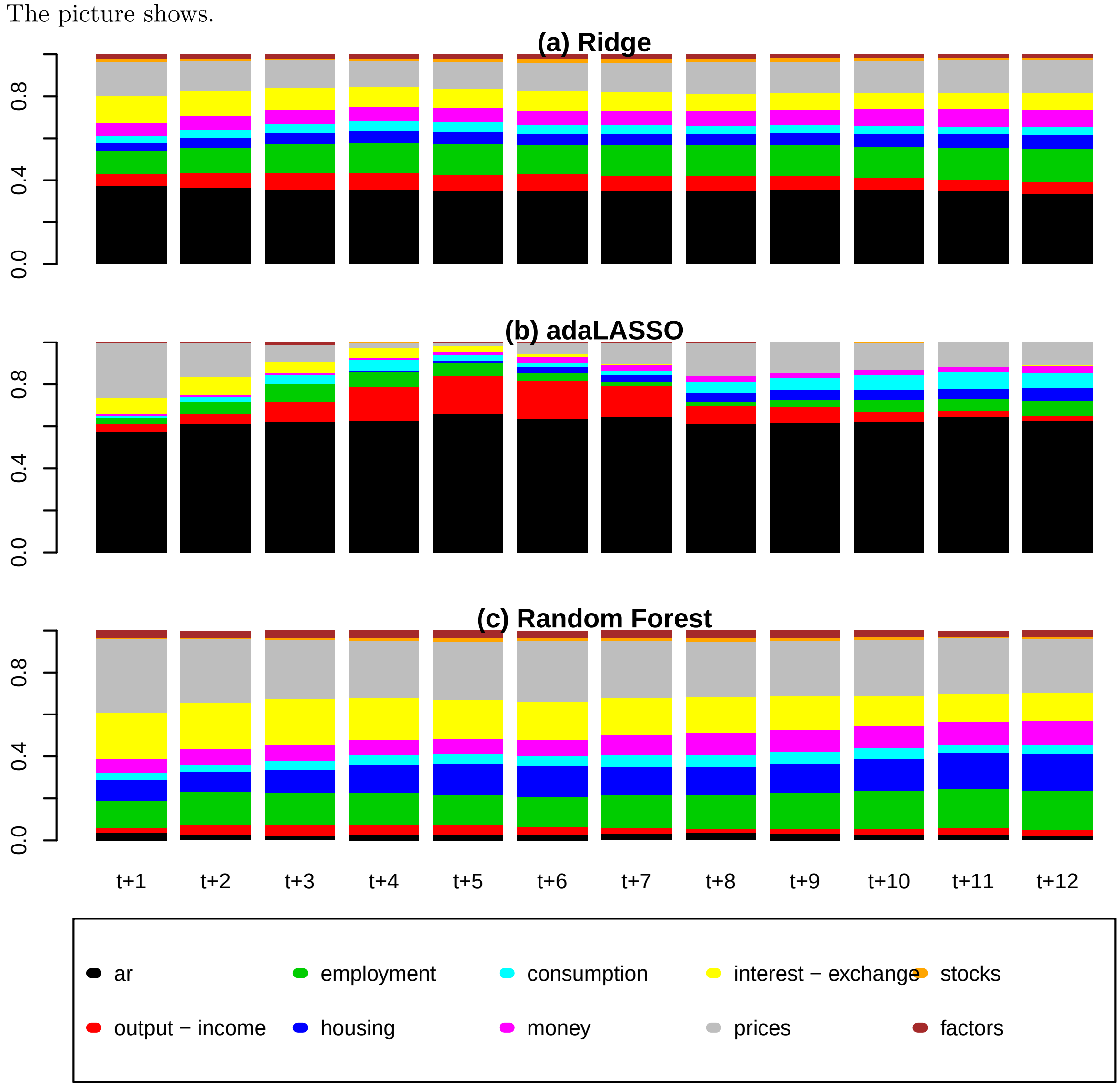
Table 12. Conditional Giacomini-White Test for Superior Predictive Ability: Macroeconomic Uncertainty

The table reports the p-value of the Giacomini & White's (2005) test of superior predictive ability of the Random Forest model with respect to each one of the alternative models considered in the paper. The test is for squared loss functions and is conducted for low versus high macroeconomic uncertainty periods.

	RW	**AR**	**UCSV**	**adaLASSO**	**adaElnet**	**Ridge**	**Boosting**	**RF/OLS**	**adaLASSO/RF**
					1-month-ahead				
Low	$-1.28\times10^{-6***}$ (3.75×10^{-7})	$-4.89\times10^{-7***}$ (1.36×10^{-7})	$-6.79\times10^{-7***}$ (2.50×10^{-7})	-6.82×10^{-8} (9.38×10^{-8})	-1.06×10^{-7} (8.98×10^{-8})	-2.87×10^{-8} (9.91×10^{-8})	$-1.13\times10^{-6***}$ (3.29×10^{-7})	9.40×10^{-8} (1.01×10^{-7})	-2.14×10^{-8} (9.46×10^{-8})
High	$-2.59\times10^{-6***}$ (1.45×10^{-6})	$-1.30\times10^{-6**}$ (6.33×10^{-7})	$-2.93\times10^{-6**}$ (1.46×10^{-6})	3.50×10^{-7} (5.30×10^{-7})	1.22×10^{-7} (3.92×10^{-7})	-8.12×10^{-8} (4.63×10^{-7})	$-2.36\times10^{-6**}$ (1.16×10^{-6})	1.11×10^{-6} (9.50×10^{-7})	-2.44×10^{-7} (3.71×10^{-7})
					3-month-ahead				
Low	$-2.52\times10^{-6***}$ (8.53×10^{-7})	$-2.86\times10^{-7**}$ (1.70×10^{-7})	-5.54×10^{-7} (4.05×10^{-7})	-1.58×10^{-7} (1.40×10^{-7})	-1.47×10^{-7} (1.33×10^{-7})	-1.72×10^{-7} (1.27×10^{-7})	$-1.39\times10^{-6***}$ (3.44×10^{-7})	-1.27×10^{-7} (1.06×10^{-7})	$-2.61\times10^{-7**}$ (1.36×10^{-7})
High	$-1.41\times10^{-5***}$ (4.64×10^{-6})	$-4.02\times10^{-6**}$ (2.25×10^{-6})	$-4.02\times10^{-6**}$ (1.78×10^{-6})	-1.37×10^{-6} (7.31×10^{-7})	-1.16×10^{-6} (7.84×10^{-7})	-4.75×10^{-7} (5.51×10^{-7})	-1.04×10^{-6} (7.19×10^{-7})	-7.26×10^{-7} (5.35×10^{-7})	-1.23×10^{-7} (5.76×10^{-7})
					12-month-ahead				
Low	$-3.37\times10^{-6***}$ (7.17×10^{-7})	$-1.09\times10^{-6***}$ (2.74×10^{-7})	$-1.14\times10^{-6***}$ (3.91×10^{-7})	$-1.09\times10^{-6***}$ 3.45×10^{-7}	$-1.32\times10^{-6***}$ (3.28×10^{-7})	-2.91×10^{-7} (2.75×10^{-7})	$-6.05\times10^{-7}**$ (3.05×10^{-7})	-4.37×10^{-7} (3.58×10^{-7})	-4.31×10^{-7} (2.46×10^{-7})
High	$-1.55\times10^{-5***}$ (5.23×10^{-6})	$-2.18\times10^{-6**}$ (1.07×10^{-6})	$-3.78\times10^{-6***}$ (1.38×10^{-6})	-5.67×10^{-7} (5.27×10^{-7})	-5.96×10^{-7} (5.13×10^{-7})	-4.10×10^{-7} (5.24×10^{-7})	$-2.19\times10^{-6***}$ (8.03×10^{-7})	-8.01×10^{-7} (4.67×10^{-7})	-4.18×10^{-7} (5.38×10^{-7})
					accumulated				
Low	-2.32×10^{-5} (2.46×10^{-5})	$-9.19\times10^{-5***}$ (2.11×10^{-5})	-1.81×10^{-5} (2.25×10^{-5})	$-3.91\times10^{-5***}$ (1.39×10^{-5})	$-4.34\times10^{-5***}$ (1.49×10^{-5})	-2.09×10^{-5} (1.43×10^{-5})	$-1.65\times10^{-4*}$ (8.84×10^{-5})	$-2.24\times10^{-5*}$ (1.32×10^{-5})	$4.13\times10^{-6*}$ (-1.32×10^{-5})
High	-2.30×10^{-4} $1.24\times10^{-4*}$	$-4.39\times10^{-4**}$ 2.21×10^{-4}	-1.26×10^{-4} 8.67×10^{-5}	$7.35\times10^{-5**}$ -1.55×10^{-4}	$7.35\times10^{-5**}$ -1.51×10^{-4}	$-9.59\times10^{-5**}$ 5.06×10^{-5}	$1.21\times10^{-4**}$ -2.64×10^{-4}	$-1.56\times10^{-4**}$ 8.15×10^{-5}	-1.28×10^{-5} 1.66×10^{-5}

FIGURE 7. Variable importance

The picture shows.



Appendix A. Variable Description

In this section, we present a description of the dataset used in this paper. Tables 13–20 describe the data and the transformations that were applied to each variable. Each table considers one of the eight different sectors in which the variables are grouped. The column tcode denotes the following data transformation for a series x: (1) no transformation; (2) Δx_t; (3) $2\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$; and (7) $\Delta(x_t/x_{t-1} - 1)$. The FRED column gives mnemonics in FRED followed by a short description. The comparable series in global insight is given in the column GS.

Table 13. Data Description: Output and Income

The column tcode denotes the following data transformation for a series x: (1) no transformation; (2) Δx_t; (3) $2\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$; (7) $\Delta(x_t/x_{t-1} - 1)$. The FRED column gives mnemonics in FRED followed by a short description. The comparable series in Global Insight is given in the column GS.

Group 1: Output and income

	id	tcode	fred	description	gsi	gsi:description
1	1	5	RPI	Real Personal Income	M_14386177	PI
2	2	5	W875RX1	Real personal income ex transfer receipts	M_145256755	PI less transfers
3	6	5	INDPRO	IP Index	M_116460980	IP: total
4	7	5	IPFPNSS	IP: Final Products and Nonindustrial Supplies	M_116460981	IP: products
5	8	5	IPFINAL	IP: Final Products (Market Group)	M_116461268	IP: final prod
6	9	5	IPCONGD	IP: Consumer Goods	M_116460982	IP: cons gds
7	10	5	IPDCONGD	IP: Durable Consumer Goods	M_116460983	IP: cons dble
8	11	5	IPNCONGD	IP: Nondurable Consumer Goods	M_116460988	IP: cons nondble
9	12	5	IPBUSEQ	IP: Business Equipment	M_116460995	IP: bus eqpt
10	13	5	IPMAT	IP: Materials	M_116461002	IP: matls
11	14	5	IPDMAT	IP: Durable Materials	M_116461004	IP: dble matls
12	15	5	IPNMAT	IP: Nondurable Materials	M_116461008	IP: nondble matls
13	16	5	IPMANSICS	IP: Manufacturing (SIC)	M_116461013	IP: mfg
14	17	5	IPB51222s	IP: Residential Utilities	M_116461276	IP: res util
15	18	5	IPFUELS	IP: Fuels	M_116461275	IP: fuels
16	19	1	NAPMPI	ISM Manufacturing: Production Index	M_110157212	NAPM prodn
17	20	2	CUMFNS	Capacity Utilization: Manufacturing	M_116461602	Cap uti

Table 14. Data Description: Labor Market

The column tcode denotes the following data transformation for a series x: (1) no transformation; (2) Δx_t; (3) $2\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$; (7) $\Delta(x_t/x_{t-1} - 1)$. The FRED column gives mnemonics in FRED followed by a short description. The comparable series in Global Insight is given in the column GS.

Group 2: Labor market

	id	tcode	fred	description	gsi	gsi:description
1	21*	2	HWI	Help-Wanted Index for United States		Help wanted indx
2	22*	2	HWIURATIO	Ratio of Help Wanted/No. Unemployed	M_110156531	Help wanted/unemp
3	23	5	CLF16OV	Civilian Labor Force	M_110156467	Emp CPS total
4	24	5	CE16OV	Civilian Employment	M_110156498	Emp CPS nonag
5	25	2	UNRATE	Civilian Unemployment Rate	M_110156541	U: all
6	26	2	UEMPMEAN	Average Duration of Unemployment (Weeks)	M_110156528	U: mean duration
7	27	5	UEMPLT5	Civilians Unemployed - Less Than 5 Weeks	M_110156527	U ¡ 5 wks
8	28	5	UEMP5TO14	Civilians Unemployed for 5-14 Weeks	M_110156523	U 5-14 wks
9	29	5	UEMP15OV	Civilians Unemployed - 15 Weeks & Over	M_110156524	U 15+ wks
10	30	5	UEMP15T26	Civilians Unemployed for 15-26 Weeks	M_110156525	U 15-26 wks
11	31	5	UEMP27OV	Civilians Unemployed for 27 Weeks and Over	M_110156526	U 27+ wks
12	32*	5	CLAIMSx	Initial Claims	M_15186204	UI claims
13	33	5	PAYEMS	All Employees: Total nonfarm	M_123109146	Emp: total
14	34	5	USGOOD	All Employees: Goods-Producing Industries	M_123109172	Emp: gds prod
15	35	5	CES1021000001	All Employees: Mining and Logging: Mining	M_123109244	Emp: mining
16	36	5	USCONS	All Employees: Construction	M_123109331	Emp: const
17	37	5	MANEMP	All Employees: Manufacturing	M_123109542	Emp: mfg
18	38	5	DMANEMP	All Employees: Durable goods	M_123109573	Emp: dble gds
19	39	5	NDMANEMP	All Employees: Nondurable goods	M_123110741	Emp: nondbles
20	40	5	SRVPRD	All Employees: Service-Providing Industries	M_123109193	Emp: services
21	41	5	USTPU	All Employees: Trade, Transportation & Utilities	M_123111543	Emp: TTU
22	42	5	USWTRADE	All Employees: Wholesale Trade	M_123111563	Emp: wholesale
23	43	5	USTRADE	All Employees: Retail Trade	M_123111867	Emp: retail
24	44	5	USFIRE	All Employees: Financial Activities	M_123112777	Emp: FIRE
25	45	5	USGOVT	All Employees: Government	M_123114411	Emp: Govt
26	46	1	CES0600000007	Avg Weekly Hours : Goods-Producing	M_140687274	Avg hrs
27	47	2	AWOTMAN	Avg Weekly Overtime Hours : Manufacturing	M_123109554	Overtime: mfg
28	48	1	AWHMAN	Avg Weekly Hours : Manufacturing	M_14386098	Avg hrs: mfg
29	49	1	NAPMEI	ISM Manufacturing: Employment Index	M_110157206	NAPM empl
30	127	6	CES0600000008	Avg Hourly Earnings : Goods-Producing	M_123109182	AHE: goods
31	128	6	CES2000000008	Avg Hourly Earnings : Construction	M_123109341	AHE: const
32	129	6	CES3000000008	Avg Hourly Earnings : Manufacturing	M_123109552	AHE: mfg

Table 15. Data Description: Housing

The column tcode denotes the following data transformation for a series x: (1) no transformation; (2) Δx_t; (3) $2\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$; (7) $\Delta(x_t/x_{t-1} - 1)$. The FRED column gives mnemonics in FRED followed by a short description. The comparable series in Global Insight is given in the column GS.

Group 3: Housing

	id	tcode	fred	description	gsi	gsi:description
1	50	4	HOUST	Housing Starts: Total New Privately Owned	M_110155536	Starts: nonfarm
2	51	4	HOUSTNE	Housing Starts, Northeast	M_110155538	Starts: NE
3	52	4	HOUSTMW	Housing Starts, Midwest	M_110155537	Starts: MW
4	53	4	HOUSTS	Housing Starts, South	M_110155543	Starts: South
5	54	4	HOUSTW	Housing Starts, West	M_110155544	Starts: West
6	55	4	PERMIT	New Private Housing Permits (SAAR)	M_110155532	BP: total
7	56	4	PERMITNE	New Private Housing Permits, Northeast (SAAR)	M_110155531	BP: NE
8	57	4	PERMITMW	New Private Housing Permits, Midwest (SAAR)	M_110155530	BP: MW
9	58	4	PERMITS	New Private Housing Permits, South (SAAR)	M_110155533	BP: South
10	59	4	PERMITW	New Private Housing Permits, West (SAAR)	M_110155534	BP: West

Table 16. Data Description: Consumption, Orders and Inventories

The column tcode denotes the following data transformation for a series x: (1) no transformation; (2) Δx_t; (3) $2\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$; (7) $\Delta(x_t/x_{t-1} - 1)$. The FRED column gives mnemonics in FRED followed by a short description. The comparable series in Global Insight is given in the column GS.

