Forecast Combination and Interpretability Using Random Subspace*

Boris Kozyrev†

October 5, 2024

Abstract

This paper investigates forecast aggregation via the random subspace regressions method (RSM) and explores the potential link between RSM and the Shapley value decomposition (SVD) using the US GDP growth rates. This technique combination enables handling high-dimensional data and reveals the relative importance of each individual forecast. First, it is possible to enhance forecasting performance in certain practical instances by randomly selecting smaller subsets of individual forecasts and obtaining a new set of predictions based on a regression-based weighting scheme. The optimal value of selected individual forecasts is also empirically studied. Then, a connection between RSM and SVD is proposed, enabling the examination of each individual forecast's contribution to the final prediction, even when there is a large number of forecasts. This approach is model-agnostic (can be applied to any set of predictions) and facilitates understanding of how the aggregated prediction is obtained based on individual forecasts, which is crucial for decision-makers.

Keywords: Forecasting, Forecast combination, Random Subset, Shapley Value Decomposition, Forecast Combination Puzzle

JEL: C22, C45.

---

*The author gratefully acknowledge Raffaella Giacomini, Katja Heinisch, Oliver Holtemöller, Gregor von Schweinitz, and conference and seminar participants at 6th Vienna Workshop on High-Dimensional Times Series in Macroeconomics and Finance (Institute for Advanced Studies), 8th RCEA Time Series Econometrics Workshop (Brunel University), 25th Federal Forecasters Conference (U.S. Treasury), NASM 2024 (Vanderbilt University), IAAE 2024 (University of Macedonia), 44th International Symposium On Forecasting (Dijon), ESIF-AIML2024 (Cornell University) for useful discussions and comments.

†Halle Institute for Economic Research (IWH), boris.kozyrev@iwh-halle.de

1 Introduction

Nowcasting and short-term forecasting of current economic activity is crucial for decision-makers such as firms and governments. Over the last few decades, the amount of data used for (improving) macroeconomic forecasting has been drastically increasing. This means one may have hundreds of different forecasts of a given macroeconomic indicator, such as GDP. These individual predictions may be derived from different methods or data sources. A natural question arises: how do we combine and/or select numerous forecasts to produce a final prediction?

Another research aspect has lately focused on unraveling the black box critique of machine learning (ML) models. Often, it is nearly impossible to study the relationship between the dependent and independent variables within the ML framework or to explain why certain estimations are made. In particular, the "interpretability" aspect is crucial for decision-makers in economics, when some sort of a structure based on economic theory or intuition is desired for specific reasons.

Not only do economists and policymakers care about precision, but they strive to comprehend the driving forces behind a given phenomenon. Forecasting models are supposed to motivate some economic or public policies instead of producing an intractable yet often precise estimation (Burgess et al., 2013). Additionally, even the "best" single model can sometimes produce imprecise estimations. In theory, understanding a model's mechanics allows for identifying and explaining its odd behaviors, a task often unfeasible for some ML models.

The noticeable increase in using predictive ML in macroeconomics has driven researchers to try adding some "interpretability" by introducing a universal method that explains the relative importance of any given feature. The Shapley value decomposition (SVD) is one of the most popular tools, allowing for decomposition into individual variable contributions. Despite its popularity, it has some disadvantages. For example, SVD handles high-dimensional data poorly and demands considerable computing time. Computational burdens force researchers to restrict their attention to fewer indicators or shift to specific models for which shortcuts have been developed.

In this article, a framework is proposed that connects two seemingly independent fields: forecast combination and model interpretability. First, a forecast combination approach is introduced based on the random subspace regression method (RSM) proposed by Boot and Nibbering (2019), which makes it possible to work within a linear framework even if the number of covariates (in this case, individual forecasts) is huge. This aggregation scheme produces a new set of forecasts based on the initial values. RSM helps smooth the distribution of predictions if there are severe outliers. The forecasting performance of this aggregation method is then compared to standard benchmarks.

Additionally, a synergy between RSM and SVD is explored. Because SVD is easily computed in a standard linear framework, one can decompose the prediction obtained by RSM into each forecast contribution using well-known shortcuts. Furthermore, this union of the methods allows researchers to handle even high-dimensional data. Thus, with this framework, it is possible to circumvent the challenges of using many variables while dealing with SVD. The main SVD advantage – the difference between the final and the historical average predictions is fairly distributed among the covariates – is fulfilled.

The contribution to the literature is twofold. First, the use of RSM as a forecast combination method is introduced, unlike previous studies where RSM has been used solely to produce forecasts. When used for forecast aggregation, RSM proves to be a powerful tool, demonstrating superior forecasting performance compared to both AR(1) models and the median of a given set of forecasts. The number of individual forecasts included in one draw plays a critical role. Second, the proposed connection between RSM and SVD allows for the easy computation of each individual forecast's contribution to the final prediction, even when the number of forecasts is large. This enables real-time analysis of the forecast decomposition, providing valuable insights into the dynamics of the target variable, which is crucial for policymakers. Both RSM and SVD are model-agnostic approaches, applicable to any arbitrarily large set of forecasts.

The paper is organized as follows. Section 2 provides the current state of the forecast combination research avenues. Section 2.3 describes RSM and how it can be used to combine forecasts. Section 3 is devoted to a feature selection problem and highlights the Shapley decomposition of forecasts and its implementation within RSM. Section 4 describes the nowcasting model used for the study. Section 5 includes an empirical application of all the above-mentioned concepts by exploiting the FRED-MD database. Finally, Section 6 concludes and provides an outlook for future research.

2 Forecast Combination

Researchers from various disciplines have analyzed different methods of forecast aggregation. In many practical cases, multiple forecasts of the same variable have been obtained either by various methods or by using different sources of information (Timmermann, 2006). It is assumed that combining different individual predictions can improve accuracy. The aggregation strategy enables the incorporation of much more available information, which may be more fruitful than seeking a "best" model. For a review of the extensive literature on forecast combinations, see Wang et al. (2023). This section describes methods dealing with point forecast combinations based on linear combinations. The focus here is on this class of forecast aggregation because, as shown below, it is possible to merge the latter with RSM. To conclude, a forecast combination puzzle is discussed.

2.1 Linear Combinations of Individual Forecasts

One of the most straightforward strategies for combining various forecasts is to assign weights to each individual prediction: the more precise the forecast, the greater the weight. There are numerous different weighting schemes. Thus, it sounds feasible to construct a linear combination of the individual forecasts and treat the OLS coefficients as weights.

More formally, suppose an N-dimensional vector of h-step ahead forecasts $\hat{\boldsymbol{y}}_{T+h|T}$ is given. Additionally, an n-dimensional vector of past observations of the targeted variable $\boldsymbol{y_T}$ is observed. Under this framework, regression-based weights can be defined as follows (Granger and Ramanathan, 1984):

$$y_{T+h} = w_{T+h|T,0} + \boldsymbol{w'_{T+h|T}} \hat{\boldsymbol{y}}_{T+h|T} + \epsilon_{T+h}, \tag{1}$$

where $\hat{\boldsymbol{y}}_{T+h|T}$ are past individual forecasts.

Some restrictions may still be imposed on $\boldsymbol{w'_{T+h|T}}$. For instance, a constant term can be omitted, and/or the weights can be adjusted so that their sum is one. In this study, the unrestricted specification of Equation 1 is chosen, thus the weights correspond to unrestricted OLS estimations.

The use of regression-based weights for forecast combinations has a long history. Both advantages and limitations of this approach have already been examined in the last century. One of the critiques is shown in De Menezes et al. (2000). While exploiting the unrestricted regression, one should address potential multicollinearity and serial correlation in forecast errors. These challenges are assumed to be mitigated when combining this approach with RSM, as discussed later.

Upon establishing the unrestricted regression-based weighting method, another concern arises. What if the total number of $\hat{\boldsymbol{y}}_{T+h|T}$ is exceptionally huge? In this case, a simple OLS estimation becomes impractical, particularly when the number of forecasts (N) exceeds the number of observations in the training set (n). It makes sense to use some variable selection methods, such as Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani, 1996), to end up only with a few individual forecasts.

Typically, when dealing with high-dimensional data (in the form of $N > n$), LASSO or other penalized regressions have been extensively used for dimension reduction in many areas, including economics. However, sometimes, it is necessary to use as much data as possible without excluding variables by assuming non-sparsity. RSM can manage this level of selection 'strictness' by utilizing all the available data.

