CPB Netherlands Bureau for
Economic Policy Analysis

The added value of machine learning for macroeconomic forecasting in the Netherlands

We create a novel dataset, NL-MD, featuring key macroeconomic indicators for the Netherlands.
Although the dataset is rich in features, it has a relatively small number of observations.
We use this dataset to examine how the machine learning features data richness, non-linearities, cross-validation, and shrinkage improve macroeconomic predictions compared to standard econometric models for the Netherlands.

Our pseudo-out-of-sample forecasting experiment shows that in a limited data environment, careful tuning of models is essential, as the optimal combination of machine learning features can be highly specific to the variable under prediction.

CPB Discussion Paper

Erik van de Winkel, Casper Vedder

July 2024

Doi: https://doi.org/10.34932/2486-RB71

Is Machine Learning Beneficial for Macroeconomic Forecasting with Limited Observations?

Evidence from the Netherlands

Casper Vedder & Erik van de Winkel
CPB Netherlands Bureau of Economic Policy Analysis

July 10, 2024

Abstract

In this paper, we explore how machine learning can improve macroeconomic forecasting for the Netherlands, a small open economy with fewer observations compared to, e.g., the United States. We introduce the NL-MD, a novel dataset featuring key macroeconomic indicators for the Netherlands, similar to existing datasets like the FRED-MD. Following the approach of Goulet Coulombe et al. (2022) on the FRED-MD, we examine how the machine learning features of data richness (i.e., a large number of variables), non-linearities, cross-validation, and shrinkage improve predictions for the Netherlands. Out-of-sample experiments across four horizons and seven variables lead to the following main takeaways: First, the relative performance of machine learning features varies notably depending on the specific variable. Second, although the best model for a variable-horizon combination is often data rich, the average data rich model does not perform significantly better than the average data poor model, likely due to the lack of observations in this experiment. Third, on average, non-linear models do not perform significantly better than traditional econometric models. However, when considering only the data rich setting, non-linear models significantly improve predictive performance, indicating that non-linear models are relatively efficient at handling a large number of features. Our main conclusion is that in a limited data environment, careful tuning of models is essential, as the optimal combination of machine learning features can be highly specific to the variable under prediction.

JEL Classification: C53, C55, E27
Keywords: Machine Learning, Big Data, Forecasting

Note: We would like to express our gratitude to Dick van Dijk, Mark Kattenberg, Bas Scheer and the members of the Data Science program of the CPB for their valuable insights and comments.

1 Introduction

Every month, tremendous amounts of data are released by statistical agencies and private institutions all over the world. Using various econometric models, forecasters can extract information from this data to make forecasts that inform policymakers at governments or central banks. In recent decades, this surge in data has sparked an interest in combining traditional macro-econometric methods with machine learning to distill as much information as possible from the data. However, in contrast with microeconomics[1], the data environment of macroeconomics is relatively scarce, having limited observations per observable. Despite this, macro-forecasters have successfully applied and tailored (machine learning) methods to this challenging setting, equipping the forecaster with a large arsenal of methods (see e.g., Bok et al. (2018), Bolhuis and Rayner (2020)).

There has been extensive research on the topic of macro-forecasting and machine learning (see e.g., Bolhuis and Rayner (2020), Chu and Qureshi (2023), Hall et al. (2018)). Determining the most effective methods for specific datasets often relies on empirical work, leading to a large literature of studying and comparing specific statistical models on various datasets, often in the form of horse race papers (see e.g. Ahmed et al. (2010), Giannone et al. (2021), Granziera and Sekhposyan (2019), Kotchoni et al. (2019), Lee et al. (1993), Medeiros et al. (2021), Stock and Watson (1998, 2012)). However, with the ever-increasing number of statistical models available, the task of comparing each individual model quickly becomes a daunting and overwhelming task. In this light, Goulet Coulombe et al. (2022) abstracts away from individual models and instead focuses on the *features* that lead to better predictive accuracy by performing a meta-analysis on a large space of models evaluated on a large dataset, thereby trying to answer *how* machine learning can improve macroeconomic forecasting. Specifically, Goulet Coulombe et al. (2022) consider an experiment where four features are given as 'treatments' to a reference model. They consider the following features: non-linearities, cross-validation, shrinkage, and the loss function, while controlling for the big data aspect. They find convincing evidence that the predictive power of the best-performing models is well explained by their non-linearity feature.

In this work, we follow the approach by Goulet Coulombe et al. (2022) and apply their methodology to a novel, large dataset containing the most important indicators of the Dutch economy, the NL-MD. Similar datasets already exist for the United Kingdom (Coulombe et al., 2021b), the United States (McCracken and Ng, 2016), and Canada (Fortin-Gagnon et al., 2022), and these datasets have made reproducible macroeconomic research more accessible. The NL-MD consists of data on 168 variables, with data taken from Statistics Netherlands, the European Central Bank, and Yahoo Finance. The NL-MD is not a

[1] For a review of machine learning methods applied in microeconomics, see Athey and Imbens (2019).

balanced dataset, with most data available between 2010 and 2023. Compared to the United Kingdom and the United States, we only have a limited number of observations available for the Netherlands. Additionally, the Dutch economy stands apart from larger economies like the United Kingdom and the United States due to its relatively small and open nature.[2] As a consequence, results obtained from those economies do not necessarily generalize. Academically, exploring the circumstances under which the results do generalize becomes crucial, as it provides further insights into the usefulness of machine learning for macroeconomic forecasting. Moreover, this study holds significant policy relevance, as it provides Dutch forecasters with valuable insights into the most promising methods worth pursuing.

While much of the macroeconomic forecasting literature is predominantly centered around English-speaking countries, particularly the United States, there has been a growing literature around smaller economies and the euro area. For example, Jansen et al. (2016) and Hindrayanto et al. (2016) compare several statistical models on data from several European countries (including the Netherlands) and find that factor models consistently perform well. Kant et al. (2022) forecast Dutch GDP in a horserace containing several machine learning models, finding that a random forest performs best. A similar result was found by Gogas et al. (2022) by forecasting unemployment in the euro area. Looking beyond the euro area, Maehashi and Shintani (2020) applied several machine learning and factor models to Japanese macroeconomic data, finding similar conclusions as Goulet Coulombe et al. (2022). Lastly, Richardson et al. (2018) conducted a similar exercise using data from New Zealand, corroborating the previous results that machine learning forecasts perform consistently better. While these studies provide valuable insights, they are often tailored to specific machine learning models. Consequently, it becomes challenging to discern the precise contributions of different components within the machine learning framework to overall performance. This is a question we try to address in our paper.

To draw conclusions from our dataset, we follow the approach of Goulet Coulombe et al. (2022) closely. We choose seven dependent variables for which we make forecasts: the AEX index, CPI, deposit interest rate, value index of exports, production index of industry, number of sold existing homes, and the unemployment rate. These seven variables together represent key facets of the Dutch economy. In our work, we study four machine learning features: data richness, non-linearities, cross-validation, and shrinkage. For data richness, we compare a data poor environment (independent variables are only lags of the dependent variables) to a data rich environment (independent variables also include other NL-MD variables). For non-linearities, we compare traditional linear (factor) models to kernel ridge regression and a random forest. For cross-validation, we compare a k-fold cross-validation approach, which does not account for the time structure of the data, to a pseudo-out-of-sample cross-validation approach, which

[2] The open nature of the Dutch economy is reflected in the NL-MD by the inclusion of many trade variables. However, due to a lack of data availability, there are not necessarily many foreign variables in the data.

does take this into account. For shrinkage, we compare standard factor models to ElasticNet, Lasso, and Ridge regressions. In total, we consider 42 models that differ in terms of the four features. Using these models and variables, we run pseudo-out-of-sample forecast experiments using four different forecast horizons ($1, 3, 6, 12$ months ahead) on a test period from 2015M01 until 2023M12. As the starting points of all variables in the NL-MD are not the same, the starting points of the seven dependent variables also differ: the starting points of the training period range between 1998M01 and 2010M01. Conclusions are drawn by running regressions of a dummy variable for the feature on the out-of-sample R^2.

Our conclusions are as follows: **1)** The relative performance of machine learning features is quite variable-dependent. Therefore, finding general conclusions that consistently hold true is challenging. **2)** While the best model for a given variable-horizon combination is often a data rich model, the average data rich model does not perform significantly better than the average data poor model; this is likely due to the lack of observations in the dataset. **3)** There is no significant negative effect from using k-fold cross-validation compared to POOS cross-validation. **4)** Averaged over all variables, non-linearities do not perform significantly better compared to linear models. Nevertheless, when only considering data rich models, the treatment effect of non-linear models is significant and positive, indicating that non-linear models can efficiently deal with a large amount of features. **5)** On average, shrinkage methods do not perform significantly better than traditional factor models. However, for several variables, they can improve predictions significantly. Especially when combined with a PCA-rotated dataset. Taking all of this into account, our main conclusion is that in environments with limited data, careful tuning of models is essential, as the optimal combination of machine learning features is quite specific to the variable being predicted.

Our contribution to the literature is twofold. First, we have developed the NL-MD. This macroeconomic dataset consists of the main macroeconomic indicators of the Netherlands and makes research on the Dutch economy more easily available and reproducible. Additionally, the NL-MD complements existing datasets for the United Kingdom (Coulombe et al., 2021b), the United States (McCracken and Ng, 2016), and Canada (Fortin-Gagnon et al., 2022). Although the NL-MD will not be released monthly at this time, the specific data used for this study (vintage February 28, 2024) will be made available. Second, we replicate the work of Goulet Coulombe et al. (2022) for a completely different dataset. Whereas Goulet Coulombe et al. (2022) focuses on the United States, we focus on the Netherlands. Compared to the United States, the Dutch dataset is not balanced and consists of far fewer observations (Goulet Coulombe et al. (2022) have training data going back to 1970, whereas we have training data going back at most to 1998). Our study reveals that the findings from Goulet Coulombe et al. (2022) do not necessarily generalize to different environments. While Goulet Coulombe et al. (2022) conclude that a data rich environment enhances forecasting performance, our results indicate that this may not hold true when the number of training observations is limited. Thus, we refine the conclusions of Goulet Coulombe et al. (2022) by demonstrating that their results vary with different datasets. Third, we extend their work by incorporating an alternative data selection method based on the variance inflation factor (VIF).

This paper is organized as follows: in Section 2 we describe our novel dataset. Then, in Section 3 we introduce our predictive modelling setup. Section 4 introduces the specific features of machine learning and the models we use to produce forecasts. Section 5 discusses the out-of-sample experiment we conduct. The results are then presented in Section 6. Finally, Section 7 concludes.

2 Data

To ensure that our arsenal of machine learning models performs optimally, we require a large dataset for training. For this purpose, we use the NL-MD: a newly constructed dataset. We pre-process this dataset to construct six distinct feature matrices. This section first describes the NL-MD, and the variables we will predict.

2.1 Dataset

In recent years, several standardised monthly datasets with a wide range of macroeconomic indicators have become available. For example, the FRED-MD for the United States (McCracken and Ng, 2016), the CAN-MD for Canada (Fortin-Gagnon et al., 2022), and the UK-MD for the United Kingdom (Coulombe et al., 2021b). As these datasets are publicly available and regularly updated (with a new vintage of data published every month), they make research on the macroeconomy more accessible and reproducible. We follow in this line by constructing the NL-MD: a large monthly dataset describing the most important facets of the Dutch macroeconomy. In contrast to the other datasets, at this point we do not publish and update the NL-MD each month. However, the specific vintage of our dataset will be available on GitHub[3]. In this research, quasi-real-time data with a vintage date of February 28, 2024, is used.

For the composition of variables in the NL-MD, we largely follow the composition of the FRED-MD, CAN-MD, and UK-MD. We supplemented this selection with additional variables based on expert input. The data in the NL-MD originates from three different sources: Statistics Netherlands' database StatLine, the Statistical Data Warehouse of the European Central Bank, and Yahoo Finance.[4] Therefore, the

[3] `https://github.com/CPB-data-science/NLMD`

[4] StatLine: `https://opendata.cbs.nl/statline`; Statistical Data Warehouse: `https://data.ecb.europa.eu/`; Yahoo

composition of the NL-MD also depends on the data availability from these three sources.

In total, the NL-MD consists of 168 variables. The variables in the NL-MD can be divided in seven categories. Table 1 shows for each category the source of the variables and its size. Appendix A gives an overview of all the variables in the NL-MD.

Just like in the FRED-MD, CAN-MD, and UK-MD, we determine a transformation code for each variable in the NL-MD (these codes can be found in Appendix A). The transformation ensures that the dataset used is stationary. To determine the transformation code for a variable, we followed these steps. First, if all elements of a series are positive, we take its logarithm. Second, we perform an Augmented Dickey-Fuller test (Dickey and Fuller, 1979) on the series. If the series is stationary at the 5 percent level, the data is not transformed. Otherwise, the first- and second-differences of the series are calculated, and the Augmented Dickey-Fuller test is applied again. The first series to exhibit stationarity is adopted. Before making forecasts on the data, the data is scaled to have a mean of zero and unit variance.

As can be seen in Figure 1a, the NL-MD is not a balanced dataset. Several variables start later than the first variable. When the latest starting period is used, the dataset contains more than 150 variables, as can be seen in Figure 1b. However, in our analysis we always start at an earlier date, as this provides more observations. In our analysis we prioritize observations over variables, as we already have a limited training period. This trade-off is made clear by considering Figure 1b. As a consequence, we never use more than about 100 variables in our analysis.

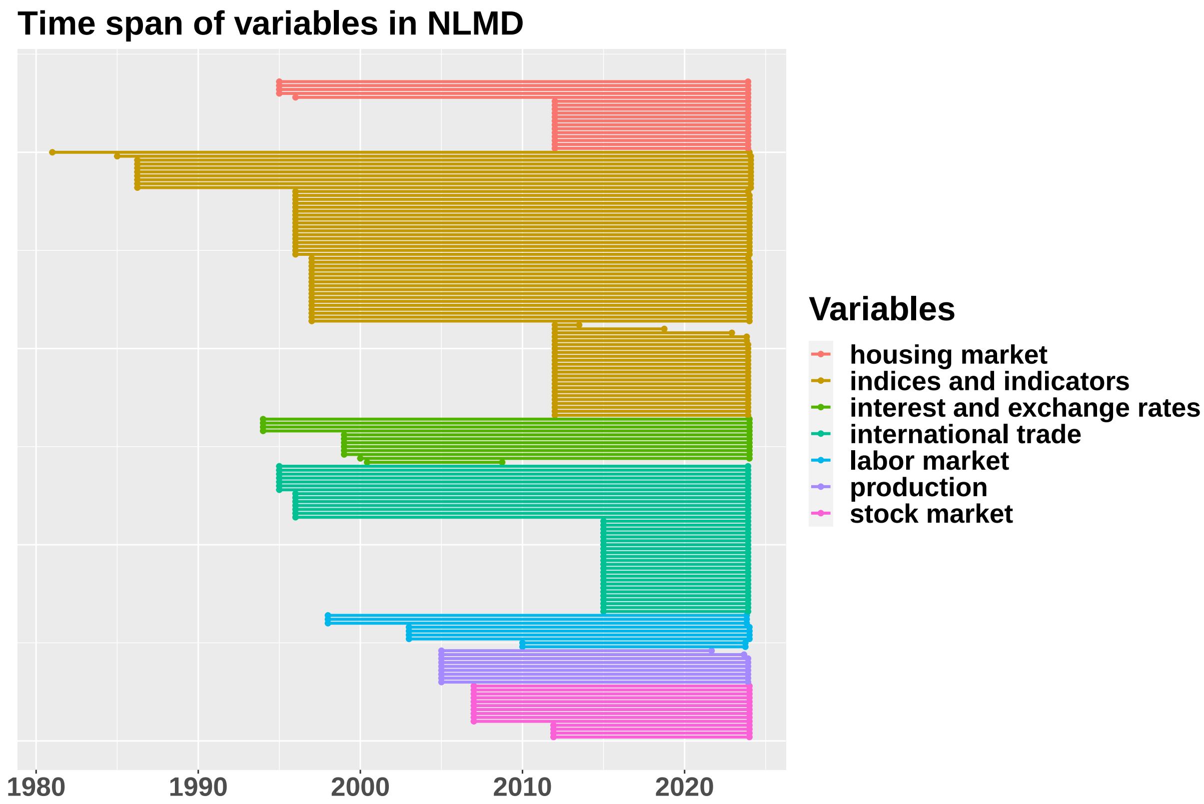
2.2 Variables of interest

We analyze seven key indicators of the Dutch economy, each one corresponding to one of the seven categories spanning the NL-MD. This spans the full breadth of the Dutch economy, thereby giving us a broad overview of the possible applications of machine learning to macro-forecasting. These indicators

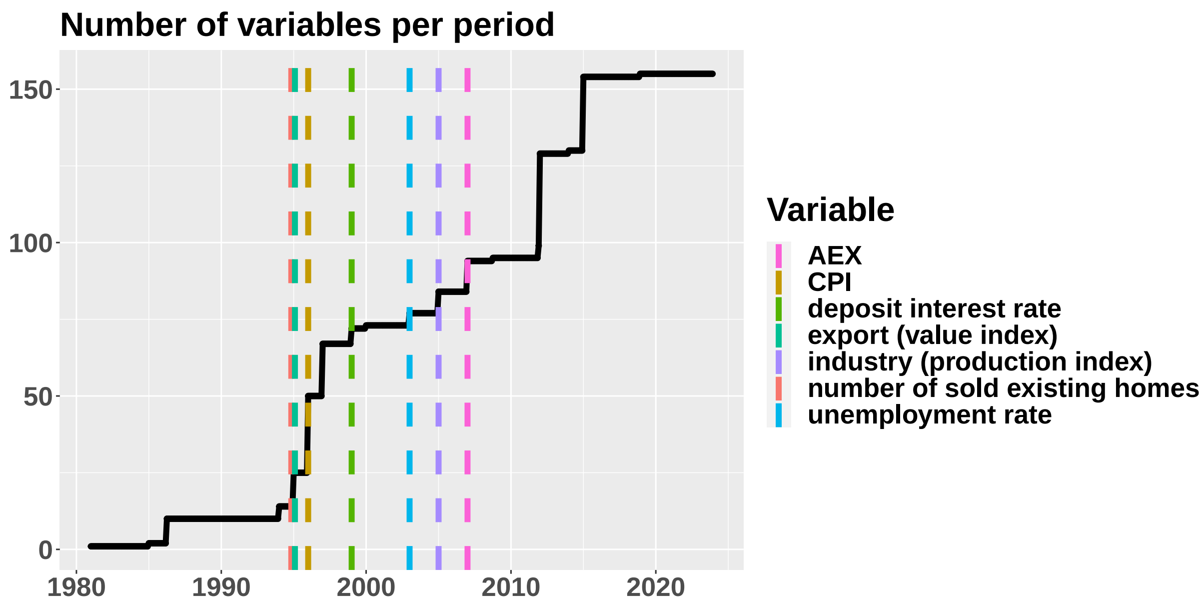
Finance: `https://finance.yahoo.com/`.

Table 1: Short description of the NL-MD dataset. CBS refers to Statistics Netherlands, ECB refers to European Central Bank and Yahoo refers to Yahoo Finance.

Category:	**Data source:**	**Number of variables:**
housing market	CBS	18
indices and indicators	CBS	68
interest and exchange rates	ECB	12
international trade	CBS	38
labor market	CBS	9
production	CBS	9
stock market	Yahoo	14



(a)



(b)

Figure 1: The two figures describe the vintage (2024-02-28) of the NL-MD that is used in this research. In the upper panel, we show for each variable in the NL-MD the time span, i.e., the periods for which we have data on each variable. The different colours correspond to the seven different categories in the NL-MD. In the lower panel, we show for each period how many variables we have data on when we construct a dataset that ends at 2023M12. The vertical lines in the lower panel indicate the points in time from which data for the variables of interest becomes available. The colours used in the lower panel correspond to those used in the upper panel.

exhibit varying behaviours, which could impact forecasting performance. Where possible, we use the same variables in this research on Dutch data as Goulet Coulombe et al. (2022) used in his research on data from the United States. Specifically, we focus on the AEX, CPI, deposit interest rate, export (value index), industry (production index), number of sold existing homes and unemployment rate variables.[5] The series are shown in Figure 2. We note that some of these series have seasonal trends, for example number of sold existing homes. However, we do not explicitly account for this in this study.

The variables under prediction, and the start of the training period corresponding to them, are shown in Table 2. The start of the training period is two years after their minimal period in the data (this point can be seen in Figure 1b). This is needed to be able to include lags up to 12 periods, and predict 12 periods forward. Besides, Table 2 shows for each variable which transformation is used when making predictions.

Table 2: The variables under prediction with their start training period and transformation.

Category:	**Data source:**	**Start training period:**	**Transformation:**
housing market	number of sold existing homes	1998M01	first-difference log
indices and indicators	CPI	1991M01	first-difference log
interest and exchange rates	deposit interest rate	2002M01	first-difference level
international trade	export (value index)	1998M01	first-difference log
labor market	unemployment rate	2006M01	first-difference log
production	industry (production index)	2008M01	first-difference log
stock market	AEX	2010M01	first-difference log

3 Forecasting setup

Machine learning offers a flexible way of predictive modelling. This can be understood from our modelling framework. Following Goulet Coulombe et al. (2022), we predict a (scaled and transformed) target variable y_t over a horizon h by estimating the following equation:

$$y_{t+h} = g_h(Z_t, \tau) + \varepsilon_{t+h}. \tag{1}$$

Here, $g_h(.)$ is a possibly very flexible function of the feature matrix $Z_t = f_Z(H_t)$ and the hyperparameters τ. Its specific parameters will be estimated using either ordinary or regularised least squares. H_t is the original data matrix that is transformed by the function f_Z. The specific feature matrices and how these are obtained from our initial dataset are discussed in Section 3.1. In this work, we will consider predictions over a horizon of 1, 3, 6 and 12 months.

[5] Arguably, the AMX would be more reflective of the Dutch economy than the AEX, as the AEX mostly consists of large multinationals. However, the AEX data is available for a significantly longer period. For the period both variables are available, we find that they are correlated with a correlation coefficient of 0.9551. Therefore, we choose to include the AEX series.

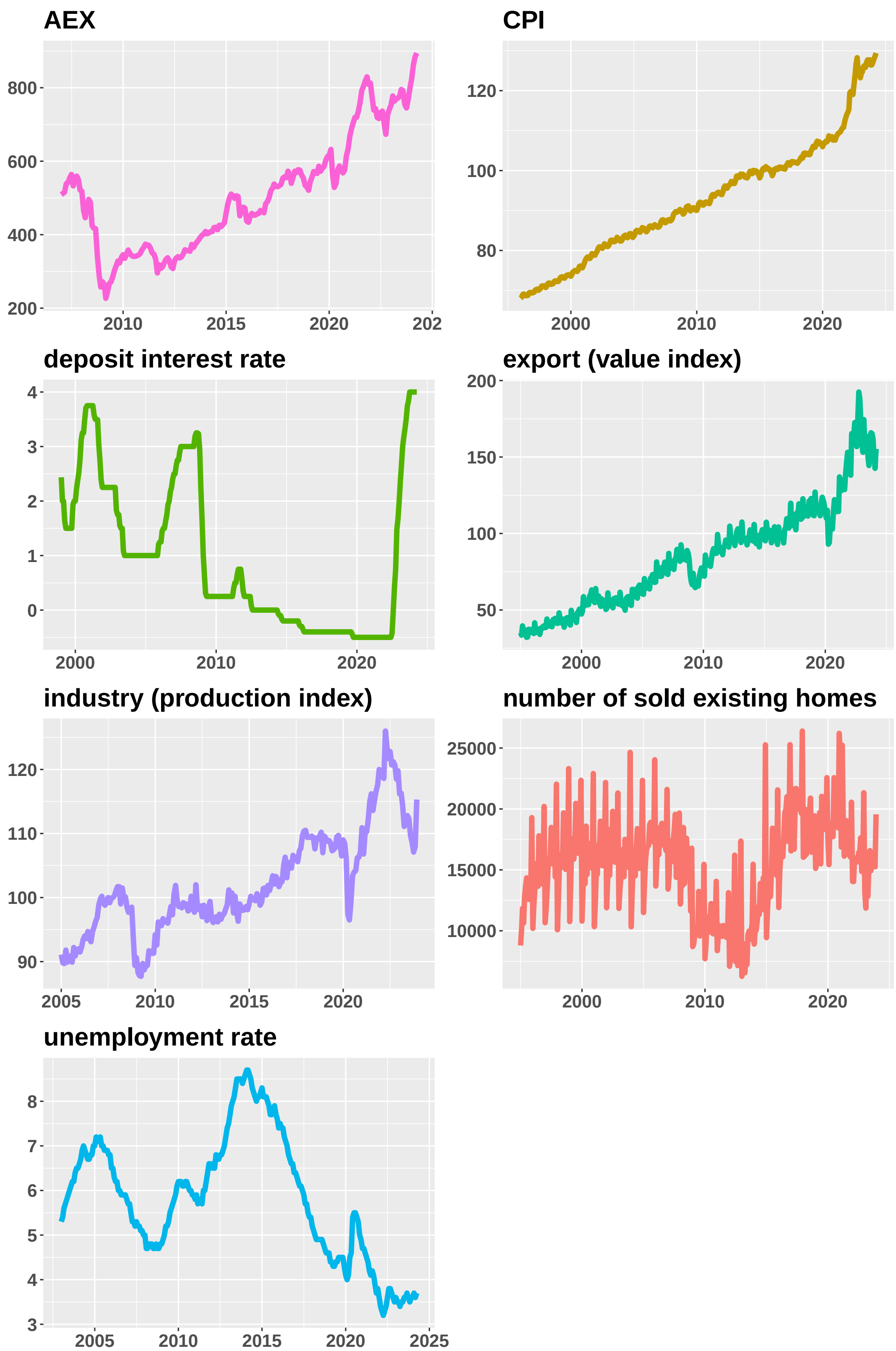


Figure 2: Seven variables that are used as dependent variables in the main analysis.

We define Y_t as the level of the target variable. Analogous to Goulet Coulombe et al. (2022), if Y_t is strictly positive, we take its log. To make sure all series are stationairy, we take its first difference. This means we forecast either the differences or growth rate of a series. We test whether our series are stationary using an Augmented Dicky-Fuller test (Dickey and Fuller, 1979).[6] Specifically, we predict the average growth rate or difference over a horizon h. This leads to two possible specifications for y_{t+h}:

$$y_{t+h} = \begin{cases} \frac{1}{h} \log \frac{Y_{t+h}}{Y_t} & \text{if } Y_t > 0, \\ \frac{1}{h} \left(Y_{t+h} - Y_t \right) & \text{if } Y_t \leq 0. \end{cases} \tag{2}$$

Note that this differs from the path average approach for $h > 1$, where forecasts are made separately for each step up to the horizon. The reason we adopt the direct approach is mainly due to computational reasons. While the two methods are equivalent for OLS-based techniques, Coulombe et al. (2021a) found evidence that the path average approach performs better when regularization (such as ridge regression) and non-linearities are involved. Finally, we scale our target to have mean zero and unit variance. After estimating Equation 1, we convert the forecasted variable in differences, $\hat{y}_{t+h}$, back to the original levels, resulting in a forecast denoted as $\hat{Y}_{t+h}$.

One of the strengths of the machine learning framework lies in the flexibility of the function $g_h(.)$ in Equation 1, allowing for a reduction in the 'true model' approximation error. However, using a flexible function also risks overfitting, which needs to be kept in check by suitable regularization. We can gain some intuition about the forecasting error by decomposing it into three terms:

$$y_{t+h} - \hat{y}_{t+h} = \underbrace{(g_h^*(Z_t) - g_h(Z_t))}_{\text{approximation error}} + \underbrace{(g_h(Z_t) - \hat{g}_h(Z_t))}_{\text{estimation error}} + \underbrace{\varepsilon_{t+h}}_{\text{irreducible error}}, \tag{3}$$

$g_h^*(Z_t)$ as the true model, unknown to the forecaster, and $\hat{g}_h(Z_t)$ as the fitted model. Their differences with g_h, the model selected by the forecaster, define the approximation and estimation errors. Finally, ε_{t+h} is the irreducible error term, representing some fundamental uncertainty.

The specific model g_h is chosen from the set G containing all possible allowed models by solving by the following minimization problem:

$$\min_{g_h \in G} \left\{ (y_{t+h} - g_h(Z_t, \tau))^2 + \text{pen}(\tau, g_h) \right\}. \tag{4}$$

[6]All series are stationary at the 5% level after differencing, according to the Augmented-Dickey-Fuller test. However, the AEX and industry (production index) variables also pass the Augmented-Dickey-Fuller test at the 5% level when in log-levels. However, after further inspection, we decided to focus on the differenced series in our main analysis, aligning with the common assumption in the literature that these variables are I(1). For completeness, the results with these variables in log-levels are provided in Appendix E.1

In the case of machine learning, G can be quite large, to allow for flexible model fitting. The first term in Equation 4 is an error term, for which we use a simple least squares function. The second term is a regularization term penalizing too much complexity to avoid overfitting. The penalty depends on the specific model g_h and a hyperparameter τ, dictating how much model complexity is penalised. Note that hyperparameters cannot be obtained from the minimization problem and need to be obtained from suitable cross-validation.

This also introduces the three aspects of machine learning we are interested in. First, the form of g_h, specifically if it is non-linear. Second, the regularization; this can be through an added penalty such as in Equation 4, by introducing randomness in the model or by restricting the feature matrix Z_t. Lastly, hyperparameter tuning/cross-validation. The hyperparameters can be obtained in several ways, which can significantly alter its predictive performance. Our workflow is visualised in Figure 3.

$$H_t \xrightarrow{\quad f_Z(H_t)\quad} Z_t \xrightarrow{\quad \text{Fitting/CV} \quad} g_h(Z_t, \tau) \xrightarrow{\qquad\qquad} \text{Prediction}$$

Figure 3: The workflow of our models. We start with a large dataset H_t, using this we construct a feature matrix Z_t on which a specific model $g_h(Z_t, \tau)$ is trained, this includes regular fitting and-cross validation. This model is then used to make a prediction.

3.1 Data environment and feature matrices

Machine learning is usually associated with having a larger dataset containing a large number of features. To distinguish the effect of the large dataset on our results and to address whether adding a large number of features is beneficial, we follow Goulet Coulombe et al. (2022) and define two data environments: data poor and data rich. In our data rich environments, we distinguish between several strategies to reduce the amount of variables and the risk of overfitting: 1) We employ two independent feature engineering techniques - using factors, where we rotate the data with principal component analysis[7] (PCA), and the variance inflation factor (VIF) - to construct a feature set with maximal information. Detailed descriptions of these methods are provided in Appendix B. 2) We enhance the loss function to penalise the selection of too many features (shrinkage). This method is commonly regarded as a machine learning technique and is discussed further in Section 4.2.1.[8] Comparing 1) and 2) will give insights in how shrinkage methods can add to the traditional econometric arsenal. Comparing VIF and factors will give insights whether rotating the data increases predictive performance, as this often leads to a decrease of interpretability.

[7]While there are more sophisticated methods available, we use the standard PCA method, as this is still widely used in forecasting.

[8]In principle, we also use a random forest as a predictive model, which in itself already includes regularization by means of bagging and random feature selection. We do not consider this independently.

For our purpose, we construct six feature matrices Z, with rows Z_t, from a set of lags, $\{y_{t-j}\}_{j=0}^{p_y}$, and the set of (potentially lagged and rotated) variables that are included in the NL-MD, which we denote as $\{X_{t-j}\}_{j=0}^{p_x}$. With p_x and p_y the largest lags included.

