BANK OF ENGLAND

Staff Working Paper No. 923

Forecasting UK GDP growth with large survey panels

Nikoleta Anesti, Eleni Kalamara and George Kapetanios

May 2021

BANK OF ENGLAND

Staff Working Paper No. 923

Forecasting UK GDP growth with large survey panels

Nikoleta Anesti,[(1)] Eleni Kalamara[(2)] and George Kapetanios[(3)]

Abstract

By employing large panels of survey data for the UK economy, we aim at reviewing linear approaches for regularisation and dimension reduction combined with techniques from the machine learning literature, like Random Forests, Support Vector Regressions and Neural Networks for forecasting GDP growth at monthly frequency for horizons from one month up to two years ahead. We compare the predictive content of surveys with text based indicators from newspaper articles and a standard macroeconomic data set and extend the empirical evidence on the contribution of survey data against text indicators and more traditional macroeconomic time series in predicting economic activity. Among the linear models, the Ridge and the Partial Least Squares models report the largest gains consistently for most of the forecasting horizons, and for the non-linear machine learning models, the SVR performs better at shorter horizons compared to the Neural Networks and Random Forest that seem to be more appropriate for longer-term forecasting. Text based indicators appear to favour more the use of non-linear models and the expansion of the information set with macroeconomic time series does not appear to add much more predictive power. The largest forecasting gains are overwhelmingly concentrated at the shorter horizons for the majority of models and datasets which provides further empirical support that non-linear machine learning models appear to be more useful during the Great Recession.

Key words: Forecasting, survey data, text indicators, machine learning.

JEL classification: C53, C55.

(1) Bank of England. Email: Nikoleta.Anesti@bankofengland.co.uk
(2) Bank of England. Email: Eleni.Kalamara@kcl.ac.uk
(3) Bank of England. Email: George.Kapetanios@bankofengland.co.uk

The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England or its committees. We are grateful to the conference participants in the International Symposium of Forecasting, Conference on Real-Time Data Analysis, Methods, and Applications and Bank of Canada for helpful comments.

The Bank's working paper series can be found at www.bankofengland.co.uk/working-paper/staff-working-papers

Bank of England, Threadneedle Street, London, EC2R 8AH
Email enquiries@bankofengland.co.uk

ISSN 1749-9135 (on-line)

1 Introduction

Extensive research undertaken by academics, central bankers and financial market participants has been focused on finding the best set of predictors to be used in monitoring and forecasting macroeconomic conditions, as they form the basis of informed economic and policy decisions and boost resilience to episodes of crises and recessions.

Many economic research institutions and central banks regularly produce and publish forecasts of economic variables. For example, the Bank of England publishes quarterly Monetary Policy Reports that set out the assessment of economic conditions and projections of key macroeconomic variables that the Monetary Policy Committee uses to make interest rate decisions.

Most traditional and widely used forecasting models for economic variables rely on fitting data to a pre-specified relationship between the input variables (indicators) and the output (target) variable. These models thereby make an implicit assumption of a stochastic process underlying the true relationship between the target variable and the indicators. A different approach to statistical analysis and forecasting more specifically is offered by machine learning algorithms, that make (almost) no assumption about the underlying relationship between the variables under consideration, but instead they rely on an algorithm to find a function that best describes the relationship between the indicators and the output data. While machine learning methods have been used in the past (for example Swanson and White (1997), Stock and Watson (1999), Nakamura (2005),Tersvirta et al. (2006), Marcellino (2008), De Mol et al. (2008)), it's only rather recently that a considerable number of studies have applied machine learning methods in the context of macroeconometrics, including macroeconomic forecasting [1]. While machine learning methods have been subject to the "black-box critique", implying a limited ability to interpret the factors that have been driving the forecasts, as Varian (2014) among others argues, growing amounts of data and complex non-linear economic relationships suggest the use of machine learning approaches in the context of macroeconomic forecasting. Meanwhile, studies like Joseph (2019) and Zhao and Hastie (2021) have contributed towards the interpretability of machine learning models.

[1] Ahmed et al. (2010), Stock and Watson (2012), Ng (2014), Smeekes and Wijler (2018), Diebold and Shin (2019), Coulombe et al. (2020), Medeiros et al. (2021), Joseph et al. (2021)

In this paper, we review standard linear approaches for regularisation and dimension reduction, like Lasso, Ridge and Elastic Net Regressions, Principal Components (PC) and Partial Least Squares (PLS) and provide empirical evidence on the predictive ability of these methods to forecast UK GDP growth up to 2 years ahead at a monthly frequency. We also use the same forecasting environment as a laboratory to test non-linear techniques drawn from the machine learning literature. In particular, we apply a set of supervised algorithms in the context of macroeconomic forecasting, more specifically Random Forests, Support Vector Machines (SVM) and Neural Networks.

The contribution of the paper is three-fold. First, we contribute to the extensive literature that explores the use of machine learning models in macroeconomic forecasting and more specifically in predicting monthly UK GDP growth from one month ahead up to 2 years using data from 2000 to 2018. We find that among the linear models the PLS, the Ridge and the Elastic Net based forecasts have better predictive content compared to both benchmark specifications, a Principal Component Regression (PCR) and a simple AR(1) model, for up to one-year ahead. Among the non linear models evaluated in this exercise, we find that the random forests, followed by the SVR yield the most accurate predictions in terms of RMSFE over the same period.

Secondly, in order to test the predictive ability of these linear and non-linear methods, we collect large panels of survey data for the UK economy from businesses and consumers that include questions about the current and future state of the economy, current and future orders, labour market prospects, consumer views on current and future financial situation over the next year. Although survey data have been widely used by central banks to forecast economic conditions (Anesti et al., 2017), mainly due to their timeliness compared to official releases that are published with substantial delays, we proceed with collecting information at a more disaggregate level. We take advantage of the richer information set that usually accompanies the survey releases, rather than just focusing on the headline balances. The motivation behind this is that different indicators might be more or less useful at different points in time or at different phases of the business cycle. Consequently, all of them might contain useful information for predicting economic activity at some point and as we are interested in testing linear and non-linear techniques that can handle large information sets, we can accommodate much more potentially useful predictors than in “traditional” forecasting models.

Recently, a new strand of research has been undertaken that explores text as an alternative high frequency data source to answer economic and policy related questions (see Gentzkow et al. (2019); Bholat et al. (2015), for a review). For this reason, we contrast the predictive content of the disaggregated survey balances with 15 text-based indicators suggested by Kalamara et al. (2020) that aim to capture uncertainty and sentiment in the UK economy. These metrics are calculated by applying existing text analytics methods to newspaper articles and modified appropriately to obtain valuable information in real time. As such, we create an alternative dataset that contains 'soft' information extracted from newspaper articles and conduct an extensive out-of-sample evaluation exercise to provide further empirical evidence on the use of surveys and text indicators for predicting economic activity by exploring different linear and non-linear machine learning techniques. Additionally, we examine the usefulness of “soft” information from business and consumer surveys, as opposed to information from text indicators and a wider macroeconomic and financial time series in the spirit of the Stock and Watson (2002c) datasets and we find that the richer information set only marginally improves our predictions for longer horizons forecasts, whereas for the shorter horizons (1 and 3 steps ahead) the surveys-only models outperform both the text and macro based forecasts.

Finally, we examine the role of the Great Recession as a potential source of unusually large forecast errors in our sample period and we find that the performance of the models is substantially affected by this period and it appears that the non-linear specifications are better suited to capture downturns in the data compared to their linear counterparts.

The rest of the paper is organised as follows. We describe the different models in Section 2 and explain how we choose the tuning parameters in Section 2.4. Section 3 provides a summary of the dataset we employed and Section 4 describes the main features of the out-of-sample forecasting exercises and reports on the empirical results. Appendices A and B contain a detailed description of the datasets and additional forecasting results.

2 Overview of the Models

In this Section, we discuss some linear methods, including two different classes to handle the high dimensionality of the dataset we consider. In the first class, we fit a model involving all V predictors but the estimated coefficients are shrunken towards to zero relative to the least squares estimate (shrinkage methods). In the second class, we project the V predictors into a K-dimensional subspace where $K < V$. This is achieved by computing K different linear combinations or projections of the variables. Then, these K projections are used as predictors to fit a linear regression model by least squares (dimensionality reduction methods).

Given the complexity of the economic system and the vast availability of the data, the linearity assumption is potentially restrictive, hence a variety of methods originating from the machine learning literature are becoming increasingly popular. Hence we also review some supervised machines algorithms here i.e, random forest regressions, neural networks and support vector regressions that allow us to incorporate the non-linearities but also to exploit the entire span of the independent variables without imposing that they all carry useful information for the prediction of the target variable (variable selection).