2.2 Forecast Combination Puzzle

Although various forecast combination methods have been proposed over the past few decades, the simple average (thus assigning equal weights to each forecast) tends to perform similarly to (or even better than) more sophisticated aggregation approaches. This fact has perplexed researchers, who have been trying to find a prolific answer to this paradox. Stock and Watson (2004) studied this phenomenon. In their paper, the term "forecast combination puzzle" was introduced.

Indeed, one should expect some gains in terms of forecast accuracy when using more rigorous tools for forecast combination than just averaging. Moreover, theoretical results for some approaches have been formally derived, proving this initial claim. However, often, a simple mean overshadows other weighting schemes. Again, Wang et al. (2023) present an excellent summary of why the "forecast combination puzzle" holds empirically is presented. To summarize, one possible explanation is that the weights are misspecified or wrongly computed due to, for instance, structural changes. Alternatively, benefits from using a sophisticated combination may be relatively small. Thus estimation per se overwhelms the potential gains.

However, if individual forecasts come from different sources or models, the variance can be high due to some outliers. If the number of outliers is high and they are not equally distributed around zero, then averaging over this forecasting set may be inefficient. Although taking a median can arise as a potential method to produce a final prediction, as used in Stock and Watson (1999), it does not indicate how varied the individual predictions are in the set.

Using RSM for forecast combination allows for relying on the mean even in the presence of outliers in the initial set. This is achieved by obtaining a new set of forecasts based on the initial one, which is supposed to be distributed so that taking a mean makes sense and provides good empirical performance.

2.3 Random Subspace

Initially, RSMs appeared in the field of ML but had little application in macroeconomic forecasting. Ho (1998) introduced this approach to train tree models. Later, Bay (1998) extended the same logic to the nearest neighbor classifiers. Wichitaksorn et al. (2023) applied RSM to feature selection in a logistic regression framework. Recently, random or complete[1] subspace methods have become a powerful tool used in macroeconomics (Elliott et al. (2013), Kotchoni et al. (2019), Pick and Carpay (2022)) and finance (Meligkotsidou et al. (2021)). Some other examples of papers utilizing subspace methods can be found in Boot and Nibbering (2020).

RSM is a randomized reduction method in which a small number of indicators are (randomly) selected to estimate a lower dimensional problem. This procedure is repeated sufficiently to produce many approximations to the initial – sometimes infeasible to calculate directly – model. Finally, the forecasts from each low-dimensional model are combined (averaged). This allows for reducing forecast variance while utilizing all available information.

More formally, suppose that the data-generating process (DGP) is defined as follows:

$$y_{t+1} = x_t'\beta + \epsilon_{t+1}, \tag{2}$$

for $t = 1, \cdots, T$, x_t' is a vector of predictors in $\mathbb{R}^N$, $\epsilon_{t+1} \sim iid(0, \sigma^2)$.

Initially, this model has been proposed to use observed time series to predict the target. Instead, this approach is used for the forecast combination. To draw a parallel with previous sections and align the notation, in Equation 2, x_t' corresponds to the N-dimensional set of individual forecasts. In contrast, the n-dimensional vector y_{t+1} is the final prediction of the independent variable.

As discussed previously, when the number of estimated coefficients is large, it is either impossible to run OLS, or there is a high variance in the latter, which can significantly decrease the forecast accuracy. To overcome both potential challenges, following Boot and Nibbering (2019), it is possible to make a projection of the N-dimensional vector x_t onto a k-dimensional subspace using a permutation matrix, denoted $R_i \in \mathbb{R}^{N \times k}$, such that:

$$\tilde{x}_t' = x_t' R_i \tag{3}$$

A permutation matrix R_i is responsible for randomly selecting (without replacement) a subset of k predictors of all N covariates. To illustrate this idea, imagine a simple example

[1]Complete subspace methods imply combining (averaging) a given set of predictors forecasts from all possible linear regressions

where there are $N = 5$ total number of variables and $k = 3$ of those are drawn. One possible R_i can look like as follows (Boot and Nibbering, 2020):

$$R_i = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \tag{4}$$

Often, a scaling factor of $\sqrt{\frac{N}{k}}$ is premultiplied to satisfy specific desirable statistical properties (Drineas et al., 2006). However, these properties seem to be redundant in this setting, so this factor will be omitted from now on[2].

Notably, there are different methods to generate R_i. Instead of selecting a subset of covariates, weighted averages can be used to construct a different set of predictions. Weights are chosen randomly from a normal distribution. Moreover, the proposed specification of RSM implies that the selection is carried out using a uniform distribution (the probability that a given variable is selected is equal across the covariate set). It is also feasible to use importance sampling probabilities or leverage scores (Mahoney et al., 2011)[3].

Assigning an equal probability to each covariate can be partially explained by the following logic. Suppose there is a concern that, for a given model, randomly selected covariates might be less relevant than the rest of the pool. At the same time, most (if not all) of the time series used for macroeconomic predictions are correlated. This fact hints at two possible conclusions. First, a "less" relevant covariate is likely correlated with one that has more predictive power. Moreover, repeating this procedure – drawing variables randomly and averaging results over each iteration – makes the model less prone to "uninformative" variable selection.

Conditional on a drawn R_i, a k-dimensional vector of OLS coefficients $\hat{\beta}^i$ can be obtained as well as a single aggregated forecast value of $\hat{y}_{t+1}$. Both statistics are of interest. First, by averaging different $\hat{y}_{t+1}$, RSM could benefit forecast aggregation by providing better predictions compared to more standard metrics. Later, the forecast performance of this method is demonstrated and compared to benchmarks. Second, to set up the connection between RSM and SVD, a subspace estimation of an OLS coefficient for a given variable x_j is required. To achieve this, the mean of the OLS coefficients for a variable x_j across all iterations was used

[2]For instance, a similar approach has been used in **?**.

[3]It is worth mentioning that leverage scores are commonly used when randomly selecting rows rather than columns.

when x_j was selected. That is, suppose that x_j was chosen n_{x_j} times. Thus, the following estimation can be obtained:

$$\hat{\beta}_{x_j} = \frac{1}{n_{x_j}} \sum_{r=1}^{n_{x_j}} \hat{\beta}_{x_j}^r, \tag{5}$$

where $\hat{\beta}_{x_j}^r$ corresponds to the OLS coefficient for a r-th iteration, when x_j is selected.

3 Feature Importance

One common assumption is that ML can improve various research tasks, including forecasting while handling high-dimensional data. However, often, the results obtained are somewhat difficult to explain. Even though extensive literature suggests that ML is superior to more traditional approaches, the inability to explain final estimations (the so-called black box critique) has been a significant obstacle to adopting ML (Molnar, 2020).

Consequently, various approaches have been developed to interpret different ML estimations. Generally, the latter can be split into several subcategories:

- Model-specific/ Model-agnostic
- Local / Global

The first item refers to model selection. A valuable feature of model-agnostic interpretation approaches is that they can be applied to any ML model, unlike their counterpart (model-specific). Thus, model-agnostic methods are significantly more flexible. In Ribeiro et al. (2016a), the advantages of using model-agnostic models are summarized below:

- Model flexibility: The interpretation method is used with any ML model regardless of complexity.
- Explanation flexibility: One can choose a specific form of explanation, depending on the nature of the problem.
- Representation flexibility: The explanation system should be capable of using various feature representations.

Interpretability models can also be either local or global. Local interpretation methods tend to explain individual features or predictions. However, global ones are exploited when describing the average behavior of an ML model, which is why they are particularly useful when a modeler wants to understand the general mechanisms in the data (Molnar, 2020). However, as mentioned in Buckmann et al. (2022), ML models are (highly) nonlinear, meaning a global measure can lead to inconclusive results when evaluated across differing domains of the input space.

Thus, nearly every interpretation method falls into the intersection of the two previously mentioned subcategories (model-agnostic/model-specific and local/global). For example, the partial dependence plot illustrates the marginal effect of one or two features on the predicted outcome (Friedman, 2001). This approach is model-agnostic and global. However, neural network layer visualization, as used in Zintgraf et al. (2017), on the other hand, is model-specific (can be applied only to deep neural networks) and global.