- *Data poor.* The data poor environment only consists of lags of the target variable, which we can formally state as the set $Z_t = \{y_{t-j}\}_{j=0}^{p_y}$. In our model the amount of lags is always tuned by cross-validation.

- *Data rich, diffusion indices, ARDI.* In this case, we follow the standard factors approach. The factors F_t are selected using standard PCA and then subsequently lagged. This results in the following feature matrix $Z_t = \{y_{t-j}\}_{j=0}^{p_y} \cup \left\{F_{t-j}^n\right\}_{j=0,n=1}^{j=p_x,n=K}$. The superscript n denotes the selected factor, while K indicates the total amount of factors used. This feature matrix has hyperparameters $\{p_y, p_x, K\}$.

- *Data rich, VIF.* We select features from the feature set $\{X_t\}$ using the VIF approach. The selected features are then lagged. This results in the following feature matrix $Z_t = \{y_{t-j}\}_{j=0}^{p_y} \cup \{V_t^n\}_{n=1}^{j=p_x}$, where V^n is the nth selected variable by the VIF method. We select up to 10 variables. This feature matrix has hyperparameters $\{p_y, p_x\}$.

- *Data rich, B1.* Here, we focus on the simplest way of including data and simply combine all the covariates and lags in a simple manner: $Z_t = \{y_{t-j}\}_{j=0}^{p_y} \cup \{X_t\}$. In this case, we study if some shrinkage regularization (Ridge, ElasticNet, Lasso) deals with the problem of overfitting the unstructured data provided.

- *Data rich, B2.* We rotate X_t using PCA and keep all the factors. This set is then combined with the lags of the dependent variable. This yields the following feature matrix: $Z_t = \{y_{t-j}\}_{j=0}^{p_y} \cup \{F_t\}$. Again, we use this feature matrix in combination with shrinkage regularization. Comparing B1 and B2 gives us insight in whether sparsity can emerge in a rotated dataset.

- *Data rich, B3.* We take the full set of lagged variables, $\{y_{t-j}\}_{j=0}^{p_y} \cup \{X_{t-j}\}_{j=0}^{p_x}$, and rotate this using PCA. Similar to B2, we keep all the factors and use shrinkage regularization methods on this dataset.

These six feature matrices then form the basic building blocks of our models. The first three (data poor and data rich, diffusion indices and VIF) will be the standard data environments of our models. Comparing the models trained on these can give us an indication of the effect of additional data. The latter three (B1, B2, B3) will be used in combination with shrinkage regularization to compare traditional factor models with shrinkage techniques.

The specific grid we use when selecting our lags and factors is given in Appendix C. Furthermore, a thorough discussion of several forms of data pre-processing and feature engineering and its effect on the forecasting performance can be found in Coulombe et al. (2021a).

4 Models and features

In this section, we introduce the models and their distinguishing features. First, we present our workhorse models, which serve as the foundation of our analysis. These models are then enhanced with specific machine learning features. To accurately assess the impact of each machine learning feature, we ensure that the machine learning models remain, apart from the machine learning feature, closely aligned with the workhorse models as much as possible, with the only difference being the addition of the respective feature. For each model, we use a feature matrix Z in combination with the specific model. Both the feature matrix and models have hyperparameters to be tuned. Appendix C gives an overview of our grid choices and the packages used in our modelling set-up. Furthermore, a list of all our models and their including features is shown in Table 3. The setup follows Goulet Coulombe et al. (2022) closely.

4.1 Traditional models

As a counterpart, or baseline, for our machine learning models, we use three more traditional 'workhorse' macroeconomic models. For our data poor environment, we use a standard autoregresive model (AR) and for our data rich environment we will apply the auto regressive model with diffusion indices (ARDI) from Stock and Watson (2002b). Additionally we employ a simple linear model on the VIF feature matrix (ARVIF). In principle, all these models are linear regressions on the feature matrices Z. The AR, ARDI and ARVIF models then differ in the feature matrix used. Therefore, all these models can be expressed as follows,

$$y_{t+h} = Z_t'\beta + \varepsilon_{t+h}. \tag{5}$$

Here, Z_t determines the model, if it is the data poor dataset, we run a simple AR model, while the diffusion indices dataset leads to the ARDI model and the VIF dataset leads to our ARVIF model. While our notation is appealing for its simplicity, the ARDI and AR models are usually expressed differently in econometric literature. The standard expressions are shown in Appendix D.

4.2 Three aspects of Machine Learning

To compare the added marginal value of machine learning to traditional econometric methods, we define three aspects of machine learning. These can be added as 'treatments' to our base models. Three aspects we investigate are alternative shrinkage, non-linearities and cross-validation.

For a large overview of machine learning models, see Hastie et al. (2009). For a more complete discussion of macroeconomic forecasting and big data, see Bok et al. (2018).

4.2.1 Shrinkage

As an alternative to the traditional factor approach or data selection using the VIF, one could opt to instead use shrinkage regularization. In this approach, a penalty term is added to the loss function, analogous to the penalty term in Equation 4. Of course, there are many other ways to regularize the amount of variables, as either factors or shrinkage are only two methods in the wide arsenal of tools available. For example, a random forest uses aggregation of bootstrapped samples and random feature selection, which naturally leads to a stable prediction without overfitting (Breiman, 2001, Hastie et al., 2009). However, for our purposes we wish to compare the traditional models with shrinkage techniques, defined as a penalty term added to the loss function that penalizes model complexity. To properly isolate the added value of shrinkage as compared to traditional methods, we cannot use PCA regularization, as we wish to compare the shrinkage methods to this regularization method. Therefore, our shrinkage models use the B1, B2 and B3 feature matrices.

ElasticNet, Lasso, Ridge We will consider three specific cases of ElasticNet regularization (Zou and Hastie, 2005). ElasticNet is a linear regression model that minimizes the following loss function:

$$\min_{\beta}\left\{\sum_{t}\left(y_{t+h}-Z_t'\beta\right)^2+\lambda\sum_{k}\left(\frac{1-\alpha}{2}\beta_k^2+\alpha|\beta_k|\right)\right\}. \tag{6}$$

The model incorporates two hyperparameters: the strength of penalization, λ, and the mixture term, α. Both parameters are tuned in the elastic net case. However, one could alternatively set α to 0, resulting in ridge regression (Hoerl and Kennard, 1970), or set α to 1, leading to lasso regression (Tibshirani, 1996). Lasso regressions are popular for dimension reduction, as they shrink less important variables to zero, unlike ridge regression, where all variables receive similar shrinkage. However, for lasso to be effective, the feature set must exhibit emerging sparsity, meaning the signal should depend on only a few variables. If this is not the case, it is generally more suitable to use ridge. Comparing ElasticNet, ridge and lasso for B1, B2, and B3 provides insight into whether rotating the dataset can induce sparsity.

By comparing these models with the traditional factor model, we can estimate the added value of shrinkage over the conventional factor approach. Similar analyses by Stock and Watson (2012) and Goulet Coulombe et al. (2022) found that, on average, shrinkage methods provide similar or lower forecasting performance compared to factor models.

4.2.2 Non-linearities

In forecasting, linear models are still very popular. This is not surprising given their simplicity and interpretability. However, a specific advantage of machine learning is in the flexibility of their functional form (Hastie et al., 2009). Recent research has shown that the non-linear part of machine learning can improve macro-forecasts (Exterkate et al., 2016, Gogas et al., 2022, Goulet Coulombe et al., 2022, Kant et al., 2022, Maehashi and Shintani, 2020), especially in crisis situations (Coulombe et al., 2021b). Therefore, we opt to include two non-linear models in our setup: a kernel ridge regression (KRR) and a random forest (RF).

Kernel Ridge Regression To introduce non-linearities in a model, one might be tempted to consider Equation 5 and modify it to include multivariate functions of predictors. However, this can quickly lead to overfitting and computational issues as such a model has a large amount of parameters. A potentially smarter way is to use the kernel trick to introduce non-linearities. In our description of the model below, we largely follow Exterkate et al. (2016). First, we introduce a function $\phi(Z_t)$, which transforms the N features into M transformed features. We call the vector of features for a time t $z_t = \tilde{Z}_t = \phi(Z_t)$. We then assume that our regression equation is linear in z:

$$y_{t+h} = z_t'\gamma + \varepsilon_{t+h}. \tag{7}$$

Using ridge regression, as described in Section 4.2.1, we can then obtain the following solution for the coefficients:

$$\hat{\gamma} = \left(\tilde{Z}'\tilde{Z} + \lambda I_M\right)^{-1} \tilde{Z}'y. \tag{8}$$

With $\tilde{Z}$ a $T \times M$ matrix of transformed features and y the vector of observations of the target. To have a flexible non-linear function, it is often needed that $M \gg N$[9], this adds to the problem that in some cases our predictor set N is already larger than our amount of observations T. The matrix $\tilde{Z}'\tilde{Z}$ has dimensions of $M \times M$, which can cause computational issues (see, e.g. Exterkate et al. (2016)). This is where the

[9] A simple example of this would be a Taylor series, to describe a relatively complicated function describing until order d one needs M proportional to N^d.

kernel trick comes in. The kernel trick exploits the fact that for such datasets it can be much more useful to work with T-dimensional data than with M-dimensional data. Specifically, we can rewrite Equation 8 as $\hat{\gamma} = \tilde{Z}' \left(\tilde{Z}\tilde{Z}' + \lambda I_T \right)^{-1} y$, meaning we can write the prediction as $\hat{y}_{t+h} = z_t' \tilde{Z}' \left(\tilde{Z}\tilde{Z}' + \lambda I_T \right)^{-1} y$ (see e.g. Exterkate et al. (2016)). Now, we define the $T \times T$ matrix $K = \tilde{Z}\tilde{Z}'$ and the vector $k_t = \tilde{Z} z_t$, such that we can write our prediction as:

$$\hat{y}_{t+h} = k_t' \left(K + \lambda I_T \right)^{-1} y. \tag{9}$$

This is completely equivalent to using Equation 8, but computationally more tractable. However, to obtain more computational savings it is important that K and k can be computed easily. For this, note that all elements of K and k_t have the following form: $\phi(a)'\phi(b)$, for some vectors a and b. Thus, instead of choosing the basis functions directly, it is sufficient to choose a mapping $\kappa(a, b) = \phi(a)'\phi(b)$, where one can choose a mapping that is computationally tractable.[10]

In our case we pick a radial basis function kernel:

$$\kappa(a, b) = \exp\left(-\frac{||a - b||^2}{2\sigma^2} \right). \tag{10}$$

The kernel function can be understood as a measure of similarity between two covariates. This choice of kernel is quite standard and has been shown to perform well in macroeconomic forecasting (Exterkate et al., 2016, Sermpinis et al., 2014). This model is then trained on the data poor, VIF and the ARDI feature matrices.

Random Forest Tree-based methods, such as decision trees, partition the feature space into a series of rectangles or hyper-rectangles. Each of these partitions, also known as end nodes or leaf nodes, is assigned a predictive value. This value can range from a simple constant to a more sophisticated calculation. Usually, the average within the node is used. This then leads to the following prediction,

$$\hat{y}_{t+h} = \sum_{m=1}^{M} c_m I_m \left(Z \in R_m \right). \tag{11}$$

Here, $I_m \left(Z \in R_m \right)$ is one for the features Z that lie in region R_m, and zero otherwise, c_m is a corresponding constant and M is the amount of end nodes. While a single tree is intuitive and easy to interpret, it is also prone to overfitting and unstable. A potential fix for both of these problems comes in the form of Random Forests, which were introduced by Breiman (2001). To obtain a random forest, a large number of trees is trained on several subsamples of the data. Furthermore, the available features that can be

[10] One can note that picking the identity mapping $\kappa(a, b) = a'b$ results in a normal ridge regression.

considered at each split is also randomized. This model is trained on the data poor and the ARDI and VIF feature matrices.

In principle, both non-linear models could work well with a more unrestricted feature matrix as the extra regularization coming from PCA or VIF is not necessarily needed. However, Coulombe et al. (2021a) gives some evidence that some forms of feature engineering do improve the predictive performance. Our reason for not giving the models access to the unrestricted dataset is more practical: we want to be able to see the specific effect of the non-linear feature. Therefore, we want to keep the KRR and RF models as much aligned with the ARDI and ARVIF models as possible.

4.2.3 Hyperparameter tuning

Choosing an optimal set of hyperparameters is crucial for generating reliable out-of-sample predictions. It is important to identify the right level of model complexity to prevent overfitting on the training data. There are several ways of approaching this. Among others, there are information criteria based approaches such as the AIC and BIC (Granger and Jeon, 2004), similarity-based approaches (Dendramis et al., 2020), and approaches making use of k-fold and Monte Carlo cross-validation (Fonseca-Delgado and Gomez-Gil, 2013). These approaches are in principle all based on estimating and minimizing the out-of-sample forecasting error.

In this case we limit our scope to two cross-validation methods for hyperparameter tuning. Cross-validation has the benefit that it is applicable to all models, and generally gives good results (Hastie et al., 2009). The standard approach in machine learning is using k-fold cross-validation, where the training data is randomly partitioned in k groups, which are iteratively used as training and test data. From a data perspective, this is quite efficient, as for each fold all data are used. However, for time series, the situation is more complicated, as randomly partitioning the data leads to training the model on slices later than the test slices. This is illustrated in Figure 4. Here, the red points are the test slices that the model is evaluated on, after being fitted on the blue slices. This can potentially be problematic and lead to overfitting. Likewise, within our framework, residual non-stationarity and serial correlation frequently exist, aspects that k-fold validation fails to address (Bergmeir and Benítez, 2012).

To avoid these problems, forecasters often adopt pseudo-out-of-sample cross-validation (POOS-CV). This is illustrated in Figure 5. In this way, the temporal structure is taken into account by removing points at the end of the dataset systematically. However, while this scheme is potentially cleaner, it does mean that less data is used for the cross-validation and more so, that the amount of data in the training slices varies significantly. In our already data sparse environment, this could also lead to a reduced performance.

In the end, the trade-off between the two is an empirical one, depending on the data. Both Bergmeir et al. (2018) and Goulet Coulombe et al. (2022) investigated this trade-off on datasets with about 300 observations, and found mild evidence in favour of k-fold. We will investigate if their results also hold on Dutch macroeconomic data.

In our implementation of k-fold we use five folds. In our POOS cross-validation implementation we use an initial window (i.e., the minimum amount of points used as a training slice) of 0.75 times the available training observations.

4.3 Overview of models

In total we have $2*(4+2*4+3*3) = 42$ models, for each variable-horizon combination. All models have a k-fold and a POOS cross-validation version, explaining the overall factor two. The first four corresponds to the data poor models, of which two have non-linear features. All of these models have two (ARDI, VIF) data rich counterparts, relating to to the $2*4$ term. Additionally, in the data rich case we also have three alternative shrinkage models (Lasso, Ridge, ElasticNet), utilizing the three different feature matrices available for them (B1, B2, B3), matching with the last term.

As a benchmark, only used for our RMSE tables, we use a simple AR model with the hyperparameters

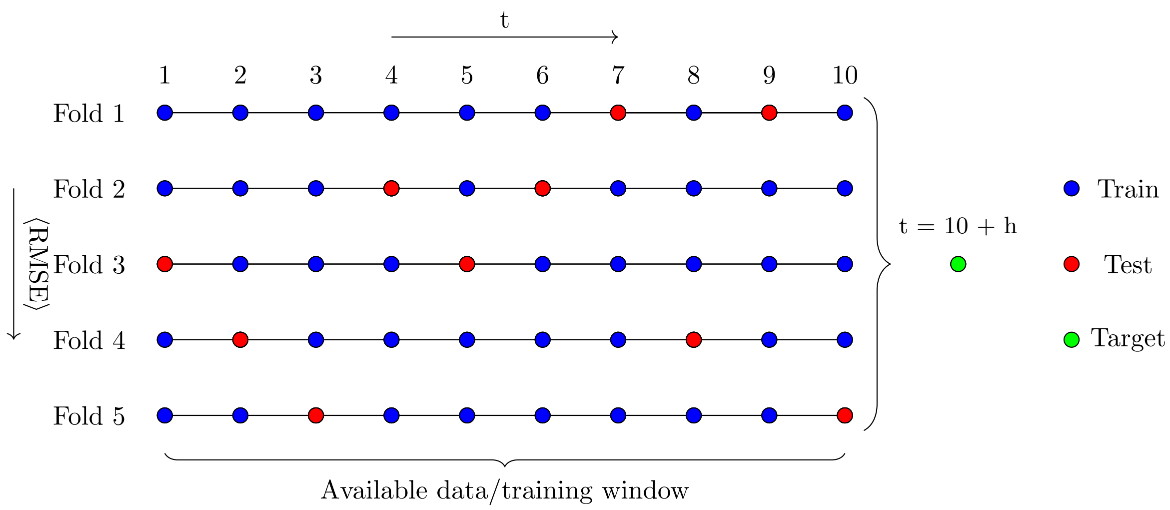


Figure 4: This figure gives an example of k-fold cross-validation, using 5 folds. The available data is randomly partitioned into 5 folds of equal size. For each set of hyperparameters and for each fold the model is fitted to the train data and evaluated on the test data. Then, for each set of hyperparameters, the average RMSE over the folds is calculated. The set with the lowest RMSE is adopted and used to forecasts the target.

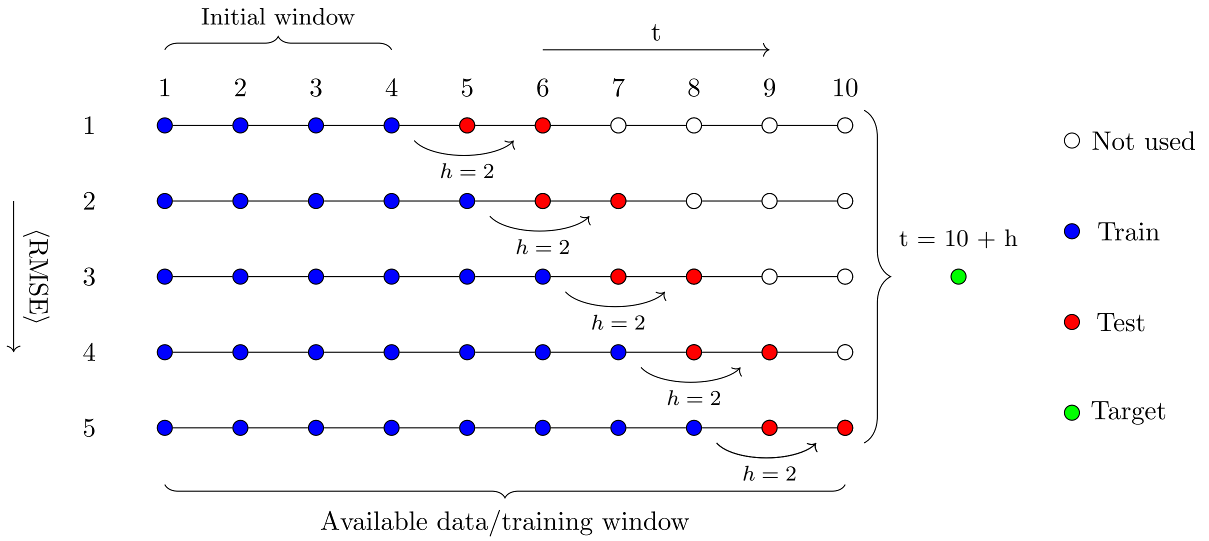


Figure 5: This figure gives an example for pseudo-out-of-sample cross-validation (POOS) for a horizon of two. During the cross-validation, POOS takes the timestructure into account as it only uses points from the past for predicting the test points. The total number of 'folds' is determined by the difference between the total window and the initial window. Each fold involves expanding the initial window by one data point until all available data has been utilized. In each fold, the Root Mean Squared Error (RMSE) is computed. These RMSE values are then averaged, and the hyperparameters yielding the lowest average RMSE are selected.

selected by a Bayesian Information Criterion (BIC). Bringing our total amount of models to 43. These models are written down in Table 3. We evaluate these models for seven variables and four different horizons, bringing the total number of models for our meta analysis on $4 * 7 * 43 = 1204$.

5 Evaluation and out-of-sample experiment

In this section we describe the strategy we use to evaluate which models lead to the best performance. First, we consider the experiment we conduct to evaluate our models. Second, we introduce the identification strategy we use to identify which aspects of the models work well. This approach closely resembles the approach used by (Goulet Coulombe et al., 2022).

5.1 Pseudo-out-of-sample experiment

To investigate the performance of our models we perform a pseudo-out-of-sample experiment. In this experiment we mimic the actual forecasting practice as much as possible. We start with an initial window, on which the model is trained. Using the fitted parameters, we forecast one horizon into the future. Then,

Table 3: For each variable and for each horizon we train 43 models. These models are shown in this table with their corresponding features, as described in Section 4.2. In total there are 16 models containing non-linearities and 18 containing alternative shrinkage. Finally, each model has both a POOS and a k-fold version.

Description in Section **Models**	**Feature 1:** 4.2.2 Non-linearities	**Feature 2:** 4.2.1 Regularization	**Feature 3:** 4.2.3 Cross-validation	**Data-environment:** 3.1 Feature matrix
Data poor				
AR, BIC (BM)			BIC	Poor
AR, k-fold			k-fold	Poor
AR, POOS			POOS	Poor
RR AR, k-fold		Ridge	k-fold	Poor
RR AR, POOS		Ridge	POOS	Poor
RF AR, k-fold	x		k-fold	Poor
RF AR, POOS	x		POOS	Poor
KRR, k-fold	x	Ridge	k-fold	Poor
KRR, POOS	x	Ridge	POOS	Poor
Data rich				
ARDI, k-fold		PCA	k-fold	ARDI
ARDI, POOS		PCA	POOS	ARDI
RR ARDI, k-fold		Ridge/PCA	k-fold	ARDI
RR ARDI, POOS		Ridge/PCA	POOS	ARDI
RF ARDI, k-fold	x	PCA	k-fold	ARDI
RF ARDI, POOS	x	PCA	POOS	ARDI
KRR ARDI, k-fold	x	Ridge/PCA	k-fold	ARDI
KRR ARDI, POOS	x	Ridge/PCA	POOS	ARDI
ARVIF, k-fold		VIF	k-fold	VIF
ARVIF, POOS		VIF	POOS	VIF
RR ARVIF, k-fold		Ridge/VIF	k-fold	VIF
RR ARVIF, POOS		Ridge/VIF	POOS	VIF
RF ARVIF, k-fold	x	VIF	k-fold	VIF
RF ARVIF, POOS	x	VIF	POOS	VIF
KRR ARVIF, k-fold	x	Ridge/VIF	k-fold	VIF
KRR ARVIF, POOS	x	Ridge/VIF	POOS	VIF
Lasso (B1), k-fold		Lasso	k-fold	B1
Lasso (B1), POOS		Lasso	POOS	B1
Lasso (B2), k-fold		Lasso/PCA	k-fold	B2
Lasso (B2), POOS		Lasso/PCA	POOS	B2
Lasso (B3), k-fold		Lasso/PCA	k-fold	B3
Lasso (B3), POOS		Lasso/PCA	POOS	B3
Ridge (B1), k-fold		Ridge	k-fold	B1
Ridge (B1), POOS		Ridge	POOS	B1
Ridge (B2), k-fold		Ridge/PCA	k-fold	B2
Ridge (B2), POOS		Ridge/PCA	POOS	B2
Ridge (B3), k-fold		Ridge/PCA	k-fold	B3
Ridge (B3), POOS		Ridge/PCA	POOS	B3
ElasticNet (B1), k-fold		EN	k-fold	B1
ElasticNet (B1), POOS		EN	POOS	B1
ElasticNet (B2), k-fold		EN/PCA	k-fold	B2
ElasticNet (B2), POOS		EN/PCA	POOS	B2
ElasticNet (B3), k-fold		EN/PCA	k-fold	B3
ElasticNet (B3), POOS		EN/PCA	POOS	B3

we add an extra point to the window and do this exercise again,[11] forecasting the next point on the line. This process is repeated until every point of the out-of-sample period is forecasted. The predictions are then compared to the realisations. This iterative process is shown in Figure 6. Our out-of-sample period ranges from January 2015 until December 2023. Similar to Goulet Coulombe (2020) and Sermpinis et al. (2014) we re-optimize our hyperparameters every two years. This is due to computational reasons.[12] Due to the limited amount of data available, we have decided to include the period containing both the COVID-19 pandemic and energy crisis in our sample. However, we included a robustness check in Appendix E.4.

Ideally, one would use *real-time vintage data*, i.e. data that was available at the time of the vintage as opposed to the data available today (which potentially has been subject to revisions). However, we do not have vintage data available for a satisfactory long period. In our experiment, we use data of the same vintage. We do not expect the results to change significantly because, although revisions introduce some uncertainty, other sources of uncertainty are typically much larger, as shown by Diron (2008) for standard regression models and Elbourne et al. (2015) for structural models.

[11] This procedure is known as the 'expanding window' approach, in contrast to a 'fixed window' approach. In the latter case the total size of the window is kept fixed. This can take less time to calculate, however, it generally leads to worse predictions.

[12] We also note that some experimenting with optimizing every year made no notable difference.

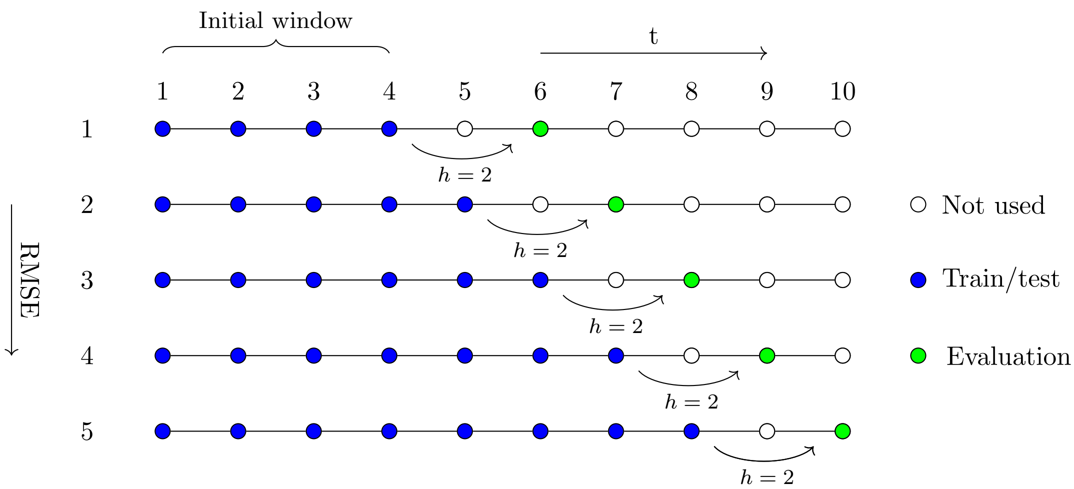


Figure 6: A schematic overview of our pseudo-out-of-sample experiment. In this case, the first four points are the intitial window, while the last six points are the out of sample period. Iteratively, a forecast is made, compared to the realisation and then the realisation is added to the expanding window to be used in the next forecast.

5.2 Evaluating the effects

To test whether the machine learning features have a significant impact on the forecasting results we first define the same R-squared as Goulet Coulombe et al. (2022) to get comparable coefficients between variables:

$$R^2_{t,h,v,m} = 1 - \frac{e^2_{t,h,v,m}}{\frac{1}{T}\sum_{i \in OOS}\left(y_{i,v} - \bar{y}_v\right)^2}, \tag{12}$$

where t, h, v, m denote the time, horizon, variable and model under consideration and $\bar{y}_v = \frac{1}{T}\sum_{i \in OOS} y_{i,v}$ the average over the out-of-sample period OOS . Using this R^2, as opposed the squared error $e^2_{t,h,v,m} = \left(\hat{y}_{t,h,v,m} - y_{t+h,v}\right)^2$, has the benefit that it yields coefficients that can be compared between variables. Note that the R^2 ranges from $-\infty$ to 1, where 1 means a perfect prediction.

For each feature of machine learning we run an independent regression on a specific subset of models $\mathcal{M}_f$. For data rich, we compare the models trained on the ARDI and ARVIF matrices to the models trained on the AR matrix. For non-linearities, $\mathcal{M}$ contains the linear AR, ARDI and ARVIF models, as wel as the KRR and RF models. For regularization, it contains the ARDI model and B-models. Finally, for cross-validation the whole model set is used. We run the following fixed-effects regression:

$$\forall m \in \mathcal{M}_f : \quad R^2_{t,h,v,m} = \alpha_f + \phi_{t,v,h} + u_{m,t,v,h}, \tag{13}$$

with ϕ a fixed effects term, taking into account fixed effects for t, v and h: demeaning the prediction target. u is the error term and α_f is a dummy for the feature under study f. This regression is then used to study the significance and size of the effect as compared to the null hypothesis that the feature has no effect on the R^2.

To isolate more specific partial effects, this term can be interacted with other terms, for example, with horizon- or variable-specific dummies.

Our pseudo-out-of-sample period consists of 108 observations, multiplying this with our 1148 models[13] gives us a total dataset of 127,008 observations. We also distinguish between the backtransformed and transformed data. The backtransformed data gives the predicted variable in levels, while the transformed data gives the variable transformed according to Equation 2. For each of these observations we calculate an R^2, according to Equation 12.

[13]The number is different from the previously mentioned number as the AR BIC is not used in this part.

6 Results

In this section we describe the results of our experiment. First, we give an overview of our panel-dataset. Then, we give a brief overview of the general performance of our models. Finally, we describe the specific effects of certain features of machine learning.

6.1 R^2 distribution

We consider the distribution of the R^2's in both the transformed and backtransformed datasets. Hereby, we define an outlier as $R^2 \notin (-1, 1)$. Table 4 shows summary statistics of the R^2's, and Figure 7 shows the distribution of the R^2's in both dataset, whereby we make a distinction between outliers and non-outliers. More detailed information about the number of outliers in each specific model can be found in Table 10 in Appendix F. It is noteworthy that, unsurprisingly, a significant proportion of outliers occurred during the COVID period in our sample.

A clear observation is that the backtransformed predictions have a higher average R^2 and less outliers. This is not a surprise, as in our definition of the R^2 the predictive error is compared to the average difference with the average. For the transformed variables, the series is more likely to be stationairy, and therefore the average should be a better prediction than for the variables in levels. This is reflected in a lower R^2.

For our main analysis, we use the transformed predictions to calculate the R^2's and winsorize the R^2's at the 5th and 95th percentiles. This is to make our analysis more robust to outliers. A forecaster would be hesitant to accept significant outliers in the predictions without further study or expert opinion. We choose to analyze the R^2-data of the errors of the transformed predictions because, in this case, the error is compared to the average of a stationary series. This provides a more meaningful statistic than comparing it to the average of a series in levels containing a unit root. Therefore, we highlight the results of the transformed predictions. However, we included a robustness check for data transformation

Table 4: Summary statistics of the R^2, as defined in Equation 12, over our full dataset.