Group 4: Consumption, orders, and inventories

	id	tcode	fred	description	gsi	gsi:description
1	3	5	DPCERA3M086SBEA	Real personal consumption expenditures	M_123008274	Real Consumption
2	4*	5	CMRMTSPLx	Real Manu. and Trade Industries Sales	M_110156998	M&T sales
3	5*	5	RETAILx	Retail and Food Services Sales	M_130439509	Retail sales
4	60	1	NAPM	ISM : PMI Composite Index	M_110157208	PMI
5	61	1	NAPMNOI	ISM : New Orders Index	M_110157210	NAPM new ordrs
6	62	1	NAPMSDI	ISM : Supplier Deliveries Index	M_110157205	NAPM vendor del
7	63	1	NAPMII	ISM : Inventories Index	M_110157211	NAPM Invent
8	64	5	ACOGNO	New Orders for Consumer Goods	M_14385863	Orders: cons gds
9	65*	5	AMDMNOx	New Orders for Durable Goods	M_14386110	Orders: dble gds
10	66*	5	ANDENOx	New Orders for Nondefense Capital Goods	M_178554409	Orders: cap gds
11	67*	5	AMDMUOx	Unfilled Orders for Durable Goods	M_14385946	Unf orders: dble
12	68*	5	BUSINVx	Total Business Inventories	M_15192014	M&T invent
13	69*	2	ISRATIOx	Total Business: Inventories to Sales Ratio	M_15191529	M&T invent/sales
14	130*	2	UMCSENTx	Consumer Sentiment Index	hhsntn	Consumer expect

TABLE 17. Data Description: Money and Credit

The column tcode denotes the following data transformation for a series x: (1) no transformation; (2) Δx_t; (3) $2\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$; (7) $\Delta(x_t/x_{t-1} - 1)$. The FRED column gives mnemonics in FRED followed by a short description. The comparable series in Global Insight is given in the column GS.

Group 5: Money and credit

	id	tcode	fred	description	gsi	gsi:description
1	70	6	M1SL	M1 Money Stock	M_110154984	M1
2	71	6	M2SL	M2 Money Stock	M_110154985	M2
3	72	5	M2REAL	Real M2 Money Stock	M_110154985	M2 (real)
4	73	6	AMBSL	St. Louis Adjusted Monetary Base	M_110154995	MB
5	74	6	TOTRESNS	Total Reserves of Depository Institutions	M_110155011	Reserves tot
6	75	7	NONBORRES	Reserves Of Depository Institutions	M_110155009	Reserves nonbor
7	76	6	BUSLOANS	Commercial and Industrial Loans	BUSLOANS	C&I loan plus
8	77	6	REALLN	Real Estate Loans at All Commercial Banks	BUSLOANS	DC&I loans
9	78	6	NONREVSL	Total Nonrevolving Credit	M_110154564	Cons credit
10	79*	2	CONSPI	Nonrevolving consumer credit to Personal Income	M_110154569	Inst cred/PI
11	131	6	MZMSL	MZM Money Stock	N.A.	N.A.
12	132	6	DTCOLNVHFNM	Consumer Motor Vehicle Loans Outstanding	N.A.	N.A.
13	133	6	DTCTHFNM	Total Consumer Loans and Leases Outstanding	N.A.	N.A.
14	134	6	INVEST	Securities in Bank Credit at All Commercial Banks	N.A.	N.A.

Table 18. Data Description: Interest and Exchange Rates

The column tcode denotes the following data transformation for a series x: (1) no transformation; (2) Δx_t; (3) $2\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta\log(x_t)$; (6) $\Delta^2\log(x_t)$; (7) $\Delta(x_t/x_{t-1}-1)$. The FRED column gives mnemonics in FRED followed by a short description. The comparable series in Global Insight is given in the column GS.

Group 6: Interest and exchange rates

	id	tcode	fred	description	gsi	gsi:description
1	84	2	FEDFUNDS	Effective Federal Funds Rate	M_110155157	Fed Funds
2	85*	2	CP3Mx	3-Month AA Financial Commercial Paper Rate	CPF3M	Comm paper
3	86	2	TB3MS	3-Month Treasury Bill:	M_110155165	3 mo T-bill
4	87	2	TB6MS	6-Month Treasury Bill:	M_110155166	6 mo T-bill
5	88	2	GS1	1-Year Treasury Rate	M_110155168	1 yr T-bond
6	89	2	GS5	5-Year Treasury Rate	M_110155174	5 yr T-bond
7	90	2	GS10	10-Year Treasury Rate	M_110155169	10 yr T-bond
8	91	2	AAA	Moody's Seasoned Aaa Corporate Bond Yield		Aaa bond
9	92	2	BAA	Moody's Seasoned Baa Corporate Bond Yield		Baa bond
10	93*	1	COMPAPFFx	3-Month Commercial Paper Minus FEDFUNDS		CP-FF spread
11	94	1	TB3SMFFM	3-Month Treasury C Minus FEDFUNDS		3 mo-FF spread
12	95	1	TB6SMFFM	6-Month Treasury C Minus FEDFUNDS		6 mo-FF spread
13	96	1	T1YFFM	1-Year Treasury C Minus FEDFUNDS		1 yr-FF spread
14	97	1	T5YFFM	5-Year Treasury C Minus FEDFUNDS		5 yr-FF spread
15	98	1	T10YFFM	10-Year Treasury C Minus FEDFUNDS		10 yr-FF spread
16	99	1	AAAFFM	Moody's Aaa Corporate Bond Minus FEDFUNDS		Aaa-FF spread
17	100	1	BAAFFM	Moody's Baa Corporate Bond Minus FEDFUNDS		Baa-FF spread
18	101	5	TWEXMMTH	Trade Weighted U.S. Dollar Index: Major Currencies		Ex rate: avg
19	102	* 5	EXSZUSx	Switzerland / U.S. Foreign Exchange Rate	M_110154768	Ex rate: Switz
20	103	* 5	EXJPUSx	Japan / U.S. Foreign Exchange Rate	M_110154755	Ex rate: Japan
21	104	* 5	EXUSUKx	U.S. / U.K. Foreign Exchange Rate	M_110154772	Ex rate: UK
22	105	* 5	EXCAUSx	Canada / U.S. Foreign Exchange Rate	M_110154744	Ex rate: Canada

Table 19. Data Description: Prices

The column tcode denotes the following data transformation for a series x: (1) no transformation; (2) Δx_t; (3) $2\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta\log(x_t)$; (6) $\Delta^2\log(x_t)$; (7) $\Delta(x_t/x_{t-1}-1)$. The FRED column gives mnemonics in FRED followed by a short description. The comparable series in Global Insight is given in the column GS.

Group 7: Prices

	id	tcode	fred	description	gsi	gsi:description
1	106	6	WPSFD49207	PPI: Finished Goods	M110157517	PPI: fin gds
2	107	6	WPSFD49502	PPI: Finished Consumer Goods	M110157508	PPI: cons gds
3	108	6	WPSID61	PPI: Intermediate Materials	M_110157527	PPI: int matls
4	109	6	WPSID62	PPI: Crude Materials	M_110157500	PPI: crude matls
5	110*	6	OILPRICEx	Crude Oil, spliced WTI and Cushing	M_110157273	Spot market price
6	111	6	PPICMM	PPI: Metals and metal products	M_110157335	PPI: nonferrous
7	112	1	NAPMPRI	ISM Manufacturing: Prices Index	M_110157204	NAPM com price
8	113	6	CPIAUCSL	CPI : All Items	M_110157323	CPI-U: all
9	114	6	CPIAPPSL	CPI : Apparel	M_110157299	CPI-U: apparel
10	115	6	CPITRNSL	CPI : Transportation	M_110157302	CPI-U: transp
11	116	6	CPIMEDSL	CPI : Medical Care	M_110157304	CPI-U: medical
12	117	6	CUSR0000SAC	CPI : Commodities	M_110157314	CPI-U: comm.
13	118	6	CUUR0000SAD	CPI : Durables	M_110157315	CPI-U: dbles
14	119	6	CUSR0000SAS	CPI : Services	M_110157325	CPI-U: services
15	120	6	CPIULFSL	CPI : All Items Less Food	M_110157328	CPI-U: ex food
16	121	6	CUUR0000SA0L2	CPI : All items less shelter	M_110157329	CPI-U: ex shelter
17	122	6	CUSR0000SA0L5	CPI : All items less medical care	M_110157330	CPI-U: ex med
18	123	6	PCEPI	Personal Cons. Expend.: Chain Index	gmdc	PCE defl
19	124	6	DDURRG3M086SBEA	Personal Cons. Exp: Durable goods	gmdcd	PCE defl: dlbes
20	125	6	DNDGRG3M086SBEA	Personal Cons. Exp: Nondurable goods	gmdcn	PCE defl: nondble
21	126	6	DSERRG3M086SBEA	Personal Cons. Exp: Services	gmdcs	PCE defl: service

Table 20. Data Description: Stock Market

The column tcode denotes the following data transformation for a series x: (1) no transformation; (2) Δx_t; (3) $2\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta\log(x_t)$; (6) $\Delta^2\log(x_t)$; (7) $\Delta(x_t/x_{t-1}-1)$. The FRED column gives mnemonics in FRED followed by a short description. The comparable series in Global Insight is given in the column GS.

Group 8: Stock Market

	id	tcode	fred	description	gsi	gsi:description
1	80*	5	S&P 500	S&P's Common Stock Price Index: Composite	M_110155044	S&P 500
2	81*	5	S&P: indust	S&P's Common Stock Price Index: Industrials	M_110155047	S&P: indust
3	82*	2	S&P div yield	S&P's Composite Common Stock: Dividend Yield		S&P div yield
4	83*	5	S&P PE ratio	S&P's Composite Common Stock: Price-Earnings Ratio		S&P PE ratio
5	135*	1	VXOCLSx	VXO		

Appendix B. Models

For all models, with the exception of the RW and UCSV specifications, we include a dummy for the November 2008, when a huge deflation was observed.

B.1. **Benchmark Models.** The first benchmark is the RW model, where for $h = 1, \ldots, 12$, the forecasts are computed as follows:

$$\widehat{\pi}_{t+h|t} = \pi_t. \tag{3}$$

For the accumulated twelve-month forecast, we consider the following equation:

$$\widehat{\pi}_{t+1:t+12|t} = \pi_{t-11:t}, \tag{4}$$

where $\pi_{t-11:t}$ is the accumulated inflation over the previous twelve months.

The second benchmark is the autoregressive (AR) model of order p, where p is determined by the Bayesian information criterion (BIC) and the parameters are estimated by OLS. The forecast equation is

$$\widehat{\pi}_{t+h|t} = \widehat{\phi}_{0,h} + \widehat{\phi}_{1,h}\pi_t + \ldots + \widehat{\phi}_{p,h}\pi_{t-p+1}. \tag{5}$$

There is a different model for each horizon. The accumulated forecasts are computed by aggregating the individual forecasts.

Finally, the third benchmark is the UCSV model, which is described as follows:

$$\begin{aligned} \pi_t =& \tau_t + e^{h_t/2}\varepsilon_t, \\ \tau_t =& \tau_{t-1} + u_t, \\ h_t =& h_{t-1} + v_t, \end{aligned} \tag{6}$$

where $\{\varepsilon_t\}$ is a sequence of independent and normally distributed random variables with zero mean and unit variance and $\varepsilon_t \sim \mathsf{N}(0,1)$, u_t and v_t are also normal with zero mean and variance given by inverse-gamma priors. $\tau_1 \sim \mathsf{N}(0, V_\tau)$ and $h_1 \sim \mathsf{N}(0, V_h)$, where $V_\tau = V_h = 0.12$. The model is estimated by Markov chain Monte Carlo (MCMC) methods. The h-steps-ahead forecast is computed as $\widehat{\pi}_{t+h} = \widehat{\tau}_{t|t}$.

For accumulated forecasts, the UCSV is estimated with the twelve-month inflation as the dependent variable.

B.2. **Shrinkage.** In this paper, we estimate several shrinkage estimators for linear models where $T_h(\boldsymbol{x}_t) = \boldsymbol{\beta}_h'\boldsymbol{x}_t$ and

$$\widehat{\boldsymbol{\beta}}_h = \arg\min_{\boldsymbol{\beta}} \left[\sum_{t=1}^{T-h} \left(\pi_{t+h} - \boldsymbol{\beta}'\boldsymbol{x}_t\right)^2 + \lambda \sum_{i=1}^{n} p(\beta_i; \omega_i, \alpha) \right], \tag{7}$$

where $p(\beta_i;\omega_i,\alpha)$ is a penalty function that depends on the penalty parameter λ and on a weight $\omega_i > 0$. We consider different choices for the penalty functions as described below.

B.2.1. *Ridge Regression (RR):.* RR shrinkage was proposed by Hoerl & Kennard (1970*b*,*a*) and consists of the following penalty function:

$$\lambda \sum_{i=1}^{n} p(\beta_i;\omega_i,\alpha) := \lambda \sum_{i=1}^{n} \beta_i^2. \tag{8}$$

RR has the advantage of having an analytical solution that is easy to compute and shrinks the irrelevant variables to zero. However, given the geometry of the penalty, the coefficients rarely reach exactly zero for any size of λ. Therefore, RR is not an sparsity-inducing method.

One interesting fact about RR is its relation to principal component (factor) models. Let $\boldsymbol{X}$ be the centered $T \times n$ predictor matrix and consider its singular value decomposition $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}'$ with $\boldsymbol{S}$ being a diagonal matrix with diagonal elements s_i, $i = 1,\ldots,n$.

The RR estimates of inflation are given by

$$\boldsymbol{\pi}_{\mathsf{ridge}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\mathsf{ridge}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{U}\mathsf{diag}\left(\frac{s_i^2}{s_i^2+\lambda}\right)\boldsymbol{U}'\boldsymbol{y},$$

whereas for the factor model with k factors are given by

$$\boldsymbol{\pi}_{\mathsf{PC}} = \boldsymbol{X}_{\mathsf{PC}}\widehat{\boldsymbol{\beta}}_{\mathsf{PC}} = \boldsymbol{U}\mathsf{diag}(\underbrace{1,\ldots,1}_{k\,\text{ones}},\underbrace{0,\ldots,0}_{n-k\,\text{zeroes}})\boldsymbol{U}'\boldsymbol{y}.$$

However, this parallel to factor models does not hold exactly in our implementation as the variable set for the RR is larger that the one for the principal component factor construction as it includes four lags of each variable, autoregressive terms and the factors as well. Nevertheless, the comparison is useful to understand the potential differences in performance between RR and factor alternatives.

B.2.2. *Least Absolute Shrinkage and Selection Operator (LASSO):.* LASSO was originally proposed by Tibshirani (1996). LASSO is similar to RR but penalizes the ℓ_1 norm of the coefficients as follows:

$$\lambda \sum_{i=1}^{n} p(\beta_i;\omega_i,\alpha) := \lambda \sum_{i=1}^{n} |\beta_i|. \tag{9}$$

LASSO shrinks the irrelevant variables to zero and has some good properties both in variable selection and goodness of fit. In order to achieve consistent variable selection, LASSO requires the irrepresentable condition[9] (IRC) to be satisfied (Zhao & Yu 2006).

[9]The irrepresentable condition imposes some restrictions on the correlation structure between the relevant and the irrelevant variables. In other words, the correlation between the two groups is bounded and must be small.

However, even if the IRC is not satisfied, LASSO still has the variable screening property, i.e., LASSO selects the relevant variables with high probability, but it may also select some extra variables.

B.2.3. *Adaptive LASSO (adaLASSO):.* adaLASSO was proposed by Zou (2006), who showed that the inclusion of some additional information regarding the importance of each variable could considerably improve the results. The adaLASSO does not need the IRC to have variable selection consistency and also has oracle properties, i.e., it not only selects the correct set of variables with high probability, but the coefficient distribution of these variables is also the same as the OLS estimation using only the correct set of variables. adaLASSO uses the same penalty as LASSO with the inclusion of a weighting parameter that comes from a first-step model that can be LASSO or even OLS:

$$\lambda \sum_{i=1}^{n} p(\beta_i; \omega_i, \alpha) := \lambda \sum_{i=1}^{n} \omega_i |\beta_i|, \tag{10}$$

where $\omega_i = |\beta_i^*|^{-1}$ and β_i^* are the coefficients from the first-step model. Finally, LASSO has some good properties for high-dimensional data. LASSO can handle many more variables than observations and works well in nonGaussian environments and under heteroskedasticity (Medeiros & Mendes 2016).

B.2.4. *Elastic Net (ElNet).* Elastic net (ElNet) is a generalization that includes LASSO and RR as special cases. ElNet is a convex combination of the ℓ_1 and the ℓ_2 norms (Zou & Hastie 2005). ElNet also does regularization and selects the most relevant variables. Since its penalty is between that of LASSO and RR, ElNet normally selects more variables than LASSO, at least for the same value of λ. The ElNet penalty is defined as follows:

$$\lambda \sum_{i=1}^{n} p(\beta_i; \omega_i, \alpha) := \alpha\lambda \sum_{i=1}^{n} \beta_i^2 + (1-\alpha)\lambda \sum_{i=1}^{n} |\beta_i|; \tag{11}$$

where $\alpha \in [0, 1]$. We also consider an adaptive version of ElNet (adaElNet). This version works in the same way as the adaptive LASSO, i.e., we estimate a first-step model and use it to calculate the weights ω_i.

B.3. **Factor Models.** Factor models using principal components are very popular approaches to avoid the curse of dimensionality when the number of predictions is potentially large. The idea is to extract common components from all variables, thus reducing the model dimension.

In the present paper, factors are computed as principal components of a large set of variables $\boldsymbol{z}_t$ such that $\boldsymbol{F}_t = \boldsymbol{A}\boldsymbol{z}_t$, where $\boldsymbol{A}$ is a rotation matrix and $\boldsymbol{F}_t$ is the vector of the principal components. Consider equation (1). In this case, $\boldsymbol{x}_t$ is given by π_{t-j},

$j = 0, 1, 2, 3$ plus $\boldsymbol{f}_{t-j}$, $j = 0, 1, 2, 3$, where $\boldsymbol{f}_t$ is the a vector with the first four principal components of $\boldsymbol{z}_t$. The assumptions and the theory behind factor models and when can we treat factors as observed variables can be found in Bai & Ng (2002, 2006, 2008).