In this paper, the main focus is on the interpretation metrics, which are both model-agnostic and local. This intersection allows one to obtain the contribution of each individual prediction to the final forecast while maintaining the advantage of not being restricted to a specific model type. Two potential candidates are Local surrogate models (LIME) (Ribeiro et al., 2016b) and Shapley values. The latter is chosen because, as described later, Shapley values guarantee that the difference between the final prediction and the historical average prediction is fairly distributed among the covariates, which is not always true for LIME. This property is vital when using Shapley values for macroeconomic forecasting. Moreover, Shapley values are associated only with the regression coefficients in linear frameworks. Consequently, it is straightforward to connect Shapley values to RSM.

3.1 Shapley Value Decomposition

Upon applying the concept of feature importance to a macroeconomic forecasting setting, one may explicitly ask how important a given variable is to producing a predicted value of a target series. Interestingly, one possible answer to this question comes from a different field of study: game theory.

Shapley et al. (1953) introduced a solution to the problem of fairly dividing a joint pay-off across all individual players in a cooperative game. For almost six decades, this concept has had nothing to do with forecasting literature. This connection remained unexplored until Strumbelj and Kononenko (2010) established a connection between players and variables used for forecasting, as well as the pay-off and a final prediction.

Following Buckmann et al. (2022), it is possible to define a contribution of variable k in observation x_i and model f:

$$\phi_k^S(x_i; f) = \sum_{x' \subseteq \mathcal{C}(x) \setminus \{k\}} \frac{|x'|!(|N| - |x'| - 1)!}{|N|!} [f(x_i | x' \cup \{k\}) - f(x_i | x')], \tag{6}$$

where $\mathcal{C}(x) \setminus \{k\}$ corresponds to the set of all possible combinations of variables excluding k, and $|x'|$ is the number of variables included in the combination.

Like initial idea, SVD represents a weighted sum of marginal contributions of a given variable (player) to all possible combinations excluding this variable. In this study, SVD is used for forecast combinations. It means that a given variable, in this case, is an individual forecast. Therefore, the contribution of an individual forecast to the final prediction can be obtained.

Buckmann et al. (2022) described obstacles when estimating Equation 6. There are two major issues. As stated earlier, the number of possible combinations with l variables is 2^l. Thus, it grows exponentially. With sufficiently large l, evaluation of all possible combinations seems infeasible. Additionally, estimating the term $f(x_i|x')$ is typically challenging (unless a feature independence is assumed). These problems, in principle, can be avoided if tree-based (see Lundberg et al. (2020)) or linear frameworks are used. In the following subsection, SVD in a linear setting is described and a connection is established between SVD and RSM.

3.2 SVD and RSM

In this subsection, the connection between RSM and SVD is introduced. In a linear framework:

$$\hat{f}(x) = \beta_0 + \beta_{x_1} x_1 + \cdots + \beta_{x_N} x_N \tag{7}$$

The contribution ϕ_j of a j-th feature x_j on the total prediction $\hat{f}(x)$ is the difference between the feature effect minus the average effect. More formally, one could express it as follows:

$$\phi_j(\hat{f}) = \beta_{x_j} x_j - E(\beta_{x_j} X_j) = \beta_j [x_j - E(X_j)] \tag{8}$$

Let us combine this decomposition with RSM. Notably, to obtain ϕ_j, one only misses the OLS coefficient β_{x_j}. However, directly obtaining β_{x_j} may be complicated in a high-dimensional case. RSM can assist in addressing this issue.

To illustrate this idea, suppose a variable x_j was selected 50 times within RSM. This implies that one has 50 different estimations of β_{x_j}, each based on different subsamples. However, the term $[x_j - E(X_j)]$ remains the same across all the iterations. Thus, it is possible to obtain an averaged contribution of ϕ_j by averaging across different realizations of estimated β_{x_j}, as depicted below:

$$\bar{\phi_j} = \hat{\beta}_{x_j} [x_j - E(X_j)], \tag{9}$$

where $\hat{\beta_{x_j}}$ is the average OLS coefficient corresponding to different draws, defined in Equation 5.

Recall that under a standard linear framework, SVD possesses a valuable characteristic – the difference between the final prediction and the historical average predictions is fairly distributed among the covariates (individual forecasts). Maintaining this property is crucial when combining SVD with RSM, as it is not inherently fulfilled by default.

The explanation is based on the following observation: $\hat{\beta}_{x_j}$ and the final prediction $\hat{f}(x)$ are averaged across different numbers. That is, $\hat{\beta}_{x_j}$ is divided by the n_{x_j}, the number of times a variable has been selected. Meanwhile, $\hat{f}(x)$ is divided by M, where M is the number of iterations. To satisfy the property, one should calculate the following value:

$$\phi_j^{RSM} = \frac{n_{x_j}}{M}\bar{\phi_j} \tag{10}$$

It is possible to interpret ϕ_j^{RSM} as an (average) Shapley value corresponding to the variable x_j in the RSM aggregation framework. It is worth observing that if a variable x_j is not selected in the r iteration using the approach, this implies that the corresponding coefficient $\beta_{x_j}^r = 0$. This means that for any x_j, there are n_{x_j} non-zero $\beta_{x_j}^r$ and $M - n_{x_j}$ amount of coefficients, which equal to zero.

4 Data and Methods

For the empirical application, the monthly dataset FRED-MD provided by McCracken and Ng (2016) to nowcast the quarterly US GDP growth rate was used.[4] In total, this resulted in 117 monthly indicators, which are used in the study. The evaluation period is Q4 1988–Q1 2023 (138 quarters). Finally, an out-of-sample forecast is performed.

One of the main advantages of the proposed methodology is that it is model-agnostic. This indicates that the nature of the set of predictions is not critical. However, a relatively large number of forecasts must be generated. For example, the Survey of Professional Forecasters (SPF) does not fulfill the desired property as the participants may change from one wave to another. As a result, a history of predictions for a given forecaster could often be too short.

Following Andreou et al. (2013), it is possible to consider only one indicator at a time to predict a target variable. Under these settings, by exploiting bivariate methods, one can obtain

[4] All three months within a quarter of each time series are assumed to be known. It means, in this paper, the ragged edge problem, which appears with differences in publication lags among the variables is not considered. This assumption seems less restrictive for the US case than for other countries.

as many forecasts as there are independent variables. In principle, any model can be used to generate bivariate predictions. For this purpose, a nowcasting model based on a generalized regression neural network (GRNN) is applied, as described in Martínez et al. (2022).

There are several reasons for selecting this method. First, GRNN is non-parametric, easy to compute, and generally more flexible than usual parametric approaches. Moreover, the initial distribution of forecasts, as shown in Section 5.2, is characterized by having outliers. The presence of outliers may cause issues while combing forecasts by assigning equal weights (i.e., by taking a mean). Even if outliers are discarded, the prevalence of indicators from the same group can result in a skewed distribution. RSM can be particularly effective in coping with these scenarios. Finally, It is shown that a median of the set generated by GRNN can produce a reliable GDP nowcast for Germany (Holtemöller and Kozyrev, 2024). However, no such results are available for the US data.

First, the initial model is estimated to study the forecasting performance of the GRNN approach. Later, a slightly modified variable selection method is exploited, as described in Medeiros et al. (2021), to study the effect of reducing a set of forecasts on forecast quality. Then, RSM was applied to carry out forecast aggregation. Finally, based on the RSM estimations, SVD and its dynamics over time were analyzed.

GRNN is a variation of a radial basis neural network proposed by Specht et al. (1991). It relies on the idea that a prediction for a given data point x_i can be computed as a weighted average of all previous values based on their proximity to x_i. This model shares some similarities with the Nadaraya-Watson Gaussian Kernel Regression estimator. To obtain a prediction, only two parameters need to be defined: the number of lags (d) and a smoothing parameter (σ).

Suppose a training set consisting of n training patterns $\{x_1, x_2, \ldots, x_n\}$ and corresponding targets $\{y_1, y_2, \ldots, y_n\}$ are given. A corresponding weight w_i could be expressed as follows:

$$w_i = \frac{exp(-\frac{||x-x_i||^2}{2\sigma^2})}{\sum_{l=1}^{n} exp(-\frac{||x-x_l||^2}{2\sigma^2})}, \tag{11}$$

where $||\cdot||$ is the Euclidian distance.

The forecast is a weighted sum of training target outputs:

$$\hat{y} = \sum_{i=1}^{n} w_i y_i, \tag{12}$$

where y_i is the target output for training data x_i.