	Transformed dataset	**Backtransformed dataset**
Statistic		
mean	-0.08933	0.6654
minimum	-173.9564	-31.6680
5th percentile	-3.9116	-0.5296
median	0.7810	0.9586
95th percentile	0.9987	0.9999
maximum	1.0000	1.0000
standard deviation	2.7393	1.0773
number of observations	127,008	127,008

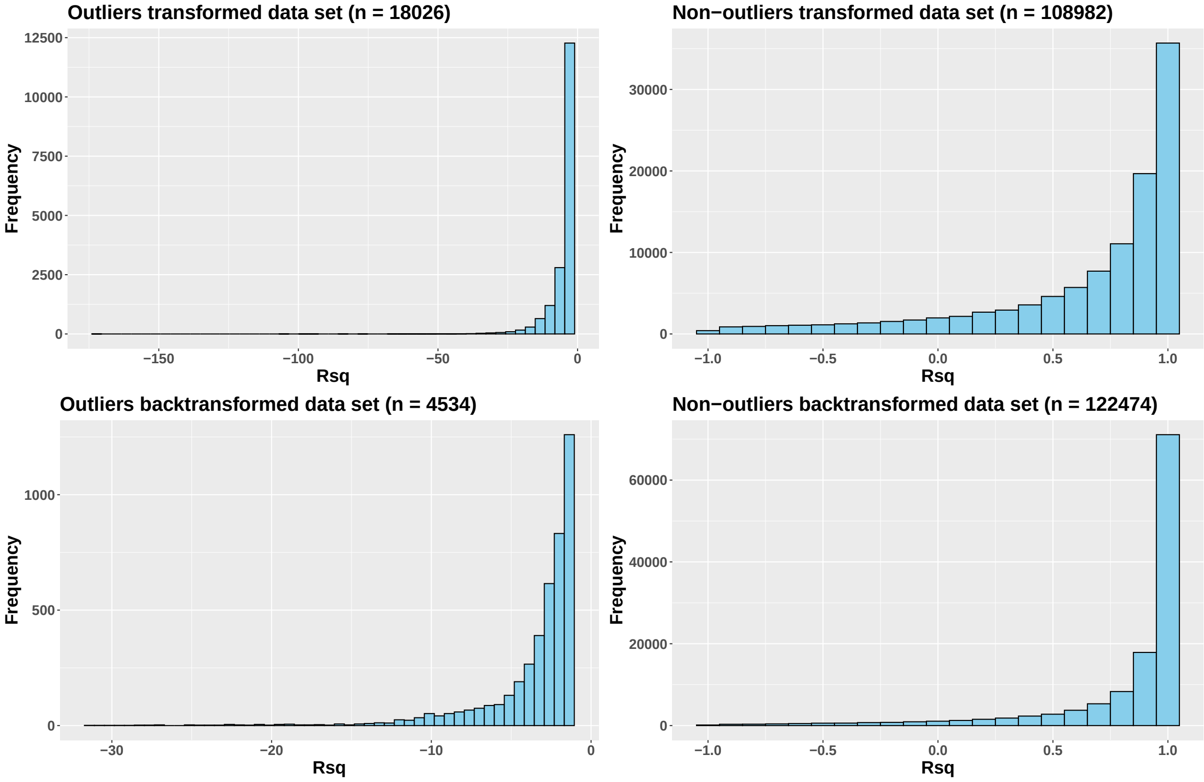


Figure 7: For each variable, horizon and model, we determine for each time in the pseudo-out-of-sample period the R^2 according to Equation 12. In this figure the distribution of R^2 in both the transformed (i.e. predictions in transformed format) and backtransformed (i.e. predictions in level format) dataset is shown. We make a distinction between outliers and non-outliers: a value of R^2 below -1 is considered to be an outlier, a value of R^2 between -1 and 1 is considered to be a non-outlier.

(Appendix E.2) and winsorizing (Appendix E.3).

6.2 General performance

We consider the general performance of all models. Appendix G summarises the predictive performance of each individual model. For each variable-horizon combination the best performing model is shown in Table 5. The result is a mixed bag of (mostly) data rich models. In 22/28 cases the best model is trained on a data rich feature matrix. From the 22 data rich models, seven are trained on a B_i matrix, 10 on the ARDI matrix and five on the ARVIF matrix.[14] Although it has to be kept in mind here that we also have significantly more data rich models in our arsenal. Then, out of the 28 best models, nine have non-linearities (three times KRR, six times RF). From the 19 linear models, 10 are standard linear models: five ARDI models, one ARVIF and four AR models. These are supplemented with several shrinkage models: seven ridge models, one Lasso and one ElasticNet. 22 out of 28 models use k-fold instead of POOS cross-validation, giving a clear sign that k-fold can work very well. However, while these models

[14] We note that in our regressions for the data rich treatment effect, we do not take the B_i matrices into account, as these are not directly comparable to a data poor model.

Table 5: The best performing model for each variable-horizon combination, according to their OOS-RMSE. The dots indicate several features the models can have. Red indicates data rich, where as opposed to our regression study, we also include the B_i feature matrix. Green indicates shrinkage and blue non-linearity's. The models with with an orange dot use k-fold cross validation, those without POOS.

Variable	h = 1	h = 3	h = 6	h = 12
AEX	AR 🟠	Ridge (B2) 🔴🟢🟠	Ridge (B2) 🔴🟢	AR 🟠
CPI	KRR ARDI 🔴🔵🟠	Ridge (B3) 🔴🟢🟠	AR 🟠	ARDI 🔴🟠
Deposit interest rate	RR ARDI 🔴🟢🟠	ARDI 🔴🟠	RR ARVIF 🔴🟢🟠	RF ARVIF 🔴🔵🟠
Export (value index)	ARVIF 🔴🟠	RF AR 🔵	KRR ARVIF 🔴🔵🟠	Ridge (B2) 🔴🟢🟠
Industry (production index)	ARDI 🔴🟠	KRR AR 🔵	RF ARDI 🔴🔵	RF ARDI 🔴🔵🟠
Number of sold existing homes	AR 🟠	RF ARVIF 🔴🔵🟠	RF ARDI 🔴🔵	Ridge (B3) 🔴🟢🟠
Unemployment rate	ARDI 🔴🟠	ARDI 🔴🟠	Lasso (B2) 🔴🟢	ElasticNet (B2) 🔴🟢🟠

might perform well for this specific variable-horizon combination, the general performance of the model can vary quite significantly. It is therefore not valid to conclude that these models are the overall best models.

6.3 Overall overview

In Table 6 the results of regression 13 on transformed predictions are shown. It shows the aggregate treatment effects over all variables and horizons. We find no significant effects. Compared to the findings of Goulet Coulombe et al. (2022), we note several differences in the sign of the treatment effect of the data rich and shrinkage features, although these are not significant, as we have to keep in mind that we have less statistical power due to a significantly shorter out-of-sample period.

To analyse the discrepancy for these features with Goulet Coulombe et al. (2022), we have done a similar study on the FRED-MD. This is shown in Appendix H. Here, we compare the treatment effects for the full FRED-MD to the treatment effect on a restricted version of the FRED-MD, that mimics our NL-MD dataset in terms of observations. We find a positive data rich treatment effect in the unrestricted case, and a null treatment effect in the restricted case. For the shrinkage treatment effect, we see that the

Table 6: This table shows the regression results corresponding to the regression in Equation 13. For the four different features under study (feature matrix, non-linearities, shrinkage, and cross-validation) we show the value of α_f including its 95% confidence interval. We are controlling for variable, time and horizon. For each variable the training period starts on a different date (see Table 2). For all variables the test period is from 2015M01 to 2023M12. These results are based on the transformed dataset with applied winsorizing. These coefficients, the treatment effects, can be interpreted as the change in out-of-sample R^2 due to adding the feature.

Feature:	**Overall treatment effect:**
Feature matrix (data poor (0) vs. data rich (1))	-0.0346 $(-0.0829, 0.0138)$
Non-linearities (linearities (0) vs. non-linearities (1))	0.00470 $(-0.0412, 0.0506)$
Shrinkage (no shrinkage (0) vs. shrinkage (1))	0.0420 $(-0.0388, 0.123)$
Cross-validation (POOS (0) vs. k-fold (1))	-0.0111 $(-0.0470, 0.0247)$

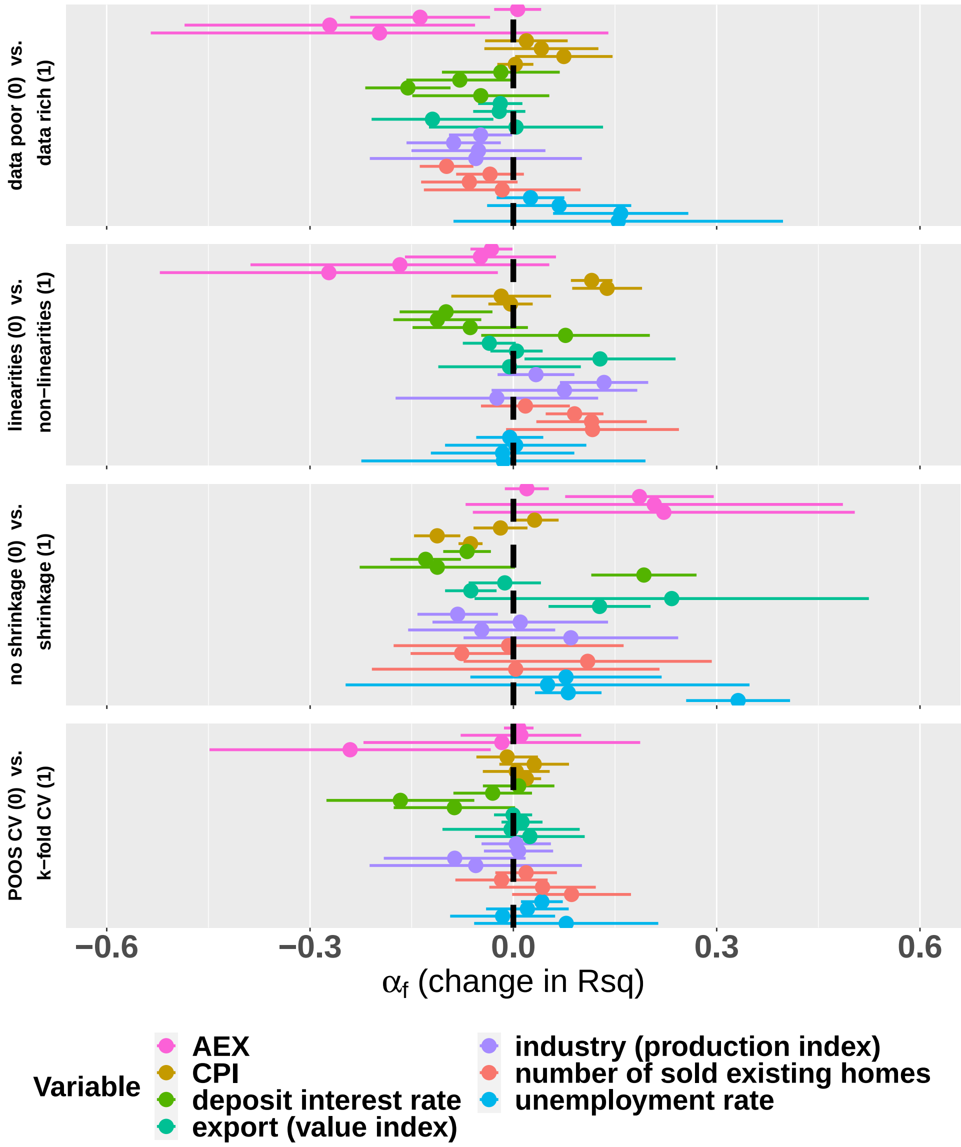
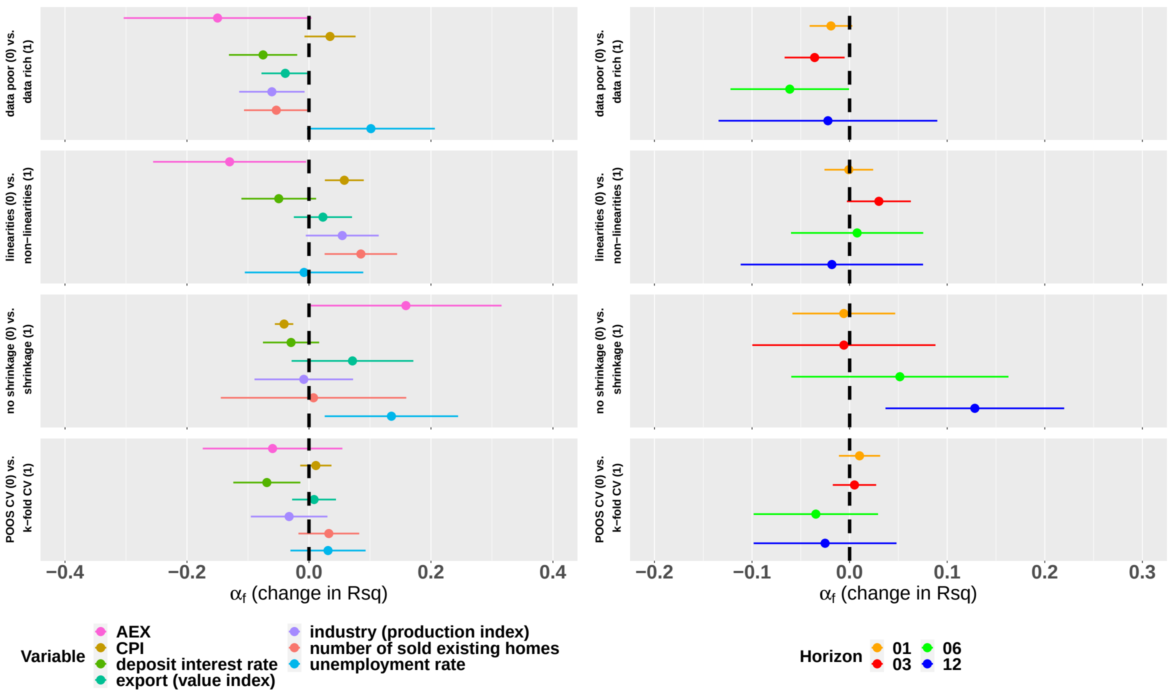


Figure 8: This figure shows the regression results corresponding to the regression in Equation 13. For the four different features under study (feature matrix, non-linearities, cross-validation, and shrinkage) we show the value of α_f including its 95% confidence interval. We are only controlling for time. For each variable, the training period starts on a different date (see Table 2). For all variables the test period is from 2015M01 to 2023M12. These results are based on the transformed dataset with winsorizing applied. The different colours correspond to the seven dependent variables that are being studied. For a specific colour, the four different horizons are shown as independent lines, where the horizon is increasing going down the figure (i.e. from $h = 1$, to $h = 3$, to $h = 6$, to $h = 12$).



(a) Controlling for time and horizon.

(b) Controlling for time.

Figure 9: The two figures show the regression results corresponding to the regression in Equation 13. For the four different features under study (feature matrix, non-linearities, cross-validation, and shrinkage) we show the value of α_f including its 95% confidence interval. In the left panel we are controlling for time and horizon (the different colours correspond to the seven dependent variables that are being studied). In the right panel we are controlling for time and variable (the different colours correspond to the four different horizons that are being studied). For each variable the training period starts on a different date (see Table 2). For all variables the test period is from 2015M01 to 2023M12. These results are based on the transformed dataset with applied winsorizing.

treatment effect on the full FRED-MD is negative, while for the restricted FRED-MD the effect is null. Although the error bars are large in the restricted case, this gives some evidence to our premise that the treatment effects are linked to the number of observations. Furthermore, we note that Goulet Coulombe et al. (2022) finds a more convincing treatment effect for non-linearities. This seems to be mainly driven by a difference in relative performance of the data poor non-linear models and is further discussed in Section 4.2.2.

In Figure 8, we show for each variable-horizon combination which machine learning feature contributes to its performance. In Figure 9, we show the aggregated treatment effects per variable (Figure 9a) and per horizon (Figure 9b). These figures show that the average increase in R^2 given the ML feature is relatively stable across horizons, but varies over variables. We see that for shrinkage the treatment effect increases with horizon, where for a forecast horizon of 12 the effect becomes statistically significant.[15] For the variables, it becomes clear from Figure 8 that there is a lot of heterogeneity between treatment effects, indicating that finding the best model is often a variable specific question.

Our main conclusion is thus that many effects seem comparable, with some notable differences between the variables. Thus, determining a general approach for identifying the most opportune features of machine learning proves to be challenging. Our results, with a limited dataset, for a small and open economy, indicate that in this setting any approach should be tailored to the variable under consideration. A similar conclusion was made by Coulombe et al. (2021b) for the UK and can be drawn from the results for Canada in Goulet Coulombe et al. (2022). We give further evidence that, when applied to a limited dataset in terms of observations, shrinkage models perform relatively well, whereas data rich models perform comparatively worse. In the following section, we analyze each feature independently to identify specific strengths and weaknesses.

6.4 A closer look at the features

In this section we take a closer look at the models and data-features that drive the treatment effects of the ML features.

6.4.1 Data richness

We consider the effect of the data rich feature by contrasting the models trained on the data rich ARDI/VIF environments with the models trained on the data poor environment, taking fixed effects into account. In this section we split the results into different subgroups and consider the difference with Goulet Coulombe et al. (2022).

The discrepancy in the data rich treatment effect compared to Goulet Coulombe et al. (2022) is likely due to the smaller number of observations in our analysis. To check this, we have trained our models on the FRED-MD as well (Appendix H). Comparing the treatment effect on the full FRED-MD, to a restricted FRED-MD, in terms of observations. The FRED-MD has balanced data going back to 1960, while our dataset only becomes comparable in the number of variables to the FRED-MD around the 2000-2010's (see Figure 1). When restricting the training period artificially for the FRED-MD data, the positive effect observed for data rich for the unrestricted dataset disappears. Based on this evidence, we

[15] In the case of non-linearities, the treatment effect is mainly stable because the non-linear treatment effect on data poor models decreases as a function of horizon, while it increases for data rich models.

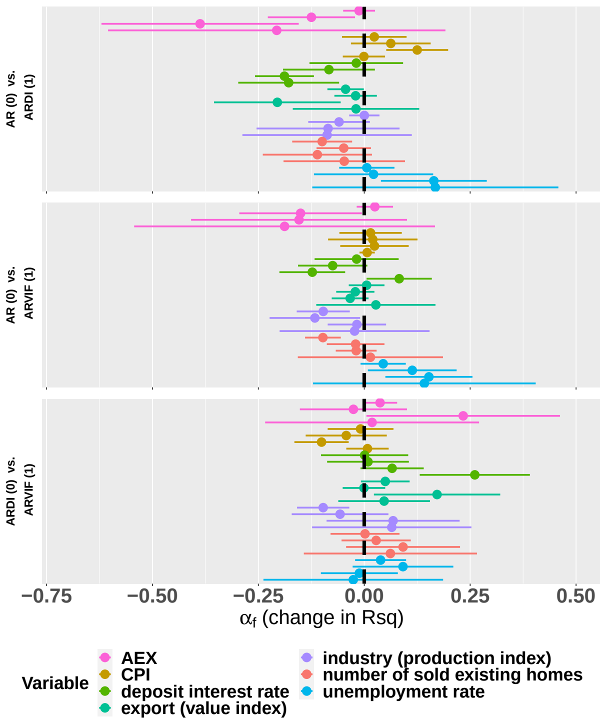


Figure 10: This figure shows the regression results corresponding to the regression in Equation 13, interacted with a specific feature matrix dummy. We show the value of α_f including its 95% confidence interval. We are only controlling for time. For each variable, the training period starts on a different date (see Table 2). For all variables the test period is from 2015M01 to 2023M12. These results are based on the transformed dataset with winsorizing applied. The different colours correspond to the seven dependent variables that are being studied. For a specific colour, the four different horizons are shown as independent lines, where the horizon is increasing going down the figure (i.e. from $h = 1$, to $h = 3$, to $h = 6$, to $h = 12$).

conclude that the null effect for the data rich treatment in our analysis is likely driven by the limited number of observations. This argument is strengthened by the results of Goulet Coulombe et al. (2022) on the CAN-MD dataset, where the effect of the data rich environment is less convincing. The analyses in this study also has a limited amount of observations as compared to the FRED-MD.

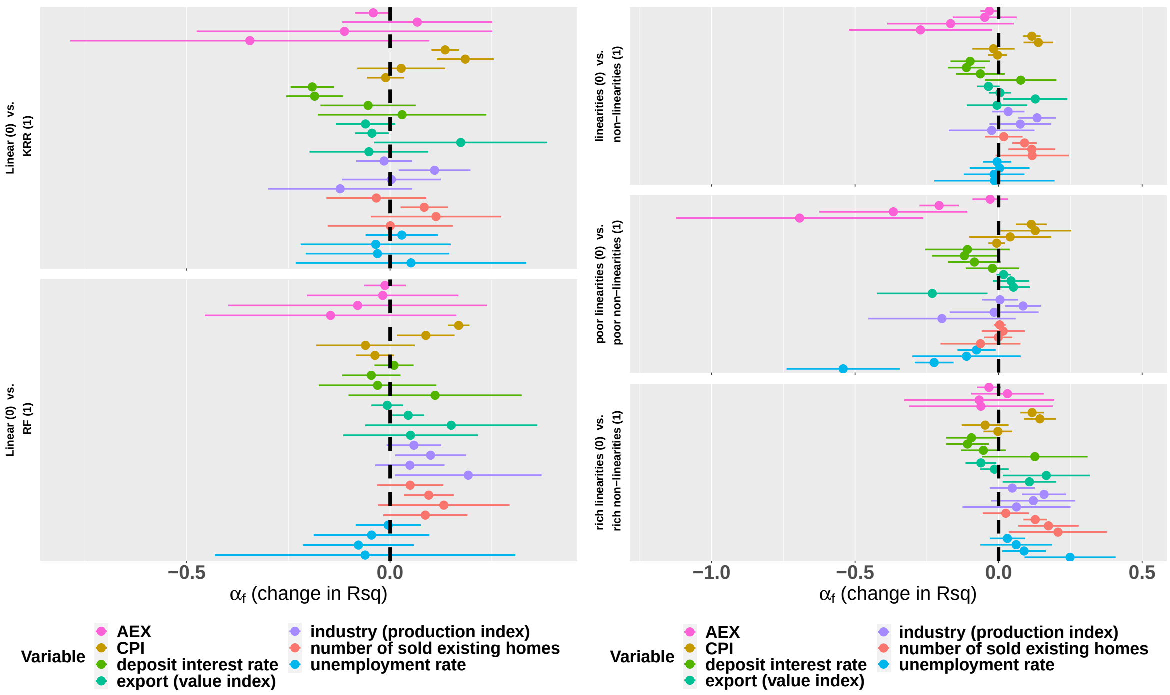
To see whether the VIF or the ARDI environment drives the result, we compare the models trained on the ARDI and ARVIF matrices to the models trained on the AR matrix. This is shown in Figure 10. Here, we see no large differences between the feature matrices. However, the AR and ARVIF models seem to perform slightly better than the ARDI model. This suggests that, with our dataset, using the factors regularization method is sub-optimal, as it does not beat the relatively straightforward VIF selection, or even a simple feature matrix built on lags. Again, this is a result of our limited dataset. To check our premise that the factor environment works sub-optimally due to our lack of observations, we did a similar analysis on the FRED-MD (McCracken and Ng, 2016), specifically comparing a longer training period to a restricted shorter one. This analysis is shown in Appendix H. In the unrestricted case, we note that, apart from the positive treatment effect for data rich, factors also perform better than the VIF data selection method. Thereby giving evidence for the fact that for factors to work optimally, a longer dataset than the NL-MD is needed. This is in agreement with several other sources, which often have data with more observations available and find positive effects for factor models (Coulombe et al., 2021a, Hindrayanto et al., 2016, Jansen et al., 2016).

6.4.2 Non-linearities

For the non-linearity feature, we contrast the non-linear models with the linear AR, ARDI and ARVIF models. The treatment coefficient can thus be interpreted as the average increase of predictive power by adding non-linearities.

Overall, the treatment effect of non-linearities is null (Table 6). In our arsenal we have two non-linear models: the random forest and the kernel ridge regression. In Figure 11, we compare the two non-linear models on different feature matrices to see which one drives the results. In the left panel we show the KRR and RF separately. In the right panel, we show the treatment effects for non-linearities per data environment.

The KRR and RF perform relatively comparable, with the RF having a slight edge (Figure 11a). Considering non-linear treatment effects per variable, most individual variables show a (mildly) positive effect, except for the AEX and deposit interest rate variables. When only considering non-linear models



(a) (b)

Figure 11: The two figures show the regression results corresponding to the regression in Equation 13. For two models containing non-linearities we show the value of α_f including its 95% confidence interval. In the left panel we compare the RF and KRR models in general. In the right panel we compare non-linearities for data rich and data poor environments specifically. We are only controlling for time. For each variable, the training period starts on a different date (see Table 2). For all variables the test period is from 2015M01 to 2023M12. These results are based on the transformed dataset with winsorizing applied. The different colours correspond to the seven dependent variables that are being studied. For a specific colour, the four different horizons are shown as independent lines, where the horizon is increasing going down the figure (i.e. from $h = 1$, to $h = 3$, to $h = 6$, to $h = 12$).

that are data rich (Figure 11b), the treatment effect notably increases. As in this case, the effect is positive and significant when aggregated over all variables, indicating that non-linear methods can deal with a large amount of features in a relatively efficient manner, leading to a better predictive performance. However, in the data poor case, the effect is negative and significant. This pattern also holds for the KRR and RF independently, both perform relatively better in a data rich environment.

While there is some recent literature finding positive effects of non-linearities for forecasting (Gogas et al., 2022, Goulet Coulombe et al., 2022, Kant et al., 2022, Maehashi and Shintani, 2020, Medeiros et al., 2021), there is also some evidence against it (Marcellino, 2008, Stock and Watson, 1998). Yet, the papers finding negative results are mainly focused on neural networks or interacted regressions. In general, forests perform relatively well in the literature. Therefore, our results are slightly surprising, as the RF does not outperform the other models consistently in our case. However, our results are not directly at odds with these papers. Most of the aforementioned papers are horserace papers, having a different set-up. Additionally, we do find a significant and positive treatment effect for non-linear models that are data rich.

6.4.3 Cross-validation

The treatment effects of k-fold cross-validation are mostly indecisive. From the seven variables we consider, only deposit interest rate has a significant treatment effect, in favour of POOS cross-validation. Therefore, it is usually acceptable to use k-fold cross-validation over POOS cross-validation. This is in line with the conclusions of Goulet Coulombe et al. (2022) and Bergmeir et al. (2018). This has practical benefits. POOS cross-validation can be time consuming, due to its computational time scaling with the amount of observations. In principle, one could reduce the number of training slices in POOS cross-validation to make the computational time comparable to k-fold cross-validation. However, we have not investigated the impact of this approach on the model's predictive performance, and it could potentially lead to a decrease in performance. Using k-fold has another practical benefit. Many off-the-shelf ML models use k-fold cross-validation as their standard method, making it easy to apply these models to an macro-forecasting environment without the need to explicitly tailor its cross-validation setup.

6.4.4 Shrinkage

To estimate the performance of shrinkage relative to the standard factor model, we contrast the shrinkage models with the ARDI model. Thus, the treatment effect is the average increase in R^2, when going from

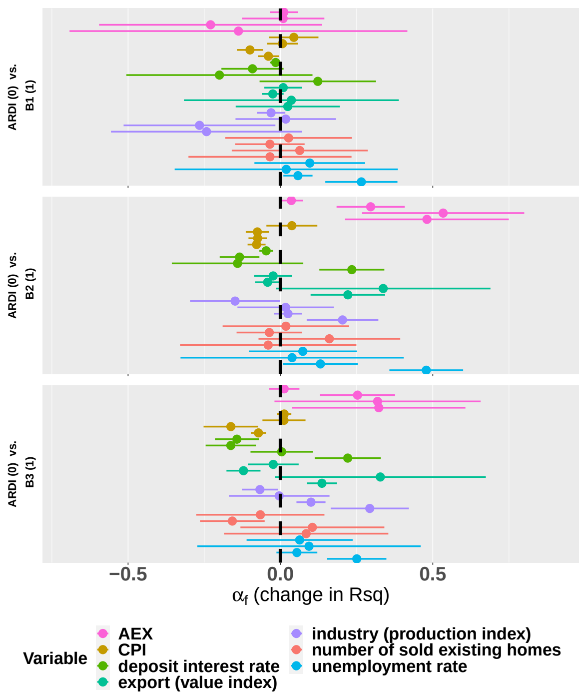


Figure 12: This figure shows the regression results corresponding to the regression in Equation 13. For three features related to shrinkage we show the value of α_f including its 95% confidence interval. We are only controlling for time. For each variable, the training period starts on a different date (see Table 2). For all variables the test period is from 2015M01 to 2023M12. These results are based on the transformed dataset with winsorizing applied. The different colours correspond to the seven dependent variables that are being studied. For a specific colour, the four different horizons are shown as independent lines, where the horizon is increasing going down the figure (i.e. from $h = 1$, to $h = 3$, to $h = 6$, to $h = 12$).

the ARDI to the shrinkage model.

From the overall results, i.e. Table 6, we know that on average the ARDI models do not outperform the shrinkage models significantly. The aggregate effect differs from the results on the FRED-MD, but it is more in line with the results on the CAN-MD, a dataset with limited observations as well (Goulet Coulombe et al., 2022). To test whether, relative to other models, the predictions of shrinkage models improve as a function of observations we turn to our FRED-MD analysis (Appendix H). This analysis provides additional evidence supporting our premise, indicating that shrinkage models perform relatively well with limited observations compared to factor models.

To compare the predictive performance of the three feature matrices, we have illustrated the partial effects of the B1, B2, and B3 feature matrices. This is shown in Figure 12. Overall, the B2 and B3 models perform comparably and often better than the benchmark, whereas the B1 models underperform relative to B2 and B3 models. This indicates that rotating the dataset is helpful in terms of predictive performance, suggesting that there is some emerging sparsity due to the factor approach that improves predictions. However, this is mainly so for the AEX, unemployment rate and export variables.

Table 7: This table shows the regression results corresponding to the regression in Equation 13 for two different training period settings. For the four different features under study (feature matrix, non-linearities, shrinkage, and cross-validation) we show the value of α_f including its 95% confidence interval. We are controlling for variable, time and horizon. We show the results for the maximum training period setting (for each variable the training period starts on a different date, see Table 2) and the short training period setting (for each variable the training period starts on 2010M01). For all variables the test period is from 2015M01 to 2023M12. These results are based on the transformed dataset with applied winsorizing. These coefficients, the treatment effects, can be interpreted as the change in out-of-sample R^2 due to adding the feature.

Feature:	**Maximum training:**	**Short training:**
Feature matrix (data poor (0) vs. data rich (1))	-0.0346 $(-0.0829, 0.0138)$	-0.0495 $(-0.0977, -0.00120)$
Non-linearities (linearities (0) vs. non-linearities (1))	0.00470 $(-0.0412, 0.0506)$	0.0356 $(-0.0193, 0.0904)$
Shrinkage (no shrinkage (0) vs. shrinkage (1))	0.0420 $(-0.0388, 0.123)$	0.0410 $(-0.0374, 0.119)$
Cross-validation (POOS (0) vs. k-fold (1))	-0.0111 $(-0.0470, 0.0247)$	-0.00306 $(-0.0447, 0.0385)$

6.5 Different train period

Arguably, a more fair comparison between the variables is the case where all variables start at the same time. Therefore, we have also trained our models on a sample where all variables have access to the same feature matrix, with variables starting at January 2010. This then also gives information about the effect of extra training data, as compared to the original sample.

We show the results with a shorter training period in Table 7. The table gives an indication of the effects of limiting the sample. The values of the maximum and shorter training period are relatively comparable, and the overall pattern is similar. We highlight several differences between the short and long training period.

The average data rich treatment effect increases slightly when using the longer training period. This is to be expected, as the bottle neck in our case is the amount of observations. There is some movement in the shrinkage case, but the overall treatment effect stays roughly the same. Furthermore, the average non-linear treatment effect increases slightly for the shorter training period. Possibly, this trend could be attributed to the greater availability of independent variables at a later starting date (see Figure 1b). In Appendix I we show the best models for each horizon for the short training period. Here, we see that a larger fraction (13/28) is now non-linear as compared to Table 5, which we attribute to our premise that non-linear models can flexibly handle the increased availability of features. Finally, even though there are less observations available, most of the best models are still data rich (22/28).