2.1 Shrinkage Methods

2.1.1 Ridge Regression

Shrinkage methods (or sparse regressions) have been suggested to produce effective estimates using different penalisation schemes. The main idea is that all the coefficients of the variables which are not part of the true model approach or become 0. Let C be a high dimensional matrix of predictors with $M \times V$ dimensions. We assume that we are interested in predicting y_i, the attribute of our interest, from the predictors c_{ij} where $i \in \{1, \ldots M\}$ and $j \in \{1, 2 \ldots V\}$. The target variable y_i is observed. The main aim of the methods discussed below, is to reduce the dimensions of the x_i matrix, producing regressors of smaller dimensions that are linear combinations of the original regressors and can be used for inference. Ridge regression penalises the residual sum of squares (RSS) with the sum of squared coefficients. This forces the coefficients with a minor contribution in the model to shrink substantially and approach zero, but never become exactly zero. Under our framework, the optimisation problem can be written as :

$$\beta^{Ridge} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i}^{M} (y_i - a - \sum_{j}^{V} \beta x_{ij})^2 + \lambda \sum_{j}^{V} \beta_j^2 \right\} \tag{1}$$

$$\tag{2}$$

for given values of α and $\lambda \geq 0$. The ridge regression coefficients estimates can substantially change when multiplying a given predictor by a constant, due to the sum of squared coefficients term in the penalty part of the ridge regression objective function. Therefore, it is typical to centre first the values of predictors and do not include the constant term. The parameter λ stands for the penalty imposed to coefficients and controls its overall magnitude (often called, tuning parameter).

It is worth mentioning that as $\lambda \to 0$, $\beta^{\hat{Ridge}} \to \beta^{\hat{OLS}}$ which is the no penalty case. Also, notice that as $\lambda \to \infty$ then $\beta^{\hat{Ridge}} \to 0$. Selecting a good value for λ is critical; there are several ways to choose the penalty parameter. Most commonly, researchers have been using cross-validation method that minimises the cross-validated squared error risk (or directly the MSE as suggested by Kapetanios et al. (2019)). As with least squares, ridge regression seeks coefficient estimates that fit the data well, by making the RSS small. However, the second term $\lambda \sum_j^V \beta_j^2$ is small when $\beta_1, \ldots \beta_V$ is close to zero and so it has the effect of shrinking the estimates of β_j towards zero.

2.1.2 Least Absolute Shrinkage and Selection Operator (LASSO)

Ridge Regression does have a distinct disadvantage; it includes all the V parameters in the model. The LASSO is a relatively recent alternative to the Ridge Regression that overcomes this obstacle (Tibshirani, 1996). LASSO regression penalises the sum of squared residuals with the L1 norm, i.e. the sum of absolute coefficients. In this case, some of the coefficients are set exactly to 0 which tends to give more parsimonious results. The LASSO estimators $\beta^{\hat{LASSO}}$ are then computed by solving the following optimisation problem:

$$\beta^{LASSO} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i}^{M} (y_i - a - \sum_{j}^{V} \beta x_{ij})^2 + \lambda \sum_{j}^{V} |\beta_j| \right\} \tag{3}$$

As with Ridge regression, the LASSO shrinks the coefficient estimates towards to zero. However, in the case of the LASSO, the L1 penalty has the effect of forcing some of the coefficient estimates to be equal to zero when the tuning parameter is sufficiently large. Hence, LASSO performs also, variable selection. We say that LASSO yields sparse models, i.e. models that involve only a subset of the variables. Just as in Ridge regression, we centre the values of the parameters and do not include the constant term. Cross-Validation again is preferred for the selection of the tuning parameter λ. The L1 LASSO penalty makes the solutions nonlinear in the y_i and there is no closed form, unlike ridge regression.

2.1.3 Elastic Net

A slightly different approach, called Elastic Net (Zou and Hastie, 2005) combines the L1 and L2 norms penalties. This method enjoys both the shrinkage of the coefficients (Ridge Regressions) and the variable selection (LASSO). The "naive" estimators of elastic net, $\beta^{naiveEN}$ are computed by solving the problem:

$$\beta^{naiveEN} = \min_{\beta}\{\sum_{i}^{M}(y_i - a - \sum_{j}^{V} x_{ij}\beta)^2 + \lambda_1 \sum_{j}^{V} \beta_j^2 + \lambda_2 \sum_{j}^{V} |\beta_j|\} \tag{4}$$

The naive version of elastic net method finds an estimator in a two-stage procedure: first for each fixed λ_2 it finds the ridge regression coefficients and then does a LASSO type shrinkage. This kind of estimation incurs a double amount of shrinkage which leads to increased bias and poor predictions. However, using the correction factor $1 + \lambda_2$ the prediction performance is improved and the elastic net estimators are given by:

$$\beta^{EN} = (1 + \lambda_2)\beta^{naiveEN} \tag{5}$$

The idea of implementing two penalty schemes is based on the aim to include all the true regressors in the model, even if they are strongly correlated (Zou and Hastie, 2005). The L1 norm penalty typically would select only one of the correlated parameters which may miss in terms of interpretation of the model.

2.2 Dimensionality Reduction Methods

2.2.1 Principal Component Regression (PCR)

Principal Component Regression (PCR) is the most popular factor model used for dimensionality reduction and was first introduced in macroeconomic forecasting by Stock and Watson (2002c). Let X be the design matrix with dimension $M \times V$ where potentially $V >> M$. To reduce the dimensions of X matrix we assume that there also exists a vector $K \times 1$ of finite latent factors $F = (F_1, F_2, \dots F_K)'$ which controls the common trends of x_i's. In matrix form:

$$x_i = \Phi' F_i + e_i \tag{6}$$

Φ is the loading matrix with dimensions of $K \times M$ and indicates the relationship between x_i and the unobserved factors. Finally, e_i is the vector of zero mean $I(0)$ errors and accounts for the unexplained part of F_i. The estimation of Φ and F_i is based on the solution of the following minimisation problem:

$$V(K) = \min_{\Phi, F_i} \frac{1}{N} \sum_{i=1}^{N} (x_i - \phi_i' F_i)^2 \tag{7}$$

where ϕ_i' denotes the vector loadings of Φ matrix. Stock and Watson (2002c) suggest a non-unique solution by implementing an eigenvalue-eigenvector decomposition. The K largest eigenvalues of the variance covariance matrix of $X'X$ are assumed to denote the rows of Φ which in turn provide the estimates of the K factors, i.e. $\hat{F_K} = \sum_{j=1}^{V} \phi_{j,K} X_j$. Usual practice that identifies the factors up to a rotation is the normalisation of the data so that they have zero mean and unit variance, before applying PCR. The principal components regression (PCR) approach involves constructing the first K principal components, and then using these components as the predictors in a linear regression model. The key idea is that often a small number of principal components suffice to explain most of the variability in the data, as well as the relationship with y_i. In other words, we assume that the directions in which the $x_i's$ show the most variation, are the directions that are associated with y_i .

2.2.2 Partial Least Squares (PLS)

PCR does not guarantee that all linear combinations that best explain the predictors will also be the best choice to use for predicting the response variables, y_i. Partial least squares regression (PLS) makes use of the dependent variable y_i to identify new features that not only approximate the old features well, but also that are related to the response. Wold (1985) first introduced the Partial Least Square method (PLS). PLS is a dimensionality reduction technique that estimates multiple regressions under a large but finite number of regressors. PLS is similar to Principal Component Analysis (PCA) in that we estimate factors that are linear combinations of the x_{it} covariates and then the obtained factors are used in the regression instead of the x_{it}. A significant difference is that PLS estimates factors are estimated by maximising both the variability of the y_i and the covariates x_{it}, while PCA only considers the variability of the covariates. There are many ways to define PLS that have much in common. Broadly speaking, PLS approach seeks to find linear combinations that help explain both the dependent variable and the regressors.

2.3 Machine Learning Models

This Section describes the set of non linear machine learning forecasting models and discuss their basic properties. Even though the models provide with a different framework, they all fall into the following general decomposition: we consider a vector of responses $Y = \{Y_1, Y_2, \ldots Y_N\}^T$, a $n \times p$ design matrix X, A $k \times 1$ vector of β^0 and a vector of identically distributed errors ε:

$$Y = g\left(X, \beta^0\right) + \varepsilon \tag{8}$$

However, a major concern when estimating such complex models is overfitting. There are a variety of splitting schemes that one could follow to perform model selection using cross validation procedures (Arlot and Celisse, 2010). In this application we choose K-fold cross validation to tune the tuning parameters for each model. For example, these can include the structure of the neural nets, the number of the trees for random forests and or the regularisation parameter i.

2.3.1 Support Vector Regression (SVR)

The Support Vector Machine (SVM) technique was originally introduced as a classification method based on the idea of using support vectors to represent the class boundaries in the classification problems (Vapnik, 1998). The model has recently gain attention on the economics and finance communities as it offers nice statistical properties and can handle and capture non-linearities in the in the data (Xiang-rong et al., 2010; Wang et al., 2012).

For simplicity we show the case of linear functions. The model is structured as follows:

Let $x_t = [x_{1t}, \dots x_{Nt}]'$ be the vector of covariates and y_t be the target variable. All appropriate transformations are applied to the data, in advance.