Some adjustments are proposed to apply a univariate GRNN approach to mixed-frequency and multivariate settings. First, GRNN cannot predict an observation that is out of range of previous observations. Thus, the additive and exponential data transformations are introduced, see Martínez et al. (2022). Then, actual (level) values of monthly time series from the FRED-MD $y_{t,n}$, $n = 1, \ldots, N$ are aggregated to match the quarterly frequency of GDP. The aggregated time series is denoted as $y^Q_{t,n}$, and a set of the aggregated time series $S_{y_t} = \{y^Q_{t,1}, y^Q_{t,2}, ..., y^Q_{t,N}\}$. Additionally, $\tilde{y}^Q_{t,n}$ is defined as follows:

$$\tilde{y}^Q_{t,n} = \frac{y^Q_{t,n}}{GDP_t} \quad \forall y^Q_{t,n} \in S_{y_t} \tag{13}$$

Each element $\tilde{y}^Q_{t,n} \in S_{\tilde{y}}$ is separately predicted by a univariate GRNN model.[5] The obtained nowcast and a set of all "bivariate" nowcasts are denoted as $\hat{\tilde{y}}^Q_{t,n}$ and $S_{\hat{\tilde{y}}} = \{\hat{\tilde{y}}^Q_{t,1}, \hat{\tilde{y}}^Q_{t,2} ... \hat{\tilde{y}}^Q_{t,N}\}$, respectively. Lastly, it is possible to retrieve GDP nowcasts by using the following relation:

$$\widehat{GDP}_{t,n} = \frac{y^Q_{t,n}}{\hat{\tilde{y}}^Q_{t,n}} \tag{14}$$

Thus, one ends up with N different "bivariate" GDP nowcasts for a given quarter t. All the nowcasts are collected in a set $S_{GDP} = \{\widehat{GDP}_{t,1}, \widehat{GDP}_{t,2} ... \widehat{GDP}_{t,N}\}$.

To summarize, the following algorithm is proposed:

- Obtain any set of individual forecasts (in this case, by applying a univariate GRNN model);
- Apply RSM with different k and fixed M.
- Estimate ϕ^{RSM}_j to analyze the contribution of each individual forecast to the final prediction.

5 Empirical Results

5.1 GRNN Nowcasting and Variable Selection

First, it is investigated whether nowcasting the GDP using GRNN could improve forecast precision compared to an AR(1) benchmark. The forecasting performance of the models is evaluated by using two measures, i.e., Relative Mean Absolute Forecast Error ($rMAFE$) and Relative Root Mean Squared Forecast Error ($rRMSFE$). Both metrics are defined as follows[6]:

[5] The selection of d, σ, and how to deal with non-stationary data is described in Martínez et al. (2022).

[6] Note, when nowcasting $h = 0$

$$
\begin{aligned}
rMAFE_h &= \frac{\sum_{i=1}^{N} |y_{i,t+h} - \hat{y}_{i,t+h|t}^{GRNN}|}{\sum_{i=1}^{N} |y_{i,t+h} - \hat{y}_{i,t+h|t}^{AR(1)}|} = \frac{MAFE_h^{GRNN}}{MAFE_h^{AR(1)}} \\
rRMSFE_h &= \frac{\sqrt{\sum_{i=1}^{N} (y_{i,t+h} - \hat{y}_{i,t+h|t}^{GRNN})^2}}{\sqrt{\sum_{i=1}^{N} (y_{i,t+h} - \hat{y}_{i,t+h|t}^{AR(1)})^2}} = \frac{RMSFE_h^{GRNN}}{RMSFE_h^{AR(1)}},
\end{aligned}
\tag{15}
$$

The second and the third columns in Table 1 depict the forecasting performance relative to AR(1) of the GRNN approach. As shown, both $rMAFE$ and $rRMSFE$ are below one, indicating that GRNN outperforms AR(1) regarding forecast accuracy. The notable difference between $rMAFE$ and $rRMSFE$ means that the AR(1) is more sensitive to outliers or that there are significant deviations from standard behavior. Intuitively, this could be explained by the relatively large evaluation period, which includes a few periods when GDP dynamics were far from being "normal" (to name some, the financial crisis of 2008, the COVID recession, and the recovery afterward).

Model	Amount	rMAFE	rRMSFE
Multiplicative	Full	0.710	0.429
Additive	Full	**0.696**	0.428
Multiplicative	Reduced	0.699	**0.426**
Additive	Reduced	0.700	0.455

Table 1: The Best RSM Aggregated Models (ranked by rMAFE)

Even though the introduced approach can handle sufficiently large datasets, it may seem particularly instrumental in eliminating some of the most irrelevant (redundant) forecasts. For instance, a similar idea is described in Ding et al. (2023), where a superior subgroup of forecasts was to be discovered. However, this path is not followed in this study because the relationships between variables may not hold over time. It is impossible to know in advance what kind of data would be useful to predict the GDP growth rate within a quarter. This means that relying on the historical accuracy of the individual forecast does not guarantee its relevance, especially during different shocks. Therefore, only irrelevant forecasts for a given quarter are eliminated.

The variable selection method used in this study is similar to that employed in Bai and Ng (2008) for target predictors and in Medeiros et al. (2021) for complete subset regressions. Given the close relationship between complete subset regressions and RSM, using similar variable selection methods is reasonable. The variable selection could be described as follows:

1. Run bivariate regressions of each individual forecast (without any additional lags or another covariate) on the observed values.

2. Store the t-statistics of each variable and rank them by absolute value.

3. Select only those variables that are significant at the 1% level.

The last step is modified compared to Medeiros et al. (2021). When using complete subset regressions, having a small number of indicators is crucial. Which is why, initially, only $\tilde{n}$ variables with the highest t-statistics were chosen. However, since RSM overcomes the problem of high-dimensional data, it is possible to be less strict and select every statistically significant forecast. This procedure is applied for each period in the evaluation set. That means that the number of forecasts used under the variable selection procedure differs occasionally.

The third and the fourth columns in Table 1 show the forecasting performance of the GRNN approach relative to AR(1) based on the reduced dataset. The forecasting performance is almost unchanged. That is why, it is possible to alter the amount of forecasts while preserving or even improving the forecasting performance.

5.2 RSM Aggregation Performance

As shown previously, taking a median of the individual forecasts is useful when nowcasting using GRNN. However, can better estimations be achieved using a less “naive” approach? Figure 1 presents a histogram of GRNN predictions (for Q1 2023) to illustrate the intuition and why it may be the case.

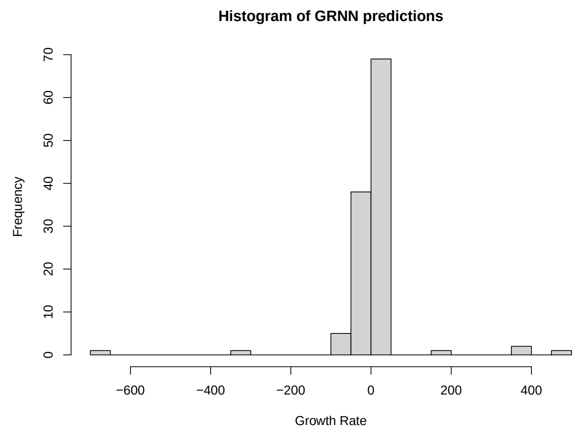

Figure 1: Histogram of GRNN predictions

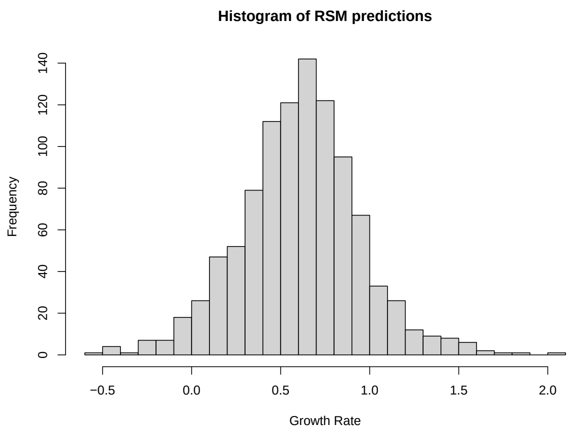

Figure 2: Histogram of RSM predictions

Figure 1 shows that even though the majority of the individual forecasts lie within the “usual” GDP growth rate bounds (e.g., between -2 and 2), there are some outliers. Due to these outliers, taking a mean of the set becomes uninformative. Although the median works empirically better than the selected benchmark, as described in the previous section, there appears to be room to improve the forecasts. To further clarify, Figure 2 depicts the distribution of the forecasts using RSM for the same set of individual forecasts.