7 Conclusion and discussion

In this paper we have applied the framework of Goulet Coulombe et al. (2022) to a novel Dutch dataset, the NL-MD, that contains a large amount of predictors describing the Dutch macroeconomy. However, while this dataset is large in predictors, it has a relatively small number of observations. We considered how several 'features' of machine learning affect the predictive performance for forecasting seven variables over four horizons in this challenging environment. We consider four features: data richness, non-linearities, cross-validation, and shrinkage. We obtained the following results.

The relative performance of machine learning features is quite variable dependent. With a dataset with limited observations, such as the NL-MD, the specific approach should therefore be tailored to the variable under prediction (Figure 8).

While the best model for a given variable-horizon combination is often a data rich model (Table 5), the average data rich model does not perform significantly better than the average data poor model (Table 6). Within the data rich environment, we compared two data compression mechanisms: a standard factor approach (Stock and Watson, 2002a), and a simple variable selection mechanism based on the variance inflation factor (VIF). In general, the simple VIF and AR models perform at least as well as the factors approach (Figure 10). While these results might appear surprising, they can be explained by the small number of observations. As a robustness check, we have also done our analysis on the FRED-MD, comparing the treatment effect on the full FRED-MD, to a restricted FRED-MD, in terms of observations. When restricting the training period artificially for the FRED-MD data, the positive effect observed for data rich using the unrestricted dataset disappears. In the unrestricted training period case the factors perform better than the VIF method as well (Appendix H), but for the restricted case there is a null effect. Therefore, the difference in data rich treatment effect is likely due to the limited number of observations in our dataset.

Averaged over all variables, non-linearities have no significant treatment effect relative to their linear counterparts (Table 6). Nevertheless, for some variables the treatment effect is positive (Figures 9a). Furthermore, when only considering data rich models, the treatment effect is positive and significant (Figure 11b). This shows that non-linear models can handle a large amount of features relatively well. While our findings do not show a definitive negative impact of non-linearities, they also do not provide an clear positive outcome. This is mainly driven by the lacking relative performance of the data poor non-linear models. This result is somewhat surprising, as the literature suggests that our non-linear models, KRR and RF, generally perform well (Exterkate et al., 2016, Goulet Coulombe et al., 2022, Kant et al., 2022, Medeiros et al., 2021).

There are almost no significant negative effects from using k-fold cross-validation (Figure 8). Therefore, the choice of method depends on which is most suitable for practical implementation. In most cases, including ours, the computational time of POOS cross-validation scales with the length of the data, whereas k-fold cross-validation, with its random sampling, provides an efficient way to keep computational time limited. Further research could explore comparing the methods when using the same number of folds, as in this case the computational time is similar. Additionally, it could investigate weighting the training slices with more observations more heavily in POOS cross-validation.

On average, shrinkage methods perform comparable to the traditional models. However, for several specific variables the effects are significantly positive (Figure 8), specifically when using with a dataset rotated using PCA (Figure 12). Compared to standard econometric models, shrinkage methods seem to perform relatively well on a limited dataset, as the treatment effect increases when we restrict the amount of observations (Appendix H).

Taking all of this into account, our main conclusion is that in environments with limited data, careful tuning of models is essential, as the optimal combination of machine learning features is quite specific to the variable being predicted.

Our results are derived from regressions run on out-of-sample R^2's for several models. As we have seen, numerous R^2 values can be deemed outliers. These outliers are the effect of outliers in the forecasts. To address this concern, literature suggests trimming. Trimming is a method in which a threshold is set for forecast values, such that forecasts do not exceed certain bounds. This approach might have mitigated the effects of outliers on our findings. Although in this research we winsorized our data to address outliers, future research could consider trimming to adjust the effect of outliers before running regressions. Another extension could be in improving our factor model. We apply a relatively simple factor approach, using standard PCA. However, possible extensions include Kalman filtering techniques (Bok et al., 2018), scaled PCA (Huang et al., 2022), and supervised PCA (Bair et al., 2006).

Furthermore, for the non-linear models, we only considered one tree-based model and one kernel-based model, but the general space of non-linear models is much larger. For instance, examining the performance of support vector machines, boosted trees, or neural networks would be of significant interest. Our findings indicate that in a data-sparse environment, specific tailoring of models is necessary. Off-the-shelf machine learning models generally do not account for the macroeconomic context of the data. Therefore, it might be beneficial to develop specialized macro machine learning models that are tuned to effectively navigate this environment. Existing examples of such tailored models include the macro-forest by Goulet Coulombe (2020) and the hemispherical neural net by Goulet Coulombe (2022).

Our results may have been influenced by the limited out-of-sample (POOS) test period. Additionally, the inclusion of the COVID-19 period within this test period could have impacted our findings. We reflect on this in Appendix E.4. Excluding the COVID-19 period from the analysis was not feasible, as it would have resulted in too few observations for proper regressions. On a larger dataset, it would be interesting to examine the effects of the business cycle (Goulet Coulombe et al., 2022, Scheer, 2022) or crisis situations (Coulombe et al., 2021b).

Lastly, in this research, we did not include a setting in which experts select the independent variables used in the analysis (this would have resulted in a different feature matrix setting). Experts often possess a deep understanding of the data and underlying patterns. In a data scarce environment, machine learning models might benefit from some assistance. Therefore, it could be advantageous to pre-select the most informative variables. Another approach would be to use alternative variable selection methods. Since the VIF method is based on linear models, it might not be optimal in combination with tree-based methods. In this case, there could be greater synergy with the tree-based Boruta method (Kursa et al., 2010). It would be interesting to explore whether this approach would enhance the performance of our non-linear models.

References

Nesreen K Ahmed, Amir F Atiya, Neamat El Gayar, and Hisham El-Shishiny. An empirical comparison of machine learning models for time series forecasting. Econometric Reviews, 29(5-6):594–621, 2010.

Susan Athey and Guido W Imbens. Machine learning methods that economists should know about. Annual Review of Economics, 11:685–725, 2019.

Eric Bair, Trevor Hastie, Debashis Paul, and Robert Tibshirani. Prediction by supervised principal components. Journal of the American Statistical Association, 101(473):119–137, 2006.

Christoph Bergmeir and José M Benítez. On the use of cross-validation for time series predictor evaluation. Information Sciences, 191:192–213, 2012.

Christoph Bergmeir, Rob J Hyndman, and Bonsoo Koo. A note on the validity of cross-validation for evaluating autoregressive time series prediction. Computational Statistics & Data Analysis, 120:70–83, 2018.

Brandyn Bok, Daniele Caratelli, Domenico Giannone, Argia M Sbordone, and Andrea Tambalotti. Macroeconomic nowcasting and forecasting with big data. Annual Review of Economics, 10:615–643, 2018.

Marijn A Bolhuis and Brett Rayner. Deus ex machina? A framework for macro forecasting with machine learning. International Monetary Fund, 2020.

Leo Breiman. Random forests. Machine Learning, 45:5–32, 2001.

Jiehong Cheng, Jun Sun, Kunshan Yao, Min Xu, and Yan Cao. A variable selection method based on mutual information and variance inflation factor. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 268:120652, 2022.

Ba Chu and Shafiullah Qureshi. Comparing out-of-sample performance of machine learning methods to forecast US GDP growth. Computational Economics, 62(4):1567–1609, 2023.

Philippe Goulet Coulombe, Maxime Leroux, Dalibor Stevanovic, and Stéphane Surprenant. Macroeconomic data transformations matter. International Journal of Forecasting, 37(4):1338–1354, 2021a.

Philippe Goulet Coulombe, Massimiliano Marcellino, and Dalibor Stevanović. Can machine learning catch the COVID-19 recession? National Institute Economic Review, 256:71–109, 2021b.

Yiannis Dendramis, George Kapetanios, and Massimiliano Marcellino. A similarity-based approach for macroeconomic forecasting. Journal of the Royal Statistical Society Series A: Statistics in Society, 183 (3):801–827, 2020.

David A Dickey and Wayne A Fuller. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American statistical association, 74(366a):427–431, 1979.

Marie Diron. Short-term forecasts of euro area real GDP growth: an assessment of real-time performance based on vintage data. Journal of forecasting, 27(5):371–390, 2008.

A Elbourne, K Grabska, H Kranendonk, and J Rhuggenaath. The effects of CBS revisions on CPB forecasts. CPB background document, 2015.

Peter Exterkate, Patrick JF Groenen, Christiaan Heij, and Dick van Dijk. Nonlinear forecasting with many predictors using kernel ridge regression. International Journal of Forecasting, 32(3):736–753, 2016.

Jeremy Ferwerda, Jens Hainmueller, and Chad J. Hazlett. Kernel-based regularized least squares in R (KRLS) and Stata (krls). Journal of Statistical Software, 79(3):1–26, 2017. doi: 10.18637/jss.v079.i03.

Rigoberto Fonseca-Delgado and Pilar Gomez-Gil. An assessment of ten-fold and Monte Carlo cross validations for time series forecasting. In 2013 10th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), pages 215–220. IEEE, 2013.

Olivier Fortin-Gagnon, Maxime Leroux, Dalibor Stevanovic, and Stéphane Surprenant. A large canadian database for macroeconomic analysis. Canadian Journal of Economics/Revue canadienne d'économique, 55(4):1799–1833, 2022.

Jerome Friedman, Robert Tibshirani, and Trevor Hastie. Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software, 33(1):1–22, 2010. doi: 10.18637/jss.v033.i01.

Domenico Giannone, Michele Lenza, and Giorgio E Primiceri. Economic predictions with big data: The illusion of sparsity. Econometrica, 89(5):2409–2437, 2021.

Periklis Gogas, Theophilos Papadimitriou, and Emmanouil Sofianos. Forecasting unemployment in the euro area with machine learning. Journal of Forecasting, 41(3):551–566, 2022.

Philippe Goulet Coulombe. The macroeconomy as a random forest. Journal of Applied Econometrics, 2020.

Philippe Goulet Coulombe, Maxime Leroux, Dalibor Stevanovic, and Stéphane Surprenant. How is machine learning useful for macroeconomic forecasting? Journal of Applied Econometrics, 37(5): 920–964, 2022.

Phillipe Goulet Coulombe. A neural phillips curve and a deep output gap, 2022.

Clive Granger and Yongil Jeon. Forecasting performance of information criteria with many macro series. Journal of Applied Statistics, 31(10):1227–1240, 2004.

Eleonora Granziera and Tatevik Sekhposyan. Predicting relative forecasting performance: An empirical investigation. International Journal of Forecasting, 35(4):1636–1657, 2019.

Aaron Smalter Hall et al. Machine learning approaches to macroeconomic forecasting. The Federal Reserve Bank of Kansas City Economic Review, 103(63):2, 2018.

Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. The elements of statistical learning: data mining, inference, and prediction, volume 2. Springer, 2009.

Irma Hindrayanto, Siem Jan Koopman, and Jasper de Winter. Forecasting and nowcasting economic growth in the euro area using factor models. International Journal of Forecasting, 32(4):1284–1305, 2016.

Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1):55–67, 1970.

Dashan Huang, Fuwei Jiang, Kunpeng Li, Guoshi Tong, and Guofu Zhou. Scaled PCA: A new approach to dimension reduction. Management Science, 68(3):1678–1695, 2022.

W Jos Jansen, Xiaowen Jin, and Jasper M de Winter. Forecasting and nowcasting real GDP: Comparing statistical models and subjective forecasts. International Journal of Forecasting, 32(2):411–436, 2016.

Dennis Kant, Andreas Pick, and Jasper de Winter. Nowcasting GDP using machine learning methods. 2022.

Rachidi Kotchoni, Maxime Leroux, and Dalibor Stevanovic. Macroeconomic forecast accuracy in a data-rich environment. Journal of Applied Econometrics, 34(7):1050–1072, 2019.

Kuhn and Max. Building predictive models in R using the caret package. Journal of Statistical Software, 28(5):1–26, 2008. doi: 10.18637/jss.v028.i05. URL https://www.jstatsoft.org/index.php/jss/article/view/v028i05.

Miron B Kursa, Aleksander Jankowski, and Witold R Rudnicki. Boruta–a system for feature selection. Fundamenta Informaticae, 101(4):271–285, 2010.

Tae-Hwy Lee, Halbert White, and Clive WJ Granger. Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests. Journal of Econometrics, 56 (3):269–290, 1993.

Kohei Maehashi and Mototsugu Shintani. Macroeconomic forecasting using factor models and machine learning: an application to Japan. Journal of the Japanese and International Economies, 58:101104, 2020.

Massimiliano Marcellino. A linear benchmark for forecasting GDP growth and inflation? Journal of Forecasting, 27(4):305–340, 2008.

Michael W McCracken and Serena Ng. FRED-MD: A monthly database for macroeconomic research. Journal of Business & Economic Statistics, 34(4):574–589, 2016.

Marcelo C Medeiros, Gabriel FR Vasconcelos, Álvaro Veiga, and Eduardo Zilberman. Forecasting inflation in a data-rich environment: the benefits of machine learning methods. Journal of Business & Economic Statistics, 39(1):98–119, 2021.

Adam Richardson, Thomas Mulder, et al. Nowcasting New Zealand GDP using machine learning algorithms. 2018.

Bas Scheer. Addressing unemployment rate forecast errors in relation to the business cycle. Technical report, CPB Netherlands Bureau for Economic Policy Analysis, 2022.

Erwan Scornet. Tuning parameters in random forests. ESAIM: Proceedings and Surveys, 60:144–162, 2017.

Georgios Sermpinis, Charalampos Stasinakis, Konstantinos Theofilatos, and Andreas Karathanasopoulos. Inflation and unemployment forecasting with genetic support vector regression. Journal of Forecasting, 33(6):471–487, 2014.

James H Stock and Mark W Watson. A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series, 1998.

James H Stock and Mark W Watson. Forecasting using principal components from a large number of predictors. Journal of the American statistical association, pages 1167–1179, 2002a.

James H Stock and Mark W Watson. Macroeconomic forecasting using diffusion indexes. Journal of Business & Economic Statistics, pages 147–162, 2002b.

James H Stock and Mark W Watson. Dynamic factor models. 2011.

James H Stock and Mark W Watson. Generalized shrinkage methods for forecasting using many predictors. Journal of Business & Economic Statistics, 30(4):481–493, 2012.

Robert Tibshirani. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology, 58(1):267–288, 1996.

Marvin N. Wright and Andreas Ziegler. ranger: A fast implementation of random forests for high dimensional data in C++ and R. Journal of Statistical Software, 77(1):1–17, 2017. doi: 10.18637/jss.v077.i01.

Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B: Statistical Methodology, 67(2):301–320, 2005.

A Variables NL-MD

Table 8: This table gives information about the variables in the NL-MD. The vintage of the dataset used in the research is 2024-02-28.

Variable name	Description	Period	Transformation	Source
Category: housing market (18 variables)				
wnm voorraad woningen Nederland	The total number of residential objects with a housing function in the Netherlands.	2012M01-2023M12	6	CBS 81955NED
wnm voorraad niet woningen Nederland	The total number of residential objects without a housing function in the Netherlands.	2012M01-2023M12	6	CBS 81955NED
wnm voorraad woningen Noord Nederland	The total number of residential objects with a housing function in the northern part of the Netherlands.	2012M01-2023M12	5	CBS 81955NED
wnm voorraad niet woningen Noord Nederland	The total number of residential objects without a housing function in the northern part of the Netherlands.	2012M01-2023M12	4	CBS 81955NED
wnm voorraad woningen Oost Nederland	The total number of residential objects with a housing function in the eastern part of the Netherlands.	2012M01-2023M12	5	CBS 81955NED
wnm voorraad niet woningen Oost Nederland	The total number of residential objects without a housing function in the eastern part of the Netherlands.	2012M01-2023M12	6	CBS 81955NED

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
wnm voorraad woningen West Nederland	The total number of residential objects with a housing function in the western part of the Netherlands.	2012M01-2023M12	5	CBS 81955NED
wnm voorraad niet woningen West Nederland	The total number of residential objects without a housing function in the western part of the Netherlands.	2012M01-2023M12	5	CBS 81955NED
wnm voorraad woningen Zuid Nederland	The total number of residential objects with a housing function in the southern part of the Netherlands.	2012M01-2023M12	6	CBS 81955NED
wnm voorraad niet woningen Zuid Nederland	The total number of residential objects without a housing function in the southern part of the Netherlands.	2012M01-2023M12	5	CBS 81955NED
wnm bouwvergunning woningen	The number of newly to be built homes for which a building permit has been granted.	2012M01-2023M12	5	CBS 83668NED
wnm bouwvergunning wooneenheden	The number of newly to be built residential units for which a building permit has been granted.	2012M01-2023M12	1	CBS 83668NED
wnm bouwvergunning recreatiewoningen	The number of newly to be built recreational homes for which a building permit has been granted.	2012M01-2023M12	4	CBS 83668NED
wnm huizenprijsindex koopwoningen bestaand	Housing price index of the stock of existing homes.	1995M01-2023M12	6	CBS 83906NED

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
wnm huizenprijsindex inflatie koopwoningen bestaand	Change in the housing price index of the stock of existing homes.	1996M01-2023M12	1	CBS 83906NED
wnm aantal verkochte koopwoningen bestaand	The total number of sold existing homes.	1995M01-2023M12	5	CBS 83906NED
wnm gemiddelde verkoopprijs koopwoningen bestaand	The average selling price of sold existing homes.	1995M01-2023M12	5	CBS 83906NED
wnm totale verkoopprijs koopwoningen bestaand	The total value of selling prices of sold existing homes.	1995M01-2023M12	5	CBS 83906NED
Category: indices and indicators (68 variables)				
ind producentenvertrouwen sector C	Producer confidence in the industrial sector.	1985M01-2024M01	1	CBS 81234ned
ind faillissementen	The number of declared bankruptcies.	1981M01-2024M01	5	CBS 82242NED
ind CPI totaal	Consumer Price Index (CPI) for all households, over all expenditures.	1996M01-2024M01	5	CBS 83131NED
ind CPI afgeleid totaal	Consumer Price Index (CPI), excluding the effect of changes in rates of product-related taxes and subsidies, for all households, over all expenditures.	1996M01-2024M01	5	CBS 83131NED
ind CPI inflatie totaal	Change in the Consumer Price Index (CPI) for all households, over all expenditures.	1997M01-2024M01	2	CBS 83131NED

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
ind CPI afgeleid inflatie totaal	Change in the Consumer Price Index (CPI), excluding the effect of changes in rates of product-related taxes and subsidies, for all households, over all expenditures.	1997M01-2024M01	1	CBS 83131NED
ind CPI woninghuur	Consumer Price Index (CPI) for all households on actual rent paid by tenants.	1996M01-2024M01	5	CBS 83131NED
ind CPI afgeleid woninghuur	Consumer Price Index (CPI), excluding the effect of changes in rates of product-related taxes and subsidies, for all households on actual rent paid by tenants.	1996M01-2024M01	5	CBS 83131NED
ind CPI inflatie woninghuur	Change in the Consumer Price Index (CPI) for all households on actual rent paid by tenants.	1997M01-2024M01	4	CBS 83131NED
ind CPI afgeleid inflatie woninghuur	Change in the Consumer Price Index (CPI), excluding the effect of changes in rates of product-related taxes and subsidies, for all households on actual rent paid by tenants.	1997M01-2024M01	4	CBS 83131NED
ind CPI goederen	Consumer Price Index (CPI) for all households on goods.	1996M01-2024M01	5	CBS 83131NED

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
ind CPI afgeleid goederen	Consumer Price Index (CPI), excluding the effect of changes in rates of product-related taxes and subsidies, for all households on goods.	1996M01-2024M01	5	CBS 83131NED
ind CPI inflatie goederen	Change in the Consumer Price Index (CPI) for all households on goods.	1997M01-2024M01	1	CBS 83131NED
ind CPI afgeleid inflatie goederen	Change in the Consumer Price Index (CPI), excluding the effect of changes in rates of product-related taxes and subsidies, for all households on goods.	1997M01-2024M01	1	CBS 83131NED
ind CPI energie	Consumer Price Index (CPI) for all households on energy including other fuels.	1996M01-2024M01	5	CBS 83131NED
ind CPI afgeleid energie	Consumer Price Index (CPI), excluding the effect of changes in rates of product-related taxes and subsidies, for all households on energy including other fuels.	1996M01-2024M01	5	CBS 83131NED
ind CPI inflatie energie	Change in the Consumer Price Index (CPI) for all households on energy including other fuels.	1997M01-2024M01	1	CBS 83131NED

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
ind CPI afgeleid inflatie energie	Change in the Consumer Price Index (CPI), excluding the effect of changes in rates of product-related taxes and subsidies, for all households on energy including other fuels.	1997M01-2024M01	1	CBS 83131NED
ind CPI voeding	Consumer Price Index (CPI) for all households on food including alcoholic beverages and tobacco.	1996M01-2024M01	5	CBS 83131NED
ind CPI afgeleid voeding	Consumer Price Index (CPI), excluding the effect of changes in rates of product-related taxes and subsidies, for all households on food including alcoholic beverages and tobacco.	1996M01-2024M01	5	CBS 83131NED
ind CPI inflatie voeding	Change in the Consumer Price Index (CPI) for all households on food including alcoholic beverages and tobacco.	1997M01-2024M01	1	CBS 83131NED
ind CPI afgeleid inflatie voeding	Change in the Consumer Price Index (CPI), excluding the effect of changes in rates of product-related taxes and subsidies, for all households on food including alcoholic beverages and tobacco.	1997M01-2024M01	1	CBS 83131NED
ind CPI diensten	Consumer Price Index (CPI) for all households on services (general index excluding goods).	1996M01-2024M01	5	CBS 83131NED

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
ind CPI afgeleid diensten	Consumer Price Index (CPI), excluding the effect of changes in rates of product-related taxes and subsidies, for all households on services (general index excluding goods).	1996M01-2024M01	5	CBS 83131NED
ind CPI inflatie diensten	Change in the Consumer Price Index (CPI) for all households on services (general index excluding goods).	1997M01-2024M01	5	CBS 83131NED
ind CPI afgeleid inflatie diensten	Change in the Consumer Price Index (CPI), excluding the effect of changes in rates of product-related taxes and subsidies, for all households on services (general index excluding goods).	1997M01-2024M01	2	CBS 83131NED
ind CPI kern	Consumer Price Index (CPI) for all households (general index excluding energy, food, alcoholic beverages, and tobacco).	1996M01-2024M01	5	CBS 83131NED
ind CPI afgeleid kern	Consumer Price Index (CPI), excluding the effect of changes in rates of product-related taxes and subsidies, for all households (general index excluding energy, food, alcoholic beverages, and tobacco).	1996M01-2024M01	5	CBS 83131NED

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
ind CPI inflatie kern	Change in the Consumer Price Index (CPI) for all households (general index excluding energy, food, alcoholic beverages, and tobacco).	1997M01-2024M01	5	CBS 83131NED
ind CPI afgeleid inflatie kern	Change in the Consumer Price Index (CPI), excluding the effect of changes in rates of product-related taxes and subsidies, for all households (general index excluding energy, food, alcoholic beverages, and tobacco).	1997M01-2024M01	5	CBS 83131NED
ind HICP	Harmonised Index of Consumer Prices (HICP) of the Netherlands.	1996M01-2024M01	5	CBS 83135NED
ind MUICP	Monetary Union Index of Consumer Prices (MUICP).	1996M01-2024M01	5	CBS 83135NED
ind EICP	European Index of Consumer Prices (EICP).	1996M01-2023M12	5	CBS 83135NED
ind HICP inflatie	Change in the Harmonised Index of Consumer Prices (HICP) of the Netherlands.	1997M01-2024M01	2	CBS 83135NED
ind MUICP inflatie	Change in the Monetary Union Index of Consumer Prices (MUICP).	1997M01-2024M01	1	CBS 83135NED
ind EICP inflatie	Change in the European Index of Consumer Prices (EICP).	1997M01-2023M12	1	CBS 83135NED

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
ind consumentenvertrouwen	Indicator of consumer sentiment regarding the developments of the Dutch economy and their own financial situation (consumer confidence).	1986M04-2024M02	2	CBS 83693NED
ind economisch klimaat	Indicator of consumer sentiment regarding the developments of the Dutch economy.	1986M04-2024M02	1	CBS 83693NED
ind koopbereidheid	Indicator of consumer sentiment regarding their own financial situation (purchase readiness).	1986M04-2024M02	2	CBS 83693NED
ind economische situatie verleden	Indicator providing information on consumers' confidence and opinions regarding the developments of the Dutch economy over the past 12 months.	1986M04-2024M02	2	CBS 83693NED
ind economische situatie toekomst	Indicator providing information on consumers' confidence and opinions regarding the developments of the Dutch economy over the next 12 months.	1986M04-2024M02	1	CBS 83693NED
ind financiele situatie verleden	Indicator providing information on consumers' opinions regarding their own financial situation over the past 12 months.	1986M04-2024M02	2	CBS 83693NED
ind financiele situatie toekomst	Indicator providing information on consumers' opinions regarding their own financial situation over the next 12 months.	1986M04-2024M02	1	CBS 83693NED

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
ind grote aankopen	The balance of the percentage of positive and negative answers to the question: “When you look at the overall economic situation, do you consider it a favorable or unfavorable time for major purchases, such as furniture, electronics, etc.?”	1986M04-2024M02	2	CBS 83693NED
ind PPI afzetprijzen binnenland delfstoffenwinning	Producer Price Index (PPI) of the prices charged by Dutch producers of crude oil and natural gas to domestic customers.	2012M01-2023M12	4	CBS 83935NED
ind PPI afzetprijzen buitenland delfstoffenwinning	Producer Price Index (PPI) of the prices charged by Dutch producers of crude oil and natural gas for export.	2012M01-2018M10	4	CBS 83935NED
ind PPI invoerprijzen delfstoffenwinning	Producer Price Index (PPI) of the prices paid by Dutch producers and traders for the import of crude oil and natural gas.	2012M01-2023M12	5	CBS 83935NED
ind PPI verbruiksprijzen delfstoffenwinning	Producer Price Index (PPI) of the prices charged by Dutch producers of crude oil and natural gas to domestic customers, adjusted for the import prices paid.	2012M01-2023M12	5	CBS 83935NED
ind PPI afzetprijzen binnenland voeding	Producer Price Index (PPI) of the prices charged by Dutch food producers to domestic customers.	2012M01-2023M12	5	CBS 83935NED

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
ind PPI afzetprijzen buitenland voeding	Producer Price Index (PPI) of the prices charged by Dutch food producers for export.	2012M01-2023M12	5	CBS 83935NED
ind PPI invoerprijzen voeding	Producer Price Index (PPI) of the prices paid by Dutch producers and traders for the import of food products.	2012M01-2023M12	6	CBS 83935NED
ind PPI verbruiksprijzen voeding	Producer Price Index (PPI) of the prices charged by Dutch food producers to domestic customers, adjusted for the import prices paid.	2012M01-2023M12	6	CBS 83935NED
ind PPI afzetprijzen binnenland chemie	Producer Price Index (PPI) of the prices charged by Dutch producers of chemical products to domestic customers.	2012M01-2023M12	5	CBS 83935NED
ind PPI afzetprijzen buitenland chemie	Producer Price Index (PPI) of the prices charged by Dutch producers of chemical products for export.	2012M01-2023M12	6	CBS 83935NED
ind PPI invoerprijzen chemie	Producer Price Index (PPI) of the prices paid by Dutch producers and traders for the import of chemical products.	2012M01-2023M12	6	CBS 83935NED
ind PPI verbruiksprijzen chemie	Producer Price Index (PPI) of the prices charged by Dutch producers of chemical products to domestic customers, adjusted for the import prices paid.	2012M01-2023M12	6	CBS 83935NED

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
ind PPI afzetprijzen binnenland basismetalen	Producer Price Index (PPI) of the prices charged by Dutch producers of primary form metals to domestic customers.	2012M01-2023M12	4	CBS 83935NED
ind PPI afzetprijzen buitenland basismetalen	Producer Price Index (PPI) of the prices charged by Dutch producers of primary form metals for export.	2012M01-2023M12	6	CBS 83935NED
ind PPI invoerprijzen basismetalen	Producer Price Index (PPI) of the prices paid by Dutch producers and traders for the import of primary form metals.	2012M01-2023M12	6	CBS 83935NED
ind PPI verbruiksprijzen basismetalen	Producer Price Index (PPI) of the prices charged by Dutch producers of primary form metals to domestic customers, adjusted for the import prices paid.	2012M01-2023M12	6	CBS 83935NED
ind PPI afzetprijzen binnenland machines	Producer Price Index (PPI) of the prices charged by Dutch producers of machinery, appliances, and equipment (n.e.c.) to domestic customers.	2012M01-2023M12	6	CBS 83935NED
ind PPI afzetprijzen buitenland machines	Producer Price Index (PPI) of the prices charged by Dutch producers of machinery, appliances, and equipment (n.e.c.) for export.	2012M01-2023M12	6	CBS 83935NED
ind PPI invoerprijzen machines	Producer Price Index (PPI) of the prices paid by Dutch producers and traders for the import of machinery, appliances, and equipment (n.e.c.).	2012M01-2023M12	6	CBS 83935NED

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
ind PPI verbruiksprijzen machines	Producer Price Index (PPI) of the prices charged by Dutch producers of machinery, appliances, and equipment (n.e.c.) to domestic customers, adjusted for the import prices paid.	2012M01-2023M12	6	CBS 83935NED
ind PPI afzetprijzen binnenland gas elektriciteit	Producer Price Index (PPI) of the prices charged by Dutch producers of electricity, gas, steam, and hot water to domestic customers.	2012M01-2023M11	6	CBS 83935NED
ind PPI afzetprijzen buitenland gas elektriciteit	Producer Price Index (PPI) of the prices charged by Dutch producers of electricity, gas, steam, and hot water for export.	2012M01-2013M07	4	CBS 83935NED
ind PPI invoerprijzen gas elektriciteit	Producer Price Index (PPI) of the prices paid by Dutch producers and traders for the import of electricity, gas, steam, and hot water.	2012M01-2022M12	4	CBS 83935NED
ind PPI verbruiksprijzen gas elektriciteit	Producer Price Index (PPI) of the prices charged by Dutch producers of electricity, gas, steam, and hot water to domestic customers, adjusted for the import prices paid.	2012M01-2023M11	6	CBS 83935NED

Category: interest and exchange rates (12 variables)

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
rwk Amerikaanse dollar euro	Exchange rate between US dollar and euro.	1999M01-2024M01	5	ECB EXR.M.USD.EUR.SP00.A (M) and EXR.Q.USD.EUR.SP00.A (Q)
rwk Britse pond euro	Exchange rate between British pound and euro.	1999M01-2024M01	5	ECB EXR.M.GBP.EUR.SP00.A (M) and EXR.Q.GBP.EUR.SP00.A (Q)
rwk Japanse yen euro	Exchange rate between Japanese yen and euro.	1999M01-2024M01	5	ECB EXR.M.JPY.EUR.SP00.A (M) and EXR.Q.JPY.EUR.SP00.A (Q)
rwk Chinese yuan euro	Exchange rate between Chinese yuan and euro.	2000M01-2024M01	5	ECB EXR.M.CNY.EUR.SP00.A (M) and EXR.Q.CNY.EUR.SP00.A (Q)

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
rwk Euribor 1 maand	The Euro Interbank Offered Rate (Euribor) with a maturity of 1 month.	1994M01-2024M01	2	ECB FM.M.U2.EUR.RT.MM.EURIBOR1MD_.HSTA (M) and FM.Q.U2.EUR.RT.MM.EURIBOR1MD_.HSTA (Q)
rwk Euribor 3 maanden	The Euro Interbank Offered Rate (Euribor) with a maturity of 3 months.	1994M01-2024M01	2	ECB FM.M.U2.EUR.RT.MM.EURIBOR3MD_.HSTA (M) and FM.Q.U2.EUR.RT.MM.EURIBOR3MD_.HSTA (Q)