B.3.1. *Target Factors.* To improve the forecasting performance of factor models, Bai & Ng (2008) proposed targeting the predictors. The idea is that if many variables in $\boldsymbol{z}_t$ are irrelevant predictors of π_{t+h}, factor analysis using all variables may result in noisy factors with poor forecasting ability. The target factors are regular factor models with a pretesting procedure to select only relevant variables to be included in the factor analysis. Let $z_{i,t}$, $i = 1, \dots, q$ be the candidate variables and $\boldsymbol{w}_t$ a set of fixed regressors that will be used as controls in the pretesting step. We follow Bai & Ng (2008) and use $\boldsymbol{w}_t$ as AR terms of π_t. The procedure is described as follows.

(1) For $i = 1, \dots, q$, regress π_{t+h} on $\boldsymbol{w}_t$ and $z_{i,t}$ and compute the t statistics for the coefficient corresponding to $z_{i,t}$.
(2) Sort all t statistics calculated in step 1 in descending order.
(3) Choose a significance level α and select all variables that are significant using the computed t statistics.
(4) Let $\boldsymbol{z}_t(\alpha)$ be the selected variables from steps 1–3. Estimate the factors $\boldsymbol{F}_t$ from $\boldsymbol{z}_t(\alpha)$ by principal components.
(5) Regress π_{t+h} on $\boldsymbol{w}_t$ and $\boldsymbol{f}_{t-j}$, $j = 0, 1, 2, 3$, where $\boldsymbol{f}_t \subset \boldsymbol{F}_t$. The number of factors in $\boldsymbol{f}_t$ is selected using the BIC. Bai & Ng (2008) also selected the number of lagged factors using the BIC.

The same procedure was used by Medeiros & Vasconcelos (2016). The authors showed that in most cases, target factors slightly reduce the forecasting errors compared to factor models without targeting.

B.3.2. *Factor Boosting.* The optimal selection of factors for predictive regressions is an open problem in the literature. Even if the factor structure is clear in the data, it is not obvious that only the most relevant factors should be included in the predictive regression. We adopt the boosting algorithm as proposed by Bai & Ng (2008) to select the factors and the number of lags that must be considered in the predictive regression for inflation. Define $\boldsymbol{z}_t \in \mathbb{R}^q$, the set of all n factors computed from the original n variables plus four lags of each factor. Therefore, $q = 5n$.

The algorithm is defined as follows:

(1) Let $\Phi_{t,0} = \bar{\pi}$ for each t, where $\bar{\pi} = \frac{1}{t}\sum_{i=1}^{t} \pi_i$.
(2) For $m = 1, \dots, M$:
    (a) Compute $\widehat{u}_t = \pi_t - \Phi_{t-h,m-1}$.

(b) For each candidate variable $i = 1, \ldots, q$, regress the current residual on $z_{i,t}$ to obtain $\widehat{b}_i$ and compute $\widehat{e}_{t,i} = \widehat{u}_t - z_{i,t}\widehat{b}_i$. Calculate $SSR_i = \widehat{\boldsymbol{e}}_i'\widehat{\boldsymbol{e}}_i$.

(c) Select i_m^* as the index of the variable which delivers the smallest SSR and define $\widehat{\phi}_{t,m} = z_{i_m^*,t}\widehat{b}_{i_m^*}$,

(d) Update $\widehat{\Phi}_{t,m} = \widehat{\Phi}_{t,m-1} + v\phi_{t,m}$, where v is the step length. We set $v = 0.2$.

(3) Stop the algorithm after the Mth iteration or when the BIC starts to increase.

B.4. **Ensemble Methods.** Ensemble forecasts are constructed from a (weighted) average of the predictions of an ensemble of methods. In this section, we describe the techniques considered in this paper.

B.4.1. *Bagging.* The term "bagging" comes from bootstrap aggregation, which was proposed by Breiman (1996). The idea is to combine forecasts from several unstable models. Normally, there is much more to gain from combinations of models if they are very different. The first source of instability is generated by re-estimating the model using bootstrap samples, and the second source comes from a pretesting step prior to the estimation, which for each bootstrap sample selects a subset of variables based on their statistical significance. The bagging steps are as follows:

(1) For each bootstrap sample b, run a regression with all candidate variables and select those with $|t| \geq c$, where c is a pre-defined critical value.
(2) Estimate a new regression only with the variables selected in the previous step.
(3) The coefficients from the second regression are finally used to compute the forecasts **on the actual sample**.
(4) Repeat the first three steps for B bootstrap samples and compute the final forecast as the average of the B forecasts.

We used $B = 100$. Note that in our case, the number of observations may be smaller than the number of variables, which makes the regression in the first step unfeasible. We solve this issue by introducing a new source of instability in the pretesting step. For each bootstrap sample we randomly divide all variables in groups and run the pretesting step for each one of the groups.

B.4.2. *Complete Subset Regressions.* CSR was developed by Elliott et al. (2013, 2015). The motivation for developing CSR was that selecting the optimal subset of $\boldsymbol{x}_t$ to predict π_{t+h} by testing all possible combinations of regressors is computationally very demanding, and in most cases, even unfeasible. Supposing that we have n candidate variables, the CSR selects a number $q \leq n$ and computes all combinations of regressions using only q variables. The forecast of the model will be the average forecast of all regressions in the subset.

CSR deals well with a small number of candidate variables. However, for large sets, the number of regressions to be estimated increases very fast. For example, with $n = 25$ and $q = 4$, we need to estimate $12,650$ regressions. As the number of candidate variables is much larger, we adopt a pretesting procedure similar to that used with the target factors. We start fitting a linear regression of π_{t+h} on each of the candidate variables (including lags) and save the t-statistics of each variable[10]. The t-statistics are ranked by absolute value, and we select the $\tilde{n}$ variables that are more relevant in the ranking. The CSR forecast is calculated on these variables. We used $\tilde{n} = 25$ and $q = 4$.

B.4.3. *Jackknife Model Averaging.* JMA is a different way to combine forecasts from several small models. Instead of using the naive average of the forecasts, JMA uses leave-one-out cross-validation to estimate optimal weights. The procedure we followed is that of Hansen & Racine (2012) with some adjustments for time series as discussed in Zhang et al. (2013).

Suppose we have M candidate models that we want to average from and write the forecast of each model as $\widehat{\pi}_{t+h}^{(m)}$, $m = 1, \ldots, M$. Set the final forecast as

$$\widehat{\pi}_{t+h} = \sum_{m=1}^{M} \omega_m \widehat{\pi}_{t+h}^{(m)},$$

where $0 \leq \omega_m \leq 1$ for all $m \in \{1, \ldots, M\}$ and $\sum_{m=1}^{M} \omega_m = 1$.

The JMA procedure is as follows:

(1) For each observation of $(\boldsymbol{x}_t, \pi_{t+h})$:
   (a) Estimate all the candidate models leaving the selected observation out of the estimation. Since we are in a time series framework with lags in the model, we also removed four observations before and four observations after $(\boldsymbol{x}_t, \pi_{t+h})$.
   (b) Compute the forecasts from each model for the observations that were removed in the previous step.
(2) Choose the weights that minimize the cross-validation errors subject to the constraints previously described.

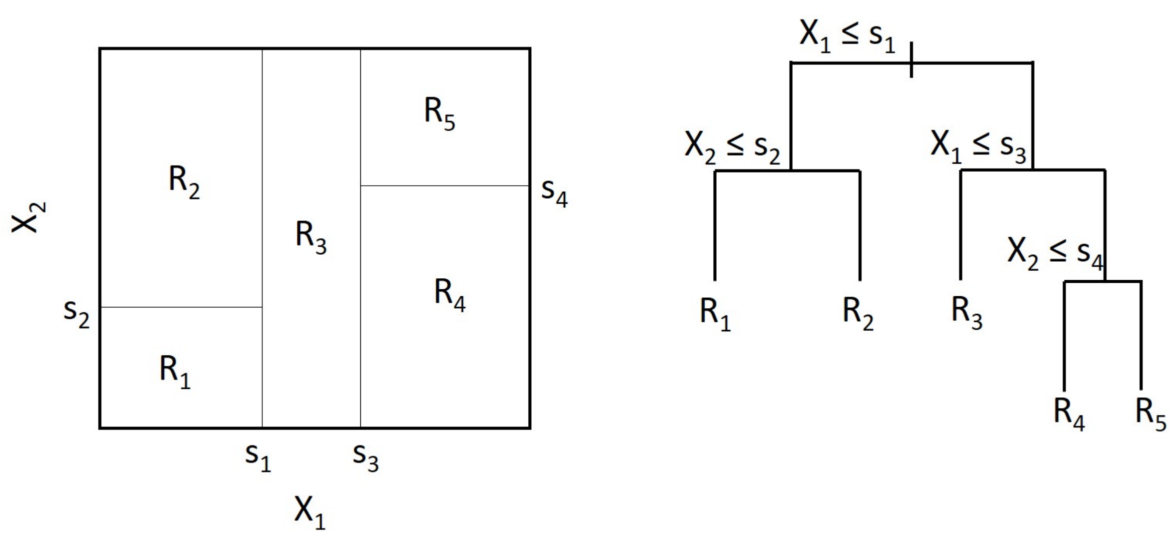
The minimization problem above is quadratic and has the restriction that w must be positive and sum to 1. The problem does not have a closed solution but can be easily solved using the quadprog package (Berwin et al. 2013) in R. Given our set of candidate variables, each candidate model in the JMA has four autoregressive lags of the inflation and four lags of one candidate variable.

[10] We did not use a fixed set of controls, $\boldsymbol{w}_t$, in the pretesting procedure like we did for the target factors.

B.5. **Regression Trees and Random Forests.** The RF methodology was initially proposed by Breiman (2001) as a solution to reducing the variance of regression trees and is based on bootstrap aggregation (bagging) of randomly constructed regression trees. In turn, regression trees are flexible nonparametric predictive models that recursively partition the set of explanatory variables, $\mathbb{X}$, into subsets, each modeled using regression methods; see Breiman (1996).

To understand how a regression tree works, an example from Hastie et al. (2001) is useful. Consider a regression problem in which X_1 and X_2 are explanatory variables, each taking values in some given interval, and Y is the dependent variable. We first split the space into two regions, at $X_1 = s_1$, and then, the region to the left (right) of $X_1 = s_1$ is split at $X_2 = s_2$ ($X_1 = s_3$). Finally, the region to the right of $X_1 = s_3$ is split at $X_2 = s_4$. As illustrated in the right plot of Figure 8, the end result is a partitioning of $\mathbb{X}$ into five regions: R_m, $m = 1, \ldots, 5$. In each region R_m, we assume that the model predicts Y with a constant c_m, which could be estimated, for example, as the sample average of realizations of Y that "fall" within region R_m. A key advantage of this recursive binary partition is that it can be represented as a single tree, as illustrated in the left plot of Figure 8. Each region corresponds to a terminal node of the tree.

Figure 8. Example of a regression tree. Reproduction of part of Figure 9.2 in Hastie et al. (2001).



Now we turn to the question as to how to choose splitting variables and split points, i.e., how to grow a tree, when there are p explanatory variables. Let $x_t = (x_{1,t}, x_{2,t}, \ldots, x_{p,t})$, for $t = 1, \ldots, T$, where $x_{i,t}$ is the realization of variable X_i in period t.

We proceed backwards. Suppose that after choosing the splitting variables and split points, we reach M regions. If we adopt the sum of squared errors as our minimization criterion, the prediction of Y at T, $\hat{c}_m$, is simply the average of previous realizations y_t such that x_t belongs to R_m. Algebraically, for $m = 1, ..., M$,

$$\hat{c}_m = \arg\min \sum_{t=1}^{T} \mathbf{I}(x_t \in R_m)(y_t - c_m)^2 = \frac{\sum_{t=1}^{T} \mathbf{I}(x_t \in R_m) y_t}{\sum_{i=1}^{T} \mathbf{I}(x_t \in R_m)}, \tag{12}$$

where $\mathbf{I}$ is the indicator function.

The idea is to use the sum of squared errors to inform how to grow the regression tree. To begin, consider a splitting variable j and a split point s to partition $\mathbb{X}$ into two regions, namely, $R_1(j,s) = \{X|X_j \leq s\}$ and $R_2(j,s) = \{X|X_j > s\}$. Then, seek the pair (j,s) that solves

$$\min_{j,s} \left[\min_{c_1} \sum_{t=1}^{T} \mathbf{I}(x_t \in R_1(j,s))(y_t - c_1)^2 + \min_{c_2} \sum_{t=1}^{T} \mathbf{I}(x_t \in R_2(j,s))(y_t - c_2)^2 \right].$$

Once the best split is found, we proceed iteratively, repeating this process on each of the resulting regions.

A natural question arises: when should we stop this process? A very large tree might overfit the data, which would be highly unstable. However, a tree that is too small might not capture a complex nonlinear relation between variables in the data. One possibility to address this trade-off is the *cost-complexity pruning* method described in Hastie et al (2009). Instead, we follow the RF method, which applies the essential idea of bagging, i.e., RF reduces the variance by averaging many noisy and unbiased models to (very large) regression trees. The drawback is the loss of interpretability.

An RF is a collection of regression trees, each specified in a bootstrapped subsample of the original data. Suppose there are B bootstrapped subsamples. For each subsample, obtain a prediction for Y by applying a modified version of the aforementioned splitting iterative process until a prespecified minimum number of observations, say five, is reached in any resulting region. In particular, the modification is to select q variables at random from the p explanatory variables at each step of the process. Finally, simply average the predictions of Y across the B bootstrapped subsamples. Since we are dealing with time series, bootstrapped samples are calculated using block bootstrapping.

The main advantages of the RF method are twofold: RF can handle both a very large number of explanatory variables and complex nonlinear relationships between variables.

B.6. **Hybrid Linear-Random Forests Models.** RF/OLS and adaLASSO/RF deserve some special attention because these are adaptations made specifically to answer how important the variable selection is and the nonlinearity in forecasting the US inflation. RF/OLS is estimated using the following steps:

(1) For each bootstrap sample b:
   (a) Grow a single tree with k nodes (we used $k = 20$) and save the $N \leq k$ split variables,
   (b) Run an OLS on the selected splitting variables,
   (c) Compute the forecast $\widehat{y}^b_{t+h}$.

(2) The final forecast will be $\hat{y}_{t+h} = B^{-1} \sum_{b=1}^{B} \widehat{y}^b_{t+h}$ where B is the number of bootstrap samples.

The main objective of the RF/OLS is to check the performance of a linear model using variables selected from the RV. If the results are very close to the full RF, we understand that nonlinearity is not an issue, and the RF is superior solely because of variable selection. However, if we see some improvement compared to other linear models, especially bagging[11], but if RF/OLS is still less accurate than RF, we have evidence that both nonlinearity and variables selection play an important role.

The second adapted model is LASSO/RF, where we use the adaptive LASSO for variable selection and then estimate a fully grown RF with the variables selected by adaptive LASSO. If LASSO/RF performs similarly to RF, we understand that the variable selection in RF is irrelevant, and the only thing that matters is the nonlinearity. LASSO/RF and RF/OLS together create an "if and only if" situation where we test the importance of variable selection and nonlinearity from both sides. Our results point to the middle case where nonlinearity and variable selection are both important. The two adapted models perform very well compared to other linear specifications, but RF is more accurate than both. In other words, the good performance of RF is driven by both variable selection and nonlinearity.

[11]Bagging and RF are bootstrap-based models, the first of which is linear and the second is nonlinear.

Appendix C. Additional Results

C.1. **Variable Selection: Word Clouds.** This Appendix presents the variable selection for several models as *word clouds*. In the present context, a word cloud is an image composed of the names of variables selected by a specific model across the estimation windows in which the size of each word indicates its frequency or importance. The names displayed in the clouds are as defined in the third column of Tables 13–20. These names represent FRED mnemonics. The clouds also indicate the degree of sparsity of each model. For instance, the smaller the cloud is, the more sparse the model is.

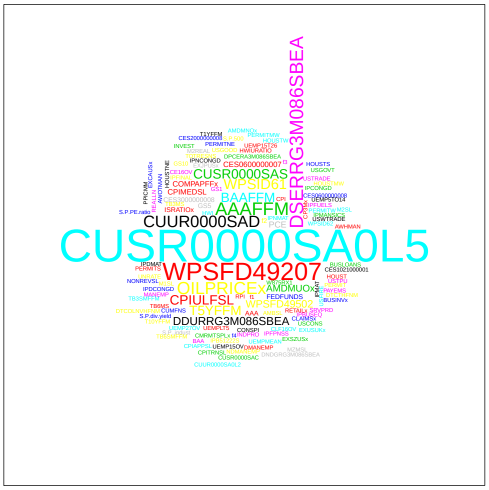
Figures 9 and 10 display the word clouds for the linear model estimated with the adaLASSO method for the first and second subsamples, respectively. In each figure, panel (a) shows the cloud for one-month-ahead models ($h = 1$), panel (b) presents the results for the three-month horizon ($h = 3$), and panels (c) and (d) consider the cases for $h = 6$ and $h = 12$, respectively. A number of findings emerge from the word clouds. First, as expected, the adaLASSO method delivers very sparse methods, and this did not change much according to the subsample considered. Second, the models across different horizons, as shown before, are quite different. For example, in the first subsample and for $h = 1$, the three variables that stand out from the cloud are CUSR0000SAOL5 (CPI: all items less medical care), WPSFD49207 (PPI: finished goods), and DSERRG3M086SBEA (PCE: Services). However, for $h = 12$, the most important variables are AMDMUOx (unfilled orders for durable goods) and HOUSTMW (Housing starts, Midwest). Finally, the pool of selected variables also changes from the first to the second sample, specially for $h = 1$. In this case, oil prices turn out to be one of the most relevant variables.