Let the model be

$$y_t = \boldsymbol{\beta}^{0\prime} x_t + \varepsilon_t. \tag{9}$$

The estimation of the $\boldsymbol{\beta}^0$ is done by formulating the following optimization problem in the primal weight space of the unknown coefficients, $\boldsymbol{\beta}^{0\prime}$:

$$\operatorname{argmin} L(\boldsymbol{\beta}, \xi_t, \xi_t^*) = \frac{1}{2} \|\boldsymbol{\beta}\|^2 + C \sum_{t=1}^{T} (\xi_t + \xi_t^*)$$

$$\textbf{s.t} = \begin{cases} y_t - \boldsymbol{\beta}' x_t \leq \epsilon + \xi_t \\ \boldsymbol{\beta}' x_t - y_t \leq \epsilon + \xi_t^* \\ \xi_t, \xi_t^* \geq 0 \quad \mathrm{t} = 1, \dots \mathrm{T} \end{cases}$$

where the parameter C tunes the trade-off between the "flatness" (complexity) of the estimated model and the amount of error ϵ, that is tolerated. ξ_t, ξ_t^* are the slack variables introduced by Vapnik (1995) to cope with the otherwise infeasible constraints of the optimization problem. Furthermore, the problem in its dual formulation can be written as follows, depending only on the dual variables α and α^*:

$$\max_{\boldsymbol{\alpha}, \boldsymbol{\alpha}^*} \left[-\frac{1}{2} \sum_{t,j=1}^{T} (\alpha_t - \alpha_t^*) \left(\alpha_j - \alpha_j^*\right) x_t' x_j - \epsilon \sum_{t=1}^{T} (\alpha_t + \alpha_t^*) + \sum_{t=1}^{T} (\alpha_t + \alpha_t^*) \, y_t \right]$$

$$\text{subject to: } \sum_{t=1}^{T} (\alpha_t - \alpha_t^*) = 0 \text{ and } \alpha_t, \alpha_t^* \geq 0.$$

2.3.2 Tree Models and Random Forests

Tree models is a non-parametric method tailored for both regression and classification problems. Key in this framework is the ability to handle complex relations within data in an accessible and conceptually easy way. Tree models based on the idea to divide consecutively split the in-sample dataset until an assignment criterion with respect to the target variable into a "data bucket" (leaf) is reached. The general algorithm of a decision tree proceeds as follows:

The aim is to minimise the objective function within areas of the target space (buckets) conditioned on the input X. Starting with the full set X of M observations, initially we divide the regression space into two parts where the split point is chosen to achieve the best fit. Consequently, we predict the target variable in each of these sub spaces and further partition them in two other spaces. The process continues until a stopping rule is applied. The algorithm builds iteratively the relationship between the target variable and the predictors X. To fix ideas, a schematic representation of a tree model with two features is given in Figure 1. Let y_t be the variable of our interest based on two predictors $x_t^{(1)}$, $x_t^{(2)}$. The independent variable $x_t^{(1)}$ is first partitioned into to a threshold variable k_1. All the observation set for which $x_t^{(1)} \leq k_1$ is further split at $x_t^{(2)} = k_2$ while the set were $x_t^{(1)} > k_1$ is split at $x_t^{(1)=k_4}$. Therefore the feature space is decomposed into 5 different sets P_i, $i = 1 \ldots 5$.

The size of the regression trees grows exponentially when we increase the number of the input variables, following the same procedure. Generally, splitting the vector of predictors x_t to M subspaces i.e $P = \{P_1, \ldots P_M\}$, the optimal estimates of β coefficients is just the average of the estimated β in each region. The regression problem becomes:

$$y_t = \boldsymbol{g(x_t;\beta)} + \varepsilon_t. \tag{10}$$

where

$$\boldsymbol{g(x_t;\beta)} = \sum_{m=1}^{M} \beta I(x_t \in P_m) \tag{11}$$

A disadvantage of regression trees is that they are not identically distributed: they are built adaptively to reduce the bias. Growing decision trees in the above form may lead to severe over-fitting. An alternative modelling set up that overcomes this problem is the so called "Random Forest" Breiman (2001). Random forest "grows" a set of uncorrelated trees which are estimated separately. Then, the predictions of the estimated regression trees are averaged out to make a single prediction of the target variable. In particular, for a given number of trees we use a subsample of observations (bagging) and a random subset of predictors for each tree. The latter results to de-correlate the trees and hence improve forecast accuracy. Hastie et al. (2009) provides an algorithm for growing a random forest from a specific number of trees. A general drawback of random forests, as compared to single trees, is that they are hard to interpret due to the built-in randomness. Recently, Athey et al. (2019) introduced the idea of "causal forest" and provides some theoretical evidence on how someone can extract causal inference from random forests. The tuning parameter to be cross validated in this setting is the total number of trees and the number of leaves on each tree.

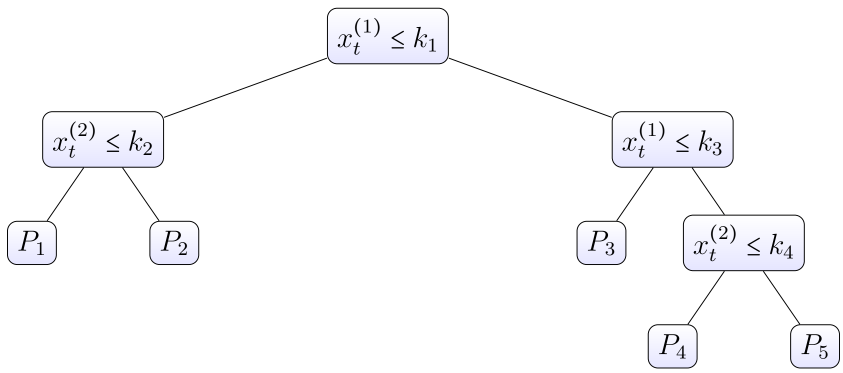


FIGURE 1: The figure displays a simple example of a regression tree for the prediction of single target y_t. The symbols $x_t^{(1)}$ and $x_t^{(2)}$ correspond to the predictors of the regression.

2.3.3 Neural Networks in Regression models

Neural networks are similar to linear and non-linear least squares regression and can be viewed as an alternative statistical approach to solving the least squares problem. Both neural networks and conventional regression analysis attempt to minimise the sum of squared errors. The bias term is analogous to the intercept term in a regression equation.

The number of input neurons is equal to the number of regressors while the output neurons represent the dependent variables. Linear regression models may be viewed as a feedforward neural network with no hidden layers and one output neuron with a linear cost function. The weights connecting the input neurons to the single output neuron are proportional to the coefficients in a linear least squares regression. Networks with one hidden layer resemble nonlinear regression models. The weights represent regression curve parameters. Figure 2 provides a schematic representation of a deep neural net with p inputs x_t, L hidden layers and one output $\hat{y}_t$.

A general definition of a multi-layer (deep) neural network follows: Let $g_1 \dots g_L$ be the activation functions of the L hidden layers of the network representing non-linear transformations of the data.

The overall G structure of the network is equal to:

$$G = g_L(g_{L-1}(\dots(g_1 \dots))) \tag{12}$$

Then, the model at hand becomes:

$$y_t = G\left(x_t, \boldsymbol{\beta}^0\right) + \varepsilon_t \tag{13}$$

where x_t is $p \times 1$, $\boldsymbol{\beta}^0$ is $k \times 1$ and contains all model parameters and G denotes the overall nonlinear mapping.

It is easy to see that the fitted model is just the hierarchical model of the form:

$$\begin{aligned}
\hat{y}_t &= \hat{b} + G(x_t, \boldsymbol{z}_t) \qquad \text{[Output layer]} \\
\boldsymbol{z}_t &= g_L(\beta_L \alpha_{L-1,t} + \alpha_L) \qquad \text{[Hidden layer L]} \\
\boldsymbol{z}_{L-1,t} &= g_{L-1}(\beta_{L-1}\alpha_{L-2,t} + \alpha_{L-1}) \qquad \text{[Hidden layer L - 1]} \\
&\vdots \\
\boldsymbol{z}_{1,t} &= g_1(\beta_1 x_t + \alpha_{L-1}) \qquad \text{[Hidden layer 1]}
\end{aligned}$$

Where β_i, α_i denote $i \dots L$ the different weights attached to each layer and the bias