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
rwk Euribor 6 maanden	The Euro Interbank Offered Rate (Euribor) with a maturity of 6 months.	1994M01-2024M01	2	ECB FM.M.U2.EUR.RT.MM.EURIBOR6MD_.HSTA (M) and FM.Q.U2.EUR.RT.MM.EURIBOR6MD_.HSTA (Q)
rwk Euribor 1 jaar	The Euro Interbank Offered Rate (Euribor) with a maturity of 1 year.	1994M01-2024M01	2	ECB FM.M.U2.EUR.RT.MM.EURIBOR1YD_.HSTA (M) and FM.Q.U2.EUR.RT.MM.EURIBOR1YD_.HSTA (Q)
rwk depositorente	Deposit interest rate.	1999M01-2024M01	2	ECB FM.D.U2.EUR.4F.KR.DFR.LEV
rwk basisherfinancieringsrente vast	Main refinancing operations (fixed rate tenders).	1999M01-2024M01	1	ECB FM.D.U2.EUR.4F.KR.MRR_FR.LEV

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
rwk basisherfinancieringsrente flexibel	Main refinancing operations (variable rate tenders)	2000M06-2008M10	4	ECB FM.D.U2.EUR.4F. KR.MRR_MBR.LE V
rwk marginale beleningsrente	Deposit facility.	1999M01-2024M01	2	ECB FM.D.U2.EUR.4F. KR.DFR.LEV
Category: international trade (38 variables)				
int invoer waarde index	Import value index.	1995M01-2023M12	5	CBS 84264NED
int invoer prijs index	Import price index.	1995M01-2023M12	5	CBS 84264NED
int invoer volume index	Import volume index.	1995M01-2023M12	5	CBS 84264NED
int invoer waarde inflatie	Change in import value index.	1996M01-2023M12	1	CBS 84264NED
int invoer prijs inflatie	Change in import price index.	1996M01-2023M12	1	CBS 84264NED
int invoer volume inflatie	Change in import volume index.	1996M01-2023M12	1	CBS 84264NED
int uitvoer waarde index	Export value index.	1995M01-2023M12	5	CBS 84264NED
int uitvoer prijs index	Export price index.	1995M01-2023M12	5	CBS 84264NED
int uitvoer volume index	Export volume index.	1995M01-2023M12	5	CBS 84264NED
int uitvoer waarde inflatie	Change in export value index.	1996M01-2023M12	1	CBS 84264NED
int uitvoer prijs inflatie	Change in export price index.	1996M01-2023M12	1	CBS 84264NED
int uitvoer volume inflatie	Change in export volume index.	1996M01-2023M12	1	CBS 84264NED

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
int ruilvoet index	Index of the ratio between the price levels of exports and imports of goods.	1995M01-2023M12	4	CBS 84265NED
int ruilvoet inflatie	Change in index of the ratio between the price levels of exports and imports of goods.	1996M01-2023M12	1	CBS 84265NED
int invoer goederen totaal	The total value of goods (total of all goods) transferred from foreign companies and individuals to Dutch companies and individuals.	2015M01-2023M12	5	CBS 85427NED
int invoer goederen voeding levende dieren	The total value of food and live animals (goods) transferred from foreign companies and individuals to Dutch companies and individuals.	2015M01-2023M12	5	CBS 85427NED
int invoer goederen dranken tabak	The total value of alcoholic beverages and tobacco (goods) transferred from foreign companies and individuals to Dutch companies and individuals.	2015M01-2023M12	4	CBS 85427NED
int invoer goederen grondstoffen	The total value of non-edible raw materials, excluding fuels (goods), transferred from foreign companies and individuals to Dutch companies and individuals.	2015M01-2023M12	5	CBS 85427NED

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
int invoer goederen brandstoffen	The total value of mineral fuels, lubricants, and similar products (goods) transferred from foreign companies and individuals to Dutch companies and individuals.	2015M01-2023M12	5	CBS 85427NED
int invoer goederen olien	The total value of animal and vegetable oils and fats (goods) transferred from foreign companies and individuals to Dutch companies and individuals.	2015M01-2023M12	5	CBS 85427NED
int invoer goederen chemische producten	The total value of chemical products (goods) transferred from foreign companies and individuals to Dutch companies and individuals.	2015M01-2023M12	5	CBS 85427NED
int invoer goederen fabricaten	The total value of manufactures primarily sorted according to raw materials (goods) transferred from foreign companies and individuals to Dutch companies and individuals.	2015M01-2023M12	5	CBS 85427NED
int invoer goederen machines	The total value of machinery and transport equipment (goods) transferred from foreign companies and individuals to Dutch companies and individuals.	2015M01-2023M12	4	CBS 85427NED

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
int invoer goederen gefabriceerd	The total value of miscellaneous manufactured goods transferred from foreign companies and individuals to Dutch companies and individuals.	2015M01-2023M12	5	CBS 85427NED
int invoer goederen niet afzonderlijk genoemd	The total value of not elsewhere specified goods transferred from foreign companies and individuals to Dutch companies and individuals.	2015M01-2023M12	5	CBS 85427NED
int invoer goederen niet fysiek	The total value of goods not physically in the Netherlands transferred from foreign companies and individuals to Dutch companies and individuals.	2015M01-2023M12	5	CBS 85427NED
int uitvoer goederen totaal	The total value of goods (total of all goods) transferred from Dutch companies and individuals to foreign companies and individuals.	2015M01-2023M12	5	CBS 85427NED
int uitvoer goederen voeding levende dieren	The total value of food and live animals (goods) transferred from Dutch companies and individuals to foreign companies and individuals.	2015M01-2023M12	5	CBS 85427NED
int uitvoer goederen dranken tabak	The total value of alcoholic beverages and tobacco (goods) transferred from Dutch companies and individuals to foreign companies and individuals.	2015M01-2023M12	4	CBS 85427NED

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
int uitvoer goederen grondstoffen	The total value of non-edible raw materials, excluding fuels (goods), transferred from Dutch companies and individuals to foreign companies and individuals.	2015M01-2023M12	4	CBS 85427NED
int uitvoer goederen brandstoffen	The total value of mineral fuels, lubricants, and similar products (goods) transferred from Dutch companies and individuals to foreign companies and individuals.	2015M01-2023M12	6	CBS 85427NED
int uitvoer goederen olien	The total value of animal and vegetable oils and fats (goods) transferred from Dutch companies and individuals to foreign companies and individuals.	2015M01-2023M12	5	CBS 85427NED
int uitvoer goederen chemische producten	The total value of chemical products (goods) transferred from Dutch companies and individuals to foreign companies and individuals.	2015M01-2023M12	5	CBS 85427NED
int uitvoer goederen fabricaten	The total value of manufactures primarily sorted according to raw materials (goods) transferred from Dutch companies and individuals to foreign companies and individuals.	2015M01-2023M12	5	CBS 85427NED

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
int uitvoer goederen machines	The total value of machinery and transport equipment (goods) transferred from Dutch companies and individuals to foreign companies and individuals.	2015M01-2023M12	5	CBS 85427NED
int uitvoer goederen gefabriceerd	The total value of miscellaneous manufactured goods transferred from Dutch companies and individuals to foreign companies and individuals.	2015M01-2023M12	4	CBS 85427NED
int uitvoer goederen niet afzonderlijk genoemd	The total value of not elsewhere specified goods transferred from Dutch companies and individuals to foreign companies and individuals.	2015M01-2023M12	5	CBS 85427NED
int uitvoer goederen niet fysiek	The total value of goods not physically in the Netherlands transferred from Dutch companies and individuals to foreign companies and individuals.	2015M01-2023M12	5	CBS 85427NED
Category: labor market (9 variables)				
arb uitkering ongeschiktheid	The total number of benefits under the disability insurance laws (WAO, WAZ, Wajong, WIA).	1998M01-2023M11	6	CBS 37789ksz
arb uitkering werkloos	The number of benefits under the unemployment insurance act (WW).	1998M01-2023M11	5	CBS 37789ksz

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
arb uitkering bijstand	The number of benefits under the work and assistance act (WWB, until 2015) or the participation act (Participatiewet, since 2015).	1998M01-2023M11	6	CBS 37789ksz
arb beroepsbevolking totaal	Labor force (total of the active and unemployed labor force).	2003M01-2024M01	5	CBS 80590ned
arb beroepsbevolking werkzaam	Active labor force (individuals engaged in paid work).	2003M01-2024M01	5	CBS 80590ned
arb beroepsbevolking werkloos	Unemployed labor force (individuals who are not engaged in paid work, have recently searched for paid work, and are immediately available for it).	2003M01-2024M01	5	CBS 80590ned
arb werkloosheidspercentage	Unemployment rate (unemployed labor force as a percentage of the labor force).	2003M01-2024M01	5	CBS 80590ned
arb cao lonen totaal	Index of collective labor agreement (CAO) hourly wages including special benefits across all sectors.	2010M01-2023M10	5	CBS 82838NED
arb cao loonkosten totaal	Index of contractual labor costs across all sectors.	2010M01-2023M10	5	CBS 82838NED
Category: production (9 variables)				
prd productie index aardolie en aardgaswinning	Production index for the extraction of crude oil and natural gas.	2005M01-2023M12	5	CBS 83838NED
prd productie index voeding	Production index for the food industry.	2005M01-2023M12	5	CBS 83838NED
prd productie index chemie	Production index for the chemical industry.	2005M01-2023M12	4	CBS 83838NED

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
prd productie index basismetalen	Production index for the basic metals industry.	2005M01-2023M12	5	CBS 83838NED
prd productie index machines	Production index for the machinery industry.	2005M01-2023M09	5	CBS 83838NED
prd productie index gas en elektriciteit	Production index for energy companies.	2005M01-2021M09	5	CBS 83838NED
prd productie index delfstoffenwinning	Production index for mining and quarrying industries.	2005M01-2023M12	5	CBS 83838NED
prd productie index nijverheid en energie	Production index for manufacturing (excluding construction) and energy.	2005M01-2023M12	5	CBS 83838NED
prd productie index industrie	Production index for industry.	2005M01-2023M12	5	CBS 83838NED
Category: stock market (14 variables)				
adm SP500 open	Opening price of the S&P 500 index.	2007M01-2024M01	4	Yahoo GSPC
adm SP500 hoog	Highest price of the S&P 500 index.	2007M01-2024M01	4	Yahoo GSPC
adm SP500 laag	Lowest price of the S&P 500 index.	2007M01-2024M01	4	Yahoo GSPC
adm SP500 sluit	Closing price of the S&P 500 index.	2007M01-2024M01	4	Yahoo GSPC
adm SP500 volume	Trading volume of the S&P 500 index.	2007M01-2024M01	5	Yahoo GSPC
adm AEX open	Opening price of the AEX index.	2007M01-2024M01	4	Yahoo AEX
adm AEX hoog	Highest price of the AEX index.	2007M01-2024M01	4	Yahoo AEX
adm AEX laag	Lowest price of the AEX index.	2007M01-2024M01	4	Yahoo AEX
adm AEX sluit	Closing price of the AEX index.	2007M01-2024M01	4	Yahoo AEX

Continued on next page

Description of NL-MD continued from previous page.

Variable name	Description	Period	Transformation	Source
adm AEX volume	Trading volume of the AEX index.	2007M01-2024M01	5	Yahoo AEX
adm AMX open	Opening price of the AMX index.	2011M12-2024M01	5	Yahoo AMX
adm AMX hoog	Highest price of the AMX index.	2011M12-2024M01	5	Yahoo AMX
adm AMX laag	Lowest price of the AMX index.	2011M12-2024M01	5	Yahoo AMX
adm AMX sluit	Closing price of the AMX index.	2011M12-2024M01	5	Yahoo AMX

B Feature selection

Prior to training our models, we perform two methods of feature selection/engineering. A traditional PCA method, to combine as much information in a small set of factors, and a selection procedure based on the mutual information and the variance inflation factor.

B.1 Factors

Using a dataset with a large amount of features, while at the same time having a limited amount of observations, can lead to severe overfitting problems when using OLS. The traditional way of tackling this problem in econometrics is by imposing a latent structure

$$X_t = \Lambda F_t + \varepsilon_t, \tag{14}$$

meaning we assume that the set of information in our feature set X_t can be explained by a (significantly smaller) amount of factors F_t. Here, we follow the approach by McCracken and Ng (2016) and estimate the factors F_t by means of principal components analysis (PCA). We then select the first k factors (ordered by eigenvalue) as features in our dataset, where we will choose k using cross-validation. Taking factors is a reasonable approach; In macroeconomic datasets we often face the situation where the number of variables significantly outnumbers the number of observations. Factors then offer a way to compress as much variation from the original features into a compressed dataset with a limited number of variables and have been quite successful in forecasting (Coulombe et al., 2021a, Goulet Coulombe et al., 2022, Stock and Watson, 2011).

B.2 Variance Inflation Factor

As an alternative to factors, we perform variable selection using the variable inflation factor (VIF). Here, we follow a similar approach as Cheng et al. (2022). First, we calculate the mutual information $I(X;Y)$ between the dependent variable and the features and order the variables from most to least mutual information,

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right), \tag{15}$$

with $p(x, y)$ the joint probability distribution of x and y, and $p(x)$ and $p(y)$ the marginal probability distributions of x and y. These are obtained after discretizing the normalised time series. Hereafter, we start with the variable having the highest mutual information and add the next one in the list. Then, we calculate the variance inflation factors (VIF) of the selected variables. The VIF of a variable i is defined as follows,

$$\text{VIF}_i = \frac{1}{1 - R_i^2}. \tag{16}$$

Here, R_i^2 is the R-squared obtained from an OLS regression with variable i as dependent variable and all other variables as independent variables. If one of the features in the set has a VIF larger than five the added variable is omitted. We repeat this process until all variables have been tested, or when we have selected ten features. Again, these features are lagged in the same manner as the ARDI model.

C Hyperparameters and packages

In this section, we show the grid of the hyperparameters, and the packages used to evaluate these models. Table 9 is split into two parts, the feature matrix, on which the models are evaluated, and the models itself. Note that the ARDI, ARVIF and AR models are not listed, this is because, apart from the hyperparameters for the feature matrix, they do not have any hyperparameters that need to be tuned.

We use the following packages in R: `glmnet` (Friedman et al., 2010), `Ranger` (Wright and Ziegler, 2017) and `KRLS` (Ferwerda et al., 2017). All the listed models are called using `Caret` (Kuhn and Max, 2008).

In the Table below, λ^* denotes the minimum value for λ for which lasso does not select any variables and N denotes the number of features in the feature matrix. The other variables are hyperparameters of the respective models. Finally, for the random forest, we fix the splitting rule on `Variance` and the number of trees on 500, which is sufficient to reach a stable prediction. The forest is not tuned, as our grid only has one value. We experimented with using a larger grid, but this made no notable difference. This is in line with the literature (Scornet, 2017).

Table 9: The hyperparameters for our models and feature matrices, with the packages used to call them.

Feature matrix/Models	**Grid**	**Package**
Feature matrices		
Data poor	$p_y \in \{1,3,6,12\}$	-
Diffusion indices (ARDI)	$p_y \in \{1,3,6,12\}$, $p_x \in \{1,3,6,12\}$, $K \in \{3,6,10\}$	-
VIF	$p_y \in \{1,3,6,12\}$, $p_x \in \{1,3,6,12\}$	-
B1	$p_y \in \{1,3,6,12\}$	-
B2	$p_y \in \{1,3,6,12\}$	-
B3	$p_y \in \{1,3,6,12\}$, $p_x \in \{1,3,6,12\}$	-
Models		
Lasso	$\lambda \in 10^{\{-5,\ldots,\log_{10} 1.5\lambda^*\}}$ (50 steps)	`glmnet`
Ridge	$\lambda \in 10^{\{-5,\ldots,\log_{10} 1.5\lambda^*\}}$ (50 steps)	`glmnet`
ElasticNet	$\lambda \in 10^{\{-5,\ldots,\log_{10} 1.5\lambda^*\}}$ (30 steps) , $\alpha \in \{0,1,2,...,10\}/10$	`glmnet`
Random Forest	`min.node.size` $= 3$, `mtry` $= \lceil \frac{N}{3} \rceil$.	`Ranger`
Kernel Ridge Regression	$\sigma^2 \in (0.5,1,1.5)N$, $\lambda \in 10^{\{-5,\ldots,1.5\lambda^*\}}$ (10 steps) .	`KRLS`

D Traditional models

In our main text we present a relatively simple way of describing the ARDI and AR models, by means of our different feature matrices Z. However, in the economic literature the AR and ARDI models are usually presented as follows:

AR.

The auto regressive model is the simplest model in our arsenal. In terms of the dependent variable y_t, we can write it down as follows

$$y_{t+h} = c + \rho(p_y)y_t + \varepsilon_{t+h}, \tag{17}$$

where c is a constant and $\rho(p_y)$ is a lag polynomial with a highest order of p_y, which is determined by cross validation.

ARDI.

For the ARDI model, we run a regression on the *data rich, diffusion indices* feature matrix. In terms of variables X_t and the dependent variable y_t, we estimate the following relation

$$\begin{aligned} y_{t+h} &= c + \rho(p_y)y_t + \rho(p_x)F_t + \varepsilon_{t+h}, \qquad (18) \\ X_t &= \Lambda F_t + u_t. \end{aligned}$$

In the second equation we impose a linear latent structure, discussed in Section 2, which we estimate by principal component analysis (PCA). In this model, F_t represents a set of K factors. The model is characterized by three essential hyperparameters: the number of lags of the dependent variable (p_y), of the factors p_x, and the number of factors K. These will be fine-tuned through cross-validation. In our base model we will estimate Equation 17 and Equation 18 using ordinary least squeares (OLS).

E Robustness

In this Appendix, we show the results of several robustness checks.

E.1 Different transformation

In our main analysis we applied the first-difference transformation to all variables. However, according to an Augmented Dickey-Fuller test, the AEX-index and production index of industry are also stationary in log-levels. We suspect that this is due to the test not having enough data to accurately detect the presence of a unit root, in combination with the messy financial crisis being in the sample, which severely impacted these two variables. As these two variables are usually expressed in growth rates, and for the reasons mentioned earlier, we decided in our main analysis to present the results of these variables using log first-differences. In Figure 13 the results are shown if they are kept in levels. In this case, the non-linear and shrinkage models perform significantly worse than their counterparts, highlighting the importance of data transformations. An elaborate study of data transformations in a macroeconomic context can be found in Coulombe et al. (2021a).

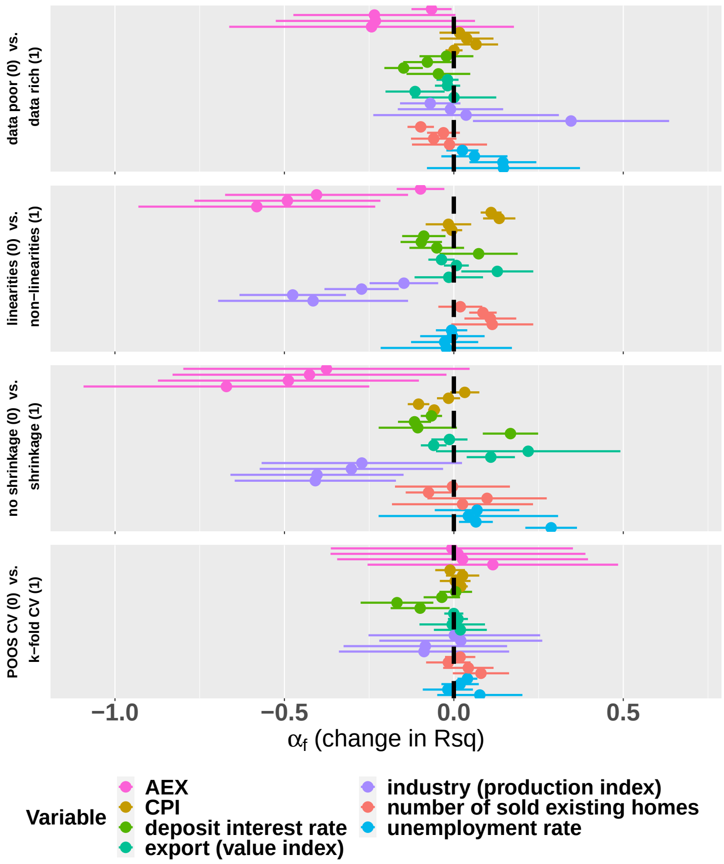


Figure 13: This figure shows the regression results corresponding to the regression in Equation 13. However, in this case the AEX and industry (production index) variables are left in log-levels. For the four different features under study (feature matrix, non-linearities, cross-validation, and shrinkage) we show the value of α_f including its 95% confidence interval. We are only controlling for time. For each variable, the training period starts on a different date (see Table 2). For all variables the test period is from 2015M01 to 2023M12. These results are based on the transformed dataset with winsorizing applied. The different colours correspond to the seven dependent variables that are being studied. For a specific colour, the four different horizons are shown as independent lines, where the horizon is increasing going down the figure (i.e. from $h = 1$, to $h = 3$, to $h = 6$, to $h = 12$).

E.2 Results with backtransformed data

In our main analysis we discuss the results on the transformed dataset. Figure 14 shows the results using the backtransformed dataset. Comparing Figure 14 (backtransformed dataset) to Figure 8 (transformed dataset) we see no significant differences, and thus, the results are robust regardless of which dataset is analyzed.

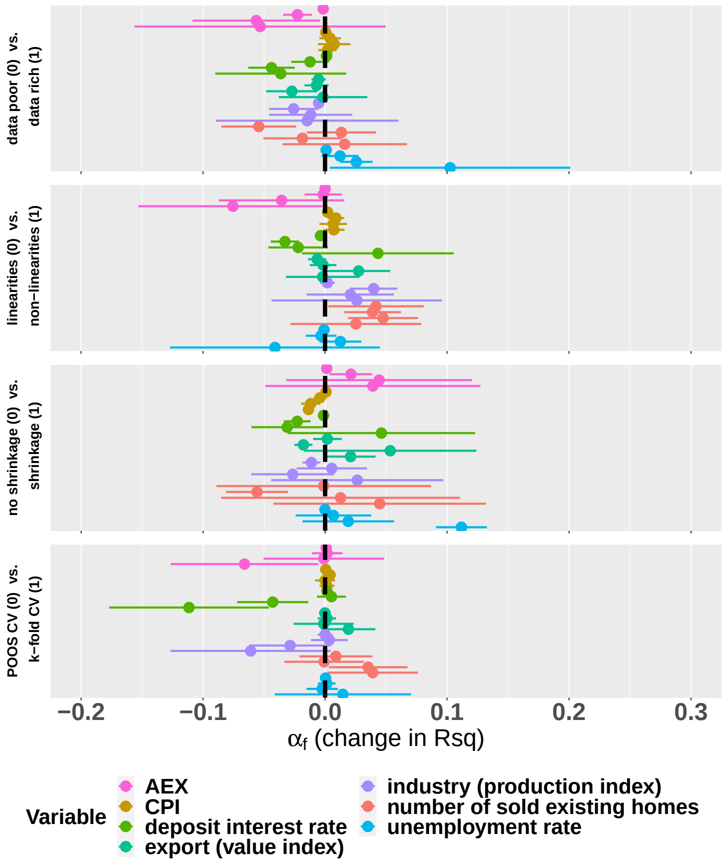


Figure 14: This figure shows the regression results corresponding to the regression in Equation 13. For the four different features under study (feature matrix, non-linearity's, cross-validation, and shrinkage) we show the value of α_f including its 95% confidence interval. We are only controlling for time. For each variable, the training period starts on a different date (see Table 2). For all variables the test period is from 2015M01 to 2023M12. These results are based on the backtransformed dataset with winsorizing applied. The different colours correspond to the seven dependent variables that are being studied. For a specific colour, the four different horizons are shown as independent lines, where the horizon is increasing going down the figure (i.e. from $h = 1$, to $h = 3$, to $h = 6$, to $h = 12$).

E.3 Results with non winsorizing the data

In our main analysis we discuss the results on a winsorized dataset. Winsorizing is applied to handle outliers in the R^2 values. Figure 15 shows the results without winsorizing. Comparing Figure 15 (without winsorizing) to Figure 8 (with winsorizing) we see a few differences. The error bars are larger, and some treatment effects are more pronounced. However, the general pattern remains consistent for most variables. Notable differences are observed in the shrinkage panel for the deposit interest rate and the unemployment rate at the largest horizon, as well as in the non-linearities for unemployment. Overall, the non-linearities treatment effect is slightly larger, but still not statistically significant.

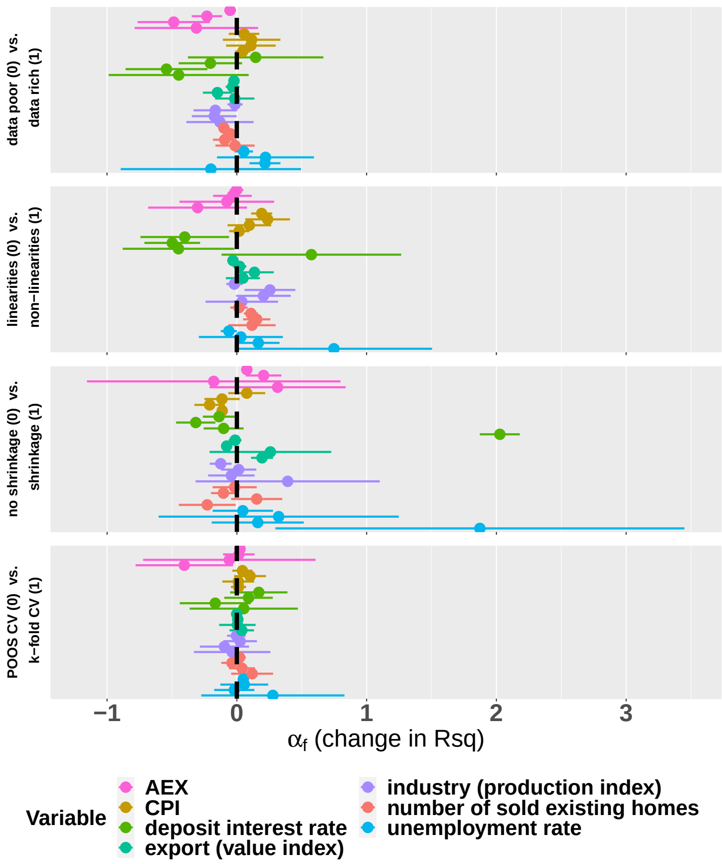


Figure 15: This figure shows the regression results corresponding to the regression in Equation 13. For the four different features under study (feature matrix, non-linearity's, cross-validation, and shrinkage) we show the value of α_f including its 95% confidence interval. We are only controlling for time. For each variable, the training period starts on a different date (see Table 2). For all variables the test period is from 2015M01 to 2023M12. These results are based on the transformed dataset with winsorizing not applied. The different colours correspond to the seven dependent variables that are being studied. For a specific colour, the four different horizons are shown as independent lines, where the horizon is increasing going down the figure (i.e. from $h = 1$, to $h = 3$, to $h = 6$, to $h = 12$).

E.4 Excluding the COVID-19 pandemic

In our main analysis, the test period is from 2015M01 to 2023M12, encompassing the COVID-19 pandemic. This pandemic significantly impacted the world economy, resulting in distinct patterns in many indicators in the NL-MD compared to previous periods. The fluctuations in these series complicate forecasting during the COVID-19 period, and studying a test period that includes this instability period may lead to unreliable results. Figure 16 shows the results of our analysis using an alternative test period from 2015M01 to 2019M12. Comparing Figure 16 (excluding COVID-19 period) to Figure 8 (including COVID-19 period), we observe differences primarily related to the deposit interest rate variable. For all variables, we notice that the effects becomes more pronounced when the COVID-19 period is excluded: treatment effects and their standard errors increase. This is likely due to the smaller number of observations compared to the results including the COVID-19 period. This is also why we include the COVID-19 period in our main analysis: excluding this period would not provide enough observations to ensure reliable results. Although the magnitude of treatment effects change when excluding the COVID-19 period, the same patterns are observable when including it, indicating that the results are robust to excluding the COVID-19 period. However, further research could investigate the impact of the COVID-19 period on our results in more detail.

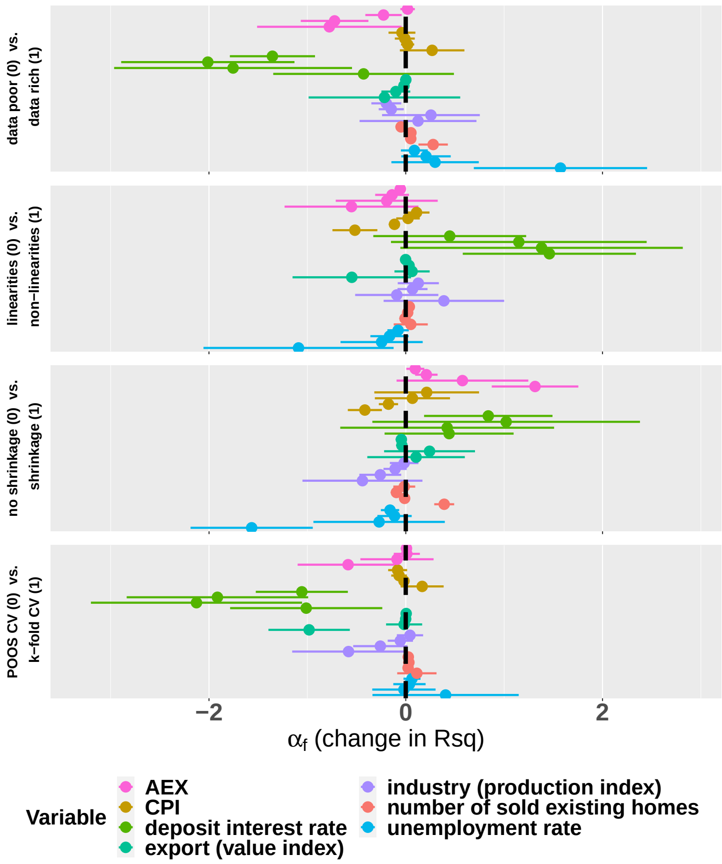


Figure 16: This figure shows the regression results corresponding to the regression in Equation 13. For the four different features under study (feature matrix, non-linearity's, cross-validation, and shrinkage) we show the value of α_f including its 95% confidence interval. We are only controlling for time. For each variable, the training period starts on a different date (see Table 2). For all variables the test period is from 2015M01 to 2019M12 (i.e. excluding the COVID-19 period). These results are based on the transformed dataset with winsorizing applied. The different colours correspond to the seven dependent variables that are being studied. For a specific colour, the four different horizons are shown as independent lines, where the horizon is increasing going down the figure (i.e. from $h = 1$, to $h = 3$, to $h = 6$, to $h = 12$).

F Outliers per model

In this section we give more information about the outliers in our dataset. We define an outlier as having an $R^2 \neq (-1, 1)$. Table 10 describes the outliers as a function of the model.

Table 10: This table describes the outliers in both the backtransformed and transformed data set for each model, horizon and variable.