Figures 11 and 12 shows the word clouds for the RF model. From the pictures it is clear that the number of important variables are much higher. As in the adaLASSO case, the variable composition changes from the first to the second subsample.

C.2. **Additional Results: Personal Consumption Expenditure (PCE).** In this section, we report forecasting results for PCE. The main message is similar to the one described in the main text: RF models outperform traditional benchmarks as well as other linear ML methods.

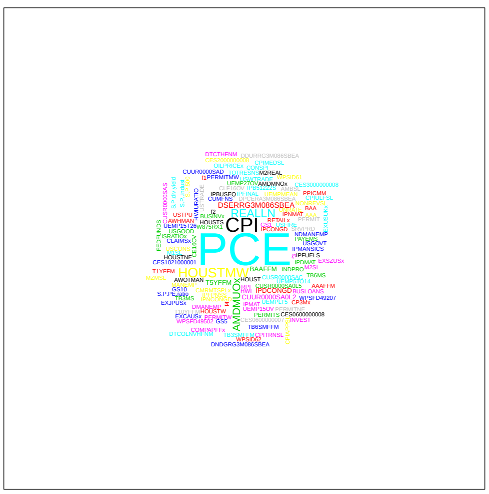
In Tables 21–23, we report for each model a number of different summary statistics across all the forecasting horizons, including the accumulated twelve-month horizon for the full out-of-sample period (1990–2015) as well as for the two subsamples considered, namely, 1990–2000 and 2001–2015. Columns (1), (2) and (3) report the RMSE, the MAE and the MAD, respectively. In columns (4), (5) and (6) we report the number of times (across horizons) each model achieved the lowest RMSE, MAE, and MAD, respectively. Columns (7)–(10) present, for square and absolute losses, the average p-values based either

FIGURE 9. Word clouds for the adaLASSO method (1990–2000).

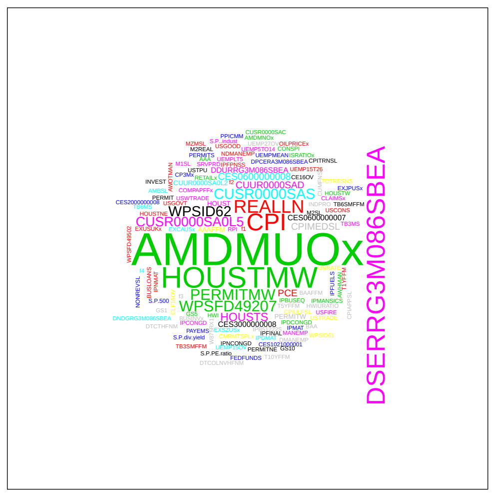


(a) h=1

(b) h=3



(c) h=6

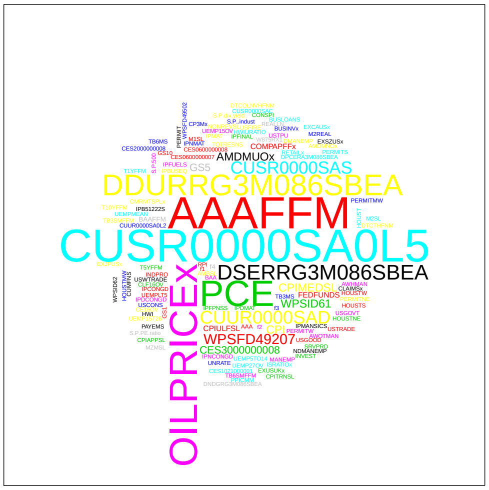


(d) h=12

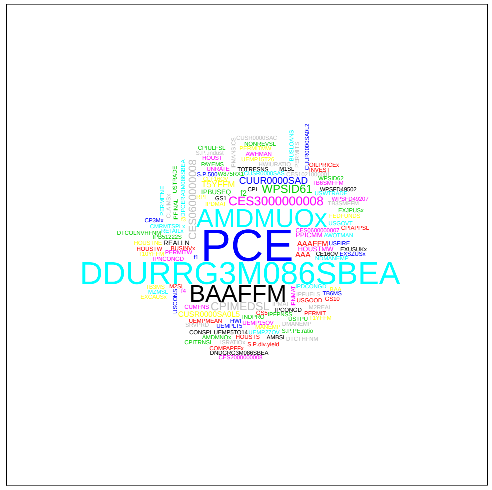
on the `range` or the $\mathsf{t_max}$ statistics as described in Hansen et al. (2011). Columns (11) and (12) show the average p-values of the SPA test proposed by Hansen (2005). Finally, columns (13) and (14) display the p-value of the multi-horizon test for superior predictive ability proposed by Quaedvlieg (2017). The superiority of the RF models is clear from the tables.

Tables 24–26 show the RMSE and, in parenthesis, the MAE for all models relative to the RW. The error measures were calculated from 132 rolling windows covering the 1990-2015 period and 180 rolling windows covering the 2001-2015 period. Values in bold show the most accurate model in each horizon. Cells in gray (blue) show the models included in the 50% MCS using the squared error (absolute error) as loss function. The MCSs

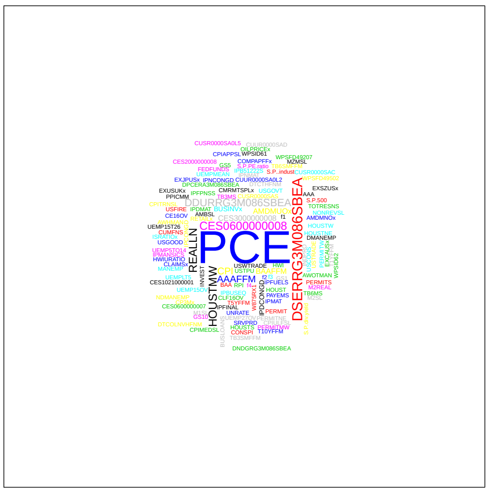
Figure 10. Word clouds for the adaLASSO method (2001–2015).



(a) h=1



(b) h=3

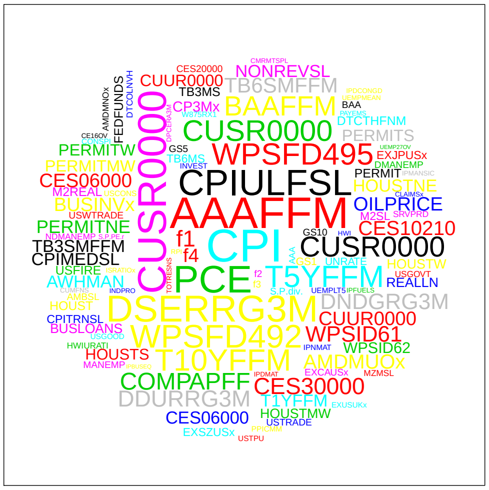


(c) h=6

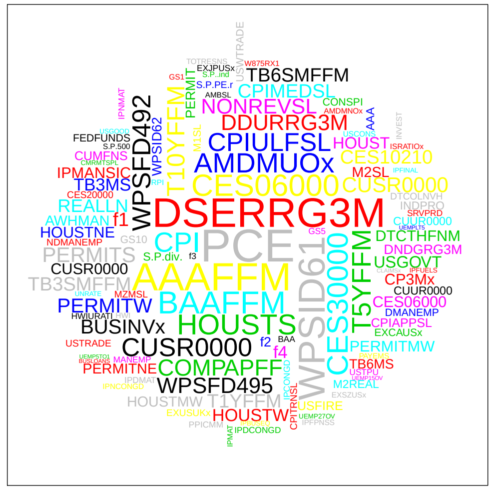
(d) h=12

were constructed based on the maximum t statistic. The last column in the table reports in how many horizons the row model was included in the MCS for square (absolute) loss. The last two rows in the table report how many models were included in the MCS for square and absolute losses. Again, the performance of the RF model is remarkable.

FIGURE 11. Word clouds for the Random Forest model (1990-2000).



(a) h=1

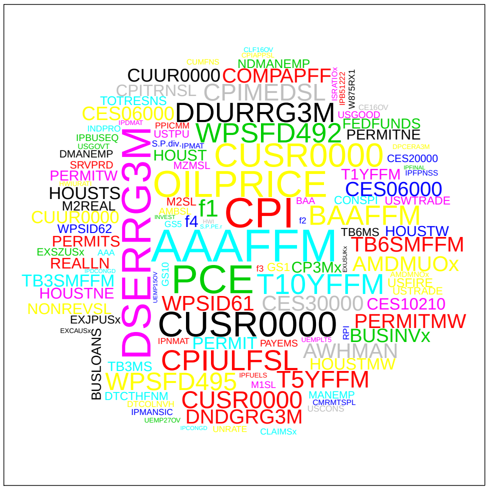


(b) h=3

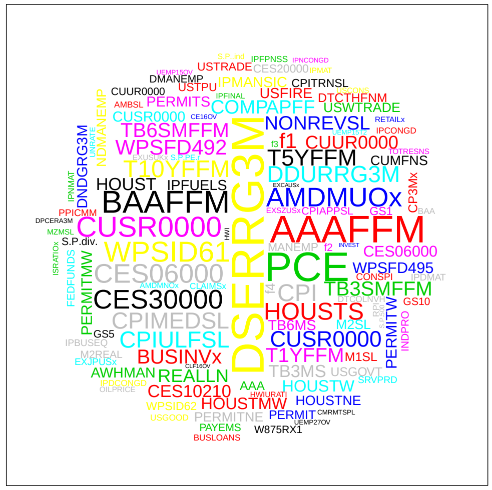
(c) h=6

(d) h=12

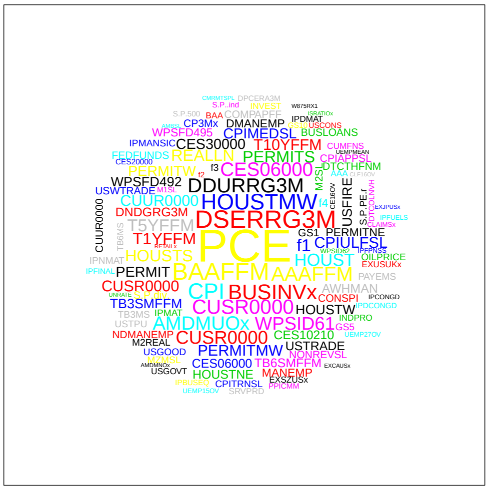
Figure 12. Word clouds for the Random Forest model (2001—2015).



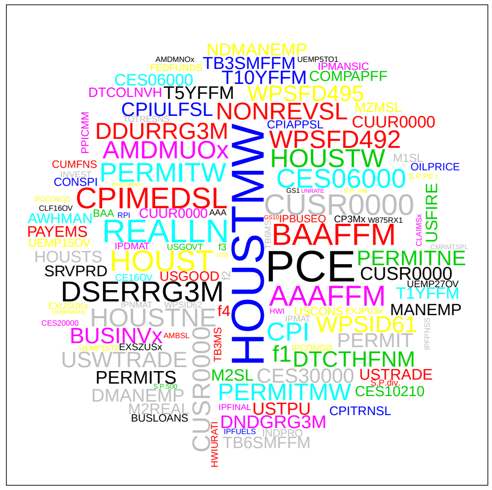
(a) h=1



(b) h=3



(c) h=6



(d) h=12

Table 21. Forecasting Results PCE: Summary statistics for the out-of-sample period from 1990–2015

The table reports, for each model, a number of different summary statistics across all the forecasting horizons, including as well the accumulated twelve-month horizon. Columns (1), (2) and (3) report the average root mean square error (RMSE), the average mean absolute error (MAE) and the average median absolute deviation (MAD). Columns (4), (5) and (6) report, respectively, the number of times (across horizons) each model achieved the lowest RMSE, MAE, and MAD. Columns (7)–(10) present, for square and absolute losses, the average p-values based either on the range or the t_max statistics as described in Hansen et al. (2011). Columns (11) and (12) show the average p-values of the Superior Predictive Ability (SPA) test proposed by Hansen (2005). Finally, columns (13) and (14) display the p-value of the multi-horizon test for superior predictive ability proposed by Quaedvlieg (2017).

	Forecasting Precision						Model Confidence Set				Sup. Pred. Ability		Multi-Horizon SPA	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	ave.	ave	ave.	# min.	# min.	# min.	ave. p.v.	ave. p.v.	ave. p.v.	ave p.v.	ave. p.v.	ave. p.v.	p.v.	p.v.
Model	RMSE	MAE	MAD	RMSE	MAE	MAD	range sq	range abs	Tmax sq	Tmax abs	sq	abs	model	test
RW	1.000	1.000	1.000	0	0	0	0.063	0.002	0.095	0.012	0.008	0.000	0.000	0.000
AR	0.835	0.825	0.802	0	0	0	0.305	0.193	0.483	0.361	0.082	0.042	0.000	0.000
UCSV	0.848	0.853	0.873	0	0	0	0.132	0.071	0.335	0.232	0.019	0.010	0.000	0.000
LASSO	0.792	0.787	0.758	0	0	1	0.254	0.286	0.603	0.570	0.153	0.188	0.623	0.623
adaLASSO	0.800	0.789	0.759	0	0	0	0.265	0.337	0.577	0.507	0.125	0.212	0.033	0.033
ElNet	0.791	0.791	0.752	0	0	1	0.266	0.264	0.603	0.554	0.163	0.174	0.997	0.997
adaElnet	0.800	0.792	0.761	0	0	0	0.250	0.226	0.581	0.480	0.106	0.118	0.001	0.001
Ridge	0.776	0.774	0.784	0	0	1	0.460	0.512	0.678	0.632	0.302	0.406	0.869	0.869
BVAR	0.837	0.852	0.813	0	0	0	0.244	0.216	0.432	0.334	0.062	0.048	0.000	0.000
Bagging	0.810	0.831	0.847	0	0	0	0.300	0.194	0.560	0.390	0.112	0.085	0.054	0.054
CSR	0.813	0.798	0.781	0	0	1	0.360	0.474	0.602	0.573	0.200	0.310	0.000	0.000
JMA	0.870	0.913	0.950	0	0	0	0.070	0.010	0.238	0.092	0.010	0.000	0.000	0.000
Factor	0.857	0.861	0.822	0	0	1	0.103	0.014	0.245	0.128	0.016	0.003	0.000	0.000
T. Factor	0.834	0.842	0.849	0	0	0	0.071	0.059	0.302	0.183	0.020	0.031	0.000	0.000
Boosting	0.859	0.894	0.926	0	0	0	0.158	0.077	0.495	0.230	0.007	0.003	0.000	0.000
RF	**0.745**	**0.736**	**0.740**	**11**	**9**	**3**	0.589	0.664	**0.960**	**0.968**	**0.928**	**0.946**	**1**	**1**
Mean	0.777	0.767	0.761	0	0	0	0.546	0.589	0.682	0.680	0.397	0.559	0.995	0.997
T.Mean	0.778	0.766	0.753	0	1	1	0.392	0.550	0.680	0.684	0.286	0.581	0.985	0.997
Median	0.779	0.766	0.749	0	0	1	0.417	0.541	0.682	0.684	0.308	0.562	0.969	0.997
RF/OLS	0.780	0.779	0.783	1	1	1	0.492	0.647	0.702	0.655	0.444	0.519	0.993	0.997
adaLASSO/RF	0.756	0.750	0.742	1	2	2	**0.658**	**0.790**	0.746	0.766	0.528	0.598	0.905	0.997

TABLE 22. Forecasting Results PCE: Summary statistics for the out-of-sample period from 1990–2000

The table reports, for each model, a number of different summary statistics across all the forecasting horizons, including as well the accumulated twelve-month horizon. Columns (1), (2) and (3) report the average root mean square error (RMSE), the average mean absolute error (MAE) and the average median absolute deviation (MAD). Columns (4), (5) and (6) report, respectively, the number of times (across horizons) each model achieved the lowest RMSE, MAE, and MAD. Columns (7)–(10) present, for square and absolute losses, the average p-values based either on the range or the t_max statistics as described in Hansen et al. (2011). Columns (11) and (12) show the average p-values of the Superior Predictive Ability (SPA) test proposed by Hansen (2005). Finally, columns (13) and (14) display the p-value of the multi-horizon test for superior predictive ability proposed by Quaedvlieg (2017).