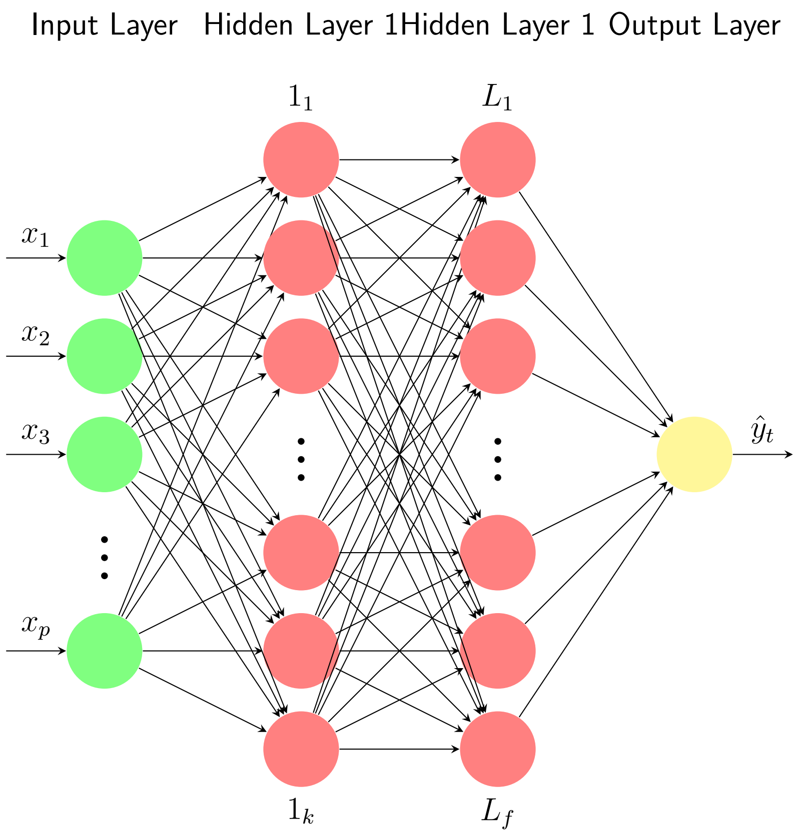


FIGURE 2: Example of a deep neural network. The figure shows the general architecture of a deep neural network for regressions. The green circles represent the input variables, i.e. the regressors in our case.The red circles denote the fully connected hidden nodes.The yellow circle represents the output of the network, i.e. the fitted value of the target variable for our purpose.

units, respectively.

The architecture of the neural net can vary with respect to the hidden layers and nodes included on each layer. For example, note that each layer can have a different number of nodes. Determining the architecture of the neural network is in fact, a model selection problem. To select the neural net specification we investigate the predictive performance of different network architectures through extensive cross-validation. We start from a shallow network with one hidden layer and expand the layers up to five, that is four hidden layers and the output layer. Similarly, we select the nodes on each layer setting up a grid from one to four nodes. We follow the existing literature and fix the activation function for each node to be the Rectified Linear Unit (ReLU), $g(z) = \max(z, 0)$ (Blake and Kapetanios, 2000; Nair and Hinton, 2010). Other common activation functions include the hyperbolic tangent (tahn), the sigmoid and the radial basis function (Blake and Kapetanios, 2000). We use stochastic gradient decent to estimate the model. Because of its compositional form, the gradient can be easily derived using the chain rule for differentiation.

2.4 Selecting the tuning parameters

A conventional practice in macroeconomic forecasting is to use some form of information criteria (e.g. AIC, BIC) to choose the tuning parameters of the model at hand. Cross validation methods have recently become a popular alternative because they can be used to any model, including those for which the derivation of information criteria is not feasible [2]. It remains a theoretical question and ongoing debate which of these methods should be used for model selection. The main difference between information criteria and cross validation methods is that the latter depends on out-of-sample performance whereas information criteria are "in-sample" statistics. [3] All of our models involve some kind of parameter selection prior to estimation. For the linear methods, we focus on the choice of the penalty imposed on the model's coefficients whereas for the non linear methods, we determine the parameters specific to the model's architecture [4].

Generally speaking, cross validation techniques consist of estimating a particular specification over the training sample and computing the forecast performance over the validation sample.

In a time-series forecasting setting, a cross-validation exercise is conducted only in the in-sample period to avoid information likeage (Kalamara et al., 2020). We opt to apply the standard 5-fold cross validation to select the tuning parameters required for each model specification. Coulombe et al. (2019) report that K-fold provides the best performance compared to a set of different cross validation procedures and information criteria.

The method is based on a resampling scheme; we use 5 folds, i.e. the in-sample set is

[2]For a review of various cross validation methods see Arlot and Celisse (2010), Bergmeir et al. (2014) and Coulombe et al. (2019)

[3] Hansen and Timmermann (2015), however, show asymptotic equivalence between test statistics for out-of-sample performance and in-sample Wald statistics.

[4]Parameter selection includes the number of nodes and layers for neural nets and the choice of kernel function and error margin for support vector regression and the maximum number of trees for random forests

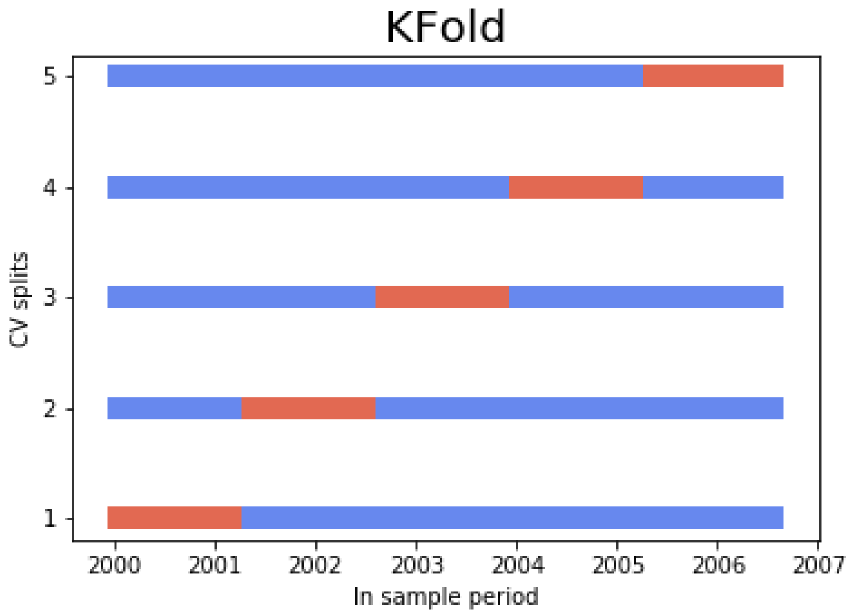


FIGURE 3: Schematic representation of 5-fold cross validation. The figure depicts the splitting strategy we perform to the in-sample period at each point in time t. The model is estimated on 80 % of the data and tested on the remaining 20%. The red area shows the test set of observations for each split.

randomly split into five disjoint subsets. For each one of the 5 subset and set of the tuning parameters considered, 4 subsets are used for estimation (training set) and the remaining corresponding observations of the in-sample set is used as a test subset (validation set) to generate forecasting errors. As a performance metric and condition to select the best tuning parameter, we consider the average mean squared error in the validation set. Once the tuning parameters are selected, each model is estimated using the whole in-sample period and used to generate out-of-sample predictions.

Figure 3 provides a schematic representation of the 5-fold cross validation method. The performance over the validation set is then given by the average performance over the 5 sub-samples. Taking the average of the data subsets helps to deal with issues that occur in regime switching models when the tuning-parameters are state-dependent (Coulombe et al., 2019).

We re-optimize every model, for every data combination and for each forecasting horizon, for every out-of-sample period. In this way, we ensures no information flow from the evaluation period to the estimation period.

3 Datasets

In this Section, we describe various aspects of the datasets we explore for forecasting UK GDP growth. We start by describing the disaggregated business and consumer survey balances used for the benchmark out of sample forecasting exercises and then we move to explain how to expand the dataset to also include text based indicators and other macroeconomic variables related to economic activity, prices, labour market statistics and some financial market data to explore if there's additional predictive content in them.

3.1 Survey Dataset

Collection and publication of official data are subject to substantial processing delays; For example in the United Kingdom the monthly index of industrial production (including manufacturing output) is published at least 40 days after the end of the month to which it refers to. Therefore, significant resources are devoted to exploit alternative sources of information in order to gauge the continuously evolving state of the real economy. As such, the use of timely and reliable information about current economic conditions is crucial for policy makers and expectations formation. Quantitative survey balances from businesses and firms are prime candidates for this and they have proven very useful for short term forecasting (Lahiri and Monokroussos, 2013; Bańbura et al., 2013; Giannone et al., 2008) due to their timeliness (they are usually published at a monthly frequency and just a couple of days after the end of the reference month) and their high correlation with GDP growth. Many policymakers and market participants take recourse to survey evidence to measure current economic conditions. This is widely evidenced by monetary policy communications, which frequently point to survey evidence when describing the current macroeconomic situation.

Their publication is usually accompanied by some discussion of what can be learnt from them about the most recent movements and short-term expected future movements in economic activity, at least in the sector to which the surveys relate to. These business surveys ask *inter alia* whether, after adjusting for normal seasonal movements, output has risen, stayed the same or fallen in recent months, what are the firms' expectations for the next quarter or the year ahead or what are the employment prospects in their business. At the same time, 'soft' information from the consumers' perspective is contained in the consumer confidence indicators published by the European Commission, where questions about the financial situation of the households or the general economic situation are included and can provide with a more timely assessment of consumers' perception of current and expected economic conditions. They have also been found to be empirically useful for forecasting movements in economic conditions.

More details on the exact definition of the survey balances can be found in the Appendix Table A.1.