	Transformed data set		**Backtransformed data set**	
	Outliers ($n = 15.813$)	**All observations ($n = 127.008$)**	**Outliers ($n = 6.811$)**	**All observations ($n = 127.008$)**
Models				
AR	774 (13%)	6,048	166 (3%)	6,048
RF AR	834 (14)	6,048	205 (3%)	6,048
RR AR	762 (13%)	6,048	166 (3%)	6,048
KRR AR	933 (15%)	6,048	230 (4%)	6,048
ARDI	953 (16%)	6,048	270 (5%)	6,048
RF ARDI	786 (13%)	6,048	172 (5%)	6,048
RR ARDI	945 (16%)	6,048	242 (3%)	6,048
KRR ARDI	875 (15%)	6,048	210 (4%)	6,048
ARVIF	823 (14%)	6,048	194 (3%)	6,048
RF ARVIF	824 (14%)	6,048	177 (3%)	6,048
RR ARVIF	792 (13%)	6,048	210 (3%)	6,048
KRR ARVIF	812 (13%)	6,048	227 (4%)	6,048
Lasso	2,693 (15%)	18,144	688 (4%)	18,144
Ridge	2,579 (14%)	18,144	707 (4%)	18,144
ElasticNet	2,641 (15%)	18,144	670 (4%)	18,144
Data				
poor (AR)	3,303 (14%)	24,192	767 (3%)	24,192
rich (ARDI)	3,559 (15%)	24,192	894 (4%)	24,192
rich (ARVIF)	3,251 (13%)	24,192	808 (3%)	24,192
rich (B1)	2,919 (16%)	18,144	827 (5%)	18,144
rich (B2)	2,456 (14%)	18,144	573 (3%)	18,144
rich (B3)	2,538 (14%)	18,144	665 (4%)	18,144
Cross-validation				
k-fold	8,997 (14%)	63,504	2,238 (4%)	63,504
POOS	9,029 (14%)	63,504	2,296 (4%)	63,504
Horizon				
1	2,640 (8%)	31,752	291 (1%)	31,752
3	4,134 (13%)	31,752	398 (1%)	31,752
6	5,069 (16%)	31,752	1,263 (4%)	31,752
12	6,183 (19%)	31,752	2,582 (8%)	31,752
Variable				
AEX	4,067 (22%)	18,144	321 (2%)	18,144
CPI	2,249 (12%)	18,144	55 (0%)	18,144
deposit interest rate	1,889 (10%)	18,144	814 (4%)	18,144
export (value index)	2,018 (11%)	18,144	248 (1%)	18,144
industry (production index)	2,749 (15%)	18,144	1,152 (6%)	18,144
number of sold existing homes	1,677 (9%)	18,144	1,784 (10%)	18,144
unemployment rate	3,377 (19%)	18,144	160 (1%)	18,144

G RMSEs

G.1 AEX

Maximum and short training period: for AEX both start 2010M01

Table 11: This table shows for the variable AEX (adm_AEX_hoog) the RMSEs of all the models that have been run for the different forecasting horizons. For this table the training period starts at 2010M01. The test period is from 2015M01 until 2023M12. For this variable a log transformation is used. The models are divided in data poor and data rich-models. The first data poor model, the AR, BIC, is the benchmark model and all other RMSEs are relative to this value. The stars indicate the results of a Diebold-Mariano test compared to the benchmark model (H_0: no difference in forecasting accuracy between the two models): $*$ significance at 10% level, $**$ significance at 5% level, and $***$ significance at 1% level.

	Variable transformed				Variable backtransformed			
Models	$h=1$	$h=3$	$h=6$	$h=12$	$h=1$	$h=3$	$h=6$	$h=12$
Data poor models								
AR, BIC (BM)	0.0317	0.0207	0.0146	0.0088	18.983	37.910	54.755	65.455
AR, k-fold	1	1	0.9541*	0.9966	1	1	0.9459	0.9943
AR, POOS	1.0275*	1.0245	0.9737	1.0102	1.0238**	1.0216	0.9620	1.0032
RF AR, k-fold	1.0365	1.0815**	1.1021*	1.3672**	1.0310	1.0840**	1.1093*	1.2560***
RF AR, POOS	1.0225	1.0955**	1.0684	1.2923**	1.0166	1.0908**	1.0783	1.2230***
RR AR, k-fold	1.0152*	0.9911	0.9524**	1.0535	1.0135**	0.9905	0.9430*	1.0211
RR AR, POOS	1.0239*	0.9969	0.9611	1.0211	1.0218**	0.9986	0.9489	1.0126
KRR AR, k-fold	1.0381	1.0871	1.2689***	1.7560***	1.0300	1.1069*	1.3130***	1.6270***
KRR AR, POOS	1.0732	1.1435**	1.0712	1.3923*	1.0714	1.1159**	1.0974	1.2456**
Data rich models								
ARDI, k-fold	1.0721	1.1346***	1.2094*	1.3721	1.0598	1.1296**	1.2022	1.1636
ARDI, POOS	1.0667*	1.1822***	1.4902*	1.5420*	1.0644	1.1983**	1.4123*	1.3194*
RF ARDI, k-fold	1.0967**	1.1328**	1.1072	1.2766*	1.1018*	1.1430*	1.0914	1.1575*
RF ARDI, POOS	1.1129**	1.1375**	1.2034***	1.2558*	1.1064*	1.1403*	1.1660***	1.1387
RR ARDI, k-fold	1.0438	1.1020*	1.2032	1.6219*	1.0540	1.0950	1.1266	1.4304*
RR ARDI, POOS	1.0567*	1.1277**	1.3831*	1.2909	1.0557	1.1191	1.2922*	1.1567
KRR ARDI, k-fold	1.0416**	1.1027**	1.4508**	1.6134**	1.0329*	1.0696*	1.2988*	1.4213**
KRR ARDI, POOS	1.0214	1.1255***	1.1583**	1.1327	1.0023	1.1128***	1.1204**	1.0475
ARVIF, k-fold	1.0826	1.1983***	1.3220*	1.4433**	1.0765	1.1956**	1.3271	1.2930***
ARVIF, POOS	1.0826	1.2499***	1.3460*	1.4053*	1.0765	1.2460***	1.3528*	1.2634**
RF ARVIF, k-fold	1.0364	1.2171***	1.4263***	1.3845**	1.0258	1.2121***	1.4417**	1.2671***
RF ARVIF, POOS	1.0096	1.2324***	1.2220**	1.4124***	1.0082	1.2223***	1.1945**	1.3274***
RR ARVIF, k-fold	1.0371	1.1601	1.0064	1.3459	1.0441	1.1790	0.9799	1.2275**
RR ARVIF, POOS	1.0413	1.1508	1.0505	1.1388*	1.0491	1.1665	1.0290	1.1225**
KRR ARVIF, k-fold	1.0169	1.1352*	1.1809**	1.5681*	1.0205	1.1094**	1.1403**	1.4671*
KRR ARVIF, POOS	1.0509	1.0165	1.0549	1.4598*	1.0683	1.0093	1.0367	1.3719*
Lasso (B1), k-fold	1.0173	1.2614**	2.1278	2.0361**	1.0184	1.2949**	1.9762	1.8839*
Lasso (B2), k-fold	1.0151	1.0191	1.1658	1.2011	1.0219	1.0228	1.2011	1.1592
Lasso (B3), k-fold	1.0401	1.0568*	1.1666***	1.2859*	1.0301	1.0497	1.1281***	1.2592
Lasso (B1), POOS	1.0708	1.1814**	2.0471*	1.3617**	1.0893	1.1621*	1.9457*	1.3212**
Lasso (B2), POOS	1.0111	1.0388	1.0666	1.0450	1.0201	1.0623	1.0842	1.0359
Lasso (B3), POOS	1.0220	1.0429	1.2315***	1.2135*	1.0254	1.0665	1.2179**	1.1454***
Ridge (B1), k-fold	1.0801	1.1506*	1.4621	1.3225	1.0939	1.1701*	1.4581	1.3036
Ridge (B2), k-fold	1.0241	0.9506	0.9911	1.0905	1.0256	0.9342	1.0054	1.0955
Ridge (B3), k-fold	1.0005	0.9821	0.9831	1.0860	1.0046	0.9790	0.9683	1.1103
Ridge (B1), POOS	1.0939	1.1329	1.4985	1.3695	1.1060	1.1537	1.4944	1.3438
Ridge (B2), POOS	1.0386	0.9661	0.9457	1.0885	1.0396	0.9693	0.9338	1.0998
Ridge (B3), POOS	1.0049	0.9835	0.9862	1.1082	1.0086	0.9810	0.9714	1.1406
ElasticNet (B1), k-fold	1.0108	1.1237	2.1770*	2.1031**	1.0133	1.1155	2.0601*	1.9256*
ElasticNet (B2), k-fold	1.0145	0.9934	1.0157	1.1874	1.0195	0.9777	1.0204	1.1553
ElasticNet (B3), k-fold	1.0006	1.0831*	1.1432**	1.2693**	0.9986	1.0931*	1.0995*	1.1878***
ElasticNet (B1), POOS	1.0741	1.1868**	2.0544*	1.2740	1.0907	1.1687*	1.9569*	1.1794
ElasticNet (B2), POOS	1.0119	1.0473	1.0294	1.0286	1.0214	1.0673	1.0412	1.0301
ElasticNet (B3), POOS	1.0525**	1.0394	1.2282***	1.2288**	1.0696**	1.0620	1.2215**	1.1652***

G.2 CPI

Maximum training period: for CPI this starts in 1999M01

Table 12: This table shows for the variable CPI (ind_CPI_totaal) the RMSEs of all the models that have been run for the different forecasting horizons. For this table the training period starts at 1999M01. The test period is from 2015M01 until 2023M12. For this variable a first-difference log transformation is used. The models are divided in data poor and data rich-models. The first data poor model, the AR, BIC, is the benchmark model and all other RMSEs are relative to this value. The stars indicate the results of a Diebold-Mariano test compared to the benchmark model (H_0: no difference in forecasting accuracy between the two models): $*$ significance at 10% level, $**$ significance at 5% level, and $***$ significance at 1% level.

	Variable transformed				Variable backtransformed			
Models	$h=1$	$h=3$	$h=6$	$h=12$	$h=1$	$h=3$	$h=6$	$h=12$
Data poor models								
AR, BIC (BM)	0.0084	0.0057	0.0036	0.0025	1.0317	2.1186	2.6812	3.5909
AR, k-fold	0.8814	0.7841	0.7928	1.0016	0.8466	0.7457	0.7624	1.0044
AR, POOS	1.0000	1.0000	1.0454	1.0023	1.0000	1.0000	1.0443	1.0017
RF AR, k-fold	0.8703*	0.9035**	0.9596	0.9959	0.8584*	0.8949**	0.9492	0.9791
RF AR, POOS	0.8681*	0.9081**	0.9963	0.9987	0.8566*	0.8995**	0.9901	0.9883
RR AR, k-fold	0.9321**	0.9465*	0.9598	0.9948	0.9251**	0.9408*	0.9533	0.9897
RR AR, POOS	0.9458**	0.9737*	1.0237	0.9922	0.9400**	0.9709*	1.0172	0.9843
KRR AR, k-fold	0.8338*	0.7510	0.8512	1.0231	0.8165*	0.7240	0.8144	1.0041
KRR AR, POOS	0.8370*	0.7529	0.8374	0.9906	0.8198*	0.7253	0.8002	0.9765
Data rich models								
ARDI, k-fold	0.8980	0.7652*	0.7965*	0.8827	0.8646	0.7461*	0.7862*	0.8740
ARDI, POOS	0.9715	0.8239	0.8354**	0.8889	0.9715	0.8126	0.8282*	0.8785
RF ARDI, k-fold	0.8742**	0.8524***	0.8978*	0.9974	0.8646**	0.8463***	0.8850	0.9926
RF ARDI, POOS	0.8691**	0.8451***	0.8722*	0.9732	0.8621**	0.8389***	0.8612*	0.9698
RR ARDI, k-fold	0.9361**	0.8507**	0.8914**	0.9700	0.9323**	0.8420*	0.8892**	0.9601
RR ARDI, POOS	0.9265**	0.8778*	0.8237*	1.0323	0.9231**	0.8737*	0.8122*	1.0248
KRR ARDI, k-fold	0.7920**	0.7397*	0.8373	0.9014	0.7768**	0.7176	0.8017	0.8879
KRR ARDI, POOS	0.7973**	0.7558	0.8303	0.9699	0.7820**	0.7322	0.7973	0.9550
ARVIF, k-fold	0.8636	0.8976	1.0210	1.0561	0.8308	0.8625	1.0241	1.0766
ARVIF, POOS	1.0043	1.0273	1.0456	1.0539	1.0025	1.0253	1.0513	1.0744
RF ARVIF, k-fold	0.8247**	0.8344**	0.9579	0.9896	0.8108**	0.8229**	0.9487	0.9762
RF ARVIF, POOS	0.8057**	0.8368**	0.9324	1.0053	0.7906**	0.8238**	0.9190	0.9914
RR ARVIF, k-fold	0.8411**	0.8316*	0.9111*	1.0028	0.8257**	0.8174*	0.9047*	1.0062
RR ARVIF, POOS	0.8702*	0.9529	0.9332	0.9884	0.8539*	0.9432	0.9337	0.9916
KRR ARVIF, k-fold	0.8349*	0.7212*	0.8386	0.9581	0.8115*	0.6993*	0.8064	0.9449
KRR ARVIF, POOS	0.8079**	0.7630	0.8270	0.9687	0.7916**	0.7384	0.7932	0.9555
Lasso (B1), k-fold	0.8900**	0.8312**	1.0019	0.9172	0.8840**	0.8245**	1.0031	0.9167
Lasso (B2), k-fold	0.8783**	0.8895**	0.9424	0.9490	0.8701**	0.8840*	0.9406	0.9527
Lasso (B3), k-fold	0.9121	0.8024	0.9573	0.9954	0.8912*	0.7784	0.9460	0.9793
Lasso (B1), POOS	0.8674**	0.8737*	0.9621	0.9162	0.8578**	0.8647*	0.9730	0.9032
Lasso (B2), POOS	0.8676**	0.9290	0.8024	0.9373	0.8581**	0.9240	0.7784	0.9199
Lasso (B3), POOS	0.9015	0.8002	0.9941	1.0117	0.8754	0.7774	0.9851	0.9933
Ridge (B1), k-fold	0.9529	0.8661	0.9382	0.8833	0.9398	0.8540	0.9421*	0.8821
Ridge (B2), k-fold	0.9333	0.8156	0.9011	0.9304	0.9155	0.7972	0.8921	0.9284
Ridge (B3), k-fold	0.8364*	0.7075*	0.7932	0.9437	0.8067*	0.6782*	0.7661	0.9317
Ridge (B1), POOS	1.0142	0.8402**	0.8894**	0.9201	1.0069	0.8287**	0.8899**	0.9245
Ridge (B2), POOS	0.9959	0.8558	0.8038	0.9147	0.9850	0.8397	0.7744	0.9080
Ridge (B3), POOS	0.8504	0.7231	0.7984	0.9518	0.8176*	0.6934	0.7697	0.9344
ElasticNet (B1), k-fold	0.8872**	0.8204**	0.9869	0.9081	0.8814**	0.8131**	0.9861	0.9065
ElasticNet (B2), k-fold	0.8906**	0.8951*	0.8111	0.9459	0.8837**	0.8895*	0.7931	0.9478
ElasticNet (B3), k-fold	0.8944	0.8452	0.9089	0.9924	0.8717*	0.8257	0.8890	0.9764
ElasticNet (B1), POOS	0.8699**	0.8698*	0.9590	0.9201	0.8602**	0.8611*	0.9705	0.9076
ElasticNet (B2), POOS	0.8682**	0.9283	0.8111	0.9361	0.8587**	0.9236	0.7867	0.9183
ElasticNet (B3), POOS	0.8984	0.8001	0.9997	1.0006	0.8731*	0.7783	0.9901	0.9821

Short training period: starts in 2010M01

Table 13: This table shows for the variable CPI (ind_CPI_totaal) the RMSEs of all the models that have been run for the different forecasting horizons. For this table the training period starts at 2010M01. The test period is from 2015M01 until 2023M12. For this variable a first-difference log transformation is used. The models are divided in data poor and data rich-models. The first data poor model, the AR, BIC, is the benchmark model and all other RMSEs are relative to this value. The stars indicate the results of a Diebold-Mariano test compared to the benchmark model (H_0: no difference in forecasting accuracy between the two models): $*$ significance at 10% level, $**$ significance at 5% level, and $***$ significance at 1% level.

	Variable transformed				Variable backtransformed			
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Data poor models								
AR, BIC (BM)	0.0079	0.0050	0.0039	0.0027	0.9339	1.7880	2.7617	3.9178
AR, k-fold	0.9436***	0.9619	0.7857	0.9849	0.9540***	0.9725	0.7874	0.9812
AR, POOS	1.1194	1.2511	1.0947	1.0042	1.1704	1.3254	1.1443	1.0036
RF AR, k-fold	0.9346	1.0294	0.9355	0.9465	0.9624	1.0705	0.9545	0.9309
RF AR, POOS	0.9280	1.0391	0.8924	0.9725	0.9552	1.0820	0.9121	0.9630
RR AR, k-fold	1.0162	1.1284	1.0382	0.9786	1.0538	1.1856	1.0818	0.9664
RR AR, POOS	1.0099	1.1936	1.0503	0.9738	1.0459	1.2593	1.0948	0.9633
KRR AR, k-fold	0.8880***	0.8655**	0.8227	0.9593	0.9064***	0.8756**	0.8165	0.9328
KRR AR, POOS	0.8975***	0.8992*	0.8099	0.9341	0.9137***	0.9091*	0.8050	0.9098
Data rich models								
ARDI, k-fold	0.9320	0.9001	0.7981	0.8445	0.9516	0.9314	0.8267	0.8376
ARDI, POOS	1.0179	1.0666	1.0249	1.0716	1.0560	1.1193	1.0940	1.0920
RF ARDI, k-fold	0.9285	0.9786	0.8518	1.0232	0.9574	1.0185	0.8759	1.0216
RF ARDI, POOS	0.9179	0.9776	0.8666	1.0315	0.9463	1.0179	0.8899	1.0412
RR ARDI, k-fold	0.9339	1.0197	0.8054	0.9889	0.9629	1.0668	0.8323	1.0020
RR ARDI, POOS	0.9704	1.0228	0.8733	1.0399	1.0007	1.0668	0.9026	1.0625
KRR ARDI, k-fold	0.8605***	0.7985**	0.7109	0.9385	0.8817***	0.8110**	0.7084	0.9109
KRR ARDI, POOS	0.8709***	0.8431**	0.7669	0.9471	0.8945***	0.8565**	0.7667	0.9234
ARVIF, k-fold	1.1052	0.9756	0.9267	0.9961	1.1392	0.9929	0.9675	0.9831
ARVIF, POOS	1.1409	1.1642	0.9628	1.0100	1.1645	1.2257	1.0047	0.9958
RF ARVIF, k-fold	0.9215	0.9566	0.8438	1.0082	0.9532	0.9897	0.8641	1.0227
RF ARVIF, POOS	0.9219	0.9333	0.8092	1.0389	0.9535	0.9658	0.8237	1.0601
RR ARVIF, k-fold	0.9445	1.0497	0.8360	1.1437	0.9692	1.0917	0.8709	1.1693
RR ARVIF, POOS	0.9609	1.1086	0.8110	1.1162	0.9792	1.1604	0.8338	1.1379
KRR ARVIF, k-fold	0.8705***	0.8203***	0.8260	0.8647	0.8880***	0.8326**	0.8279	0.8464
KRR ARVIF, POOS	0.8811***	0.7954**	0.8420	0.9576	0.8985***	0.8085**	0.8420	0.9323
Lasso (B1), k-fold	0.9558	0.9288	0.8125	0.8358	0.9902	0.9695	0.8564	0.8476
Lasso (B2), k-fold	0.9346	0.9955	0.8123	0.9720	0.9660	1.0403	0.8334	0.9694
Lasso (B3), k-fold	0.9185**	0.9038	0.8370	1.0543	0.9376	0.9267	0.8470	1.0579
Lasso (B1), POOS	0.9615	0.8741	0.8157	0.9371	0.9960	0.9037	0.8562	0.9406
Lasso (B2), POOS	0.9094	0.8947	0.7832	0.9735	0.9368	0.9249	0.8036	0.9520
Lasso (B3), POOS	0.9059**	0.8952	0.8019	0.9595	0.9230*	0.9217	0.8192	0.9409
Ridge (B1), k-fold	1.0402	0.8994	0.8267	0.9128	1.0643	0.9263	0.8671	0.9176
Ridge (B2), k-fold	0.9497	0.9605	0.8799	0.9655	0.9607	0.9837	0.9023	0.9552
Ridge (B3), k-fold	0.9442**	0.9129**	0.8037	0.9412	0.9475**	0.9158**	0.7921	0.9194
Ridge (B1), POOS	1.1386	0.9263	0.8274	0.8644	1.1674	0.9547	0.8666	0.8742
Ridge (B2), POOS	0.9745	0.9623	0.8562	0.9462	0.9884	0.9874	0.8771	0.9328
Ridge (B3), POOS	0.9469**	0.9094**	0.8025	0.9436	0.9509**	0.9118**	0.7911	0.9218
ElasticNet (B1), k-fold	0.9512	0.9353	0.8336	0.8603	0.9846	0.9735	0.8780	0.8733
ElasticNet (B2), k-fold	0.9206	0.9809	0.9521	0.9403	0.9504	1.0238	1.0000	0.9300
ElasticNet (B3), k-fold	0.9112**	0.9013	0.7682	1.0376	0.9294*	0.9265	0.7739	1.0364
ElasticNet (B1), POOS	0.9870	0.9277	0.8161	0.9401	1.0238	0.9659	0.8562	0.9434
ElasticNet (B2), POOS	0.9104	0.9899	0.8531	0.9304	0.9376	1.0352	0.8779	0.9053
ElasticNet (B3), POOS	0.9096**	0.8565	0.7799	0.9400	0.9277	0.8788	0.7943	0.9192

G.3 Deposit interest rate

Maximum training period: for deposit interest rate this starts in 2002M01

Table 14: This table shows for the variable deposit interest rate (rwk_depositorente) the RMSEs of all the models that have been run for the different forecasting horizons. For this table the training period starts at 2002M01. The test period is from 2015M01 until 2023M12. For this variable a first-difference level transformation is used. The models are divided in data poor and data rich-models. The first data poor model, the AR, BIC, is the benchmark model and all other RMSEs are relative to this value. The stars indicate the results of a Diebold-Mariano test compared to the benchmark model (H_0: no difference in forecasting accuracy between the two models): $*$ significance at 10% level, $**$ significance at 5% level, and $***$ significance at 1% level.

	Variable transformed				Variable backtransformed			
Models	$h=1$	$h=3$	$h=6$	$h=12$	$h=1$	$h=3$	$h=6$	$h=12$
Data poor models								
AR, BIC (BM)	0.0741	0.0727	0.0948	0.0958	0.0741	0.2181	0.5689	1.1500
AR, k-fold	1.0183	1.0000	0.9706	1.0000	1.0183	1.0000	0.9706	1.0000
AR, POOS	1.1156*	1.0013	0.9710	1.0004	1.1156*	1.0013	0.9710	1.0004
RF AR, k-fold	1.1015	1.2775	1.0034	0.9590	1.1015	1.2775	1.0034	0.9590
RF AR, POOS	1.0687	1.2862	1.0518	1.0680	1.0687	1.2862	1.0518	1.0680
RR AR, k-fold	0.9956	1.0619	1.0011	1.0517	0.9956	1.0619	1.0011	1.0517
RR AR, POOS	1.0154	1.0591	1.0005	1.0089	1.0154	1.0591	1.0005	1.0089
KRR AR, k-fold	1.3277*	1.3667**	1.0599	0.9179	1.3277*	1.3667**	1.0599	0.9179
KRR AR, POOS	2.4969	1.3476*	1.0863	0.9571	2.4969	1.3476*	1.0863	0.9571
Data rich models								
ARDI, k-fold	0.9953	0.9987	1.2558*	1.7312	0.9953	0.9987	1.2558*	1.7312
ARDI, POOS	1.1032	1.1446*	1.1935*	1.7025	1.1032	1.1446*	1.1935*	1.7025
RF ARDI, k-fold	1.1800	1.5940	1.8725	1.0078	1.1800	1.5940	1.8725	1.0078
RF ARDI, POOS	1.2130	2.0552	1.8024	1.0418	1.2130	2.0552	1.8024	1.0418
RR ARDI, k-fold	0.9609	1.0707	1.2107**	1.6185	0.9609	1.0707	1.2107**	1.6185
RR ARDI, POOS	1.0166	1.1426	1.1674*	1.6153	1.0166	1.1426	1.1674*	1.6153
KRR ARDI, k-fold	1.6562**	1.7111**	1.3364**	1.3311	1.6562**	1.7111**	1.3364**	1.3311
KRR ARDI, POOS	1.5688**	1.6191*	1.1990	1.0690	1.5688**	1.6191*	1.1990	1.0690
ARVIF, k-fold	1.1092	1.0906	1.0931	0.8004	1.1092	1.0906	1.0931	0.8004
ARVIF, POOS	1.0996	1.1007	1.1065	0.8060	1.0996	1.1007	1.1065	0.8060
RF ARVIF, k-fold	0.9865	1.2145	1.2918*	0.6892	0.9865	1.2145	1.2918*	0.6892
RF ARVIF, POOS	1.0175	1.2352	1.3020*	0.7003	1.0175	1.2352	1.3020*	0.7003
RR ARVIF, k-fold	1.0042	1.0912	0.9692	0.8758	1.0042	1.0912	0.9692	0.8758
RR ARVIF, POOS	1.0086	1.1012	1.1421	0.8749	1.0086	1.1012	1.1421	0.8749
KRR ARVIF, k-fold	1.4465**	1.4812*	1.2291*	1.0550	1.4465**	1.4812*	1.2291*	1.0550
KRR ARVIF, POOS	1.6196**	1.6177*	1.2000	1.0581	1.6196**	1.6177*	1.2000	1.0581
Lasso (B1), k-fold	1.0154	1.2588*	1.4489**	1.1281	1.0154	1.2588*	1.4489**	1.1281
Lasso (B2), k-fold	1.0943	1.2025**	1.4003***	0.9027	1.0943	1.2025**	1.4003***	0.9027
Lasso (B3), k-fold	1.0921	1.4592***	1.3133**	1.1662***	1.0921	1.4592***	1.3133**	1.1662***
Lasso (B1), POOS	1.0567	1.0871	1.1024	1.0594	1.0567	1.0871	1.1024	1.0594
Lasso (B2), POOS	1.2153	1.5386*	1.2070	1.0637	1.2153	1.5386*	1.2070	1.0637
Lasso (B3), POOS	1.5488**	1.5855*	1.2113	1.1004	1.5488**	1.5855*	1.2113	1.1004
Ridge (B1), k-fold	1.0096	1.3351**	1.3448**	0.8641	1.0096	1.3351**	1.3448**	0.8641
Ridge (B2), k-fold	1.1573	1.4712*	1.3281***	1.0647	1.1573	1.4712*	1.3281***	1.0647
Ridge (B3), k-fold	1.4554**	1.3824**	1.2145*	1.1063*	1.4554**	1.3824**	1.2145*	1.1063*
Ridge (B1), POOS	0.9950	1.1312*	1.2815**	1.2414	0.9950	1.1312*	1.2815**	1.2414
Ridge (B2), POOS	1.1510	1.2885*	1.1389**	1.1128	1.1510	1.2885*	1.1389**	1.1128
Ridge (B3), POOS	1.6147**	1.5969*	1.2307	1.1161	1.6147**	1.5969*	1.2307	1.1161
ElasticNet (B1), k-fold	1.0158	1.3484*	1.4234***	1.1050	1.0158	1.3484*	1.4234***	1.1050
ElasticNet (B2), k-fold	1.0632	1.1530*	1.3555***	0.8463	1.0632	1.1530*	1.3555***	0.8463
ElasticNet (B3), k-fold	1.0889	1.2763**	1.3275**	1.1672*	1.0889	1.2763**	1.3275**	1.1672*
ElasticNet (B1), POOS	1.1222	1.1544	1.0890	1.1948	1.1222	1.1544	1.0890	1.1948
ElasticNet (B2), POOS	1.1870	1.5322*	1.2133	1.0807	1.1870	1.5322*	1.2133	1.0807
ElasticNet (B3), POOS	1.5227*	1.5634*	1.2106	1.0870	1.5227*	1.5634*	1.2106	1.0870

Short training period: starts in 2010M01

Table 15: This table shows for the variable deposit interest rate (rwk_depositorente) the RMSEs of all the models that have been run for the different forecasting horizons. For this table the training period starts at 2010M01. The test period is from 2015M01 until 2023M12. For this variable a first-difference level transformation is used. The models are divided in data poor and data rich-models. The first data poor model, the AR, BIC, is the benchmark model and all other RMSEs are relative to this value. The stars indicate the results of a Diebold-Mariano test compared to the benchmark model (H_0: no difference in forecasting accuracy between the two models): $*$ significance at 10% level, $**$ significance at 5% level, and $***$ significance at 1% level.