	Forecasting Precision						Model Confidence Set				Sup. Pred. Ability		Multi-Horizon SPA	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Model	ave. RMSE	ave MAE	ave. MAD	# min. RMSE	# min. MAE	# min. MAD	ave. p.v. range sq	ave. p.v. range abs	ave. p.v. Tmax sq	ave p.v. Tmax abs	ave. p.v. sq	ave. p.v. abs	p.v. model	p.v. test
RW	1.000	1.000	1.000	0	0	0	0.095	0.028	0.335	0.346	0.008	0.003	0.000	0.000
AR	0.849	0.871	0.782	0	0	1	0.366	0.351	0.715	0.716	0.302	0.258	0.199	0.199
UCSV	0.882	0.892	0.902	0	0	0	0.407	0.495	0.740	0.773	0.204	0.377	0.228	0.228
LASSO	0.875	0.908	0.812	0	0	1	0.122	0.161	0.534	0.482	0.032	0.083	0.008	0.008
adaLASSO	0.856	0.870	0.818	0	0	0	0.173	0.350	0.650	0.730	0.080	0.211	0.008	0.008
ElNet	0.887	0.928	0.803	1	0	1	0.127	0.115	0.477	0.345	0.079	0.080	0.777	0.863
adaElnet	0.861	0.883	0.817	0	0	0	0.119	0.183	0.583	0.585	0.058	0.086	0.001	0.001
Ridge	0.828	0.850	0.827	0	1	0	0.481	0.525	0.845	0.832	0.326	0.376	0.863	0.863
BVAR	1.011	1.078	0.853	0	0	1	0.268	0.165	0.383	0.263	0.054	0.044	0.005	0.005
Bagging	0.892	0.920	0.936	0	0	0	0.448	0.634	0.773	0.755	0.215	0.293	0.177	0.177
CSR	0.825	0.841	0.839	3	4	0	0.589	0.640	0.863	0.890	0.526	0.600	0.993	0.993
JMA	1.025	1.044	1.052	0	0	0	0.107	0.131	0.271	0.268	0.016	0.029	0.000	0.000
Factor	0.949	0.985	0.866	0	0	2	0.059	0.068	0.277	0.252	0.006	0.010	0.000	0.000
T. Factor	0.929	0.959	0.876	0	0	1	0.083	0.102	0.363	0.325	0.029	0.071	0.001	0.001
Boosting	0.987	1.038	1.062	0	0	0	0.293	0.274	0.497	0.452	0.024	0.042	0.000	0.000
RF	**0.799**	**0.816**	**0.725**	**7**	**5**	**5**	0.676	0.625	**0.953**	**0.912**	**0.889**	**0.748**	**1**	**1**
Mean	0.817	0.834	0.789	0	0	0	0.658	0.668	0.863	0.882	0.638	0.642	0.961	0.993
T.Mean	0.827	0.843	0.788	0	0	0	0.303	0.380	0.836	0.876	0.250	0.284	0.067	0.067
Median	0.829	0.846	0.788	0	0	0	0.342	0.405	0.815	0.848	0.266	0.311	0.150	0.150
RF/OLS	0.815	0.837	0.845	2	3	0	**0.794**	**0.742**	0.888	0.910	0.794	0.750	0.943	0.993
adaLASSO/RF	0.835	0.852	0.790	0	0	1	0.453	0.457	0.754	0.736	0.373	0.362	0.795	0.863

TABLE 23. Forecasting Results PCE: Summary statistics for the out-of-sample period from 2001–2015

The table reports, for each model, a number of different summary statistics across all the forecasting horizons, including as well the accumulated twelve-month horizon. Columns (1), (2) and (3) report the average root mean square error (RMSE), the average mean absolute error (MAE) and the average median absolute deviation (MAD). Columns (4), (5) and (6) report, respectively, the number of times (across horizons) each model achieved the lowest RMSE, MAE, and MAD. Columns (7)–(10) present, for square and absolute losses, the average p-values based either on the range or the t_max statistics as described in Hansen et al. (2011). Columns (11) and (12) show the average p-values of the Superior Predictive Ability (SPA) test proposed by Hansen (2005). Finally, columns (13) and (14) display the p-value of the multi-horizon test for superior predictive ability proposed by Quaedvlieg (2017).

	Forecasting Precision							Model Confidence Set					Sup. Pred. Ability		Multi-Horizon SPA	
	(1)	(2)	(3)	(4)	(5)	(6)		(7)	(8)	(9)	(10)		(11)	(12)	(13)	(14)
Model	ave. RMSE	ave MAE	ave. MAD	# min. RMSE	# min. MAE	# min. MAD		ave. p.v. range sq	ave. p.v. range abs	ave. p.v. Tmax sq	ave p.v. Tmax abs		ave. p.v. sq	ave. p.v. abs	p.v. model	p.v. test
RW	1.000	1.000	1.000	0	0	0	0.090	0.002	0.090	0.004	0.011	0.000	0.000	0.000		
AR	0.832	0.805	0.788	0	0	0	0.332	0.165	0.543	0.274	0.082	0.032	0.000	0.000		
UCSV	0.841	0.837	0.842	0	0	0	0.138	0.041	0.316	0.144	0.024	0.007	0.000	0.000		
LASSO	0.771	0.733	0.708	0	1	0	0.470	0.547	0.802	0.680	0.309	0.451	0.887	0.935		
adaLASSO	0.786	0.752	0.708	0	0	**3**	0.423	0.430	0.756	0.616	0.234	0.245	0.209	0.209		
ElNet	0.767	0.730	0.708	0	0	0	0.449	0.541	0.812	0.699	0.292	0.485	0.824	0.935		
adaElnet	0.785	0.751	0.715	0	0	0	0.446	0.436	0.760	0.615	0.236	0.220	0.149	0.149		
Ridge	0.763	0.750	0.706	0	0	0	0.529	0.584	0.818	0.645	0.380	0.417	0.408	0.408		
BVAR	0.790	0.750	0.701	0	0	**3**	0.405	0.538	0.767	0.679	0.209	0.317	0.200	0.200		
Bagging	0.789	0.794	0.782	0	0	2	0.416	0.278	0.661	0.359	0.223	0.105	0.176	0.176		
CSR	0.810	0.778	0.758	0	0	0	0.431	0.399	0.657	0.518	0.188	0.188	0.013	0.013		
JMA	0.830	0.856	0.883	0	0	0	0.288	0.074	0.480	0.153	0.062	0.006	0.000	0.000		
Factor	0.834	0.805	0.765	0	0	0	0.296	0.103	0.445	0.234	0.070	0.024	0.000	0.000		
T. Factor	0.809	0.790	0.771	0	0	0	0.319	0.129	0.543	0.262	0.114	0.050	0.000	0.000		
Boosting	0.826	0.832	0.829	0	0	0	0.334	0.228	0.690	0.407	0.040	0.029	0.000	0.000		
RF	**0.733**	**0.700**	0.688	**8**	4	2	**0.768**	**0.823**	**0.934**	**0.887**	**0.863**	**0.875**	**1**	**1**		
Mean	0.768	0.736	0.710	0	0	0	0.591	0.482	0.818	0.671	0.424	0.402	0.876	0.876		
T.Mean	0.766	0.731	0.714	0	0	0	0.482	0.616	0.811	0.693	0.348	0.622	0.794	0.935		
Median	0.766	0.731	0.711	0	0	0	0.507	0.640	0.816	0.704	0.397	0.640	0.868	0.935		
RF/OLS	0.772	0.754	0.705	1	0	2	0.540	0.550	0.826	0.648	0.343	0.330	0.935	0.935		
adaLASSO/RF	0.736	0.705	**0.680**	4	**8**	1	0.706	0.741	0.878	0.893	0.774	0.746	0.910	0.935		

Table 24. Forecasting Errors for the PCE from 1990 to 2015

The table shows the root mean squared error (RMSE) and, between parenthesis, the mean absolute errors (MAE) for all models relative to the Random Walk (RW). The error measures were calculated from 132 rolling windows covering the 1990-2015 period and 180 rolling windows covering the 2001-2015 period. Values in bold show the most accurate model in each horizon. Cells in gray (blue) show the models included in the 50% model confidence set (MCS) using the squared error (absolute error) as loss function. The MCSs were constructed based on the maximum t-statistic. The last column in the table reports in how many horizons the row model was included in the MCS for square (absolute) loss. The last two rows in the table reports how many models were included in the MCS for square and absolute losses.

Personal Consumer Expenditure 1990–2015														
	Forecasting Horizon													
RMSE/(MAE)	1	2	3	4	5	6	7	8	9	10	11	12	Acc.	RMSE count (MAE count)
RW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(0)
AR	0.89	0.83	0.82	0.84	0.82	0.80	0.79	0.78	0.82	0.84	0.87	0.80	0.96	7
	(0.87)	(0.81)	(0.79)	(0.83)	(0.82)	(0.81)	(0.79)	(0.75)	(0.82)	(0.84)	(0.89)	(0.82)	(0.91)	(3)
UCSV	0.94	0.86	0.84	0.86	0.84	0.83	0.83	0.82	0.83	0.84	0.86	0.83	0.85	5
	(0.92)	(0.86)	(0.82)	(0.85)	(0.84)	(0.83)	(0.84)	(0.82)	(0.84)	(0.84)	(0.90)	(0.83)	(0.90)	(0)
LASSO	0.83	0.78	0.75	0.79	0.78	0.78	0.78	0.77	0.81	0.82	0.84	0.76	0.79	9
	(0.80)	(0.77)	(0.73)	(0.80)	(0.81)	(0.79)	(0.78)	(0.73)	(0.80)	(0.81)	(0.86)	(0.77)	(0.77)	(7)
adaLASSO	0.84	0.79	0.77	0.80	0.79	0.78	0.78	0.79	0.82	0.82	0.84	0.76	0.83	9
	(0.82)	(0.78)	(0.74)	(0.80)	(0.79)	(0.77)	(0.77)	(0.74)	(0.80)	(0.80)	(0.85)	(0.77)	(0.82)	(6)
ElNet	0.83	0.78	0.75	0.79	0.78	0.78	0.77	0.77	0.81	0.82	0.85	0.76	0.79	9
	(0.80)	(0.77)	(0.73)	(0.81)	(0.81)	(0.80)	(0.78)	(0.74)	(0.81)	(0.81)	(0.87)	(0.78)	(0.78)	(6)
adaElnet	0.84	0.79	0.76	0.80	0.79	0.78	0.78	0.78	0.82	0.82	0.84	0.76	0.83	9
	(0.83)	(0.79)	(0.74)	(0.80)	(0.80)	(0.78)	(0.77)	(0.74)	(0.80)	(0.81)	(0.86)	(0.77)	(0.81)	(5)
Ridge	0.85	0.76	0.76	0.79	0.76	0.77	0.76	0.75	0.79	0.78	0.81	0.74	0.77	13
	(0.83)	(0.75)	(0.73)	(0.78)	(0.79)	(0.78)	(0.75)	(0.72)	(0.79)	(0.78)	(0.83)	(0.75)	(0.77)	(9)
BVAR	0.90	0.80	0.79	0.81	0.79	0.80	0.81	0.82	0.86	0.87	0.90	0.83	0.90	5
	(0.89)	(0.78)	(0.79)	(0.83)	(0.83)	(0.84)	(0.83)	(0.82)	(0.89)	(0.89)	(0.94)	(0.87)	(0.89)	(3)
Bagging	0.87	0.78	0.78	0.83	0.82	0.81	0.82	0.80	0.83	0.81	0.84	0.76	0.76	8
	(0.86)	(0.76)	(0.76)	(0.86)	(0.90)	(0.86)	(0.83)	(0.80)	(0.86)	(0.83)	(0.88)	(0.78)	(0.82)	(4)
CSR	0.85	0.78	0.77	0.80	0.78	0.79	0.79	0.78	0.82	0.84	0.87	0.81	0.90	11
	(0.82)	(0.78)	(0.74)	(0.80)	(0.79)	(0.78)	(0.76)	(0.73)	(0.80)	(0.82)	(0.87)	(0.81)	(0.87)	(8)
JMA	0.95	0.88	0.84	0.89	0.87	0.84	0.88	0.85	0.92	0.87	0.91	0.82	0.79	2
	(0.94)	(0.91)	(0.85)	(0.97)	(0.97)	(0.91)	(0.91)	(0.86)	(0.97)	(0.92)	(0.95)	(0.88)	(0.83)	(1)
Factor	0.89	0.84	0.83	0.84	0.83	0.84	0.83	0.83	0.87	0.87	0.89	0.85	0.95	0
	(0.87)	(0.84)	(0.83)	(0.84)	(0.85)	(0.85)	(0.84)	(0.81)	(0.89)	(0.88)	(0.90)	(0.88)	(0.92)	(0)
T. Factor	0.90	0.83	0.81	0.81	0.81	0.81	0.80	0.80	0.83	0.84	0.85	0.82	0.92	2
	(0.91)	(0.85)	(0.82)	(0.80)	(0.83)	(0.83)	(0.81)	(0.77)	(0.84)	(0.86)	(0.87)	(0.84)	(0.90)	(1)
Boosting	0.99	0.83	0.82	0.84	0.82	0.84	0.84	0.84	0.87	0.86	0.88	0.81	0.91	7
	(1.00)	(0.83)	(0.83)	(0.89)	(0.89)	(0.89)	(0.88)	(0.85)	(0.91)	(0.88)	(0.92)	(0.83)	(1.02)	(1)
RF	0.86	**0.76**	**0.74**	**0.76**	**0.73**	**0.73**	**0.72**	0.72	**0.75**	**0.75**	**0.78**	**0.71**	**0.67**	13
	(0.82)	**(0.74)**	(0.73)	**(0.77)**	**(0.77)**	(0.75)	**(0.72)**	(0.68)	**(0.73)**	**(0.74)**	**(0.78)**	**(0.70)**	**(0.63)**	(13)
Mean	0.84	0.77	0.76	0.78	0.77	0.77	0.76	0.75	0.78	0.78	0.80	0.75	0.80	13
	(0.81)	(0.77)	(0.73)	(0.78)	(0.78)	(0.77)	(0.76)	(0.71)	(0.77)	(0.77)	(0.81)	(0.75)	(0.77)	(10)
T.Mean	0.84	0.77	0.75	0.78	0.77	0.76	0.76	0.75	0.78	0.79	0.81	0.75	0.80	13
	(0.81)	(0.76)	**(0.73)**	(0.78)	(0.78)	(0.77)	(0.75)	(0.71)	(0.77)	(0.77)	(0.82)	(0.75)	(0.77)	(10)
Median	0.83	0.77	0.75	0.78	0.77	0.77	0.76	0.76	0.78	0.79	0.81	0.75	0.79	13
	(0.81)	(0.76)	(0.73)	(0.78)	(0.78)	(0.77)	(0.75)	(0.71)	(0.77)	(0.77)	(0.82)	(0.75)	(0.77)	(10)
RF/OLS	**0.82**	0.76	0.75	0.78	0.76	0.76	0.76	0.76	0.79	0.79	0.82	0.76	0.82	13
	(0.80)	(0.76)	(0.73)	(0.78)	(0.78)	(0.78)	(0.75)	(0.72)	(0.79)	(0.79)	(0.85)	(0.78)	(0.82)	(9)
adaLASSO/RF	0.85	0.79	0.75	0.76	0.74	0.74	0.72	**0.72**	0.76	0.76	0.80	0.74	0.70	13
	(0.81)	(0.78)	(0.73)	(0.78)	(0.77)	**(0.75)**	(0.72)	**(0.68)**	(0.74)	(0.76)	(0.81)	(0.73)	(0.68)	(11)
RMSE count	13	16	13	16	16	17	8	8	16	18	7	8	18	
MAE count	(10)	(16)	(13)	(15)	(14)	(11)	(7)	(8)	(1)	(3)	(1)	(8)	(10)	

C.3. **Additional Results: CPI-Core.** In this section, we report forecasting results for the Core of the Consumer Price Index. The CPI-Core series exhibits a dynamics quite different from the other two inflation indexes reported before. More specifically there is a clear seasonal patern in the series.

Table 25. Forecasting Errors for the PCE from 1990 to 2000

The table shows the root mean squared error (RMSE) and, between parenthesis, the mean absolute errors (MAE) for all models relative to the Random Walk (RW). The error measures were calculated from 132 rolling windows covering the 1990-2015 period and 180 rolling windows covering the 2001-2015 period. Values in bold show the most accurate model in each horizon. Cells in gray (blue) show the models included in the 50% model confidence set (MCS) using the squared error (absolute error) as loss function. The MCSs were constructed based on the maximum t-statistic. The last column in the table reports in how many horizons the row model was included in the MCS for square (absolute) loss. The last two rows in the table reports how many models were included in the MCS for square and absolute losses.