The distribution mentioned in Figure 2 looks similar to a normal one with almost no severe outliers. Thus, the RSM approach can serve to smooth the initial forecasts' distribution (if necessary, as in this case). That is why one should not be concerned with taking a mean of these forecasts as they are all centered around it. Undoubtedly, decision-makers would likely prefer working with normally distributed forecasts rather than seeking various methods to exclude potential outliers.

To study RSM for forecast aggregation, 50 different models are run, varying across different characteristics. More specifically, different combinations of the following features were estimated:

- **Model specification**: GRNN Additive or GRNN Multiplicative
- **Type of data**: Level or growth rates. Since nowcasting using GRNN aims to predict the level value of a target variable, it is possible to convert level data to the corresponding growth rates. Intuitively, it is more useful to decompose the difference between the final and the historical average predictions using growth rates since many indicators (including GDP) depict a linear or exponential trend.
- **Amount of data**: Full or reduced according to a variable selection procedure, as described in Section 5.1.
- **The number of indicators** k: 5, 10, 20, 30, 40.

Table 2 presents the best RSM aggregated models across various model specifications, data selections, and values of k in terms of $rMAFE$ and $rRMSFE$. Notably, the RSM aggregation method manages to achieve lower values for $rMAFE$ and $rRMSFE$ for selected specifications.

Model	Type	Amount	Selected k	rMAFE	rRMSFE
Additive	Growth rates	Reduced	10	0.556	0.299
Multiplicative	Growth rates	Reduced	10	0.557	0.302
Additive	Growth rates	Full	10	0.561	0.321
Additive	Level	Reduced	10	0.568	0.3
Additive	Growth rates	Reduced	5	0.570	0.334
Multiplicative	Growth rates	Reduced	5	0.572	0.34
Additive	Growth rates	Reduced	20	0.581	0.332
Multiplicative	Level	Reduced	10	0.583	0.306
Additive	Level	Reduced	20	0.584	0.306

Table 2: The Best RSM Aggregated Models (ranked by rMAFE)

Based on Table 2, the results can be summarized as follows:

- **Model**: Both additive and multiplicative specifications perform relatively well. However, additive specifications seem to be slightly superior. The best model is GRNN Additive. When growth rates are used as input data, data is reduced, and $k = 10$ are taken for each iteration. However, GRNN Multiplicative with the same choices selected is the second best model with a slight decrease in both forecasting error statistics.

- **Type of data**: Forecasting the GDP growth rate based on growth data is a more appealing strategy rather than using level values. Correspondingly, this makes it possible to use growth rates because SVD is more tractable while explaining the discrepancy, as mentioned above.

- **Amount of data**: The forecasting performance based on a reduced dataset outperforms the same model trained by using the entire dataset. This potentially indicates that redundant forecasts exist, and the introduced variable selection scheme can detect and eliminate those for any quarter.

- **Selected *k***: The results suggest that taking fewer individual forecasts for RSM aggregation may increase the forecasting performance. These findings contradict those mentioned in Boot and Nibbering (2019). They suggest that the dimension of the subspace should be chosen relatively large. This discrepancy between the results can be explained by two observations. First, RSM aggregation is performed on an evaluation set, which is rather small for effectively estimating OLS using a relatively large k. The second reason may be related to the type of data used. Initially, RSM is constructed using actual observations rather than derived forecasts based on these observations. That is, on average, most of the forecasts are assumed to be close to the observed values of the independent variable. Thus, a lower amount of covariates is required to undercover the dynamics of the target variable.

5.3 OLS Coefficients

Before analyzing SVD, some points should be mentioned regarding the distribution of OLS coefficients for a given variable under RSM aggregation. To my knowledge, the statistical properties of $\hat{\beta}_{x_j}$, as defined in Equation 5, have not been well-studied. Buckland et al. (1997) derived an expression for the variance of the estimator under a couple of strong assumptions. However, no studies were conducted regarding the distribution or the bias of $\hat{\beta}_{x_j}$. In this subsection, the distributions of $\hat{\beta}_{x_j}$, based on different subsamples, are described empirically.

For this purpose, using the best-performing model, the most and the least volatile distributions, that arise from predictions of the Q1 2023 US GDP growth rate (the last period in the evaluation sample) are selected. Some other distributions can be seen in Appendix A. Figure 3 depicts the distribution of OLS coefficients of the Total Nonfarm Payroll, denoted as PAYEMS. This is a measure of the number of U.S. workers in the economy that excludes proprietors, private household employees, unpaid volunteers, farm employees, and the unincorporated self-employed[7]. Figure 4 illustrates the same distribution for Reserves of Depository Institutions[8], denoted as TOTRESNS.

[7] https://fred.stlouisfed.org/series/PAYEMS
[8] https://fred.stlouisfed.org/series/TOTRESNS

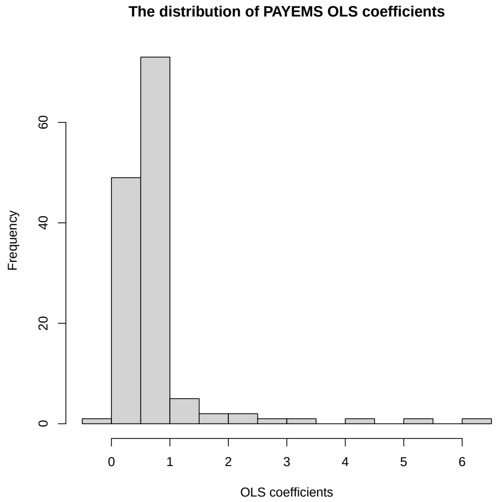

Figure 3: PAYEMS OLS coefficients

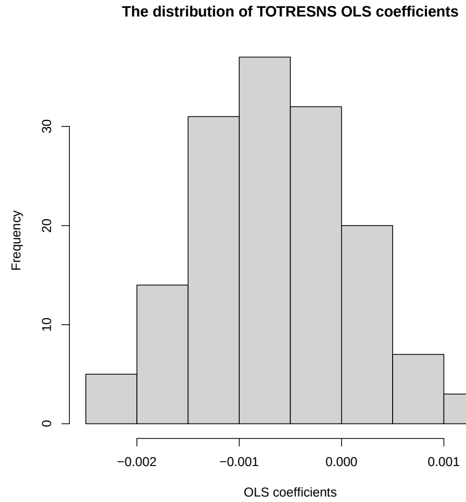

Figure 4: TOTRESNS OLS coefficients

Figure 3 shows that in some cases $\hat{\beta}_{x_j}$ may indeed be volatile. The variance of $\hat{\beta}_{x_{PAYEMS}}$ across different iterations is 0.649. However, this volatility is seemingly caused by some outliers, whose proportion to the total number of estimations is relatively small. Most of $\hat{\beta}_{x_{PAYEMS}}$, as observed, are between zero and one. However, in seven out of 137 cases, $\hat{\beta}_{x_{PAYEMS}}$ was estimated to be greater than two. This may occur when randomly selected variables, apart from PAYEMS, are less informative. In this case, PAYEMS has a greater weight than all other covariates. However, this occurs not often because, on average, all chosen variables are supposed to be more or less relevant for predicting the independent variable. Eventually, some other employment-related variables possess huge volatility regarding OLS coefficients. Thus, All Employees, Manufacturing[9], denoted MANEMP, and All Employees, Durable Goods[10], denoted DMANEMP, provide the third and the fourth most volatile OLS estimations (0.467 and 0.185, respectively).

On the other hand, one may notice that the distribution depicted in Figure 4 appears similar to Normal distribution with low variance (5.746×10^7). For most individual forecasts, the distribution of OLS coefficients behaves similarly. The average variance of OLS estimations across all variables is 0.04. This fact, along with the bell-curved distribution, suggests that, on average, almost all $\hat{\beta}^r_{x_j}$ and $\hat{\beta}^q_{x_j}$ for $r \neq q$ (OLS coefficients associated with different iterations) are strongly correlated, as expected.

In conclusion, based on the empirical evidence, $\hat{\beta}_{x_j}$ is assumed to behave in a plausible way to use them for SVD.