	Variable transformed				Variable backtransformed			
Models	$h=1$	$h=3$	$h=6$	$h=12$	$h=1$	$h=3$	$h=6$	$h=12$
Data poor models								
AR, BIC (BM)	0.0964	0.0886	0.1039	0.0950	0.0964	0.2658	0.6235	1.1407
AR, k-fold	0.8857	1.0128	0.9999	1.0000	0.8857	1.0128	0.9999	1.0000
AR, POOS	1.0603	1.0129	1.0089	1.0000	1.0603	1.0129	1.0089	1.0000
RF AR, k-fold	0.8385	1.0297	1.1103	1.0677	0.8385	1.0297	1.1103	1.0677
RF AR, POOS	0.8265	1.0598	1.1337	1.0664	0.8265	1.0598	1.1337	1.0664
RR AR, k-fold	0.8598	0.9173	1.0113	1.0368	0.8598	0.9173	1.0113	1.0368
RR AR, POOS	0.9015	0.9798	1.0032	1.0203	0.9015	0.9798	1.0032	1.0203
KRR AR, k-fold	1.2769	1.2313*	1.0499	1.1349	1.2769	1.2313*	1.0499	1.1349
KRR AR, POOS	1.4553	1.2341	1.0650	1.0800	1.4553	1.2341	1.0650	1.0800
Data rich models								
ARDI, k-fold	0.8370	0.7472*	0.7639	0.8397	0.8370	0.7472*	0.7639	0.8397
ARDI, POOS	0.9622	0.8529	0.8519	0.8587	0.9622	0.8529	0.8519	0.8587
RF ARDI, k-fold	0.7852	0.9366	0.9574	0.9925	0.7852	0.9366	0.9574	0.9925
RF ARDI, POOS	0.8822	0.9604	0.8956	1.0002	0.8822	0.9604	0.8956	1.0002
RR ARDI, k-fold	0.7973	0.7694	0.8754	0.9037	0.7973	0.7694	0.8754	0.9037
RR ARDI, POOS	0.8340	0.9252	0.9164	0.9609	0.8340	0.9252	0.9164	0.9609
KRR ARDI, k-fold	1.2357	1.2676*	1.0706	1.0071	1.2357	1.2676*	1.0706	1.0071
KRR ARDI, POOS	1.1344	1.2856*	1.0506	1.0227	1.1344	1.2856*	1.0506	1.0227
ARVIF, k-fold	0.8485	0.7856	1.0326	1.0093	0.8485	0.7856	1.0326	1.0093
ARVIF, POOS	0.9874	0.7648	1.0015	1.0056	0.9874	0.7648	1.0015	1.0056
RF ARVIF, k-fold	0.7047	0.7759*	0.9437	0.9895	0.7047	0.7759*	0.9437	0.9895
RF ARVIF, POOS	0.7135	0.8363	0.9385	0.9454	0.7135	0.8363	0.9385	0.9454
RR ARVIF, k-fold	0.6802*	0.8417	0.9490	0.9821	0.6802*	0.8417	0.9490	0.9821
RR ARVIF, POOS	0.7136	0.8467	0.8885	0.9756	0.7136	0.8467	0.8885	0.9756
KRR ARVIF, k-fold	1.1397	1.2599*	1.0504	1.0270	1.1397	1.2599*	1.0504	1.0270
KRR ARVIF, POOS	1.1753	1.2730*	1.0524	1.0237	1.1753	1.2730*	1.0524	1.0237
Lasso (B1), k-fold	0.8553	0.7909*	0.7766	1.0669	0.8553	0.7909*	0.7766	1.0669
Lasso (B2), k-fold	0.8881	1.0668	0.9232	0.9397	0.8881	1.0668	0.9232	0.9397
Lasso (B3), k-fold	0.9368	1.0240	0.9151	1.0190	0.9368	1.0240	0.9151	1.0190
Lasso (B1), POOS	0.8205	0.8988	0.9464	0.9705	0.8205	0.8988	0.9464	0.9705
Lasso (B2), POOS	0.8702	1.2779*	1.0080	1.0236	0.8702	1.2779*	1.0080	1.0236
Lasso (B3), POOS	0.9993	1.0863	0.9378	1.0160	0.9993	1.0863	0.9378	1.0160
Ridge (B1), k-fold	0.7695	0.7714*	0.7939	1.0972	0.7695	0.7714*	0.7939	1.0972
Ridge (B2), k-fold	0.9061	1.0198	1.0101	0.9997	0.9061	1.0198	1.0101	0.9997
Ridge (B3), k-fold	1.2268	1.2517*	1.0232	1.0194	1.2268	1.2517*	1.0232	1.0194
Ridge (B1), POOS	0.7691	0.7408*	0.7944	1.0350	0.7691	0.7408*	0.7944	1.0350
Ridge (B2), POOS	0.9106	1.0280	1.0092	1.0003	0.9106	1.0280	1.0092	1.0003
Ridge (B3), POOS	1.2385	1.2660*	1.0404	1.0137	1.2385	1.2660*	1.0404	1.0137
ElasticNet (B1), k-fold	0.8406	0.7631*	0.8749	1.0189	0.8406	0.7631*	0.8749	1.0189
ElasticNet (B2), k-fold	0.8765	1.0924	1.0213	1.0006	0.8765	1.0924	1.0213	1.0006
ElasticNet (B3), k-fold	0.9895	1.0562	0.9231	1.0179	0.9895	1.0562	0.9231	1.0179
ElasticNet (B1), POOS	0.8185	0.9035	0.9149	0.9783	0.8185	0.9035	0.9149	0.9783
ElasticNet (B2), POOS	0.8828	1.1276	1.0092	1.0202	0.8828	1.1276	1.0092	1.0202
ElasticNet (B3), POOS	1.0755	1.2661*	0.9628	1.0088	1.0755	1.2661*	0.9628	1.0088

G.4 Export (value index)

Maximum training period: for export (value index) this starts in 1998M01

Table 16: This table shows for the variable export (value index) (int_uitvoer_waarde_index) the RMSEs of all the models that have been run for the different forecasting horizons. For this table the training period starts at 1998M01. The test period is from 2015M01 until 2023M12. For this variable a first-difference log transformation is used. The models are divided in data poor and data rich-models. The first data poor model, the AR, BIC, is the benchmark model and all other RMSEs are relative to this value. The stars indicate the results of a Diebold-Mariano test compared to the benchmark model (H_0: no difference in forecasting accuracy between the two models): $*$ significance at 10% level, $**$ significance at 5% level, and $***$ significance at 1% level.

	Variable transformed				Variable backtransformed			
Models	$h=1$	$h=3$	$h=6$	$h=12$	$h=1$	$h=3$	$h=6$	$h=12$
Data poor models								
AR, BIC (BM)	0.0371	0.0219	0.0139	0.0093	4.7688	8.4512	11.170	15.323
AR, k-fold	1	1	0.9803	0.9540	1	1	0.9825	0.9339
AR, POOS	1	1	0.9644	0.9901	1	1	0.9502	0.9692
RF AR, k-fold	0.9859	0.8983*	0.9834	0.9452	1.0032	0.9192	1.0242	0.8797
RF AR, POOS	0.9835	0.8911*	0.9644	1.1479*	0.9995	0.9129	1.0055	1.1309
RR AR, k-fold	0.9874	0.9882	0.9682	0.9320	0.9918	0.9874	0.9737	0.9150
RR AR, POOS	0.9870	0.9908	0.9881	0.9820	0.9914	0.9905	0.9866	0.9639
KRR AR, k-fold	0.9277	0.9617	0.9784	1.1023	0.9357	0.9906	1.0175	1.0256
KRR AR, POOS	0.9339	1.0171	0.9585	1.0785	0.9447	1.0512	0.9605	1.1097
Data rich models								
ARDI, k-fold	0.9744	0.9436	1.1228	1.0771	0.9560	0.9282	1.1567	1.0306
ARDI, POOS	1.0636	0.9727	1.3885**	1.0993	1.1396	0.9559	1.3932*	1.0574
RF ARDI, k-fold	1.0081	0.9410	1.0572	0.9499	1.0169	0.9549	1.0625	0.8761
RF ARDI, POOS	0.9858	0.9307	0.9919	0.9804	1.0025	0.9444	0.9960	0.9346
RR ARDI, k-fold	1.0399	1.0466	1.1054	1.1220	1.1052	1.0355	1.0559	1.0445
RR ARDI, POOS	0.9662	1.0188	1.1059	1.0686	0.9729	1.0162	1.0654	1.0409
KRR ARDI, k-fold	1.1161	1.0448	1.0773	1.0410	1.2246*	1.1012	1.1231	1.0109
KRR ARDI, POOS	1.1855*	1.0294	0.9744	0.9244	1.3067**	1.0921	0.9839	0.9204
ARVIF, k-fold	0.9060	1.0178	1.0282	1.0654	0.8989	1.0562	1.0948	1.0438
ARVIF, POOS	0.9060	1.0289	1.0287	1.0786	0.8989	1.0709	1.1102	1.0685
RF ARVIF, k-fold	0.9557	0.9754	0.9611	0.8895	0.9687	1.0075	0.9657	0.8416
RF ARVIF, POOS	0.9478	0.9186	0.9934	0.9978	0.9643	0.9351	1.0034	0.9609
RR ARVIF, k-fold	0.9428	0.9734	1.0075	1.0321	0.9560	0.9789	0.9833	0.9672
RR ARVIF, POOS	0.9169	0.9559	1.0439	0.9934	0.9087	0.9754	1.0180	0.9688
KRR ARVIF, k-fold	1.1293	1.0359	0.9560	1.0265	1.2423**	1.1228	0.9719	0.9823
KRR ARVIF, POOS	1.0606	1.0718	0.9957	1.0464	1.1383	1.1507	1.0033	0.9895
Lasso (B1), k-fold	0.9976	0.9989	1.2338	0.9982	0.9770	0.9792	1.2271	0.9664
Lasso (B2), k-fold	1.0492	0.9983	1.1706	0.9669	1.0511	0.9832	1.2862	1.0523
Lasso (B3), k-fold	0.9798	1.0251	1.0858*	1.0353	0.9610	1.0433	1.0960*	1.0561
Lasso (B1), POOS	1.0091	0.9788	1.3275*	0.9361	0.9846	0.9704	1.3369*	0.9071
Lasso (B2), POOS	1.0310	1.0371	1.1181	1.0552	1.0204	1.0237	1.2039	1.1220
Lasso (B3), POOS	1.0132	1.0568	1.0218	0.9951	1.0081	1.0773	0.9962	0.9722
Ridge (B1), k-fold	1.0148	0.9957	1.2487	1.1006	0.9905	0.9917	1.2670	1.0567
Ridge (B2), k-fold	1.0777	1.0106	1.0765	0.8518	1.0529	1.0074	1.1424	0.8607
Ridge (B3), k-fold	1.1749**	1.0918	1.0690	0.9497	1.1664**	1.1226	1.1062	0.9614
Ridge (B1), POOS	1.0049	0.9862	1.2355	1.0925	0.9829	0.9856	1.2688	1.0511
Ridge (B2), POOS	1.0542	1.0124	1.0105	0.9720	1.0394	1.0115	1.0788	1.0211
Ridge (B3), POOS	1.1724**	1.0951	1.0603	1.0003	1.1664**	1.1191	1.1052	1.0167
ElasticNet (B1), k-fold	0.9823	0.9946	1.2783	1.1112	0.9624	0.9926	1.2939	1.0750
ElasticNet (B2), k-fold	1.0401	1.0294	1.0859	0.9139	1.0326	1.0114	1.1730	0.9403
ElasticNet (B3), k-fold	1.0386	0.9697	1.0676	1.0187	1.0555	1.0020	1.0746	1.0140
ElasticNet (B1), POOS	0.9992	0.9818	1.2875*	0.9773	0.9751	0.9769	1.3012	0.9370
ElasticNet (B2), POOS	1.0556	1.0328	1.0505	1.0515	1.0505	1.0198	1.1090	1.1183
ElasticNet (B3), POOS	1.0180	1.0512	1.0135	1.0018	1.0148	1.0719	0.9898	0.9777

Short training period: starts in 2010M01

Table 17: This table shows for the variable export (value index) (int_uitvoer_waarde_index) the RMSEs of all the models that have been run for the different forecasting horizons. For this table the training period starts at 2010M01. The test period is from 2015M01 until 2023M12. For this variable a first-difference log transformation is used. The models are divided in data poor and data rich-models. The first data poor model, the AR, BIC, is the benchmark model and all other RMSEs are relative to this value. The stars indicate the results of a Diebold-Mariano test compared to the benchmark model (H_0: no difference in forecasting accuracy between the two models): $*$ significance at 10% level, $**$ significance at 5% level, and $***$ significance at 1% level.

	Variable transformed				Variable backtransformed			
Models	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Data poor models								
AR, BIC (BM)	0.0387	0.0232	0.0146	0.0097	5.0071	9.0048	11.965	16.192
AR, k-fold	1.0000	0.9521	0.9806	0.9619	1.0000	0.9650	0.9865	0.9495
AR, POOS	1.0000	1.0259	0.9808	0.9656	1.0000	1.0526	0.9870	0.9518
RF AR, k-fold	0.9910	0.9211	0.9988	0.8660	0.9923	0.9563	1.0677	0.9179
RF AR, POOS	0.9922	0.9042	0.9371	0.8701	0.9922	0.9284	0.9932	0.9230
RR AR, k-fold	0.9843	0.9723	0.9723	0.9303	0.9876	0.9821	0.9689	0.9137
RR AR, POOS	0.9828	0.9856	0.9664	0.9499	0.9985	0.9956	0.9706	0.9348
KRR AR, k-fold	1.0469	0.9076	0.9872	0.7568*	1.0426	0.9319	1.0541	0.7452*
KRR AR, POOS	0.9430	1.0056	1.0524	0.8837	0.9511	1.1133	1.1304	0.8483
Data rich models								
ARDI, k-fold	1.1918***	0.9436	1.1793	1.1674**	1.1026	0.9417	1.1040	1.1198*
ARDI, POOS	1.1669***	1.0385	1.1564	1.1383	1.0672	1.0000	1.0946	1.1080
RF ARDI, k-fold	1.0157	0.9484	1.0661	0.9681	1.0251	0.9694	1.0837	0.9920
RF ARDI, POOS	1.0053	0.9695	1.0473	0.9662	1.0135	1.0102	1.1145	1.0205
RR ARDI, k-fold	1.0803	1.1094	1.1007	1.4198*	1.0369	1.0772	0.9870	1.3438
RR ARDI, POOS	1.1196	1.1648	1.1906	0.9888	1.0654	1.1395	1.1286	0.9866
KRR ARDI, k-fold	1.1958**	1.0735	1.0368	0.9842	1.2634**	1.1139	1.0044	0.9215
KRR ARDI, POOS	1.1899**	1.1967	0.9144	1.0883	1.2616**	1.2696	0.9170	1.0479
ARVIF, k-fold	1.0429	1.0730	0.9726	1.2271*	1.0603	1.0975	0.9776	1.1947
ARVIF, k-fold	1.3722**	1.0287	1.0891	1.3611*	1.4405**	1.0487	0.9875	1.3132*
ARVIF, POOS	1.3985**	1.1221	1.1662*	1.4055**	1.4641**	1.1731	1.1194	1.4111*
RF ARVIF, k-fold	0.9721	0.9276	1.0710	1.2678	0.9752	0.9540	1.1129	1.3524
RF ARVIF, POOS	0.9776	0.9406	1.0536	1.1082	0.9825	0.9722	1.0737	1.0857
RR ARVIF, k-fold	1.0808	1.1141*	1.0820	1.2064	1.1006	1.1953*	1.0589	1.1600
RR ARVIF, POOS	1.0396	1.1043*	1.1869	1.0940	1.0602	1.1609*	1.1418	1.0605
KRR ARVIF, k-fold	1.1418	1.0563	0.9458	0.9679	1.1802*	1.1053	0.9215	0.9673
KRR ARVIF, POOS	1.1262	1.0180	0.9235	0.9852	1.1899*	1.0859	0.8998	0.9756
Lasso (B1), k-fold	1.0170	1.1272	1.2122	1.2582	1.0177	1.0862	1.1947	1.2416
Lasso (B2), k-fold	1.0205	0.9934	1.0251	1.0289	1.0434	0.9786	1.0083	1.0479
Lasso (B3), k-fold	1.0984	0.9436	0.9555	1.0963	1.1071	0.9510	0.9065	1.1090
Lasso (B1), POOS	1.0102	1.0881	1.2806**	1.2221*	1.0145	1.0892	1.2378	1.2700*
Lasso (B2), POOS	1.0075	0.9967	1.0554	0.9847	1.0316	0.9736	1.0414	0.9972
Lasso (B3), POOS	1.1423*	0.9955	1.0672	1.0345	1.1480	1.0114	1.0338	1.0529
Ridge (B1), k-fold	1.4296***	1.0700	1.4624*	1.3884	1.4469***	1.0749	1.4576	1.3530
Ridge (B2), k-fold	1.3831***	1.0174	1.0145	0.9880	1.3714***	0.9977	1.0130	1.0049
Ridge (B3), k-fold	1.6524***	1.1592*	1.0290	0.9563	1.6511***	1.2016*	1.0062	0.9480
Ridge (B1), POOS	1.3311***	1.0520	1.3494	1.3742	1.3051***	1.0507	1.3545	1.3401
Ridge (B2), POOS	1.2953***	1.0337	0.9881	1.0136	1.2527***	0.9828	0.9682	1.0520
Ridge (B3), POOS	1.6577***	1.1663*	1.0463	0.9503	1.6527***	1.2081*	1.0359	0.9414
ElasticNet (B1), k-fold	1.0155	1.0650	1.2860	1.1975	1.0258	1.0216	1.2746	1.2456
ElasticNet (B2), k-fold	1.0196	0.9920	1.0152	1.0323	1.0433	0.9758	1.0055	1.0551
ElasticNet (B3), k-fold	1.1286	0.9554	0.9709	1.1193	1.1183	0.9600	0.9510	1.1282
ElasticNet (B1), POOS	1.0133	1.0380	1.2614*	1.2240*	1.0213	1.0041	1.2380	1.2714*
ElasticNet (B2), POOS	1.0131	0.9810	1.0204	1.0102	1.0396	0.9569	1.0104	1.0420
ElasticNet (B3), POOS	1.1393*	1.0016	1.0662	1.0289	1.1440	1.0155	1.0335	1.0506

G.5 Industry (production index)

Maximum training period: for industry (production index) this starts in 2008M01

Table 18: This table shows for the variable industry (production index) (prd_productie_index_industrie) the RMSEs of all the models that have been run for the different forecasting horizons. For this table the training period starts at 2008M01. The test period is from 2015M01 until 2023M12. For this variable a first-difference log transformation is used. The models are divided in data poor and data rich-models. The first data poor model, the AR, BIC, is the benchmark model and all other RMSEs are relative to this value. The stars indicate the results of a Diebold-Mariano test compared to the benchmark model (H_0: no difference in forecasting accuracy between the two models): $*$ significance at 10% level, $**$ significance at 5% level, and $***$ significance at 1% level.

	Variable transformed				Variable backtransformed			
Models	$h=1$	$h=3$	$h=6$	$h=12$	$h=1$	$h=3$	$h=6$	$h=12$
Data poor models								
AR, BIC (BM)	0.0192	0.0102	0.0068	0.0039	2.0872	3.2978	4.4787	5.1523
AR, k-fold	1.0000	0.9948*	0.9762	1.0000	1.0000	0.9943*	0.9735	1.0000
AR, POOS	1.0028	1.0052	1.0093	1.1438*	1.0025	1.0061	1.0114	1.1391*
RF AR, k-fold	1.0246	0.9596	0.9851	1.0320	1.0223	0.9541	0.9664	1.0416
RF AR, POOS	0.9872	0.9321*	1.0316	1.1563	0.9875	0.9255*	1.0165	1.1907
RR AR, k-fold	0.9965	0.9984	0.9826	1.0015	0.9964	0.9976	0.9788	0.9985
RR AR, POOS	1.0044	0.9975	0.9839	1.1309*	1.0044	0.9966	0.9806	1.1292*
KRR AR, k-fold	1.0645*	0.9737	1.0478	1.2296	1.0654*	0.9769	1.0435	1.2542
KRR AR, POOS	1.0392	0.9047*	0.9117	1.4448	1.0390	0.9052*	0.9155	1.4868
Data rich models								
ARDI, k-fold	0.9624	0.9812	1.0467	1.3264	0.9611	0.9722	1.0377	1.3174
ARDI, POOS	0.9822	1.0539	1.0450	1.7103	0.9825	1.0488	1.0392	1.6155
RF ARDI, k-fold	1.0605	1.0212	0.9212	0.9613	1.0619	1.0277	0.9106	0.9661
RF ARDI, POOS	1.0382	0.9853	0.8985	0.9812	1.0417	0.9962	0.8954	0.9873
RR ARDI, k-fold	0.9983	1.0506	1.2179*	1.2184	1.0018	1.0475	1.2024*	1.2138
RR ARDI, POOS	0.9789	1.0908	1.3437	1.4102	0.9760	1.0968	1.3218	1.3740
KRR ARDI, k-fold	1.0334	0.9767	1.0691	1.2078	1.0329	0.9807	1.0622	1.2241
KRR ARDI, POOS	1.0212	0.9215	0.9092**	1.2235	1.0245	0.9322	0.9140**	1.2478
ARVIF, k-fold	1.0753	1.0699	1.0155	1.1818**	1.0851	1.0800	0.9919	1.1601*
ARVIF, POOS	1.0533	1.0798	1.0574	1.1915**	1.0656	1.0933	1.0271	1.1682**
RF ARVIF, k-fold	0.9683	0.9726	1.1070	1.1155	0.9704	0.9719	1.0786	1.1191
RF ARVIF, POOS	0.9930	0.9890	1.1010	1.1247	0.9925	0.9890	1.0741	1.1298
RR ARVIF, k-fold	1.0332	1.0908	1.0340	1.0509	1.0306	1.0999	1.0053	1.0424
RR ARVIF, POOS	1.0856	1.4265**	1.2326	1.0040	1.0864*	1.4607**	1.1935	0.9973
KRR ARVIF, k-fold	1.0340	0.9452	0.9897	1.3252*	1.0347	0.9500	0.9909	1.3465*
KRR ARVIF, POOS	1.0191	0.9719	0.9667	1.3854	1.0217	0.9749	0.9687	1.3815
Lasso (B1), k-fold	0.9859	1.0264	1.3397**	1.8444**	0.9900	1.0257	1.3225**	1.8353**
Lasso (B2), k-fold	0.9812	1.0650*	1.0594	1.3524	0.9857	1.0608*	1.0417	1.3884
Lasso (B3), k-fold	1.0078	0.9777	0.9534	1.2312	1.0153	0.9889	0.9509	1.2430
Lasso (B1), POOS	0.9924	0.9728	0.9697	1.2443	0.9971	0.9810	0.9718	1.2409
Lasso (B2), POOS	0.9807	0.9639	0.9729	1.1339	0.9852	0.9720	0.9680	1.1414
Lasso (B3), POOS	1.0278	1.0246	0.9117*	1.0335	1.0310	1.0393	0.9074*	1.0348
Ridge (B1), k-fold	1.0886	1.0526	1.2994*	1.3734	1.1043	1.0602	1.2963*	1.3695
Ridge (B2), k-fold	1.1886***	1.0398	1.0447	1.1864	1.2090***	1.0457	1.0404	1.2043
Ridge (B3), k-fold	0.9825	0.9642	0.9338**	1.1860	0.9882	0.9729	0.9300**	1.2016
Ridge (B1), POOS	1.0939	1.0674	1.2326	1.6609	1.1091	1.0760	1.2338	1.6249
Ridge (B2), POOS	1.1896***	1.0492	1.0311	1.1035	1.2094***	1.0546	1.0259	1.1132
Ridge (B3), POOS	0.9817	0.9635	0.9593**	1.1839	0.9861	0.9722	0.9594*	1.2014
ElasticNet (B1), k-fold	0.9776	1.0180	1.3038**	2.0161**	0.9831	1.0206	1.2845**	2.0687**
ElasticNet (B2), k-fold	1.0504*	1.0522	1.0538	1.2208*	1.0507*	1.0490	1.0388	1.2355*
ElasticNet (B3), k-fold	0.9937	0.9985	0.9746	1.1345	0.9992	1.0108	0.9702	1.1318
ElasticNet (B1), POOS	0.9845	0.9392	0.9974	1.5153	0.9894	0.9426	0.9978	1.4736
ElasticNet (B2), POOS	1.0026	0.9884	1.0060	1.1411	1.0077	0.9967	0.9964	1.1488
ElasticNet (B3), POOS	1.0203	1.0429	0.9492	1.0308	1.0263	1.0513	0.9474	1.0342

Short training period: starts in 2010M01

Table 19: This table shows for the variable industry (production index) (prd_productie_index_industrie) the RMSEs of all the models that have been run for the different forecasting horizons. For this table the training period starts at 2010M01. The test period is from 2015M01 until 2023M12. For this variable a first-difference log transformation is used. The models are divided in data poor and data rich-models. The first data poor model, the AR, BIC, is the benchmark model and all other RMSEs are relative to this value. The stars indicate the results of a Diebold-Mariano test compared to the benchmark model (H_0: no difference in forecasting accuracy between the two models): $*$ significance at 10% level, $**$ significance at 5% level, and $***$ significance at 1% level.

	Variable transformed				Variable backtransformed			
Models	$h=1$	$h=3$	$h=6$	$h=12$	$h=1$	$h=3$	$h=6$	$h=12$
Data poor models								
AR, BIC (BM)	0.0197	0.0101	0.0070	0.0039	2.1373	3.3030	4.6298	5.2716
AR, k-fold	1.0089	0.9941	0.9821	1	1.0083	0.9931	0.9809	1
AR, POOS	1.0089	1.0514	1.0249	1.0267	1.0083	1.0519	1.0259	1.0262
RF AR, k-fold	0.9832	0.9899	0.9452	1.0676	0.9853	0.9824	0.9394	1.0732
RF AR, POOS	0.9856	0.9531*	0.9740	1.1969	0.9879	0.9436*	0.9776	1.2204
RR AR, k-fold	0.9912	1.0041	0.9689	1.0181	0.9924	1.0041	0.9669	1.0204
RR AR, POOS	0.9838	1.0407	0.9918	0.9919	0.9858	1.0403	0.9889	0.9929
KRR AR, k-fold	1.0174	0.9533	1.0488	1.2468	1.0205	0.9519	1.0567	1.2811
KRR AR, POOS	1.0251	1.0099	1.0700	1.6383	1.0269	1.0184	1.0831	1.6177
Data rich models								
ARDI, k-fold	1.0068	0.9758	1.0053	1.1564*	1.0107	0.9691	0.9939	1.1545*
ARDI, POOS	1.0471	1.0737	1.1883**	1.2352	1.0520	1.0651	1.1936**	1.2343
RF ARDI, k-fold	1.0151	1.0479	0.9993	1.0809	1.0259	1.0554	0.9983	1.0766
RF ARDI, POOS	1.0074	1.1093	1.0867	1.1683	1.0136	1.1062	1.0908	1.1647
RR ARDI, k-fold	0.9927	0.9514	0.9943	1.1708	0.9986	0.9506	0.9834	1.1760
RR ARDI, POOS	1.0733	1.0126	0.9612	1.1886	1.0814	1.0155	0.9529	1.1869
KRR ARDI, k-fold	0.9974	1.0081	0.9690	1.3800*	1.0001	1.0091	0.9668	1.3901*
KRR ARDI, POOS	0.9763	0.9412	1.0448	1.3829**	0.9771	0.9426	1.0450	1.3949**
ARVIF, k-fold	0.9963	1.1886*	1.1215*	1.3578***	0.9935	1.1917*	1.1120*	1.3559***
ARVIF, POOS	0.9883	1.1746*	1.1159*	1.3632***	0.9866	1.1765	1.1093*	1.3565***
RF ARVIF, k-fold	0.9487	1.1256	1.1802	1.6337*	0.9539	1.1085	1.1662	1.6908
RF ARVIF, POOS	0.9350	1.1656	1.1282	1.3480	0.9374	1.1353	1.1153	1.3585
RR ARVIF, k-fold	0.9857	1.0857	1.0576	1.1286**	0.9841	1.0868	1.0520	1.1219**
RR ARVIF, POOS	0.9611	1.4995**	1.0831	1.1517**	0.9630	1.5243**	1.0812*	1.1424**
KRR ARVIF, k-fold	1.0947	0.9809	0.9404	1.3559**	1.1341	0.9826	0.9426	1.3662*
KRR ARVIF, POOS	0.9565	0.9394	1.0136	1.2896	0.9595	0.9359	1.0369	1.3078
Lasso (B1), k-fold	1.0030	1.1447*	1.2274*	1.6751*	1.0056	1.1485**	1.2080	1.6363*
Lasso (B2), k-fold	0.9887	1.0530	1.0809	1.1106	0.9997	1.0645	1.0844	1.1160
Lasso (B3), k-fold	0.9767	0.9600	0.9869	1.1950	0.9839	0.9629	0.9885	1.1923
Lasso (B1), POOS	0.9595	1.3046**	1.2065	1.1969	0.9665	1.3359**	1.1801	1.1783
Lasso (B2), POOS	0.9670	1.0459	1.0589	1.2078	0.9725	1.0575	1.0693	1.2317
Lasso (B3), POOS	1.0267	1.0100	0.9835	1.1612	1.0315	1.0132	0.9824	1.1707
Ridge (B1), k-fold	1.0778	1.3590	1.1360	1.5862	1.0842	1.3493	1.1312	1.5425
Ridge (B2), k-fold	1.0827*	0.9780	0.9644	1.0994	1.0953**	0.9814	0.9642	1.1144
Ridge (B3), k-fold	0.9555	0.9675	0.9207**	1.1913	0.9625	0.9686	0.9176**	1.2099
Ridge (B1), POOS	1.0818	1.1913**	1.0977	1.6475	1.0884	1.1806*	1.0963	1.5865
Ridge (B2), POOS	1.0778*	1.1330	1.0546	1.1542	1.0903**	1.1530	1.0522	1.1844
Ridge (B3), POOS	0.9510	0.9673	0.9245**	1.1936	0.9577	0.9695	0.9217**	1.2115
ElasticNet (B1), k-fold	1.0305	1.2545	1.1303	1.5744	1.0301	1.2421	1.1144	1.5331
ElasticNet (B2), k-fold	0.9760	1.0198	0.9772	1.0623	0.9848	1.0259	0.9748	1.0676
ElasticNet (B3), k-fold	0.9686	0.9741	0.9990	1.0990	0.9764	0.9735	1.0039	1.1002
ElasticNet (B1), POOS	0.9637	1.3034**	1.2174	1.4766*	0.9703	1.3356**	1.1933	1.4511*
ElasticNet (B2), POOS	0.9699	1.1757*	1.0578	1.2173	0.9756	1.2007**	1.0535	1.2430
ElasticNet (B3), POOS	1.0458	1.0084	0.9797	1.2086	1.0501	1.0108	0.9780	1.2148

G.6 Number of sold existing homes

Maximum training period: for number of sold existing homes this starts in 1998M01

Table 20: This table shows for the variable number of sold existing homes (wnm_aantal_verkochte_koopwoningen_bestaand) the RMSEs of all the models that have been run for the different forecasting horizons. For this table the training period starts at 1998M01. The test period is from 2015M01 until 2023M12. For this variable a first-difference log transformation is used. The models are divided in data poor and data rich-models. The first data poor model, the AR, BIC, is the benchmark model and all other RMSEs are relative to this value. The stars indicate the results of a Diebold-Mariano test compared to the benchmark model (H_0: no difference in forecasting accuracy between the two models): $*$ significance at 10% level, $**$ significance at 5% level, and $***$ significance at 1% level.