3.2 Text Dataset

As a robustness check, we explore whether other types of soft information add any significant value to our results. Recently, a lot of research has been undertaken that explores text as an alternative data source and its usefulness to answer economic and policy related questions. To this end, we make use of the 15 text-based indicators suggested by Kalamara et al. (2020) that aim to capture uncertainty and sentiment in the economy. The metrics are created by applying existing text analysis methods to newspaper articles and modified appropriately to obtain valuable information in real time.

In a forecasting setting, Kalamara et al. (2020) convert UK newspaper articles [5] to time series text-metrics applying a range of text analysis methods. Then, the generated indicators are used as predictors to forecast key economic indicators. Their findings indicate that text significantly improves the out of sample forecasting performance of GDP growth relative to popular benchmarks and particularly true during periods of stress. This is an important finding because it suggests that text-based data can act as a strong complement to high frequency financial market data and to less timely, survey data.

The text analytics models used to create the indices can be found in Table 1, and are the most commonly used in the literature to extract signal from text. The dataset includes a wide range of methods from simple counts of specific term occurrences such

[5]The raw text dataset is comprised of three highly circulated newspapers in the UK, i.e. the Daily Mail, the Guardian and the Daily Mirror.

Positive and negative dictionary	Boolean	Computer science-based
Financial stability (Correa et al., 2017)	Economic Uncertainty (Alexopoulos et al., 2009)	VADER sentiment (Gilbert, 2014)
Finance oriented (Loughran and McDonald, 2013)	Monetary policy uncertainty (Husted et al., 2017)	'Opinion' sentiment (Hu and Liu, 2004; Hu et al., 2017)
Afinn sentiment (Nielsen, 2011)	Economic Policy Uncertainty (Baker et al., 2016)	Punctuation economy (this paper)
Harvard IV (used in (Tetlock, 2007))		
Anxiety-excitement (Nyman et al., 2018)		
Single word counts of "uncertain" and "econom"		
tf-idf applied to "uncertain" and "econom"		

Table 1: The three broad categories of algorithm-based text metrics used.

as the word 'uncertain' and 'economy' in each article divided by the number of words in the article, to more sophisticated approaches adopted from text-specific algorithms. The numerical scores for a particular month are found from the mean of the scores of the articles that were published in that month.

Broadly speaking, the methods fall into three main categories: dictionary-based which associate specific scores (e.g. positive or negative sentiment) and count the net score per article, boolean methods which provide a count of articles only if the terms in an article satisfy some logical condition and computer-science based which follow different algorithmic procedures drawn from computational linguistics.

3.3 Macro Dataset

In this paper, we are also interested in evaluating the informational content of a wider dataset that includes not only "hard" indicators as published by the official statistical agency, like data on production, services, prices and labour market statistics, but also some financial market data taken as monthly averages. The series are selected to represent broad categories of macroeconomic time series: real output and income, employment and hours, real retail, manufacturing and sales data, international trade, labor costs, price indexes, interest rates, stock market indicators, and foreign exchange measures for the UK Economy. This type of datasets has been widely used in the literature (Stock and Watson, 2002a,b) for the US Economy and has been empirically proven to correlate fairly well with future GDP growth. Therefore, we employ a battery of linear and non-linear models to understand the predictive content of these series: do they have any additional predictive ability over and above the panel of survey balances and if they do at which forecast horizons and which phases of the business cycle?

The variables used in all the exercises are considered at a monthly frequency, starting in January 2000 until August 2018, purely due to data availability constraints. Although we do not explicitly take into account the real time release calendar of the input series, in the design of the out-of-sample forecasting evaluation exercise we only consider the series that were available at every point in time. Our target variable for this exercise is the monthly GDP estimate as published by the Office of National Statistics, transformed to three month-on-three month growth rate. All series are seasonally adjusted.

Further details on the extended dataset can be found in the Appendix Table A.3.

4 Empirical Results

In Section 4, we first outline the design of the out of sample forecast evaluation exercise and the various specifications we considered for robustness and then we present and discuss the main empirical findings. We design the evaluation exercise in such a way so that we can assess the usefulness of various data sources individually and collectively. To begin with, we examine the usefulness of our set of models using the disaggregate survey balances compared to two standard benchmarks in the literature (Principal Components Regression (PCR) and AR(1)[6]) using linear and non-linear models. We then proceed with testing the informational content of a more extensive dataset, first using text based indicators only and then more standard macroeconomic time series, related to activity, prices, financial market data, labour market in the spirit of Stock and Watson (2002c) and report results in terms of RMSFEs estimated using the same information set every time. Finally, we combine all the "soft" information from the surveys and text based indicators and examine whether there's additional predictive content in a larger information set or simply the use of survey balances is sufficient to forecast economic activity from one-month ahead up to two-years ahead.

We conclude this part by asking whether during recessionary episodes, like the Great Recession, linear and non-linear models perform similarly or non-linear models can be

[6]While AR(1) models are simple, there's substantial literature confirming that, overall, they are particularly tough benchmarks to beat (see Carriero et al. (2019) for a review for various countries and across different time periods.)

more useful in an out-of-sample forecasting exercise.

4.1 Evaluation Design

In our benchmark specification, we use the data from January 2000 until August 2006 as training sample (estimation sample), as a result the origin of the forecast exercise is September 2006 and we end up with 145 observations for the out-of-sample evaluation period that ends in August 2018. We set a maximum of 24 months (two-years) ahead forecasts generated at every month and we compute **direct** forecasts from models estimated with expanding samples over the out-of-sample period, that is, at each forecast origin we re-estimate each model and we use all observations available up to the forecasting origin. We use the relative Root Mean Squared Forecast Errors (RMSFE) as a measure of forecasting performance against two different benchmarks, an AR(1) and a PCR models, and the Diebold and Mariano (1995) t statistics to test for equal accuracy with the Newey-West estimator with maximum order increasing with the horizon.

4.2 Empirical Results

4.2.1 Linear vs Non Linear Models

Table 2 summarises the results of the benchmark specification. In the upper panel of the table we report the relative RMSFEs against the PCR specification and in the lower panel we benchmark our results against an AR(1) model using only the survey balances as indicators, as a consequence values larger than 1 indicate that the benchmark is more accurate and values less than 1 indicate that the model under consideration is more accurate. Overall, based on our empirical results, the PCR is a harder benchmark to beat compared to an AR(1) and as such we will use it as our main model to benchmark our results for all the subsequent exercises [7].

Among the linear models, Ridge, PLS and the Elastic Net are among the ones reporting the largest gains consistently for all forecasting horizons with that go up to 27% for the Ridge regression-based forecasts against the PCR benchmark at 1 month ahead horizon. At longer horizons, the PLS and the Elastic Net regressions perform better

[7] The respective results against the AR(1) benchmark are in Appendix B.2 for completeness

Table 2: Relative RMSFE Across Different Models with Survey Data

relative to PCR	(1)	(3)	(6)	(9)	(12)	(24)
Lasso	0.991	0.977	0.961	0.940	0.950	0.990
Ridge	**0.726*****	**0.824***	0.846	0.891	0.935	1.104
Elastic	**0.817*****	**0.843***	0.843	0.868	0.908	1.065
PLS	**0.826*****	**0.841***	0.840	0.861	0.899	1.068
Random Forest	**0.879*****	0.901	0.856	0.885	0.914	1.045
SVM	**0.722*****	**0.859***	0.872	0.889	0.920	1.074
NN	0.768	0.874	0.869	0.900	0.937	1.103
relative to AR(1)	**(1)**	**(3)**	**(6)**	**(9)**	**(12)**	**(24)**
Lasso	**0.720*****	**0.826***	0.846	0.896	0.937	1.109
Ridge	**0.702*****	0.820	0.890	0.969	**1.011***	0.975
Elastic	**0.726*****	**0.716***	0.758	0.907	0.993	0.965
PLS	**0.792*****	0.789	0.830	0.939	0.998	0.976
Random Forest	**0.751*****	0.770	0.835	0.877	0.972	0.972
SVM	**0.625*****	0.850	0.891	0.970	1.010	0.985
NN	**0.687*****	0.895	0.923	0.986	1.011	0.968

Note: Top panel: Relative RMSFEs across different specifications against a **PCR** model using only *Survey* data for $h = 1$ up to $h = 24$ months ahead. Bottom panel: Relative RMSFEs across different specifications against a **AR(1)** model using only *Survey* data. *,** & *** denote rejection at 10%, 5% and 1% level of the null hypothesis of equal forecasting method accuracy of the Diebold and Mariano (1995) test against the respective benchmark models.

with gains up to 10% against the PCR at one-year ahead forecasts and the Ridge-based forecasts report gains ranging from 6% to 15% for medium term forecasts (from 6 to 12 months ahead).

The non-linear models do not appear to perform substantially better than the linear ones for most of the forecasting horizons, when using just survey balances as input variables. Within this class of models, however, SVM reports the largest gains for shorter horizons compared to the NN and Random Forest with forecasting gains 23% and 12% at one-month ahead forecasts respectively.