[9]https://fred.stlouisfed.org/series/MANEMP

[10]https://fred.stlouisfed.org/series/DMANEMP

5.4 SVD

In the previous sections, RSM aggregation is shown to improve forecasting performance per se. However, this forecast combination method also allows analysis of each individual forecast contribution to the final prediction. This may be particularly interesting for various decision-makers since SVD can determine the relative importance of a given forecast. Importantly, the more covariates are used for predictions, the less the corresponding contribution will be. Even though the proposed mechanism allows for the use of arbitrarily large numbers of individual forecasts, in many practical cases, this amount can be limited to a much smaller number (for instance, by prior knowledge).

In this study, all available forecasts (upon variable selection) are used to be more flexible. As previously mentioned, it is impossible to know in advance which indicator will be most beneficial for nowcasting GDP in a given quarter. That is why, it is expected that ϕ_j^{RSM} is typically close to zero.

Below, ϕ_j^{RSM} over the entire evaluation sample is presented for two different individual forecasts, based on industrial production: Total Index[11], denoted INDPRO, and Average Weeks Unemployed[12], denoted UEMPMEAN, to study the potential differences in the underlying dynamics. These values were obtained by the best-performing model (GRNN Additive, reduced data set of growth rates, $k = 10$). Similar plots for some other variables can be found in Appendix B.

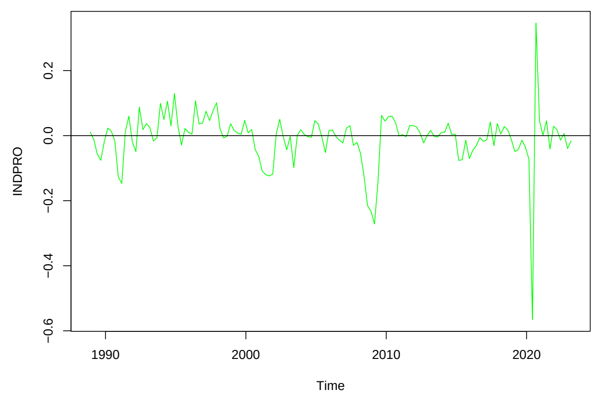

Figure 5: Average Shapley Value of INDPRO

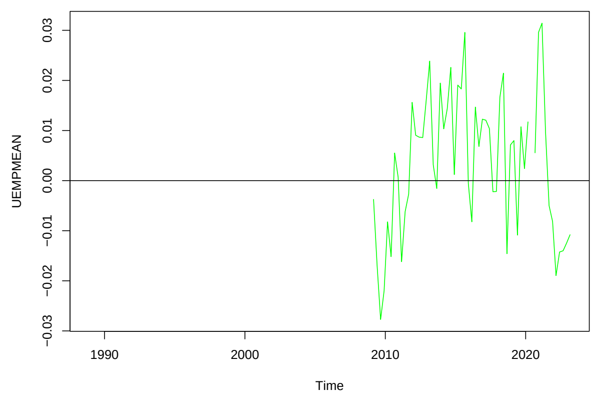

Figure 6: Average Shapley Value of UEMPMEAN

In Figure 5, ϕ_{INDPRO}^{RSM} is presented. Industrial production is a crucial variable for assessing current economic activity see, for instance, Eraslan and Götz (2021). That is why, it is unsur-

[11] https://fred.stlouisfed.org/series/INDPRO
[12] https://fred.stlouisfed.org/series/UEMPMEAN

prising that the forecasts based on this specific variable have been chosen for each quarter over the evaluation set. Recall that Shapley values, defined in Equations 9 and 10, depend on the feature effect minus the average effect. Thus, the contribution of the individual forecast is high when this difference is high (given that OLS coefficients are relatively small, which is usually the case).

A couple of extreme values become clear from the graph: they took place during 2008-2009 and the first quarters of 2020. These ϕ^{RSM}_{INDPRO} align with what was occurring during these periods. In 2009, there was a recession after the financial crisis, and in the first quarter of 2020, the COVID pandemic caused a significant decrease in production. Thus, the contribution of industrial production was negative during that time. Moreover, the negative ϕ^{RSM}_{INDPRO} observed for 2001 can be attributed to a recession that occurred in that year (Brown, 2009). Meanwhile, a rebound in economic activity was observed in the second quarter of 2020, reflected in the industrial production forecast. Overall, it is possible to notice that, on average, industrial production contributes quite significantly.

Meanwhile, Figure 6 shows $\phi^{RSM}_{UEMPMEAN}$. Compared to the previous case, the first difference is that UEMPMEAN was not selected until 2009. It may hint at the fact that previously this variable did not provide reliable forecasts. Furthermore, as shown on the y-axis, the contribution of the forecast based on UEMPMEAN is almost negligible. This indicates that UEMPMEAN typically reflects an average GDP growth rate for any given quarter, thus failing to capture various fluctuations of GDP over time.

Additionally, one can examine the ϕ^{RSM}_{j} across different covariates for some of the most challenging quarters, that is Q1 2009 and Q2 2020 to study the relative variable importance. In Figure 9, SVD for Q1 2009 is presented. The most negative contribution came from the forecasts, derived by production-related variables (IPDMAT[13] (Industrial Production: Durable Goods Materials), INDPRO, IPMAT[14] (Industrial Production: Materials), IPMANSICS[15] (Industrial Production: Manufacturing).

[13]https://fred.stlouisfed.org/series/IPDMAT
[14]https://fred.stlouisfed.org/series/IPMAT
[15]https://fred.stlouisfed.org/series/IPMANSICS

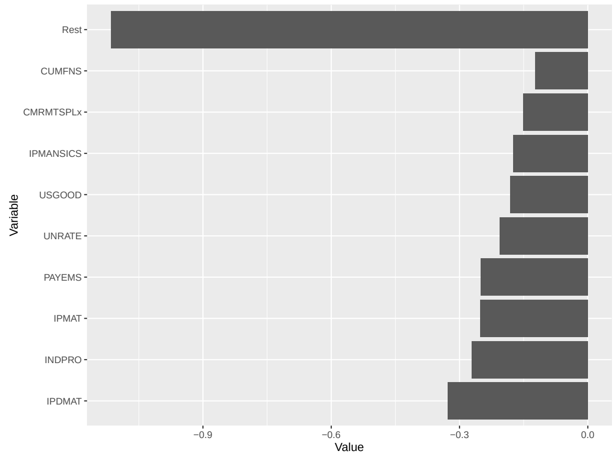

Figure 7: SVD for Q1 2009

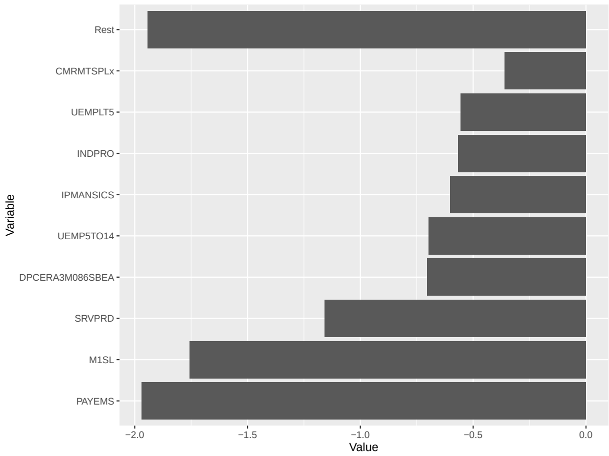

Figure 8: SVD for Q2 2020

Figure 10 presents the same decomposition for Q2 2020. Some industrial production variables are depicted here in a way that is similar to the 2009 recession. However, more labor-related forecasts negatively contributed to (PAYEMS, SRVPRD[16] (All Employees, Service-Providing), UEMP5TO14[17] (Number Unemployed for 5-14 Weeks), UEMPLT5[18] (Number Unemployed for Less Than 5 Weeks). Some other extreme values of ϕ_j^{RSM} over the evaluation set are presented in Appendix C.

6 Conclusion

In this study, an opportunity to use a forecast aggregation method based on RSM is explored, and the connection between RSM and SVD is provided. First, a potential implementation of RSM for forecast combination is analyzed. When the number of individual forecasts is relatively large, it is possible to randomly and uniformly draw much smaller subsets of them to estimate lower dimensional problems. This method circumvents high dimensionality and enables working with non-sparse data. Forecasting performance can be improved using RSM rather than taking a median of the initial set. Moreover, the introduced method allows for an average of the newly produced forecasts to be taken. Alternatively, it is possible to transform a set of individual predictions in a way that makes taking an average sensible, even if initially it does not produce a reliable estimation. Thus, this contributes to the "forecast combination puzzle" literature.