	Variable transformed				Variable backtransformed			
Models	$h=1$	$h=3$	$h=6$	$h=12$	$h=1$	$h=3$	$h=6$	$h=12$
Data poor models								
AR, BIC (BM)	0.1108	0.0389	0.0240	0.0155	1965.8	2154.7	2621.6	3174.4
AR, k-fold	1.0000	1.0000	0.9575	0.9673	1.0000	1.0000	0.9777	0.9855
AR, POOS	1.0000	1.1119*	1.0549**	1.0100	1.0000	1.0693	1.0400*	1.0177
RF AR, k-fold	1.0150	1.0245	1.0103	1.0109	1.0686	0.9783	1.0032	1.0740
RF AR, POOS	1.0206	1.0339	1.0171	1.0082	1.0784	0.9870	1.0140	1.0826
RR AR, k-fold	1.0211	0.9716	0.9758	0.9694	1.0043	0.9537*	0.9762	0.9913
RR AR, POOS	1.0636	1.0770	1.0248	0.9687	1.0350	1.0246	1.0244	0.9915
KRR AR, k-fold	1.0144	1.0383	0.9869	1.0127	1.0499	1.0194	0.9951	1.0316
KRR AR, POOS	1.0098	1.0020	1.0102	1.0649	1.0437	0.9727	1.0070	1.1201*
Data rich models								
ARDI, k-fold	1.0298	1.0887	1.2775	0.9579	1.0224	1.0241	1.1270	0.9982
ARDI, POOS	1.3491***	1.1811	1.4741	1.0455	1.3214***	1.0701	1.2329	1.0279
RF ARDI, k-fold	1.0465	1.0097	1.0066	0.9360	1.1014	0.9460	1.0011	0.9399
RF ARDI, POOS	1.0572	1.0510	0.9333	0.9932	1.1060	0.9894	0.9438	0.9971
RR ARDI, k-fold	1.2091**	1.1764	1.3362	1.1841	1.2408**	1.1147	1.1579	1.0801
RR ARDI, POOS	1.2411***	1.2067	1.2010	1.0391	1.1812**	1.1227	1.0860	0.9916
KRR ARDI, k-fold	1.0798	1.0713	0.9704	0.9273	1.0880	1.0591	0.9775	0.9751
KRR ARDI, POOS	1.3743***	0.9541	0.9776	1.0363	1.4249***	0.9238	0.9650	1.0591
ARVIF, k-fold	1.1642***	1.1274**	0.9931	0.9339	1.1490**	1.1204*	0.9823	0.9938
ARVIF, POOS	1.1642***	1.1274**	1.0408	0.9789	1.1490**	1.1204*	1.0113	1.0306
RF ARVIF, k-fold	1.0521	0.9209	0.9378	0.9201	1.0925	0.8668*	0.9335	0.9696
RF ARVIF, POOS	1.0726	0.9749	0.9599	0.9641	1.1149	0.9224	0.9589	0.9944
RR ARVIF, k-fold	1.1941**	1.2512	1.1342	0.9633	1.1646**	1.1280	1.0244	1.0061
RR ARVIF, POOS	1.1878*	1.1031	1.1920	1.2691	1.1552*	1.0316	1.0531	1.2404
KRR ARVIF, k-fold	1.2436**	1.0212	1.0180	0.9598	1.2937***	1.0085	0.9872	1.0314
KRR ARVIF, POOS	1.3056***	1.0022	1.0310	0.9409	1.4148***	0.9949	0.9979	1.0314
Lasso (B1), k-fold	1.1231*	1.1492	1.2995	1.0930	1.1100	1.0088	1.1508	1.0864
Lasso (B2), k-fold	1.1143*	1.2215*	1.1495	1.1473	1.0980	1.1128	1.1305	1.1998
Lasso (B3), k-fold	1.4151***	1.2724**	1.1131	1.0086	1.4242***	1.2087*	1.0892	1.0027
Lasso (B1), POOS	1.1655	1.0977	1.2642	1.2086	1.1384	1.0084	1.1579	1.2410
Lasso (B2), POOS	1.1711	1.0913	1.3123**	1.1874	1.1509*	1.0363	1.2746**	1.1974
Lasso (B3), POOS	1.2383**	1.3052***	1.2493	1.0035	1.2243**	1.2350**	1.1810	1.0778
Ridge (B1), k-fold	1.2072**	1.3469*	1.3261	1.0592	1.2101**	1.1861	1.1848	1.0477
Ridge (B2), k-fold	1.2400**	1.2849**	1.1246	1.0313	1.2331**	1.2140	1.1411	1.1081
Ridge (B3), k-fold	1.2552**	1.4348***	1.2749**	0.8964	1.2134**	1.3772**	1.2916*	0.9705
Ridge (B1), POOS	1.2434**	1.3360*	1.3173	1.0720	1.2077*	1.1704	1.1870	1.0671
Ridge (B2), POOS	1.2337*	1.1857	1.2057*	1.0991	1.2074**	1.1272	1.2401	1.1978
Ridge (B3), POOS	1.2463**	1.2815**	1.3295***	0.9787	1.2092**	1.2103*	1.3618**	1.0641
ElasticNet (B1), k-fold	1.1311*	1.3399	1.4602	1.2620	1.1187*	1.1069	1.2363	1.1784
ElasticNet (B2), k-fold	1.1742*	1.1470	1.0933	1.0636	1.1542*	1.0450	1.0888	1.1223
ElasticNet (B3), k-fold	1.3820**	1.2479**	1.1223	1.0227	1.4161**	1.1587*	1.0999	1.0642
ElasticNet (B1), POOS	1.1635	1.0875	1.2539	1.2007	1.1503	1.0033	1.1516	1.2354
ElasticNet (B2), POOS	1.2191*	1.0648	1.1802	1.1981	1.1920**	1.0078	1.1786	1.2602
ElasticNet (B3), POOS	1.2866***	1.1917*	1.3511*	1.0033	1.2613**	1.1635	1.2474*	1.0855

Short training period: starts in 2010M01

Table 21: This table shows for the variable number of sold existing homes (wnm_aantal_verkochte_koopwoningen_bestaand) the RMSEs of all the models that have been run for the different forecasting horizons. For this table the training period starts at 2010M01. The test period is from 2015M01 until 2023M12. For this variable a first-difference log transformation is used. The models are divided in data poor and data rich-models. The first data poor model, the AR, BIC, is the benchmark model and all other RMSEs are relative to this value. The stars indicate the results of a Diebold-Mariano test compared to the benchmark model (H_0: no difference in forecasting accuracy between the two models): $*$ significance at 10% level, $**$ significance at 5% level, and $***$ significance at 1% level.

	Variable transformed				Variable backtransformed			
Models	$h=1$	$h=3$	$h=6$	$h=12$	$h=1$	$h=3$	$h=6$	$h=12$
Data poor models								
AR, BIC (BM)	0.1191	0.0444	0.0245	0.0168	2091.5	2380.5	2734.3	3484.9
AR, k-fold	1.0000	1.0085	0.9919	0.9452	1.0000	1.0223	1.0015	0.9715
AR, POOS	1.1537*	1.1990**	1.0009	0.9680	1.1257**	1.1752**	1.0013	1.0005
RF AR, k-fold	1.0296	0.9418	1.0178	0.9346	1.0852	0.9370	1.0270	0.9996
RF AR, POOS	1.0322	0.9453	1.0279	0.9462	1.0856	0.9366	1.0373	1.0306
RR AR, k-fold	1.0549	0.9875	1.0150	0.9219	1.0177	0.9757	1.0216	0.9867
RR AR, POOS	1.0274	1.0131	1.0385	0.9566	0.9986	1.0016	1.0365	0.9963
KRR AR, k-fold	1.0427	1.0484	1.0704	0.9434	1.0875	1.1250	1.0503	1.0030
KRR AR, POOS	1.0818	1.0595	1.0427	1.0560	1.1409	1.0483	1.0402	1.0780
Data rich models								
ARDI, k-fold	1.3823***	1.3473**	1.1667	1.1833	1.3690***	1.3198***	1.1164	1.1217*
ARDI, POOS	1.3936***	1.3856**	1.5689	1.7079	1.3180***	1.2919**	1.3239*	1.3585*
RF ARDI, k-fold	1.1723	1.0437	0.9899	0.9078	1.1850*	1.0138	0.9838	0.9870
RF ARDI, POOS	1.1860*	1.0041	0.9857	0.8469	1.2527*	0.9676	0.9503	0.9212
RR ARDI, k-fold	1.3958**	1.3172*	1.2493	1.1607	1.4042***	1.2478*	1.1457	1.3432
RR ARDI, POOS	1.2620**	1.0627	1.3459	1.5797	1.2415**	1.0856	1.1469	1.4885*
KRR ARDI, k-fold	1.4305***	0.9434	1.0022	0.9359	1.4518***	0.9253	0.9874	1.0293
KRR ARDI, POOS	1.2335**	1.0025	1.0684	0.8855	1.2746**	1.0178	1.0679	0.9864
ARVIF, k-fold	1.4685***	1.4774***	1.3874**	1.3166*	1.5719***	1.5156***	1.3452***	1.6299***
ARVIF, POOS	1.4257***	1.5311***	1.5874	1.3519*	1.5260***	1.6063***	1.4089**	1.6744**
RF ARVIF, k-fold	1.1352	0.9226	0.9462	0.8891	1.1767*	0.9127	0.9410	1.0433
RF ARVIF, POOS	1.1223	0.9638	1.0084	0.8782	1.1677	0.9630	0.9962	1.0205
RR ARVIF, k-fold	1.1753	1.2454*	1.2350*	1.0236	1.1858*	1.3092**	1.1929**	1.2248
RR ARVIF, POOS	1.1116	1.0736	1.2058	1.3065	1.1226	1.0935	1.2132*	1.6190
KRR ARVIF, k-fold	1.1986	0.9806	1.0921	1.0265	1.1961	0.9589	1.0830	1.1982
KRR ARVIF, POOS	1.2611	0.9494	1.0565	0.9825	1.2546*	0.9609	1.0385	1.1239
Lasso (B1), k-fold	1.1200	1.3803	1.4063	1.0733	1.1081	1.2534*	1.2013	1.1639
Lasso (B2), k-fold	1.1687	1.1822	1.2235	0.9962	1.1593*	1.1596	1.1479	1.0813
Lasso (B3), k-fold	1.2974**	1.3028*	1.3100*	0.9663	1.3078**	1.2606*	1.3140	1.0658
Lasso (B1), POOS	1.1205	1.0439	1.3514	0.8829	1.1024	1.0282	1.1849	1.0012
Lasso (B2), POOS	1.1764**	1.0603	1.2402*	0.9063	1.1783**	1.0711	1.1886*	0.9830
Lasso (B3), POOS	1.4338**	1.1841	1.4050**	0.9421	1.4550***	1.1509	1.4530**	1.0277
Ridge (B1), k-fold	1.5304**	1.4847**	1.8112**	1.1969	1.5560***	1.3765**	1.6064***	1.2366
Ridge (B2), k-fold	1.4357**	1.1928	1.2589**	0.8794	1.4423***	1.1713	1.2169*	0.9601
Ridge (B3), k-fold	1.6764***	1.2468*	1.4371***	0.8661	1.7165***	1.2387*	1.3961***	0.9322*
Ridge (B1), POOS	1.4825**	1.3720**	1.8311*	1.1263	1.4818**	1.3170**	1.6093***	1.2042
Ridge (B2), POOS	1.4037**	1.1716	1.2485**	0.8906	1.4099**	1.1520	1.2140**	0.9743
Ridge (B3), POOS	1.6864***	1.2255*	1.4197***	0.8645	1.7175***	1.2213*	1.3881***	0.9329
ElasticNet (B1), k-fold	1.1200	1.4857*	1.4866*	1.1776	1.1147	1.3376**	1.3327*	1.2280*
ElasticNet (B2), k-fold	1.1395	1.2061*	1.1840	0.8876	1.1355*	1.2070	1.1398	0.9672
ElasticNet (B3), k-fold	1.3443***	1.3055**	1.3900**	1.0292	1.3501**	1.2710*	1.4371**	1.1795
ElasticNet (B1), POOS	1.1447	1.0540	1.4175	0.8867	1.1206	1.0362	1.2266	1.0036
ElasticNet (B2), POOS	1.1889**	1.0962	1.1838	0.8974	1.1924**	1.1097	1.1487	0.9824
ElasticNet (B3), POOS	1.3919**	1.1743	1.4078**	0.9258	1.3968**	1.1405	1.4561**	1.0120

G.7 Unemployment rate

Maximum training period: for unemployment rate this starts in 2006M01

Table 22: This table shows for the variable unemployment rate (arb_werkloosheidspercentage) the RMSEs of all the models that have been run for the different forecasting horizons. For this table the training period starts at 2006M01. The test period is from 2015M01 until 2023M12. For this variable a first-difference log transformation is used. The models are divided in data poor and data rich-models. The first data poor model, the AR, BIC, is the benchmark model and all other RMSEs are relative to this value. The stars indicate the results of a Diebold-Mariano test compared to the benchmark model (H_0: no difference in forecasting accuracy between the two models): $*$ significance at 10% level, $**$ significance at 5% level, and $***$ significance at 1% level.

	Variable transformed				Variable backtransformed			
Models	$h=1$	$h=3$	$h=6$	$h=12$	$h=1$	$h=3$	$h=6$	$h=12$
Data poor models								
AR, BIC (BM)	0.0291	0.0234	0.0197	0.0131	0.1381	0.3264	0.5692	0.8131
AR, k-fold	1.0000	0.9387	0.9929	1.0000	1.0000	0.9220	0.9960	1.0000
AR, POOS	1.0000	1.0243	1.0057	1.0183	1.0000	1.0187	1.0160	1.0228
RF AR, k-fold	1.0533	1.0921	1.0150	1.2108	1.0446	1.0985	1.0278	1.4105
RF AR, POOS	1.1112**	1.1258	1.0292	1.2379*	1.0997**	1.1450	0.9950	1.4584*
RR AR, k-fold	1.0079	1.0952	0.9860	0.9959	1.0009	1.0965	0.9880	0.9945
RR AR, POOS	1.0506	1.0225	0.9977	1.0147	1.0462	1.0185	1.0006	1.0196
KRR AR, k-fold	1.0668	1.4759**	1.0116	1.1751	1.0672	1.4660**	1.0910	1.3503
KRR AR, POOS	1.0891	1.0022	1.0106	1.3509	1.1217	1.0043	0.9571	1.5423**
Data rich models								
ARDI, k-fold	0.9813	0.8719	0.9273	1.3583	0.9565	0.8220	0.8846	1.2472
ARDI, POOS	1.1293*	1.3324	1.0647	1.7397	1.0601	1.2130	1.0820	1.7661
RF ARDI, k-fold	1.0151	0.9173	0.8842	0.8823*	1.0063	0.9291	0.8563	0.8430**
RF ARDI, POOS	1.0288	0.9255	0.8697*	0.8480*	1.0044	0.9317	0.8477	0.7949**
RR ARDI, k-fold	1.0090	1.0746	0.9870	1.3862	0.9894	1.0313	0.9262	1.3087
RR ARDI, POOS	0.9962	1.2291	1.1198	1.5895	0.9749	1.1477	1.0649	1.5621
KRR ARDI, k-fold	1.0182	1.0053	0.8710	0.9771	1.0056	0.9983	0.8885	0.9852
KRR ARDI, POOS	1.0602	0.9994	0.8277	0.9786	1.0364	0.9891	0.8102	1.0192
ARVIF, k-fold	0.9963	0.9548	0.9310	1.0282	0.9638	0.9135	0.9383	0.9794
ARVIF, POOS	1.0093	0.9828	0.9310	1.0397	0.9728	0.9482	0.9383	0.9959
RF ARVIF, k-fold	1.0278	0.9806	0.9328	1.0536	1.0058	0.9882	0.9253	1.1643
RF ARVIF, POOS	1.0472	0.9816	0.8928	1.0603	1.0241	0.9923	0.8920	1.1110
RR ARVIF, k-fold	0.9848	0.9742	0.9307	1.2473	0.9718	0.9454	0.9444	1.1887
RR ARVIF, POOS	0.9939	1.0831	0.9326	1.4150	0.9709	1.0348	0.9340	1.3809
KRR ARVIF, k-fold	0.9992	0.9156*	0.9010	0.8796*	0.9985	0.9308	0.9660	0.8672*
KRR ARVIF, POOS	1.0218	0.9699	0.8649	1.0118	1.0347	0.9642	0.8693	1.0045
Lasso (B1), k-fold	1.0149	0.9918	1.1315	1.4147	0.9856	0.9386	1.0800	1.4286
Lasso (B2), k-fold	1.0186	1.0412	0.8301	1.0503	0.9974	1.0035	0.7997	1.0046
Lasso (B3), k-fold	1.0048	0.9645	0.8874	1.0904	0.9977	0.9719	0.8758	1.0071
Lasso (B1), POOS	1.0226	1.0288	0.9651	0.9632	0.9915	0.9692	0.9045	0.9194
Lasso (B2), POOS	1.0048	0.9926	0.8259	1.1066	0.9885	0.9522	0.8074	1.0797
Lasso (B3), POOS	1.0242	0.9670	0.8820	1.3721	1.0060	0.9582	0.8862	1.4043
Ridge (B1), k-fold	1.0312	0.9955	1.0876	1.2549	0.9923	0.9394	1.0251	1.1617
Ridge (B2), k-fold	1.0778	1.0058	0.9959	0.8573	1.0645	0.9834	0.9597	0.8496
Ridge (B3), k-fold	1.0493	0.9758	0.8886	0.9653	1.0496	0.9936	0.9066	1.0087
Ridge (B1), POOS	1.0472	1.0195	1.0140	1.3915	1.0134	0.9662	0.9495	1.3071
Ridge (B2), POOS	1.0743	1.0439	0.9505	0.8835	1.0596	1.0117	0.9228	0.8919
Ridge (B3), POOS	1.0461	0.9940	0.9204	1.0087	1.0440	1.0057	0.9168	1.0227
ElasticNet (B1), k-fold	1.0091	1.0327	1.0988	1.4948	0.9821	0.9722	1.0421	1.5915
ElasticNet (B2), k-fold	1.0110	0.9724	0.8380	0.8423	0.9943	0.9503	0.8118	0.7975
ElasticNet (B3), k-fold	1.0233	0.8880	0.8473	1.0087	0.9940	0.8806*	0.8481	0.9469
ElasticNet (B1), POOS	1.0230	1.0438	1.0304	1.0397	0.9965	0.9874	0.9633	0.9728
ElasticNet (B2), POOS	1.0476	0.9929	0.8313	1.1028	1.0318	0.9539	0.8133	1.0754
ElasticNet (B3), POOS	1.0958	0.9686	0.8700	1.3799	1.0782	0.9599	0.8830	1.4187

Short training period: starts in 2010M01

Table 23: This table shows for the variable unemployment rate (arb_werkloosheidspercentage) the RMSEs of all the models that have been run for the different forecasting horizons. For this table the training period starts at 2010M01. The test period is from 2015M01 until 2023M12. For this variable a first-difference log transformation is used. The models are divided in data poor and data rich-models. The first data poor model, the AR, BIC, is the benchmark model and all other RMSEs are relative to this value. The stars indicate the results of a Diebold-Mariano test compared to the benchmark model (H_0: no difference in forecasting accuracy between the two models): $*$ significance at 10% level, $**$ significance at 5% level, and $***$ significance at 1% level.

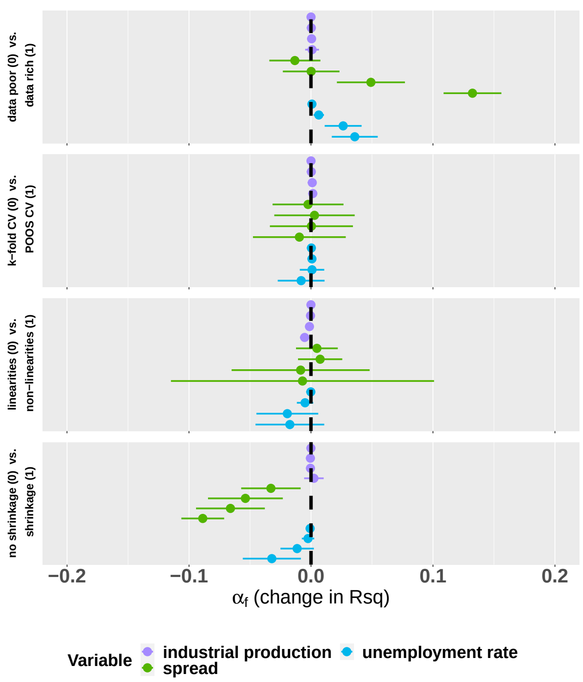
	Variable transformed				Variable backtransformed			
Models	$h=1$	$h=3$	$h=6$	$h=12$	$h=1$	$h=3$	$h=6$	$h=12$
Data poor models								
AR, BIC (BM)	0.0294	0.0232	0.0198	0.0167	0.1408	0.3248	0.5842	1.2578
AR, k-fold	1.0000	0.9593	1.0250	1.0000	1.0000	0.9512	1.0153	1.0000
AR, POOS	1.0093	1.0547	1.0309*	1.1224	1.0070	1.0793	1.0663	1.1767
RF AR, k-fold	1.0721	1.1686	1.0965	1.0932	1.0597	1.2314	1.1890	1.0508
RF AR, POOS	1.0409	1.1572	1.1440	1.1168*	1.0188	1.1697	1.2155	1.0582
RR AR, k-fold	1.0011	1.0570	0.9892	1.0853	0.9950	1.0855	1.0102	1.1199
RR AR, POOS	1.0342	1.0982	1.0217	1.0513	1.0279	1.1186	1.0507	1.0524
KRR AR, k-fold	1.0420	1.0780	1.0729	1.1295	1.0610	1.1451*	1.1921	1.1021
KRR AR, POOS	1.0976*	1.4524**	0.9597	0.9770	1.0848	1.4907***	1.0085	0.9513
Data rich models								
ARDI, k-fold	1.0147	1.0429	0.9862	0.9440	0.9433	0.9696	0.9025	0.7018
ARDI, POOS	1.0468	1.0229	1.0282	1.0575	0.9813	0.9472	0.9442	0.8131
RF ARDI, k-fold	1.0018	0.9854	0.9786	0.8355	0.9832	0.9758	0.9300	0.7303*
RF ARDI, POOS	0.9851	1.0474	1.0406	0.8355	0.9594	1.0322	1.0021	0.7216*
RR ARDI, k-fold	1.0160	0.9718	1.0762	1.0648	0.9768	0.9224	1.0070	0.8402
RR ARDI, POOS	1.0083	0.9665	0.9759	1.0949	0.9625	0.9151	0.8980	0.9065
KRR ARDI, k-fold	1.0320	1.0488	1.0162	0.8255	1.0563	1.0371	1.0490	0.7012*
KRR ARDI, POOS	1.0557	0.9972	1.0653	0.8110	1.0261	1.0080	1.0965	0.6236*
ARVIF, k-fold	1.1489	1.0937	1.0467	1.1386	1.0742	1.0074	0.9623	1.1186
ARVIF, POOS	1.1734*	1.1460*	1.0705	1.2450	1.0879	1.0593	0.9744	1.3318
RF ARVIF, k-fold	0.9884	1.1138	1.1076	0.9885	0.9852	1.0824	1.0755	0.8536
RF ARVIF, POOS	0.9899	1.1095	1.1466	0.9498	0.9618	1.0791	1.1239	0.7906
RR ARVIF, k-fold	1.0272	1.1722*	0.9894	0.9524	0.9777	1.0696	0.9671	0.8833
RR ARVIF, POOS	1.2129*	1.1275	0.9894	0.9285	1.0958	1.0274	0.9564	0.8626
KRR ARVIF, k-fold	0.9893	0.9642	0.9148	0.9821	0.9725	0.9525	0.9083	0.9472
KRR ARVIF, POOS	0.9824	0.9674	0.8767	0.9141	0.9805	0.9600	0.8546	0.7698
Lasso (B1), k-fold	1.0184	1.1767	1.2249	1.6106	0.9762	1.0811	1.1228	1.5420
Lasso (B2), k-fold	1.0334	0.9595	0.8745	0.8332	1.0062	0.9280	0.8328	0.7921*
Lasso (B3), k-fold	1.0028	1.0380	0.8852	0.7933	0.9895	1.0248	0.8228	0.6517*
Lasso (B1), POOS	1.0968*	1.4126	1.2941	0.8690	1.0325	1.2887	1.2394	0.6637
Lasso (B2), POOS	1.2141**	1.0524	0.8517	0.7638**	1.1210	0.9978	0.8180	0.7340**
Lasso (B3), POOS	1.0545	0.9902	0.9631	0.6724**	1.0525	0.9835	0.9548	0.5400**
Ridge (B1), k-fold	1.0183	1.0765	1.1262	1.7752	0.9636	1.0110	1.0604	1.8594
Ridge (B2), k-fold	1.0095	1.0547	0.9954	0.9883	0.9869	1.0418	1.0153	0.9316
Ridge (B3), k-fold	1.0321	0.9899	0.9074	0.8778	1.0284	1.0229	0.9338	0.8188
Ridge (B1), POOS	1.1607*	1.2549	1.0971	1.3730	1.0471	1.1577	1.0381	1.1910
Ridge (B2), POOS	1.0496	1.0664	1.0564	0.9341	1.0126	1.0587	1.0642	0.8848
Ridge (B3), POOS	1.0279	0.9882	0.9134	0.8781	1.0243	1.0218	0.9432	0.8207
ElasticNet (B1), k-fold	1.0130	1.0993	1.2275	1.4528	0.9563	1.0160	1.1214	1.3472
ElasticNet (B2), k-fold	1.0263	0.9574	0.9410	0.7988	1.0006	0.9250	0.9057	0.7411*
ElasticNet (B3), k-fold	1.0296	1.0084	0.8934	0.8008	1.0125	0.9857	0.8343	0.6541*
ElasticNet (B1), POOS	1.2031**	1.3051	1.2278	0.8411	1.0973	1.2000	1.1840	0.6431
ElasticNet (B2), POOS	1.0618	1.0474	0.8856	0.7536**	1.0268	0.9981	0.9063	0.7280**

H Comparison with the FRED-MD

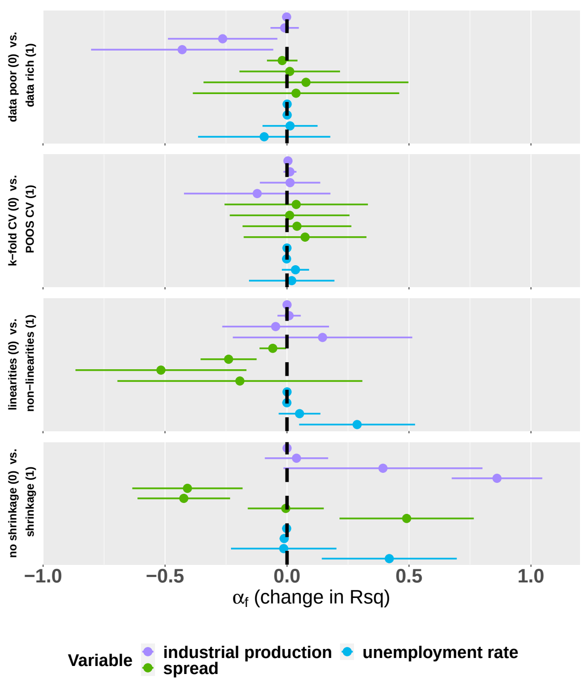
In this section we present the results of our analysis on the FRED-MD (McCracken and Ng, 2016), using the 2024M02 vintage. We stationarize all variables by applying the suggested transformations by McCracken and Ng (2016). After this, we transform all variables to have zero mean and unit variance. In this analysis, we follow the methodology described in the main text, with the exception of omitting the KRR model. Both our results and those of Goulet Coulombe et al. (2022) show that the performance of the two models is often comparable. Consequently, we have excluded the KRR model from this analysis for computational reasons.

To analyse the effects of having a longer dataset, we have done this analysis two times: 1) On the full FRED-MD, where our pseudo out of sample period ranges from 1980M01 until 2017M01, and the training period starts at 1963M01. 2) On a restricted FRED-MD. Here, our training period starts at 2005M01 and our out of sample period ranges from 2010M01 until 2017M01. For computational reasons we choose only three of the five dependent variables as used by Goulet Coulombe et al. (2022) in our analysis.

In Figure 17 we compare the results over both datasets. In contrary to our NL-MD results we see a clear positive effect for the data rich feature when the maximum training period is used (Figure 17a). This



(a) Maximum training period.



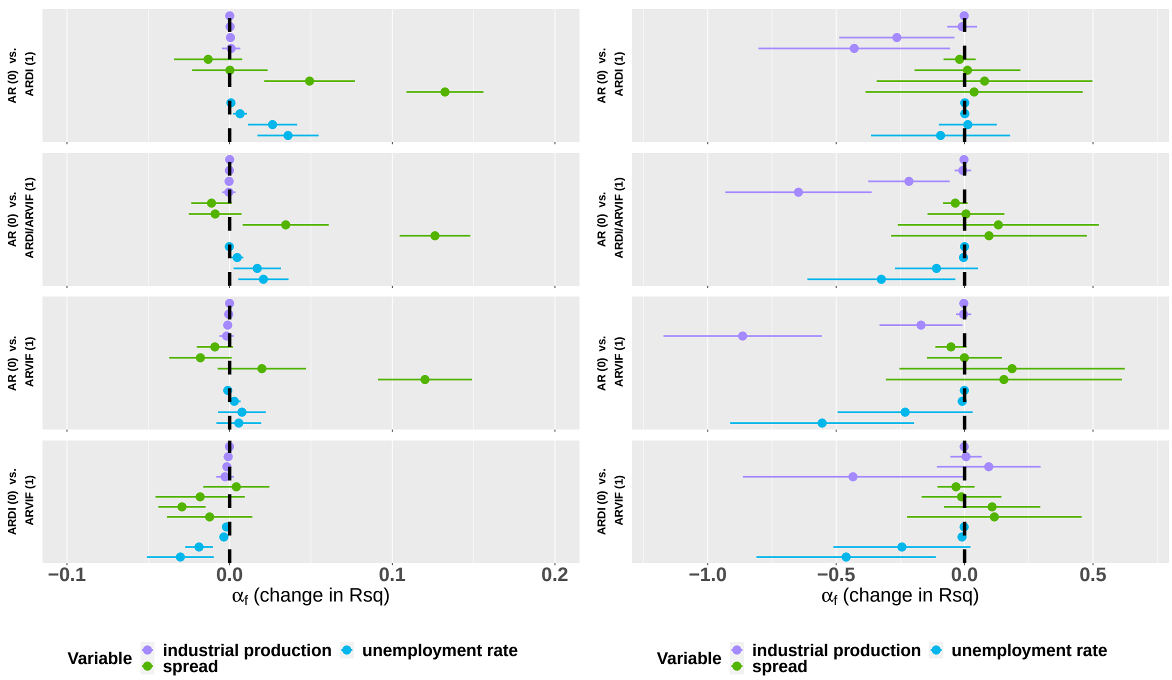
(b) Short training period.

Figure 17: Results on the FRED-MD. VIF is excluded from the analysis. The left plot has a long training and testing period (train - start POOS - test: 1963M01 - 1980M01 - 2017M12), whereas the right plot has a short training and test period (train - start POOS - test: 2005M01 - 2010M01 - 2017M12).

effect becomes stronger with horizon, as is seen in Goulet Coulombe et al. (2022). This effect disappears when we restrict the training period for the FRED (Figure 17b). This gives evidence for the fact that for a large dataset to work optimally, more observations are needed. Although some caution is needed here, as the treatment effects are relatively small. Furthermore, in both the NL-MD and the FRED-MD short training period case, the error bars are large due to the limited pseudo-out-of-sample period. Similarly, we observe a shift in the shrinkage feature when transitioning from the full sample to the restricted sample. In the full sample, the treatment effect is negative, while in the restricted sample, it is positive. This indicates that shrinkage models handle limited data relatively well compared to the standard factor model.

The other features are relatively comparable to the NL-MD (Table 6, Figure 8). They mostly show null effects.

In Figure 18, we specifically compare the VIF and factor method for the data rich sample, comparing the long and short training period. First we consider the maximum training period, Figure 18a. As expected from the literature (Coulombe et al., 2021a, Hindrayanto et al., 2016, Jansen et al., 2016), factors now perform significantly better than both the lags and the VIF method. In Figure 18b, we show the results



(a) Maximum training period.

(b) Short training period.

Figure 18: Results on the FRED-MD. The left plot has a long training and testing period (train - start POOS - test: 1963M01 - 1980M01 - 2017M12), whereas the right plot has a short training and test period (train - start POOS - test: 2005M01 - 2010M01 - 2017M12). This figure shows the differences between models trained on several feature matrices.

of the short training period. In this case, the VIF environment seems to perform relatively worse for larger horizons. However, this effect is not significant.

Our main conclusion from our FRED analysis is that to be able to optimally use a large amount of variables, without variable specific tuning, a longer training period is needed. If enough data is available, factors provide a good way for data compression. However, on smaller datasets they do not necessarily improve predictive performance.

I Best models for the short training period

Table 24: The best performing model for each variable-horizon combination, according to their OOS-RMSE. Here, we consider the models trained on the short trainings period. The dots indicate several features the models can have. Red indicates data rich, green shrinkage and blue non-linearity's. The models with with an orange dot use k-fold cross validation, those without POOS.

Variable	h = 1	h = 3	h = 6	h = 12
AEX	AR 🟠	Ridge (B2) 🔴🟢🟠	Ridge (B2) 🔴🟢	AR 🟠
CPI	KRR ARDI 🔴🔵🟠	KRR ARVIF 🔴🔵	KRR ARDI 🔴🔵🟠	Lasso (B1) 🔴🟢🟠
Deposit interest rate	RR ARVIF 🔴🟢🟠	Ridge (B1) 🔴🟢	ARDI 🔴🟠	ARDI 🔴🟠
Export (value index)	KRR AR 🔵	RF AR 🔵	KRR ARDI 🔴🔵	KRR AR 🔵🟠
Industry (production index)	RF ARVIF 🔴🔵	KRR ARVIF 🔴🔵	Ridge (B3) 🔴🟢🟠	RR AR 🟢
Number of sold existing homes	AR 🟠	RF ARVIF 🔴🔵🟠	RF ARVIF 🔴🔵🟠	RF ARDI 🔴🔵
Unemployment rate	KRR ARVIF 🔴🔵	ElasticNet (B2) 🔴🟢🟠	Lasso (B2) 🔴🟢	Lasso (B3) 🔴🟢