Another interesting feature appears in Table 2: although the majority of the models are able to maintain their gains against the benchmark specifications for longer forecasting horizons (one-year ahead), it seems that these gains are getting smaller after the first two quarters. Notably, there's no other model, apart from the Lasso regression, able to outperform the PCR specification at the two-years horizon. This finding is consistent with the predictive content of business and consumer surveys in terms of forecasting economic activity: 'soft' information from survey data is mostly relevant for short term forecasting, whereas for longer term one needs to explore official releases further. We will explore this last finding further in the next Section.

Table 3: Relative RMSFE Across Different Models and Datasets

relative to PCR	(1)	(3)	(6)	(9)	(12)	(24)
Lasso	**0.902***	0.914	0.896	0.898	0.918	1.033
Ridge	**0.904***	0.914	0.897	0.898	0.918	1.033
Elastic	0.986	0.975	0.959	0.940	0.950	0.990
PLS	**0.937*****	0.932	0.924	0.916	0.932	1.012
Random Forest	**0.930** * * *	0.941	0.936	0.927	0.940	1.014
SVM	**0.823*****	**0.882***	0.904	0.912	0.939	1.040
NN	**0.879*****	0.904	0.950	0.903	0.924	1.056

relative to PCR	(1)	(3)	(6)	(9)	(12)	(24)
Lasso	**0.959*****	0.956	0.940	0.931	0.945	0.997
Ridge	**0.855*****	0.884	0.863	0.882	0.896	1.065
Elastic	**0.946*****	0.947	0.932	0.925	0.940	1.003
PLS	**0.885*****	0.901	0.876	0.888	0.899	1.045
Random Forest	**0.933*****	0.947	0.920	0.921	0.938	1.020
SVM	**0.722*****	0.886	0.872	0.896	0.929	1.118
NN	**0.802*****	**0.895***	0.928	0.897	0.934	1.125

relative to PCR	(1)	(3)	(6)	(9)	(12)	(24)
Lasso	0.992	0.977	0.960	0.940	0.950	0.990
Ridge	**0.698*****	**0.831***	0.855	0.901	0.946	1.097
Elastic	**0.805*****	**0.837***	0.844	0.873	0.913	1.062
PLS	**0.818*****	**0.834***	0.840	0.867	0.909	1.057
Random Forest	**0.794*****	0.882	0.873	0.882	0.916	1.044
SVM	0.663	0.874	0.865	0.895	0.936	1.125
NN	0.667	0.851	0.878	0.903	0.952	1.148

Note: Top panel: Relative RMSFEs across different specifications using only *Text* data. Middle panel: Relative RMSFEs across different specifications using only *Macro* data. Bottom panel: Relative RMSFEs across different specifications combining *Survey* and *Text* data. The results are reported against a **PCR** model as a benchmark estimated using only *Text*, only Macro and combined *Survey* and *Text* data respectively. *,** & *** denote rejection at 10%, 5% and 1% level of the null hypothesis of equal forecasting method accuracy of the Diebold and Mariano (1995) test against the benchmark specification.

4.2.2 Text and other data sources

As explained in Section 3, we want to understand whether other types of 'soft' information could have predictive value for GDP forecasting. An alternative data source that we explore in this content is text-based indicators capturing uncertainty and sentiment about the economy. Text based indicators of sentiment and uncertainty might contain some of the forward looking information that is crucial for policymakers in the decision making process. As such, we are interested in exploring their predictive content and compare it with this of surveys. We explained in Section 3.2 how the measures are constructed and we now use them in an identical forecasting environment to test their predictive content. The results are summarised in the upper panel of Table 3. The performance of the models is evaluated in terms of relative RMSFE with the benchmark specification being a PCR model estimated using the same indicators for all the models.

Overall, their predictive content is fairly similar to this of survey balances: their reporting gains are mainly concentrated in the shorter horizons (1-6 months ahead) and they gradually decrease as the forecast horizon increases. Within that, on average the non linear methods perform better that the linear ones, with predictive gains ranging from 5% up to 14% for the Neural Net that is the best performing specification among both the linear and non-linear models that is in line with what Kalamara et al. (2020) find in their paper. SVM also performs better than the the PCR benchmark with predictive gains from 6% in the one-year ahead forecasts to 18% for one-month ahead. In terms of predictive content, while the surveys and the text based forecasts are pretty similar, the surveys-only models perform marginally better in shorter horizons and the text-only models are able to maintain some of their gains for longer horizons up to one-year ahead. All models, both linear and non-linear, perform better than the benchmark specification, a PCR model, and the gains range between 1% to 17%.

As explained in Section 3.3, we also considered a macroeconomic dataset in the spirit of Stock and Watson (2002c) in attempt to evaluate the informational content of additional data sources like economic activity indicators, prices and financial market data that have been traditionally used for macroeconomic forecasting. We repeated the same out of sample evaluation exercise with a different dataset and the results are summarised in the middle panel of Table 3.

The overall message is that the relative RMSFEs based on the out of sample exercise suggest that information from macroeconomic time series is only marginally more useful for some non-linear methods like the SVM and the NN for up to one-year ahead forecasts compared to the alternative of using text based indicators, although the macro-only forecasts cannot maintain their predictive gains for longer than one-year ahead against the benchmark specification. On average, the macro-only forecasts do not outperform the forecasts of text indicators suggesting that the information contained in newspaper articles is comparable, if not superior, to the traditional macroeconomic dataset that has been widely used for forecasting.

In a central bank environment, policymakers need to make decisions and weight all the available information when publishing a set of forecasts about the future path of the economy. With official data published with a substantial delay, they need to make

Table 4: Relative RMSFE Across Different Models and Datasets

relative to PCR	(1)	(3)	(6)	(9)	(12)	(24)
Lasso	**0.906*****	0.916	0.897	0.898	0.918	1.033
Ridge	**0.908*****	0.916	0.898	0.898	0.918	1.033
Elastic	0.990	0.977	0.961	0.940	0.950	0.990
PLS	**0.941*****	0.934	0.925	0.916	0.932	1.011
Random Forest	0.934	0.943	0.937	0.927	0.940	1.013
SVM	**0.827*****	**0.884***	0.906	0.912	0.939	1.040
NN	**0.883*****	0.906	0.952	0.903	0.923	1.056
relative to PCR	(1)	(3)	(6)	(9)	(12)	(24)
Lasso	**0.970*****	**0.965****	0.947	0.937	0.952	0.953
Ridge	**0.866*****	**0.892****	0.869	0.887	0.903	1.018
Elastic	**0.957*****	**0.956****	0.939	0.931	0.948	0.959
PLS	**0.896*****	**0.909****	0.883	0.893	0.906	0.999
Random Forest	**0.944*****	**0.956****	0.926	0.927	0.945	0.975
SVM	**0.730*****	**0.894****	0.878	0.902	0.936	1.068
NN	**0.812*****	**0.904****	0.935	0.903	0.942	1.075
relative to PCR	(1)	(3)	(6)	(9)	(12)	(24)
Lasso	**0.990*****	0.977	0.961	0.940	0.950	0.990
Ridge	**0.696*****	**0.831***	0.855	0.900	0.946	1.097
Elastic	0.804	**0.837***	0.844	0.873	0.913	1.062
PLS	**0.816*****	**0.833***	0.840	0.867	0.909	1.057
Random Forest	**0.792*****	0.882	0.873	0.882	0.916	1.044
SVM	0.662	0.874	0.866	0.895	0.936	1.125
NN	**0.666*****	0.851	0.879	0.903	0.952	1.148

Note: Top panel: Relative RMSFEs across different specifications using only *Text* data. Middle panel: Relative RMSFEs across different specifications using only *Macro* data. Bottom panel: Relative RMSFEs across different specifications combining *Survey* and *Text* data. **All** results are reported against a **PCR** model as a benchmark estimated using on the survey balances dataset from Section 3.1. *,** & *** denote rejection at 10%, 5% and 1% level of the null hypothesis of equal forecasting method accuracy of the Diebold and Mariano (1995) test against the benchmark specification.

decisions in real-time with incomplete information. As such, combining information from surveys and text indicators might provide them with a better understanding of the current and future economic conditions and this is in practice how policy making is performed in real time.

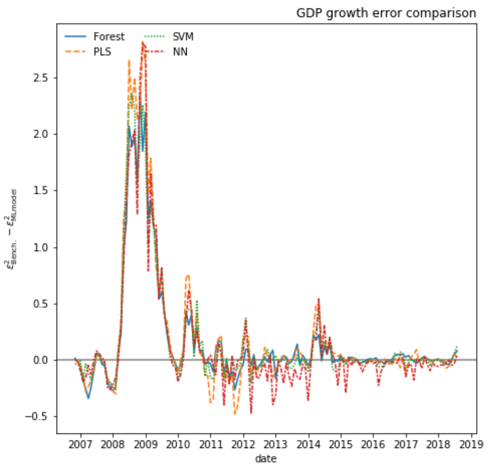
Additionally, all these methods are designed for handling large datasets, we performed the same exercise combining all the 'soft' information at hand, i.e. the surveys and the text information to understand whether we can achieve better forecasting performance with larger information set. The results are reported in the bottom panel of Table 3 against a PCR model specification that uses both surveys and text indicators. The combination of the survey and the text indicators does not seem to offer any substantial improvement on the results: the general pattern of the relative RMSFEs is closer to that of the forecasting exercise when using just the survey balances suggesting that enlarging the information set of survey balances with text indicators does not add much predictive content. For robustness, we also consider a combined dataset with the surveys, the text and the macroeconomic indicators and the overall narrative doesn't change in any substantial way. The results are summarised in Appendix B for completeness.