A potential connection between RSM and SVD is addressed. It is possible to implement SVD into the framework of RSM aggregation to study the relative importance of individual

[16] https://fred.stlouisfed.org/series/SRVPRD

[17] https://fred.stlouisfed.org/series/UEMP5TO14

[18] https://fred.stlouisfed.org/series/UEMPLT5

forecasts, thus adding some interpretability to the final estimation. By decomposing the final prediction, one can analyze each forecast contribution, which makes this method appealing to governments and institutions. Specifically, SVD for the recession, which took place in 2009, as well as the COVID-19 recession, are studied. The results indicate that both industrial production variables and different employment covariates contribute the most.

There are some additional research questions not covered in this paper. Seemingly, there is evidence that some indicators are more important than others. Thus, it may sound intuitive to assign greater probabilities to those covariates that possess more forecasting power. Then, if one can first select a smaller subset of forecasts to start with the corresponding SVD values – the relative importance – may be increased in absolute terms. Finally, more research can be conducted on the distribution of OLS coefficients within the RS method.

References

Elena Andreou, Eric Ghysels, and Andros Kourtellos. Should macroeconomic forecasters use daily financial data and how? *Journal of Business & Economic Statistics*, 31(2):240–251, 2013.

Jushan Bai and Serena Ng. Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317, 2008.

Stephen D Bay. Combining nearest neighbor classifiers through multiple feature subsets. In *ICML*, volume 98, pages 37–45. Citeseer, 1998.

Tom Boot and Didier Nibbering. Forecasting using random subspace methods. *Journal of Econometrics*, 209(2):391–406, 2019.

Tom Boot and Didier Nibbering. Subspace methods. *Macroeconomic Forecasting in the Era of Big Data: Theory and Practice*, pages 267–291, 2020.

Richard A Brown. The 2009 economic landscape: how the recession is unfolding across four us regions. *FDIC Quarterly*, 3(1):26–38, 2009.

Steven T Buckland, Kenneth P Burnham, and Nicole H Augustin. Model selection: an integral part of inference. *Biometrics*, pages 603–618, 1997.

Marcus Buckmann, Andreas Joseph, and Helena Robertson. An interpretable machine learning workflow with an application to economic forecasting. Technical report, Bank of England, 2022.

Stephen Burgess, Emilio Fernandez-Corugedo, Charlotta Groth, Richard Harrison, Francesca Monti, Konstantinos Theodoridis, Matt Waldron, et al. *The Bank of England's forecasting platform: COMPASS, MAPS, EASE and the suite of models*. SSRN, 2013.

Lilian M De Menezes, Derek W Bunn, and James W Taylor. Review of guidelines for the use of combined forecasts. *European Journal of Operational Research*, 120(1):190–204, 2000.

Zhenni Ding, Huayou Chen, and Ligang Zhou. Using shapely values to define subgroups of forecasts for combining. *Journal of Forecasting*, 42(4):905–923, 2023.

Viet Hoang Dinh, Didier Nibbering, and Benjamin Wong. Random subspace local projections. *Review of Economics and Statistics*, pages 1–33, 2024.

Petros Drineas, Ravi Kannan, and Michael W Mahoney. Fast monte carlo algorithms for matrices i: Approximating matrix multiplication. *SIAM Journal on Computing*, 36(1):132–157, 2006.

Graham Elliott, Antonio Gargano, and Allan Timmermann. Complete subset regressions. *Journal of Econometrics*, 177(2):357–373, 2013.

Sercan Eraslan and Thomas Götz. An unconventional weekly economic activity index for Germany. *Economics Letters*, 204:109881, 2021.

Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

Iffat A Gheyas and Leslie S Smith. A neural network approach to time series forecasting. In *Proceedings of the World Congress on Engineering*, volume 2, pages 1–3, 2009.

Clive WJ Granger and Ramu Ramanathan. Improved methods of combining forecasts. *Journal of forecasting*, 3(2):197–204, 1984.

Tin Kam Ho. The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence*, 20(8):832–844, 1998.

Oliver Holtemöller and Boris Kozyrev. Forecasting economic activity using a neural network in uncertain times: Monte Carlo evidence and application to the German GDP. IWH Discussion Papers 6/2024, Halle Institute for Economic Research (IWH), 2024.

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

Scott M Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. From local explanations to global understanding with explainable ai for trees. *Nature machine intelligence*, 2(1):56–67, 2020.

Michael W Mahoney et al. Randomized algorithms for matrices and data. *Foundations and Trends® in Machine Learning*, 3(2):123–224, 2011.

Francisco Martínez, Francisco Charte, María Pilar Frías, and Ana María Martínez-Rodríguez. Strategies for time series forecasting with generalized regression neural networks. *Neurocomputing*, 491:509–521, 2022.

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.

Loukia Meligkotsidou, Ekaterini Panopoulou, Ioannis D Vrontos, and Spyridon D Vrontos. Out-of-sample equity premium prediction: A complete subset quantile regression approach. *The European Journal of Finance*, 27(1-2):110–135, 2021.

Christoph Molnar. *Interpretable machine learning*. Lulu. com, 2020.

Andreas Pick and Matthijs Carpay. Multi-step forecasting with large vector autoregressions. In *Essays in Honor of M. Hashem Pesaran: Prediction and Macro Modeling*, pages 73–98. Emerald Publishing Limited, 2022.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. Model-agnostic interpretability of machine learning. *arXiv preprint arXiv:1606.05386*, 2016a.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016b.

Lloyd S Shapley et al. A value for n-person games. 1953.

Donald F Specht et al. A general regression neural network. *IEEE transactions on neural networks*, 2(6):568–576, 1991.

James H Stock and Mark W Watson. Forecasting inflation. *Journal of monetary economics*, 44(2):293–335, 1999.

James H Stock and Mark W Watson. Combination forecasts of output growth in a seven-country data set. *Journal of forecasting*, 23(6):405–430, 2004.

Erik Strumbelj and Igor Kononenko. An efficient explanation of individual classifications using game theory. *The Journal of Machine Learning Research*, 11:1–18, 2010.

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

Allan Timmermann. Forecast combinations. *Handbook of economic forecasting*, 1:135–196, 2006.

Xiaoqian Wang, Rob J Hyndman, Feng Li, and Yanfei Kang. Forecast combinations: an over 50-year review. *International Journal of Forecasting*, 39(4):1518–1547, 2023.

Nuttanan Wichitaksorn, Yingyue Kang, and Faqiang Zhang. Random feature selection using random subspace logistic regression. *Expert Systems with Applications*, 217:119535, 2023.

Luisa M Zintgraf, Taco S Cohen, Tameem Adel, and Max Welling. Visualizing deep neural network decisions: Prediction difference analysis. *arXiv preprint arXiv:1702.04595*, 2017.

Appendices

A Different Distributions of OLS estimations

The distributions of OLS estimations based on different subsamples are further explored.The focus is on the most and least volatile distributions. Figure 9 illustrates the OLS estimations of New Privately-Owned Housing Units Started: Total Units in the Midwest Census Region[19] (HOUSTMW). Meanwhile, Figure 10 represents the same statistics for CLAIMSx. Both distributions are among the least volatile ones.