Personal Consumer Expenditure 1990–2000														
	Forecasting Horizon													
RMSE/(MAE)	1	2	3	4	5	6	7	8	9	10	11	12	Acc.	RMSE count (MAE count)
RW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	3
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(2)
AR	0.84	0.80	0.86	0.82	0.79	0.80	0.83	0.84	0.90	0.86	0.95	0.92	0.82	12
	(0.86)	(0.79)	(0.88)	(0.85)	(0.79)	(0.84)	(0.89)	(0.80)	(0.89)	(0.90)	(1.00)	(0.98)	(0.84)	(11)
UCSV	0.89	0.85	0.86	0.88	0.85	0.84	0.86	0.87	0.89	0.87	0.90	0.89	1.04	12
	(0.90)	(0.84)	(0.87)	(0.88)	(0.84)	(0.85)	(0.89)	(0.86)	(0.86)	(0.87)	(0.93)	(0.89)	(1.15)	(11)
LASSO	0.83	0.83	0.86	0.89	0.85	0.88	0.89	0.89	0.96	0.88	0.95	0.88	0.78	8
	(0.83)	(0.82)	(0.87)	(0.94)	(0.89)	(0.93)	(1.00)	(0.86)	(0.96)	(0.91)	(1.02)	(0.95)	(0.81)	(6)
adaLASSO	0.84	0.84	0.86	0.85	0.81	0.83	0.86	0.88	0.93	0.83	0.90	0.87	0.82	10
	(0.84)	(0.83)	(0.87)	(0.87)	(0.82)	(0.85)	(0.92)	(0.83)	(0.91)	(0.85)	(0.94)	(0.92)	(0.85)	(11)
ElNet	**0.80**	0.83	0.87	0.90	0.86	0.89	0.92	0.91	0.95	0.88	1.00	0.91	0.80	7
	(0.81)	(0.83)	(0.89)	(0.97)	(0.92)	(0.96)	(1.02)	(0.88)	(0.96)	(0.92)	(1.08)	(0.98)	(0.83)	(2)
adaElnet	0.85	0.84	0.86	0.86	0.80	0.84	0.86	0.88	0.95	0.84	0.92	0.88	0.80	8
	(0.86)	(0.84)	(0.87)	(0.90)	(0.82)	(0.87)	(0.93)	(0.84)	(0.93)	(0.87)	(0.97)	(0.94)	(0.83)	(6)
Ridge	0.82	0.77	0.85	0.83	0.78	0.82	0.83	0.84	0.90	0.83	0.91	0.84	0.76	13
	(0.82)	**(0.75)**	(0.86)	(0.85)	(0.82)	(0.87)	(0.90)	(0.80)	(0.90)	(0.87)	(0.96)	(0.87)	(0.80)	(13)
BVAR	0.99	0.83	0.95	0.91	0.87	0.94	1.02	1.07	1.15	1.05	1.16	1.10	1.11	6
	(1.00)	(0.80)	(1.00)	(0.99)	(0.95)	(1.05)	(1.14)	(1.08)	(1.22)	(1.15)	(1.29)	(1.22)	(1.17)	(2)
Bagging	0.85	0.82	0.94	0.90	0.86	0.86	0.84	0.83	0.91	0.86	0.97	0.89	1.09	12
	(0.85)	(0.80)	(0.95)	(0.94)	(0.93)	(0.92)	(0.91)	(0.80)	(0.90)	(0.88)	(1.00)	(0.91)	(1.19)	(9)
CSR	0.83	0.83	0.86	0.81	**0.76**	**0.78**	0.80	**0.81**	0.87	0.83	0.90	0.85	0.81	13
	(0.83)	(0.81)	(0.87)	(0.83)	**(0.78)**	**(0.80)**	**(0.85)**	**(0.76)**	(0.86)	(0.86)	(0.93)	(0.89)	(0.87)	(13)
JMA	0.94	1.00	1.06	1.04	1.00	0.92	1.02	1.01	1.17	1.00	1.08	1.00	1.09	4
	(0.97)	(1.03)	(1.10)	(1.06)	(1.06)	(0.95)	(1.11)	(0.92)	(1.15)	(1.03)	(1.06)	(0.99)	(1.16)	(2)
Factor	0.89	0.89	0.96	0.88	0.86	0.92	0.91	0.92	1.05	0.98	1.05	1.04	0.96	3
	(0.91)	(0.89)	(1.01)	(0.89)	(0.87)	(0.95)	(0.98)	(0.89)	(1.09)	(1.04)	(1.10)	(1.13)	(1.04)	(1)
T. Factor	0.95	0.91	1.01	0.84	0.83	0.88	0.89	0.88	1.00	0.96	0.99	0.97	0.94	6
	(0.96)	(0.91)	(1.07)	(0.85)	(0.84)	(0.91)	(0.98)	(0.86)	(1.01)	(1.01)	(1.03)	(1.02)	(1.02)	(3)
Boosting	0.99	0.90	1.01	0.96	0.91	0.98	1.00	1.04	1.06	0.94	0.99	0.92	1.09	6
	(1.03)	(0.91)	(1.09)	(1.02)	(0.97)	(1.05)	(1.09)	(1.01)	(1.06)	(0.96)	(1.03)	(0.94)	(1.30)	(4)
RF	0.82	**0.77**	0.86	0.83	0.78	0.80	**0.80**	0.81	**0.84**	**0.78**	**0.85**	**0.79**	**0.67**	13
	(0.82)	(0.77)	(0.90)	(0.87)	(0.81)	(0.85)	(0.85)	(0.76)	**(0.82)**	**(0.80)**	**(0.90)**	**(0.82)**	**(0.63)**	(12)
Mean	0.82	0.79	0.84	0.82	0.78	0.80	0.81	0.81	0.88	0.82	0.87	0.84	0.74	13
	(0.83)	(0.79)	(0.85)	(0.84)	(0.79)	(0.83)	(0.88)	(0.77)	(0.86)	(0.84)	(0.92)	(0.88)	(0.74)	(13)
T.Mean	0.82	0.80	0.85	0.82	0.79	0.80	0.82	0.83	0.90	0.82	0.89	0.85	0.75	13
	(0.83)	(0.79)	(0.86)	(0.84)	(0.80)	(0.83)	(0.89)	(0.79)	(0.88)	(0.85)	(0.94)	(0.90)	(0.76)	(13)
Median	0.82	0.80	0.84	0.82	0.79	0.81	0.83	0.84	0.90	0.82	0.89	0.85	0.75	13
	(0.83)	(0.79)	(0.85)	(0.85)	(0.81)	(0.83)	(0.89)	(0.79)	(0.89)	(0.85)	(0.94)	(0.90)	(0.76)	(13)
RF/OLS	0.81	0.77	**0.83**	**0.80**	0.76	0.79	0.80	0.83	0.87	0.80	0.87	0.83	0.81	13
	(0.81)	(0.76)	**(0.84)**	**(0.82)**	(0.79)	(0.84)	(0.87)	(0.78)	(0.86)	(0.84)	(0.93)	(0.87)	(0.89)	(13)
adaLASSO/RF	0.81	0.89	0.91	0.87	0.81	0.82	0.84	0.81	0.85	0.79	0.90	0.84	0.71	11
	(0.82)	(0.87)	(0.94)	(0.90)	(0.83)	(0.87)	(0.89)	(0.77)	(0.84)	(0.83)	(0.94)	(0.87)	(0.71)	(10)
RMSE count	13	15	14	16	17	20	11	12	11	15	18	16	21	
MAE count	(12)	(14)	(11)	(11)	(13)	(13)	(11)	(11)	(12)	(15)	(18)	(13)	(16)	

In Tables 27–??, we report for each model a number of different summary statistics across all the forecasting horizons, including the accumulated twelve-month horizon for the full out-of-sample period (1990–2015) as well as for the two subsamples considered, namely, 1990–2000 and 2001–2015. Columns (1), (2) and (3) report the RMSE, the MAE

Table 26. Forecasting Errors for the PCE from 2001 to 2015

The table shows the root mean squared error (RMSE) and, between parenthesis, the mean absolute errors (MAE) for all models relative to the Random Walk (RW). The error measures were calculated from 132 rolling windows covering the 2001-2015 period and 180 rolling windows covering the 2001-2015 period. Values in bold show the most accurate model in each horizon. Cells in gray (blue) show the models included in the 50% model confidence set (MCS) using the squared error (absolute error) as loss function. The MCSs were constructed based on the maximum t-statistic. The last column in the table reports in how many horizons the row model was included in the MCS for square (absolute) loss. The last two rows in the table reports how many models were included in the MCS for square and absolute losses.

Personal Consumer Expenditure 1990–2015														
	Forecasting Horizon													
RMSE/(MAE)	1	2	3	4	5	6	7	8	9	10	11	12	Acc.	RMSE count (MAE count)
RW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(0)
AR	0.91	0.84	0.81	0.84	0.82	0.80	0.79	0.77	0.80	0.83	0.85	0.77	0.99	8
	(0.88)	(0.82)	(0.75)	(0.82)	(0.83)	(0.79)	(0.74)	(0.72)	(0.78)	(0.81)	(0.83)	(0.74)	(0.93)	(0)
UCSV	0.96	0.86	0.84	0.86	0.83	0.83	0.82	0.81	0.82	0.83	0.85	0.81	0.81	3
	(0.94)	(0.87)	(0.81)	(0.84)	(0.84)	(0.83)	(0.82)	(0.81)	(0.83)	(0.83)	(0.88)	(0.80)	(0.81)	(1)
LASSO	0.83	0.77	0.73	0.77	0.76	0.76	0.75	0.75	0.78	0.80	0.81	0.73	0.79	13
	(0.79)	(0.74)	(0.67)	(0.74)	(0.77)	(0.72)	(0.69)	(0.68)	(0.73)	(0.76)	(0.78)	(0.68)	(0.76)	(11)
adaLASSO	0.84	0.77	0.74	0.78	0.78	0.77	0.76	0.77	0.79	0.81	0.83	0.73	0.83	13
	(0.80)	(0.76)	(0.69)	(0.77)	(0.78)	(0.74)	(0.71)	(0.70)	(0.74)	(0.78)	(0.81)	(0.70)	(0.80)	(10)
ElNet	0.84	0.76	0.72	0.76	0.76	0.75	0.74	0.74	0.78	0.80	0.81	0.72	0.78	13
	(0.80)	(0.74)	(0.67)	(0.73)	(0.76)	(0.72)	(0.68)	(0.67)	(0.74)	(0.76)	(0.78)	(0.68)	(0.76)	(12)
adaElnet	0.84	0.78	0.74	0.78	0.79	0.77	0.76	0.76	0.79	0.81	0.82	0.73	0.83	13
	(0.81)	(0.76)	(0.68)	(0.76)	(0.78)	(0.74)	(0.71)	(0.69)	(0.75)	(0.78)	(0.81)	(0.70)	(0.80)	(9)
Ridge	0.87	0.76	0.73	0.77	0.76	0.75	0.74	0.73	0.76	0.77	0.78	0.71	0.77	13
	(0.84)	(0.74)	(0.67)	(0.75)	(0.78)	(0.75)	(0.69)	(0.68)	(0.74)	(0.74)	(0.76)	(0.70)	(0.76)	(11)
BVAR	0.85	0.79	0.75	0.78	0.77	0.77	0.76	0.76	0.79	0.81	0.82	0.76	0.85	12
	(0.83)	(0.77)	(0.70)	(0.76)	(0.77)	(0.74)	(0.69)	(0.70)	(0.75)	(0.76)	(0.78)	(0.71)	(0.80)	(12)
Bagging	0.88	0.76	0.74	0.81	0.81	0.79	0.82	0.79	0.81	0.80	0.81	0.73	0.68	11
	(0.87)	(0.75)	(0.69)	(0.82)	(0.88)	(0.83)	(0.80)	(0.80)	(0.84)	(0.80)	(0.82)	(0.72)	(0.70)	(5)
CSR	0.86	0.77	0.75	0.79	0.79	0.79	0.79	0.77	0.81	0.84	0.86	0.80	0.92	10
	(0.81)	(0.76)	(0.68)	(0.79)	(0.80)	(0.77)	(0.73)	(0.71)	(0.77)	(0.81)	(0.84)	(0.77)	(0.87)	(6)
JMA	0.96	0.84	0.78	0.85	0.83	0.82	0.84	0.82	0.86	0.83	0.86	0.77	0.72	6
	(0.92)	(0.84)	(0.75)	(0.92)	(0.93)	(0.89)	(0.82)	(0.83)	(0.89)	(0.87)	(0.91)	(0.83)	(0.73)	(1)
Factor	0.89	0.83	0.80	0.83	0.82	0.81	0.81	0.81	0.83	0.83	0.84	0.80	0.95	4
	(0.85)	(0.82)	(0.75)	(0.82)	(0.83)	(0.80)	(0.78)	(0.77)	(0.80)	(0.80)	(0.81)	(0.77)	(0.88)	(1)
T. Factor	0.88	0.80	0.76	0.80	0.80	0.79	0.78	0.78	0.79	0.80	0.82	0.78	0.91	8
	(0.88)	(0.81)	(0.72)	(0.78)	(0.82)	(0.80)	(0.74)	(0.74)	(0.77)	(0.78)	(0.80)	(0.76)	(0.86)	(2)
Boosting	1.00	0.80	0.77	0.81	0.79	0.80	0.80	0.79	0.82	0.84	0.85	0.79	0.87	11
	(0.98)	(0.79)	(0.73)	(0.82)	(0.84)	(0.82)	(0.79)	(0.78)	(0.85)	(0.84)	(0.87)	(0.78)	(0.93)	(3)
RF	0.88	0.76	**0.71**	0.74	**0.72**	**0.71**	0.70	0.70	**0.73**	**0.74**	**0.76**	**0.70**	**0.67**	13
	(0.82)	(0.73)	(0.66)	(0.73)	(0.74)	(0.71)	(0.67)	(0.65)	(0.70)	**(0.71)**	**(0.72)**	**(0.64)**	**(0.63)**	(13)
Mean	0.85	0.76	0.74	0.77	0.77	0.76	0.75	0.74	0.76	0.77	0.78	0.72	0.81	13
	(0.81)	(0.75)	(0.68)	(0.75)	(0.77)	(0.74)	(0.71)	(0.68)	(0.73)	(0.73)	(0.76)	(0.69)	(0.78)	(11)
T.Mean	0.84	0.76	0.73	0.77	0.76	0.75	0.74	0.74	0.76	0.78	0.79	0.72	0.80	13
	(0.80)	(0.75)	(0.67)	(0.75)	(0.77)	(0.73)	(0.69)	(0.67)	(0.72)	(0.74)	(0.76)	(0.68)	(0.77)	(12)
Median	0.84	0.76	0.73	0.77	0.76	0.76	0.74	0.74	0.76	0.78	0.79	0.72	0.80	13
	(0.80)	(0.75)	(0.68)	(0.75)	(0.76)	(0.74)	(0.69)	(0.67)	(0.72)	(0.74)	(0.76)	(0.68)	(0.77)	(13)
RF/OLS	**0.83**	0.76	0.74	0.78	0.76	0.76	0.75	0.74	0.77	0.78	0.81	0.74	0.82	13
	(0.80)	(0.75)	(0.68)	(0.76)	(0.78)	(0.75)	(0.70)	(0.69)	(0.76)	(0.77)	(0.81)	(0.74)	(0.80)	(11)
adaLASSO/RF	0.86	**0.75**	0.71	**0.73**	0.73	0.72	**0.69**	**0.70**	0.74	0.76	0.78	0.71	0.69	13
	(0.81)	**(0.73)**	**(0.65)**	**(0.73)**	**(0.74)**	**(0.70)**	**(0.66)**	**(0.64)**	**(0.70)**	(0.72)	(0.75)	(0.67)	(0.67)	(13)
RMSE count	15	16	18	17	20	16	15	15	18	20	17	13	16	
MAE count	(12)	(15)	(15)	(9)	(12)	(12)	(11)	(13)	(3)	(16)	(10)	(13)	(16)	

and the MAD, respectively. In columns (4), (5) and (6) we report the number of times (across horizons) each model achieved the lowest RMSE, MAE, and MAD, respectively. Columns (7)–(10) present, for square and absolute losses, the average p-values based either on the range or the $t_{m}ax$ statistics as described in Hansen et al. (2011). Columns (11) and (12) show the average p-values of the SPA test proposed by Hansen (2005). Finally, columns (13) and (14) display the p-value of the multi-horizon test for superior predictive ability proposed by Quaedvlieg (2017).

Tables 30–32 show the RMSE and, in parenthesis, the MAE for all models relative to the RW. The error measures were calculated from 132 rolling windows covering the 1990-2015 period and 180 rolling windows covering the 2001-2015 period. Values in bold show the most accurate model in each horizon. Cells in gray (blue) show the models included in the 50% MCS using the squared error (absolute error) as loss function. The MCSs were constructed based on the maximum t statistic. The last column in the table reports in how many horizons the row model was included in the MCS for square (absolute) loss. The last two rows in the table report how many models were included in the MCS for square and absolute losses.

Table 27. Forecasting Results CPI-Core: Summary statistics for the out-of-sample period from 1990–2015

The table reports, for each model, a number of different summary statistics across all the forecasting horizons, including as well the accumulated twelve-month horizon. Columns (1), (2) and (3) report the average root mean square error (RMSE), the average mean absolute error (MAE) and the average median absolute deviation (MAD). Columns (4), (5) and (6) report, respectively, the number of times (across horizons) each model achieved the lowest RMSE, MAE, and MAD. Columns (7)–(10) present, for square and absolute losses, the average p-values based either on the range or the t_max statistics as described in Hansen et al. (2011). Columns (11) and (12) show the average p-values of the Superior Predictive Ability (SPA) test proposed by Hansen (2005). Finally, columns (13) and (14) display the p-value of the multi-horizon test for superior predictive ability proposed by Quaedvlieg (2017).