A final robustness check that we performed in this Section, summarised in Table 4, is to look at our results using exactly the same benchmark in order to carefully assess and compare the predictive content of the different datasets. We used the same forecast errors implied by the models in Table 3 but now we use as a benchmark the same PCR model estimated with survey balances only. While each panel of theTable 4 corresponds to different data sources, Text, Macro and Survey combined with Text, the benchmark against which all the models are assessed in now the same in order to better compare the predictive content of the different data sources.

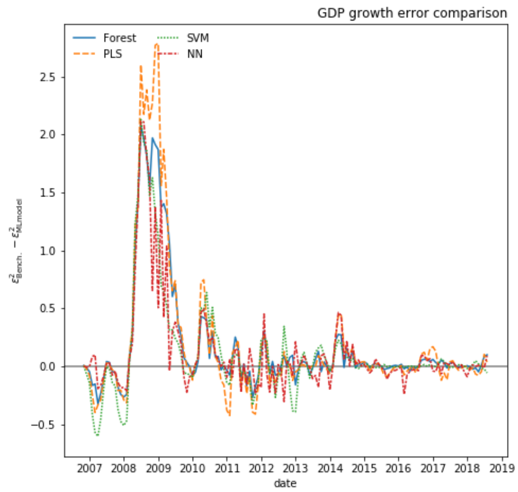
The performance of the models using text and macro series (top and middle panels of Table 4) is broadly similar and the overall message is in line with the previous exercises: the models appear to be doing reasonably well in the short run, especially in 1-3 months ahead and their predictive ability diminishes over time. However, it seems that the combination of 'soft' information from surveys and text indicators pays off somewhat in terms of predictive ability: on average there are larger gains in the shorter horizons, with the non-linear models benefitting the most from the combined information from surveys and text indicators. This finding provides with some further empirical support for the use of non linear, machine learning models in a data rich environment for forecasting GDP growth.

4.2.3 Are machine learning models more useful around recessionary episodes?

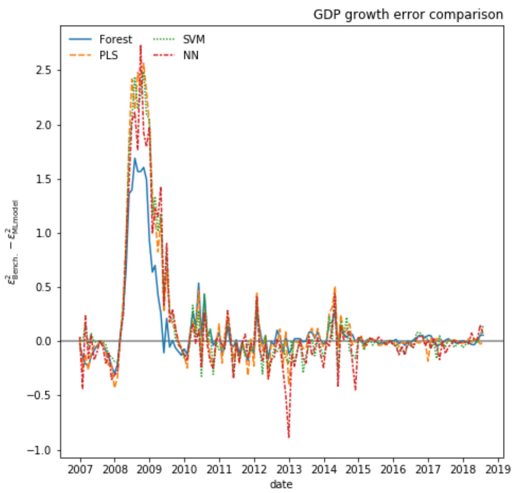
In the previous Section, we looked at the overall evaluation results across different datasets and models to assess their ability in tracking UK GDP growth. In this Section, we focus on the behaviour of the forecast errors of the models in the out of sample evaluation period against the PCR benchmark model which uses as input only survey indicators (the corresponding evaluation results are summarised in Table 4) to get a better understanding of whether linear and non linear models can be more/less useful during specific points in time. An episode of special interest remains the Great Recession as it would allows us



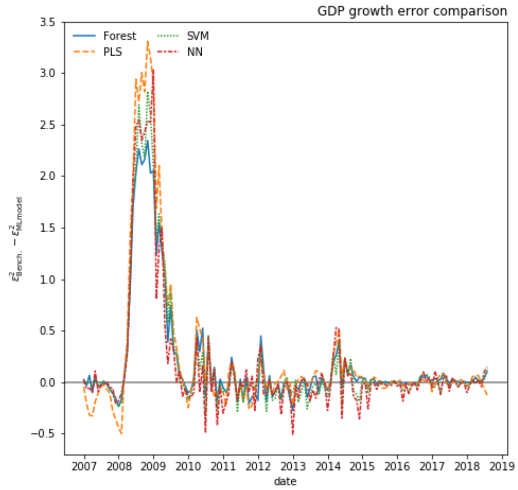
(A) Surveys



(B) Text



(C) Macro



(D) Combination of Surveys, Text and Macro data

FIGURE 4: Average mean squared error differences between the PCR benchmark and other models over h-month ahead out-of-sample forecasts, where h = 1, 3, 6, 9,12 ,24. A line above zero means that the relative model produces smaller errors than the benchmark model

to test whether highly non-linear (Machine Learning) models are able to capture turning points more timely and accurately than their linear counterparts.

Figure 4 plots the mean squared error differences between the benchmark PCR model and the 3 non-linear (SVM, NN and Random Forest) and one linear (the best performing model is the PLS in this case) models during the out of sample evaluation period. A point above zero means that the model produces smaller errors than the benchmark and a point below zero implies that the benchmark is more accurate.

A couple of interesting features stand out in Figure 4. First of all, the largest forecast errors both in absolute and in relative terms (against the benchmark specification) are reported during the Great Recession and this is to be expected, given the magnitude of the episode and the lack of a similar event in the estimation part of the sample that could have been used to train the models appropriately. However within that, the Machine Learning models perform overall better than the PLS (and the PCR) specification, chosen here specifically as it is on average the best performing linear model. Consequently, the majority of improvement in terms of forecasting errors (highest forecast gains) takes place during the Great Recession and this is also a generalised finding for all the datasets and their combination that we consider in the paper, lending some empirical evidence to the fact that the machine learning models might be more appropriate to capture large non-linearities in the data compared to their linear counterparts.

Finally, another interesting point in terms of the relative forecasting performance of the different types of datasets we examine is the direction of the forecast errors difference during the Great Recession: in all cases, both the linear and non linear models perform substantially better than the benchmark specification. The latter finding is broadly in line with the evidence found in Kalamara et al. (2020), who argue that text in combination of a non-linear machine learning model predicts better stressed times. While their exercise is based on a rather different setting and their forecasts are produced by feeding the machine learning model with high dimensional term frequency vectors, the overall message about the use of non-linear Machine Learning models during recessionary episodes is rather consistent.

5 Conclusions

In this paper, we have reviewed linear approaches for regularisation and dimension reduction combined with techniques from the machine learning literature to forecast UK GDP growth at monthly frequency for horizons from one-month up to two-years ahead. To this end, we compiled large panels of disaggregated survey data from UK businesses and consumers and explore their informational content. We also consider text based indicators from newspaper articles and a more standard macroeconomic dataset as potential predictors for the target variable and run comprehensive out of sample evaluation exercises to assess their predictive ability across models and datasets.

We provide empirical evidence on the usefulness of the three different datasets, frequently used in macroeconomic forecasting, to predict UK GDP growth from one-month up to two-years ahead. We find that survey and text-based forecasts perform similarly, however survey-based models are marginally better at shorter horizons and the text-only models can maintain some of their gains for longer horizons up to one-year ahead. We compare linear to non-linear models and we find that PLS is among the best performing linear specification and the SVR performs better on average among the machine learning models.We finally document some new evidence on the usefulness of machine learning models in forecasting economic activity during turbulent times and their ability to capture non-linearities in the data in a more accurate way.

References

Ahmed, Nesreen K., Amir F. Atiya, Neamat El Gayar, and Hisham El-Shishiny (2010) "An Empirical Comparison of Machine Learning Models for Time Series Forecasting," *Econometric Reviews*, Vol. 29, No. 5-6, pp. 594–621.

Alexopoulos, Michelle, Jon Cohen et al. (2009) "Uncertain times, uncertain measures," *University of Toronto Department of Economics Working Paper*, Vol. 352.

Anesti, Nikoleta, Simon Hayes, Andre Moreira, and James Tasker (2017) "Peering into the present: the Bank's approach to GDP nowcasting," *Bank of England Quarterly Bulletin*, Vol. 57, No. 2, pp. 122–133.

Arlot, Sylvain and Alain Celisse (2010) "A survey of cross-validation procedures for model selection," *Statist. Surv.*, Vol. 4, pp. 40–79.

Athey, Susan, Julie Tibshirani, Stefan Wager et al. (2019) "Generalized random forests," *The Annals of Statistics*, Vol. 47, No. 2, pp. 1148–1178.

Baker, Scott R, Nicholas Bloom, and Steven J Davis (2016) "Measuring economic policy uncertainty," *The Quarterly Journal of Economics*, Vol. 131, No. 4, pp. 1593–1636.