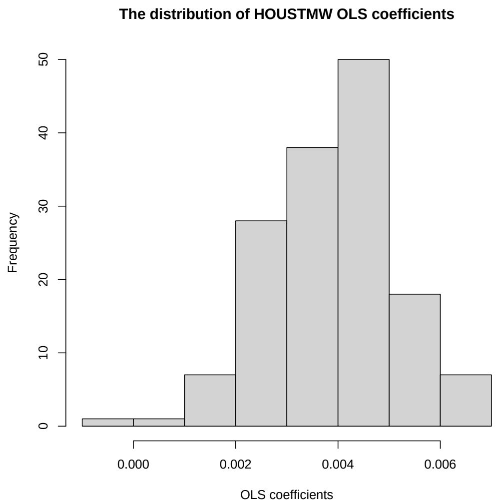

Figure 9: HOUSTMW OLS coefficients

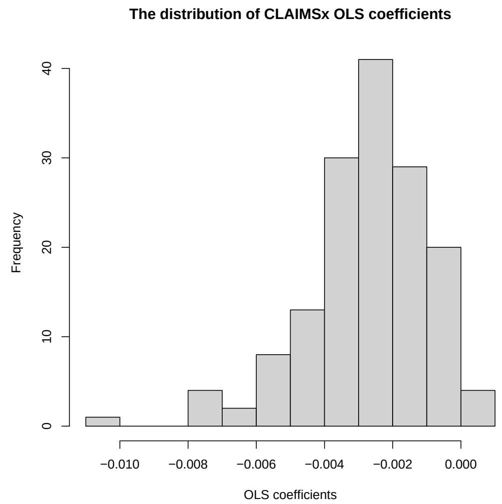

Figure 10: CLAIMSx OLS coefficients

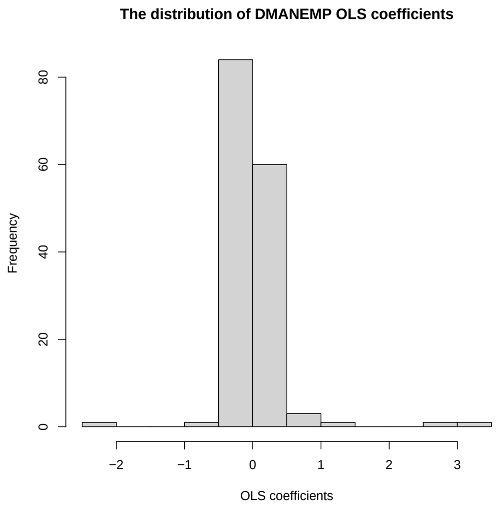

Figure 11: DMANEMP OLS coefficients

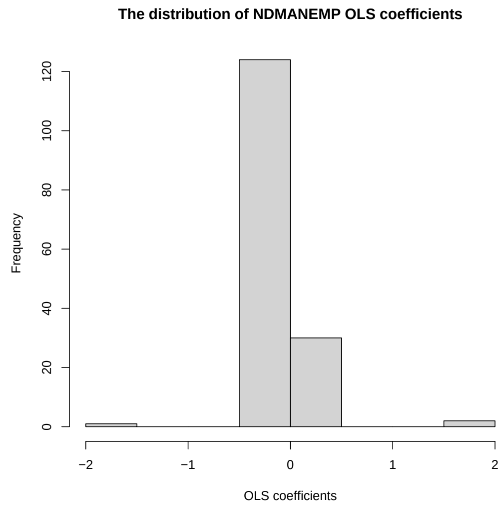

Figure 12: NDMANEMP OLS coefficients

[19]https://fred.stlouisfed.org/series/HOUSTMW

Both of these distributions belong to the least volative ones (their respective variances are 1.53×10^6 and 3.25×10^6). As argued earlier, most covariates have a similar distribution with almost no outliers, which may potentially affect the mean estimation.

Conversely, Figure 11 and Figure 12 display the distributions with the highest variance (DMANEMP and All Employees, Nondurable Goods[20] (NDMANEMP), respectively). As mentioned in the text, the most volatile distributions correspond to employment-related variables for a given period. As shown, the number of outliers is lower compared to the PAYEMS case, indicating that the role of such outliers is indeed negligible when dividing by the number of iterations.

[20]https://fred.stlouisfed.org/series/NDMANEMP

B Different ϕ_j^{RSM} over Time

Additional ϕ_j^{RSM} values are depicted to analyze the potential differences across different variables. Figure 13 presents SVD for Retail and Food Services[21] (RETAILx). Forecasts, based on this indicator, seem to be essential for nowcasting the US GDP growth rate. Consequently, this covariate has a long history of being selected in the SVD analysis. The same conclusion holds for All Employees, Retail Trade[22] (USTRADE), as shown in Figure 14.

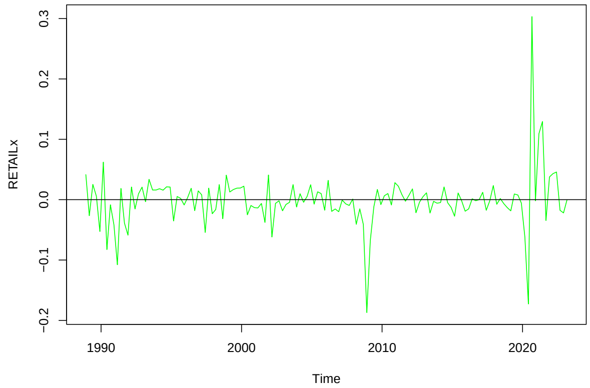

Figure 13: $\phi_{RETAILx}^{RSM}$

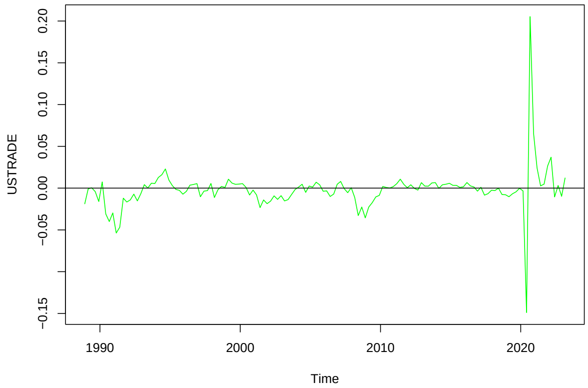

Figure 14: $\phi_{USTRADE}^{RSM}$

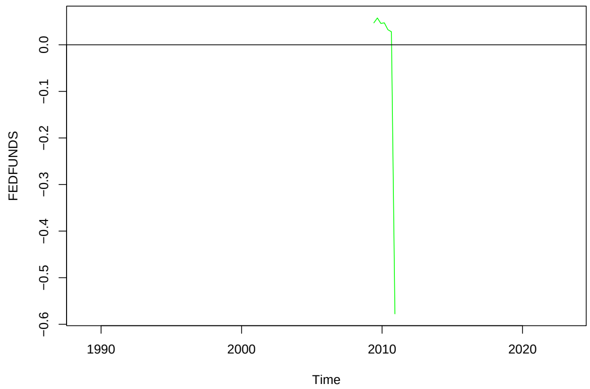

Figure 15: $\phi_{FEDFUNDS}^{RSM}$

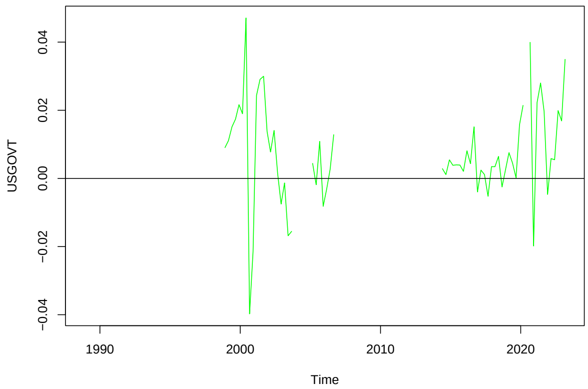

Figure 16: ϕ_{USGOVT}^{RSM}

Recall, that the variable selection method proceeds to choose a different set of covariates for a given quarter. Therefore, a given indicator may or may not be selected. This implies that if ϕ_j^{RSM} values are plotted over time, there might be gaps when this variable was not selected for nowcasting the US GDP growth rate These irregularities are depicted in Figure 15 for the SVD

[21] https://fred.stlouisfed.org/series/RSAFS
[22] https://fred.stlouisfed.org/series/USTRADE/1000

values of FEDFUNDS and in Figure 16, which presents the contributions of All Employees, Government[23] (USGOVT) are presented. It is particularly noteworthy that FEDFUNDS rarely reflects the dynamics of the US GDP growth rate according to the variable selection procedure. It was selected only for seven quarters out of 138. However, it was notably selected during the period leading up to the infamous 2011 United States debt-ceiling crisis. During that period, the contribution of FEDFUNDS seems to be significant.

[23] https://fred.stlouisfed.org/series/USGOVT

C Lowest and Highest ϕ_j^{RSM} over Time

Lowest ϕ_j^{RSM}			Highest ϕ_j^{RSM}		
Variable	Value	Period	Variable	Value	Period
PAYEMS	-1.97	Q2 2020	DPCERA3M086SBEA	0.47	Q3 2020
FEDFUNDS	-0.58	Q4 2010	MANEMP	0.13	Q1 1991
CLAIMSx	-0.35	Q1 2020	INDPRO	0.13	Q4 1994
IPDMAT	-0.32	Q1 2009	USGOOD	0.13	Q1 1993
CP3Mx	-0.24	Q2 2010	RETAILx	0.13	Q2 2022

Table 3: Some of the lowest and highest ϕ_j^{RSM}