	Forecasting Precision							Model Confidence Set					Sup. Pred. Ability		Multi-Horizon SPA	
	(1)	(2)	(3)	(4)	(5)	(6)		(7)	(8)	(9)	(10)		(11)	(12)	(13)	(14)
Model	ave. RMSE	ave MAE	ave. MAD	# min. RMSE	# min. MAE	# min. MAD		ave. p.v. range sq	ave. p.v. range abs	ave. p.v. Tmax sq	ave p.v. Tmax abs		ave. p.v. sq	ave. p.v. abs	p.v. model	p.v. test
RW	1.000	1.000	1.000	2	2	2	0.046	0.065	0.159	0.155	0.141	0.140				
AR	0.835	0.847	0.747	0	0	0	0.177	0.120	0.405	0.324	0.153	0.090				
UCSV	0.891	0.887	0.882	0	0	0	0.080	0.081	0.100	0.104	0.001	0.001				
LASSO	0.882	0.892	0.780	0	0	0	0.000	0.000	0.031	0.018	0.000	0.000				
adaLASSO	0.826	0.817	0.766	0	0	0	0.037	0.027	0.192	0.209	0.020	0.015				
ElNet	0.895	0.905	0.792	0	0	0	0.000	0.000	0.030	0.015	0.000	0.000				
adaElnet	0.833	0.830	0.778	0	0	0	0.008	0.008	0.141	0.150	0.003	0.003				
Ridge	0.881	0.890	0.835	0	0	0	0.002	0.002	0.068	0.094	0.000	0.000				
BVAR	1.126	1.143	0.937	0	0	0	0.000	0.000	0.000	0.000	0.000	0.000				
Bagging	0.787	0.765	0.746	3	3	2	0.387	0.354	0.607	0.631	0.386	0.411				
CSR	0.800	0.797	0.732	1	0	1	0.481	0.469	0.578	0.523	0.493	0.414				
JMA	0.902	0.871	0.814	0	0	0	0.013	0.042	0.107	0.178	0.002	0.010				
Factor	0.877	0.878	0.805	0	0	0	0.034	0.050	0.219	0.215	0.008	0.015				
T. Factor	0.895	0.894	0.842	0	0	0	0.067	0.092	0.268	0.237	0.011	0.013				
Boosting	1.080	1.099	1.071	0	0	0	0.001	0.000	0.011	0.005	0.000	0.000				
RF	**0.749**	**0.751**	**0.681**	**6**	**5**	**6**	**0.598**	**0.605**	**0.739**	**0.725**	**0.695**	**0.665**				
Mean	0.792	0.795	0.742	0	0	0	0.340	0.228	0.436	0.391	0.363	0.293				
T.Mean	0.801	0.806	0.737	0	0	0	0.116	0.062	0.295	0.307	0.122	0.049				
Median	0.803	0.807	0.734	0	0	0	0.125	0.083	0.293	0.314	0.123	0.080				
RF/OLS	0.777	0.771	0.746	0	0	0	0.347	0.335	0.539	0.526	0.402	0.377				
adaLASSO/RF	0.771	0.758	0.700	1	3	2	0.547	0.588	0.531	0.610	0.378	0.464				

TABLE 28. Forecasting Results CPI-Core: Summary statistics for the out-of-sample period from 1990–2000

The table reports, for each model, a number of different summary statistics across all the forecasting horizons, including as well the accumulated twelve-month horizon. Columns (1), (2) and (3) report the average root mean square error (RMSE), the average mean absolute error (MAE) and the average median absolute deviation (MAD). Columns (4), (5) and (6) report, respectively, the number of times (across horizons) each model achieved the lowest RMSE, MAE, and MAD. Columns (7)–(10) present, for square and absolute losses, the average p-values based either on the range or the t_max statistics as described in Hansen et al. (2011). Columns (11) and (12) show the average p-values of the Superior Predictive Ability (SPA) test proposed by Hansen (2005). Finally, columns (13) and (14) display the p-value of the multi-horizon test for superior predictive ability proposed by Quaedvlieg (2017).

	Forecasting Precision							Model Confidence Set					Sup. Pred. Ability		Multi-Horizon SPA	
	(1)	(2)	(3)	(4)	(5)	(6)		(7)	(8)	(9)	(10)		(11)	(12)	(13)	(14)
Model	ave. RMSE	ave MAE	ave. MAD	# min. RMSE	# min. MAE	# min. MAD		ave. p.v. range sq	ave. p.v. range abs	ave. p.v. Tmax sq	ave p.v. Tmax abs		ave. p.v. sq	ave. p.v. abs	p.v. model	p.v. test
RW	1.000	1.000	1.000	2	2	2	0.076	0.079	0.188	0.186	0.149	0.144				
AR	0.872	0.906	0.774	0	0	0	0.150	0.109	0.379	0.318	0.103	0.062				
UCSV	0.881	0.888	0.938	0	0	0	0.086	0.087	0.156	0.202	0.002	0.004				
LASSO	0.906	0.937	0.804	0	0	0	0.002	0.002	0.108	0.120	0.001	0.003				
adaLASSO	0.826	0.835	0.777	0	0	0	0.109	0.077	0.324	0.303	0.086	0.085				
ElNet	0.926	0.959	0.830	0	0	0	0.001	0.001	0.120	0.123	0.000	0.001				
adaElnet	0.838	0.857	0.796	0	0	0	0.066	0.035	0.244	0.237	0.039	0.038				
Ridge	0.871	0.894	0.871	0	0	0	0.046	0.044	0.200	0.234	0.048	0.028				
BVAR	1.252	1.327	0.867	0	0	0	0.000	0.000	0.001	0.001	0.000	0.000				
Bagging	0.768	**0.752**	0.737	**4**	**6**	**5**	0.317	0.325	0.681	0.648	0.457	0.493				
CSR	0.782	0.792	0.697	4	1	4	**0.749**	**0.745**	**0.697**	**0.687**	**0.723**	**0.652**				
JMA	0.888	0.867	0.845	0	0	0	0.091	0.215	0.294	0.359	0.043	0.070				
Factor	0.954	0.971	0.869	0	0	0	0.069	0.102	0.225	0.225	0.010	0.028				
T. Factor	1.019	1.055	0.841	0	0	0	0.044	0.056	0.183	0.148	0.006	0.006				
Boosting	1.075	1.093	1.014	0	0	0	0.031	0.060	0.095	0.138	0.001	0.004				
RF	**0.767**	0.782	0.709	3	2	1	0.521	0.478	0.687	0.617	0.594	0.542				
Mean	0.822	0.842	0.753	0	0	0	0.165	0.142	0.396	0.309	0.210	0.171				
T.Mean	0.829	0.852	0.754	0	0	0	0.082	0.071	0.383	0.271	0.056	0.066				
Median	0.828	0.850	0.751	0	0	0	0.079	0.126	0.363	0.266	0.064	0.111				
RF/OLS	0.777	0.784	0.742	0	0	0	0.301	0.309	0.543	0.479	0.342	0.370				
adaLASSO/RF	0.785	0.775	**0.733**	0	2	1	0.492	0.576	0.594	0.635	0.424	0.517				

Table 29. Forecasting Results CPI-Core: Summary statistics for the out-of-sample period from 2001–2015

The table reports, for each model, a number of different summary statistics across all the forecasting horizons, including as well the accumulated twelve-month horizon. Columns (1), (2) and (3) report the average root mean square error (RMSE), the average mean absolute error (MAE) and the average median absolute deviation (MAD). Columns (4), (5) and (6) report, respectively, the number of times (across horizons) each model achieved the lowest RMSE, MAE, and MAD. Columns (7)–(10) present, for square and absolute losses, the average p-values based either on the range or the t_max statistics as described in Hansen et al. (2011). Columns (11) and (12) show the average p-values of the Superior Predictive Ability (SPA) test proposed by Hansen (2005). Finally, columns (13) and (14) display the p-value of the multi-horizon test for superior predictive ability proposed by Quaedvlieg (2017).

	Forecasting Precision							Model Confidence Set					Sup. Pred. Ability		Multi-Horizon SPA	
	(1)	(2)	(3)	(4)	(5)	(6)		(7)	(8)	(9)	(10)		(11)	(12)	(13)	(14)
Model	ave. RMSE	ave MAE	ave. MAD	# min. RMSE	# min. MAE	# min. MAD		ave. p.v. range sq	ave. p.v. range abs	ave. p.v. Tmax sq	ave p.v. Tmax abs		ave. p.v. sq	ave. p.v. abs	p.v. model	p.v. test
RW	1.000	1.000	1.000	1	1	1	0.100	0.117	0.139	0.117	0.139	0.139				
AR	0.804	0.803	0.712	0	0	1	0.347	0.389	0.615	0.591	0.284	0.291				
UCSV	0.889	0.888	0.889	1	1	1	0.138	0.171	0.267	0.291	0.079	0.102				
LASSO	0.860	0.858	0.761	0	0	0	0.002	0.006	0.100	0.096	0.001	0.000				
adaLASSO	0.822	0.803	0.746	0	0	0	0.199	0.345	0.385	0.487	0.097	0.178				
ElNet	0.870	0.865	0.756	0	0	0	0.000	0.002	0.085	0.064	0.000	0.000				
adaElnet	0.824	0.808	0.738	0	0	0	0.102	0.288	0.328	0.448	0.049	0.145				
Ridge	0.890	0.888	0.831	0	0	0	0.007	0.011	0.104	0.098	0.000	0.001				
BVAR	1.017	1.008	0.828	0	0	0	0.002	0.000	0.032	0.014	0.001	0.000				
Bagging	0.802	0.776	0.747	2	3	1	0.572	0.535	0.702	0.710	0.410	0.418				
CSR	0.814	0.802	0.753	0	0	0	0.504	0.508	0.580	0.626	0.383	0.358				
JMA	0.914	0.875	0.785	0	0	0	0.110	0.139	0.220	0.260	0.019	0.027				
Factor	0.812	0.809	0.761	0	0	0	0.235	0.324	0.483	0.542	0.157	0.200				
T. Factor	0.781	0.774	0.757	0	1	0	0.554	0.534	0.725	0.730	0.435	0.404				
Boosting	1.087	1.105	1.127	0	0	0	0.006	0.003	0.040	0.017	0.000	0.000				
RF	**0.736**	**0.728**	**0.676**	**5**	**6**	**7**	**0.677**	**0.758**	**0.851**	**0.872**	**0.765**	**0.827**				
Mean	0.768	0.761	0.710	2	0	0	0.518	0.569	0.652	0.670	0.539	0.571				
T.Mean	0.778	0.773	0.719	0	0	0	0.227	0.264	0.563	0.544	0.209	0.249				
Median	0.782	0.776	0.719	0	0	0	0.241	0.297	0.511	0.552	0.192	0.266				
RF/OLS	0.775	0.762	0.740	0	0	0	0.653	0.630	0.760	0.759	0.591	0.543				
adaLASSO/RF	0.759	0.744	0.683	2	1	2	0.624	0.713	0.810	0.801	0.498	0.600				

Table 30. Forecasting Errors for the CPI-Core from 1990 to 2015

The table shows the root mean squared error (RMSE) and, between parenthesis, the mean absolute errors (MAE) for all models relative to the Random Walk (RW). The error measures were calculated from 132 rolling windows covering the 1990-2015 period and 180 rolling windows covering the 2001-2015 period. Values in bold show the most accurate model in each horizon. Cells in gray (blue) show the models included in the 50% model confidence set (MCS) using the squared error (absolute error) as loss function. The MCSs were constructed based on the maximum t-statistic. The last column in the table reports in how many horizons the row model was included in the MCS for square (absolute) loss. The last two rows in the table reports how many models were included in the MCS for square and absolute losses.

Consumer Price Index (Core) 1990–2015														
	Forecasting Horizon													
RMSE/(MAE)	1	2	3	4	5	6	7	8	9	10	11	12	Acc.	RMSE count (MAE count)
RW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**1.00**	**1.00**	2
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	**(1.00)**	**(1.00)**	(2)
AR	0.87	0.69	0.59	0.62	0.79	0.93	0.90	0.68	0.47	0.49	0.67	1.21	1.96	4
	(0.92)	(0.70)	(0.59)	(0.62)	(0.80)	(0.91)	(0.85)	(0.63)	(0.45)	(0.49)	(0.70)	(1.27)	(2.09)	(2)
UCSV	0.96	0.80	0.76	0.77	0.86	0.96	0.93	0.84	0.76	0.73	0.82	1.38	1.02	1
	(1.03)	(0.80)	(0.74)	(0.76)	(0.87)	(0.93)	(0.90)	(0.80)	(0.74)	(0.74)	(0.84)	(1.38)	(1.01)	(1)
LASSO	0.85	0.66	0.62	0.69	0.86	0.97	0.90	0.72	0.54	0.58	0.80	1.39	1.90	0
	(0.92)	(0.66)	(0.60)	(0.68)	(0.85)	(0.93)	(0.86)	(0.68)	(0.52)	(0.57)	(0.83)	(1.43)	(2.06)	(0)
adaLASSO	0.84	0.65	0.59	0.66	0.84	0.94	0.89	0.73	0.51	0.52	0.72	1.24	1.64	0
	(0.89)	(0.65)	(0.57)	(0.65)	(0.82)	(0.90)	(0.85)	(0.68)	(0.48)	(0.51)	(0.74)	(1.27)	(1.62)	(1)
ElNet	0.86	0.67	0.61	0.68	0.86	0.98	0.89	0.72	0.55	0.59	0.83	1.46	1.94	0
	(0.92)	(0.67)	(0.60)	(0.67)	(0.85)	(0.94)	(0.85)	(0.68)	(0.53)	(0.58)	(0.85)	(1.49)	(2.12)	(0)
adaElnet	0.85	0.66	0.60	0.68	0.83	0.94	0.90	0.72	0.50	0.52	0.73	1.26	1.66	0
	(0.90)	(0.66)	(0.58)	(0.67)	(0.82)	(0.90)	(0.85)	(0.68)	(0.48)	(0.52)	(0.75)	(1.30)	(1.69)	(0)
Ridge	0.92	0.66	0.61	0.67	0.85	0.95	0.84	0.68	0.59	0.62	0.87	1.56	1.63	0
	(0.99)	(0.67)	(0.60)	(0.66)	(0.85)	(0.93)	(0.81)	(0.65)	(0.58)	(0.63)	(0.92)	(1.62)	(1.69)	(1)
BVAR	0.91	0.71	0.70	0.79	1.01	1.13	1.01	0.83	0.75	0.80	1.10	1.93	2.97	0
	(0.97)	(0.71)	(0.67)	(0.76)	(0.99)	(1.08)	(0.96)	(0.79)	(0.73)	(0.80)	(1.15)	(1.99)	(3.26)	(0)
Bagging	**0.81**	**0.57**	**0.55**	0.67	0.91	0.96	0.81	0.65	0.48	0.50	0.71	1.28	1.33	7
	(0.81)	**(0.56)**	**(0.52)**	(0.65)	(0.87)	(0.88)	(0.77)	(0.62)	(0.46)	(0.49)	(0.73)	(1.28)	(1.31)	(7)
CSR	0.82	0.65	0.57	0.62	**0.78**	0.89	0.85	0.66	0.46	0.48	0.65	1.21	1.77	8
	(0.86)	(0.65)	(0.56)	(0.60)	(0.77)	(0.87)	(0.82)	(0.61)	(0.45)	(0.47)	(0.68)	(1.25)	(1.78)	(7)
JMA	0.91	0.68	0.68	0.75	1.05	1.06	0.96	0.74	0.57	0.55	0.82	1.45	1.50	0
	(0.92)	(0.65)	(0.65)	(0.74)	(0.98)	(0.99)	(0.91)	(0.68)	(0.53)	(0.53)	(0.81)	(1.46)	(1.48)	(2)
Factor	0.90	0.71	0.62	0.68	0.86	0.96	0.90	0.73	0.47	0.52	0.73	1.26	2.07	1
	(0.95)	(0.71)	(0.61)	(0.67)	(0.87)	(0.93)	(0.86)	(0.68)	(0.46)	(0.51)	(0.75)	(1.29)	(2.13)	(1)
T. Factor	0.90	0.72	0.64	0.69	0.86	0.97	0.92	0.75	0.47	0.51	0.71	1.25	2.25	1
	(0.94)	(0.72)	(0.62)	(0.68)	(0.86)	(0.95)	(0.88)	(0.69)	(0.47)	(0.51)	(0.75)	(1.31)	(2.25)	(2)
Boosting	0.94	0.70	0.69	0.79	1.02	1.10	0.98	0.77	0.64	0.72	0.99	1.67	3.03	0
	(1.00)	(0.70)	(0.67)	(0.78)	(1.02)	(1.07)	(0.94)	(0.75)	(0.64)	(0.73)	(1.05)	(1.74)	(3.19)	(0)
RF	0.84	0.61	0.56	0.62	0.78	**0.87**	**0.80**	**0.64**	**0.43**	**0.45**	**0.63**	1.14	1.36	9
	(0.89)	(0.61)	(0.54)	(0.60)	(0.78)	**(0.84)**	**(0.77)**	**(0.61)**	**(0.42)**	(0.45)	**(0.66)**	(1.16)	(1.45)	(10)
Mean	0.81	0.64	0.58	0.64	0.79	0.88	0.81	0.66	0.50	0.52	0.69	1.19	1.59	5
	(0.86)	(0.63)	(0.57)	(0.63)	(0.79)	(0.86)	(0.78)	(0.62)	(0.49)	(0.52)	(0.72)	(1.22)	(1.66)	(4)
T.Mean	0.83	0.64	0.58	0.63	0.79	0.89	0.83	0.67	0.49	0.51	0.70	1.21	1.64	2
	(0.88)	(0.65)	(0.57)	(0.63)	(0.79)	(0.87)	(0.80)	(0.63)	(0.47)	(0.51)	(0.72)	(1.24)	(1.73)	(3)
Median	0.83	0.65	0.58	0.64	0.79	0.89	0.84	0.68	0.49	0.51	0.70	1.22	1.64	2
	(0.88)	(0.65)	(0.57)	(0.63)	(0.79)	(0.87)	(0.80)	(0.64)	(0.47)	(0.51)	(0.72)	(1.25)	(1.73)	(3)
RF/OLS	0.82	0.63	0.57	0.64	0.80	0.89	0.81	0.65	0.46	0.48	0.66	1.19	1.51	7
	(0.88)	(0.63)	(0.55)	(0.62)	(0.80)	(0.86)	(0.77)	(0.61)	(0.44)	(0.48)	(0.69)	(1.22)	(1.48)	(5)
adaLASSO/RF	0.84	0.62	0.56	**0.61**	0.78	0.91	0.89	0.69	0.45	0.46	0.65	1.16	1.40	7
	(0.89)	(0.62)	(0.54)	**(0.60)**	**(0.76)**	(0.86)	(0.82)	(0.64)	(0.43)	**(0.45)**	(0.66)	(1.15)	(1.43)	(9)
RMSE count	3	1	5	4	8	8	4	5	8	2	5	1	2	
MAE count	(10)	(1)	(3)	(3)	(3)	(8)	(4)	(12)	(8)	(2)	(3)	(1)	(2)	

Table 31. Forecasting Errors for the CPI-Core from 1990 to 2000

The table shows the root mean squared error (RMSE) and, between parenthesis, the mean absolute errors (MAE) for all models relative to the Random Walk (RW). The error measures were calculated from 132 rolling windows covering the 1990-2015 period and 180 rolling windows covering the 2001-2015 period. Values in bold show the most accurate model in each horizon. Cells in gray (blue) show the models included in the 50% model confidence set (MCS) using the squared error (absolute error) as loss function. The MCSs were constructed based on the maximum t-statistic. The last column in the table reports in how many horizons the row model was included in the MCS for square (absolute) loss. The last two rows in the table reports how many models were included in the MCS for square and absolute losses.

Consumer Price Index (Core) 1990–2000														
	Forecasting Horizon													
RMSE/(MAE)	1	2	3	4	5	6	7	8	9	10	11	12	Acc.	RMSE count (MAE count)
RW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**1.00**	**1.00**	2
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	**(1.00)**	**(1.00)**	(2)
AR	0.88	0.69	0.59	0.62	0.79	0.96	0.91	0.71	0.50	0.54	0.71	1.26	2.16	5
	(0.92)	(0.70)	(0.58)	(0.62)	(0.82)	(0.96)	(0.90)	(0.65)	(0.48)	(0.53)	(0.76)	(1.43)	(2.38)	(3)
UCSV	0.91	0.68	0.64	0.72	0.90	1.04	0.91	0.75	0.66	0.68	0.90	1.61	1.06	1
	(0.98)	(0.67)	(0.60)	(0.69)	(0.93)	(1.02)	(0.91)	(0.70)	(0.61)	(0.70)	(0.95)	(1.75)	(1.03)	(1)
LASSO	0.85	0.65	0.61	0.67	0.87	1.01	0.91	0.73	0.59	0.63	0.87	1.44	1.91	0
	(0.91)	(0.64)	(0.58)	(0.65)	(0.88)	(0.98)	(0.90)	(0.68)	(0.56)	(0.63)	(0.92)	(1.59)	(2.18)	(1)
adaLASSO	0.85	0.63	0.56	0.63	0.83	0.97	0.90	0.75	0.56	0.56	0.77	1.27	1.39	3
	(0.90)	(0.63)	(0.53)	(0.61)	(0.83)	(0.95)	(0.90)	(0.70)	(0.53)	(0.56)	(0.80)	(1.38)	(1.47)	(3)
ElNet	0.85	0.66	0.61	0.68	0.87	1.02	0.90	0.73	0.59	0.64	0.89	1.55	2.00	0
	(0.91)	(0.65)	(0.58)	(0.66)	(0.88)	(0.99)	(0.90)	(0.67)	(0.56)	(0.65)	(0.94)	(1.70)	(2.30)	(0)
adaElnet	0.85	0.64	0.58	0.65	0.82	0.97	0.90	0.74	0.55	0.57	0.79	1.32	1.47	1
	(0.91)	(0.64)	(0.54)	(0.63)	(0.83)	(0.95)	(0.90)	(0.69)	(0.52)	(0.58)	(0.83)	(1.45)	(1.61)	(1)
Ridge	0.88	0.65	0.60	0.66	0.84	0.99	0.85	0.69	0.59	0.63	0.86	1.51	1.52	1
	(0.95)	(0.65)	(0.57)	(0.63)	(0.85)	(0.97)	(0.85)	(0.65)	(0.58)	(0.65)	(0.92)	(1.62)	(1.66)	(1)
BVAR	0.91	0.72	0.72	0.84	1.09	1.28	1.10	0.90	0.81	0.88	1.19	2.04	3.73	0
	(0.97)	(0.71)	(0.68)	(0.80)	(1.10)	(1.26)	(1.10)	(0.85)	(0.79)	(0.90)	(1.27)	(2.26)	(4.49)	(0)
Bagging	**0.77**	**0.57**	0.55	0.65	0.82	0.93	**0.79**	**0.65**	0.52	0.56	0.74	1.19	1.22	8
	(0.79)	**(0.55)**	**(0.50)**	(0.62)	(0.82)	**(0.87)**	**(0.77)**	**(0.59)**	(0.47)	(0.54)	(0.74)	(1.25)	(1.19)	(9)
CSR	0.81	0.65	**0.55**	**0.59**	**0.76**	**0.90**	0.83	0.66	0.47	0.49	0.66	1.15	1.56	10
	(0.85)	(0.64)	(0.53)	**(0.57)**	(0.77)	(0.89)	(0.82)	(0.61)	(0.45)	(0.48)	(0.67)	(1.22)	(1.68)	(10)
JMA	0.85	0.72	0.62	0.73	1.00	1.11	0.96	0.70	0.58	0.57	0.78	1.44	1.40	2
	(0.86)	(0.68)	(0.58)	(0.71)	(0.95)	(1.04)	(0.92)	(0.63)	(0.54)	(0.56)	(0.77)	(1.54)	(1.41)	(3)
Factor	0.88	0.75	0.65	0.70	0.89	1.04	0.93	0.82	0.52	0.57	0.78	1.36	2.49	1
	(0.91)	(0.74)	(0.61)	(0.68)	(0.91)	(1.03)	(0.92)	(0.77)	(0.50)	(0.57)	(0.82)	(1.48)	(2.63)	(0)
T. Factor	0.92	0.81	0.72	0.74	0.91	1.07	0.99	0.86	0.52	0.58	0.78	1.35	2.91	1
	(0.96)	(0.80)	(0.69)	(0.72)	(0.94)	(1.08)	(0.99)	(0.81)	(0.51)	(0.59)	(0.83)	(1.53)	(3.14)	(0)
Boosting	0.92	0.69	0.68	0.80	1.03	1.15	0.97	0.77	0.63	0.73	0.96	1.55	2.96	0
	(0.98)	(0.69)	(0.66)	(0.77)	(1.05)	(1.13)	(0.96)	(0.72)	(0.62)	(0.74)	(1.03)	(1.70)	(2.93)	(0)
RF	0.83	0.61	0.56	0.62	0.79	0.92	0.82	0.67	**0.46**	**0.48**	**0.64**	1.14	1.40	10
	(0.90)	(0.61)	(0.54)	(0.60)	(0.79)	(0.90)	(0.81)	(0.62)	**(0.43)**	**(0.47)**	(0.67)	(1.22)	(1.57)	(8)
Mean	0.81	0.64	0.58	0.63	0.80	0.93	0.83	0.69	0.53	0.56	0.74	1.24	1.65	4
	(0.86)	(0.63)	(0.55)	(0.61)	(0.82)	(0.92)	(0.83)	(0.64)	(0.51)	(0.56)	(0.77)	(1.35)	(1.81)	(2)
T.Mean	0.82	0.65	0.58	0.63	0.80	0.94	0.84	0.69	0.52	0.55	0.74	1.26	1.68	4
	(0.87)	(0.64)	(0.55)	(0.62)	(0.82)	(0.93)	(0.84)	(0.64)	(0.50)	(0.56)	(0.78)	(1.38)	(1.86)	(2)
Median	0.83	0.65	0.58	0.63	0.80	0.93	0.85	0.69	0.52	0.56	0.74	1.26	1.66	4
	(0.88)	(0.64)	(0.56)	(0.62)	(0.81)	(0.92)	(0.84)	(0.64)	(0.50)	(0.56)	(0.78)	(1.39)	(1.83)	(2)
RF/OLS	0.83	0.63	0.56	0.62	0.79	0.93	0.83	0.68	0.49	0.50	0.68	1.18	1.35	7
	(0.89)	(0.63)	(0.53)	(0.59)	(0.80)	(0.90)	(0.83)	(0.64)	(0.46)	(0.50)	(0.69)	(1.24)	(1.42)	(7)
adaLASSO/RF	0.85	0.64	0.55	0.61	0.77	0.93	0.91	0.72	0.49	0.49	0.66	1.17	1.37	7
	(0.93)	(0.64)	(0.51)	(0.58)	**(0.77)**	(0.89)	(0.87)	(0.65)	(0.43)	(0.47)	**(0.66)**	(1.21)	(1.39)	(8)
RMSE count	9	1	8	6	9	8	3	10	8	3	3	1	2	
MAE count	(3)	(13)	(5)	(7)	(5)	(5)	(1)	(9)	(6)	(3)	(3)	(1)	(2)	

Table 32. Forecasting Errors for the CPI-Core from 2001 to 2015

The table shows the root mean squared error (RMSE) and, between parenthesis, the mean absolute errors (MAE) for all models relative to the Random Walk (RW). The error measures were calculated from 132 rolling windows covering the 1990-2015 period and 180 rolling windows covering the 2001-2015 period. Values in bold show the most accurate model in each horizon. Cells in gray (blue) show the models included in the 50% model confidence set (MCS) using the squared error (absolute error) as loss function. The MCSs were constructed based on the maximum t-statistic. The last column in the table reports in how many horizons the row model was included in the MCS for square (absolute) loss. The last two rows in the table reports how many models were included in the MCS for square and absolute losses.

Consumer Price Index (Core) 1990–2015														
	Forecasting Horizon													
RMSE/(MAE)	1	2	3	4	5	6	7	8	9	10	11	12	Acc.	RMSE count (MAE count)
RW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**1.00**	1.00	2
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	**(1.00)**	(1.00)	(2)
AR	0.86	0.68	0.60	0.62	0.79	0.90	0.89	0.66	0.43	0.45	0.62	1.16	1.80	8
	(0.91)	(0.69)	(0.59)	(0.63)	(0.78)	(0.87)	(0.81)	(0.61)	(0.43)	(0.45)	(0.66)	(1.15)	(1.90)	(8)
UCSV	1.00	0.90	0.85	0.81	0.83	0.89	0.94	0.90	0.84	0.78	0.74	1.09	**0.99**	4
	(1.07)	(0.91)	(0.85)	(0.82)	(0.82)	(0.87)	(0.90)	(0.87)	(0.84)	(0.77)	(0.75)	(1.09)	**(1.00)**	(4)
LASSO	0.86	0.66	0.62	0.71	0.85	0.94	0.89	0.71	0.49	0.53	0.73	1.34	1.85	0
	(0.92)	(0.68)	(0.62)	(0.71)	(0.84)	(0.90)	(0.82)	(0.69)	(0.49)	(0.52)	(0.75)	(1.31)	(1.98)	(0)
adaLASSO	0.84	0.66	0.60	0.69	0.85	0.91	0.88	0.71	0.45	0.48	0.66	1.20	1.74	4
	(0.88)	(0.66)	(0.60)	(0.69)	(0.82)	(0.87)	(0.81)	(0.67)	(0.44)	(0.46)	(0.69)	(1.18)	(1.72)	(7)
ElNet	0.87	0.68	0.62	0.69	0.85	0.95	0.88	0.72	0.52	0.54	0.76	1.37	1.86	0
	(0.94)	(0.69)	(0.62)	(0.69)	(0.83)	(0.91)	(0.82)	(0.68)	(0.51)	(0.52)	(0.78)	(1.32)	(2.00)	(0)
adaElnet	0.84	0.67	0.62	0.70	0.84	0.91	0.89	0.71	0.45	0.48	0.67	1.20	1.74	3
	(0.89)	(0.68)	(0.61)	(0.70)	(0.81)	(0.87)	(0.81)	(0.67)	(0.44)	(0.47)	(0.69)	(1.17)	(1.74)	(6)
Ridge	0.95	0.68	0.62	0.68	0.86	0.92	0.84	0.68	0.58	0.61	0.87	1.61	1.67	0
	(1.01)	(0.68)	(0.62)	(0.68)	(0.85)	(0.90)	(0.78)	(0.66)	(0.58)	(0.61)	(0.92)	(1.61)	(1.70)	(1)
BVAR	0.91	0.71	0.68	0.75	0.93	1.00	0.93	0.76	0.69	0.73	1.02	1.81	2.29	0
	(0.97)	(0.70)	(0.66)	(0.73)	(0.90)	(0.95)	(0.86)	(0.74)	(0.69)	(0.73)	(1.05)	(1.78)	(2.49)	(0)
Bagging	0.84	**0.58**	**0.55**	0.69	0.98	0.99	0.82	0.66	0.45	0.46	0.68	1.36	1.38	10
	(0.83)	**(0.57)**	**(0.54)**	(0.67)	(0.92)	(0.88)	(0.77)	(0.64)	(0.44)	(0.45)	(0.71)	(1.30)	(1.39)	(10)
CSR	0.82	0.65	0.59	0.64	0.80	0.88	0.87	0.66	0.45	0.47	0.65	1.27	1.84	8
	(0.87)	(0.65)	(0.59)	(0.63)	(0.78)	(0.85)	(0.81)	(0.62)	(0.44)	(0.46)	(0.68)	(1.27)	(1.85)	(8)
JMA	0.97	0.65	0.72	0.77	1.09	1.02	0.96	0.77	0.57	0.53	0.86	1.46	1.52	1
	(0.97)	(0.63)	(0.70)	(0.77)	(1.00)	(0.95)	(0.91)	(0.72)	(0.52)	(0.50)	(0.85)	(1.40)	(1.53)	(2)
Factor	0.91	0.68	0.60	0.65	0.84	0.90	0.88	0.65	0.43	0.48	0.68	1.15	1.71	6
	(0.98)	(0.70)	(0.60)	(0.66)	(0.83)	(0.87)	(0.82)	(0.60)	(0.43)	(0.47)	(0.70)	(1.14)	(1.81)	(7)
T. Factor	0.88	0.63	0.57	0.64	0.81	0.89	0.86	0.63	0.42	0.45	0.63	1.13	1.61	10
	(0.92)	(0.65)	(0.57)	(0.64)	(0.80)	(0.85)	(0.80)	**(0.59)**	(0.43)	(0.45)	(0.67)	(1.13)	(1.70)	(10)
Boosting	0.96	0.71	0.69	0.79	1.01	1.07	0.99	0.78	0.65	0.72	1.00	1.78	2.99	0
	(1.01)	(0.71)	(0.68)	(0.79)	(1.00)	(1.03)	(0.93)	(0.78)	(0.66)	(0.73)	(1.07)	(1.78)	(3.35)	(0)
RF	0.86	0.61	0.55	0.62	0.78	**0.83**	**0.78**	**0.62**	**0.41**	0.43	**0.62**	1.13	1.32	11
	(0.88)	(0.61)	(0.54)	**(0.61)**	**(0.76)**	**(0.79)**	**(0.73)**	(0.60)	**(0.41)**	(0.43)	**(0.65)**	(1.10)	(1.38)	(12)
Mean	**0.81**	0.63	0.58	0.64	**0.78**	0.84	0.80	0.64	0.47	0.49	0.65	1.14	1.51	8
	(0.86)	(0.64)	(0.58)	(0.64)	(0.77)	(0.81)	(0.74)	(0.61)	(0.46)	(0.48)	(0.67)	(1.12)	(1.56)	(9)
T.Mean	0.83	0.64	0.58	0.64	0.78	0.86	0.82	0.66	0.46	0.48	0.65	1.16	1.58	7
	(0.88)	(0.65)	(0.58)	(0.64)	(0.77)	(0.83)	(0.76)	(0.63)	(0.45)	(0.47)	(0.67)	(1.13)	(1.64)	(8)
Median	0.83	0.64	0.58	0.64	0.79	0.86	0.83	0.66	0.46	0.48	0.65	1.16	1.60	5
	(0.88)	(0.66)	(0.58)	(0.64)	(0.77)	(0.83)	(0.77)	(0.63)	(0.45)	(0.47)	(0.67)	(1.14)	(1.66)	(8)
RF/OLS	0.82	0.63	0.57	0.65	0.81	0.86	0.79	0.63	0.44	0.46	0.65	1.20	1.57	11
	(0.86)	(0.63)	(0.57)	(0.65)	(0.80)	(0.83)	(0.73)	(0.60)	(0.42)	(0.46)	(0.68)	(1.20)	(1.51)	(10)
adaLASSO/RF	0.83	0.61	0.57	**0.62**	0.80	0.89	0.86	0.66	0.42	**0.43**	0.64	1.16	1.39	11
	(0.86)	(0.61)	(0.56)	(0.61)	(0.76)	(0.84)	(0.79)	(0.63)	(0.42)	**(0.42)**	(0.66)	(1.11)	(1.45)	(11)
RMSE count	9	6	5	11	10	14	4	10	9	7	13	9	2	
MAE count	(9)	(4)	(5)	(11)	(12)	(14)	(8)	(9)	(10)	(14)	(13)	(12)	(2)	