Bańbura, Marta, D. Giannone, M. Modugno, and L. Reichlin (2013) "Now-casting and the Real-Time Data Flow," in *Handbook of Economic Forecasting, volume 2A*: Elsevier, Chap. 4, pp. 195–237.

Bergmeir, Christoph, Mauro Costantini, and José M Benítez (2014) "On the usefulness of cross-validation for directional forecast evaluation," *Computational Statistics & Data Analysis*, Vol. 76, pp. 132–143.

Bholat, David M, Stephen Hansen, Pedro M Santos, and Cheryl Schonhardt-Bailey (2015) "Text mining for central banks."

Blake, Andrew P and George Kapetanios (2000) "A radial basis function artificial neural network test for ARCH," *Economics Letters*, Vol. 69, No. 1, pp. 15–23.

Breiman, Leo (2001) "Random forests," *Machine learning*, Vol. 45, No. 1, pp. 5–32.

Carriero, Andrea, Ana Beatriz Galvo, and George Kapetanios (2019) "A comprehensive evaluation of macroeconomic forecasting methods," *International Journal of Forecasting*, Vol. 35, No. 4, pp. 1226 – 1239.

Correa, Ricardo, Keshav Garud, Juan M Londono, Nathan Mislang et al. (2017) "Constructing a Dictionary for Financial Stability,"Technical report, Board of Governors of the Federal Reserve System (US).

Coulombe, Philippe Goulet, Maxime Leroux, Dalibor Stevanovic, and Stphane Surprenant (2019) "How is Machine Learning Useful for Macroeconomic Forecasting?," CIRANO Working Papers 2019s-22, CIRANO.

——— (2020) "How is Machine Learning Useful for Macroeconomic Forecasting?."

De Mol, Christine, Domenico Giannone, and Lucrezia Reichlin (2008) "Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?" *Journal of Econometrics*, Vol. 146, No. 2, pp. 318–328. Honoring the research contributions of Charles R. Nelson.

Diebold, Francis X. and Roberto S. Mariano (1995) "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, Vol. 13, No. 3, pp. 253–263.

Diebold, Francis X. and Minchul Shin (2019) "Machine learning for regularized survey forecast combination: Partially-egalitarian LASSO and its derivatives," *International Journal of Forecasting*, Vol. 35, No. 4, pp. 1679–1691.

Efron, Bradley, Trevor Hastie, Iain Johnstone, Robert Tibshirani et al. (2004) "Least angle regression," *The Annals of statistics*, Vol. 32, No. 2, pp. 407–499.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy (2019) "Text as data," *Journal of Economic Literature*, Vol. 57, No. 3, pp. 535–74.

Giannone, D., L. Reichlin, and D. Small (2008) "Nowcasting GDP and Inflation: The Real-Time Informational Content of Macroeconomic Data Releases," *Journal of Monetary Economics*, Vol. 55, pp. 665–676.

Gilbert, CJ Hutto Eric (2014) "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf.*

Hansen, Peter Reinhard and Allan Timmermann (2015) "Equivalence between out-of-sample forecast comparisons and Wald statistics," *Econometrica*, Vol. 83, No. 6, pp. 2485–2505.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009) "Random forests," in *The elements of statistical learning*: Springer, pp. 587–604.

Hu, Minqing and Bing Liu (2004) "Mining and Summarizing Customer Reviews," in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ?04, p. 168?177, New York, NY, USA: Association for Computing Machinery.

Hu, Guoning, Preeti Bhargava, Saul Fuhrmann, Sarah Ellinger, and Nemanja Spasojevic (2017) "Analyzing users' sentiment towards popular consumer industries and brands on Twitter."

Husted, Lucas F, John H Rogers, and Bo Sun (2017) "Monetary Policy Uncertainty."

Joseph, Andreas (2019) "Parametric inference with universal function approximators," Bank of England working papers 784, Bank of England.

Joseph, Andreas, Eleni Kalamara, Galina Potjagailo, and George Kapetanios (2021) "Forecasting UK inflation bottom up," *Unpublished Manuscript*.

Kalamara, Eleni, Arthur Turrell, Chris Redl, George Kapetanios, and Sujit Kapadia (2020) "Making text count: economic forecasting using newspaper text," *Bank of England Staff Working Papers*, Vol. 865.

Kapetanios, George, Massimiliano Marcellino, and Fabrizio Venditti (2019) "Large time-varying parameter VARs: A nonparametric approach," *Journal of Applied Econometrics*, Vol. 34, No. 7, pp. 1027–1049.

Lahiri, Kajal and George Monokroussos (2013) "Nowcasting US GDP: The role of ISM business surveys," *International Journal of Forecasting*, Vol. 29, No. 4, pp. 644–658.

Loughran, Tim and Bill McDonald (2013) "IPO first-day returns, offer price revisions, volatility, and form S-1 language," *Journal of Financial Economics*, Vol. 109, No. 2, pp. 307–326.

Marcellino, Massimiliano (2008) "A linear benchmark for forecasting GDP growth and inflation?," *Journal of Forecasting*, Vol. 27, No. 4, pp. 305–340.

Medeiros, Marcelo C., Gabriel F. R. Vasconcelos, lvaro Veiga, and Eduardo Zilberman (2021) "Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods," *Journal of Business & Economic Statistics*, Vol. 39, No. 1, pp. 98–119.

Nair, Vinod and Geoffrey E Hinton (2010) “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807–814.

Nakamura, Emi (2005) “Inflation forecasting using a neural network,” *Economics Letters*, Vol. 86, No. 3, pp. 373–378.

Ng, Serena (2014) “Viewpoint: Boosting Recessions,” *Canadian Journal of Economics*, Vol. 47, No. 1, pp. 1–34.

Nielsen, Finn Årup (2011) “A new ANEW: Evaluation of a word list for sentiment analysis in microblogs,” *arXiv preprint arXiv:1103.2903*.

Nyman, Rickard, Sujit Kapadia, David Tuckett, David Gregory, Paul Ormerod, and Robert Smith (2018) “News and narratives in financial systems: exploiting big data for systemic risk assessment,” *Bank of England Staff Working Papers*, Vol. 704.

Smeekes, Stephan and Etienne Wijler (2018) “Macroeconomic forecasting using penalized regression methods,” *International Journal of Forecasting*, Vol. 34, No. 3, pp. 408–430.

Stock, James and M. Watson (1999) *A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series*, pp. 1–44, Oxford: Oxford University Press.

Stock, James and M.W. Watson (2002a) “Forecasting Using Principal Components from a Large Number of Predictors,” *Journal of the American Statistical Association*, Vol. 97, p. 1167–1179.

——— (2002b) “Macroeconomic Forecasting Using Diffusion Indexes,” *Journal of Business and Economic Statistics*, Vol. 20, No. 2, p. 147–162.

Stock, J.H. and M.W. Watson (2002c) “Forecasting using principal components from large numbers of predictors,” *Journal of the American Statistical Association*, Vol. 97(460), pp. 1167–1179.

Stock, James H. and Mark W. Watson (2012) "Generalized Shrinkage Methods for Forecasting Using Many Predictors," *Journal of Business & Economic Statistics*, Vol. 30, No. 4, pp. 481–493.

Swanson, Norman R. and Halbert White (1997) "A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks," *The Review of Economics and Statistics*, Vol. 79, No. 4, pp. 540–550.

Tersvirta, Timo, Marcelo Medeiros, and Gianluigi Rech (2006) "Building neural network models for time series: a statistical approach," *Journal of Forecasting*, Vol. 25, No. 1, pp. 49–75.

Tetlock, P (2007) "Giving content to investor sentiment: the role of media in the stock market," *Journal of Finance*, Vol. 62(3), pp. 1139–1168.

Tibshirani, R (1996) "Regression shrinkage and selection via the lasso," *Journal of the Royal Statiscal Society*, Vol. Series B 58(1), pp. 267–288.

Vapnik, Vladimir (1995) *The nature of statistical learning theory*: Springer science & business media.

——— (1998) "Statistical Learning Theory. John Wiley&Sons," *Inc., New York*.

Varian, Hal R. (2014) "Big Data: New Tricks for Econometrics," *Journal of Economic Perspectives*, Vol. 28, No. 2, pp. 3–28, May.

Wang, Ye, Bo Wang, and Xinyang Zhang (2012) "A new application of the support vector regression on the construction of financial conditions index to CPI prediction," *Procedia Computer Science*, Vol. 9, pp. 1263–1272.

Wold, Hellman (1985) *Partial Least Squares*: Encyclopedia of statistical sciences.

Xiang-rong, Zhang, Hu Long-ying, and Wang Zhi-sheng (2010) "Multiple kernel support vector regression for economic forecasting," in *2010 International Conference on*

Management Science & Engineering 17th Annual Conference Proceedings, pp. 129–134, IEEE.

Zhao, Qingyuan and Trevor Hastie (2021) "Causal Interpretations of Black-Box Models," *Journal of Business & Economic Statistics*, Vol. 39, No. 1, pp. 272–281.

Zou, H and T Hastie (2005) "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society*, Vol. Series B 101(476), pp. 301–320.

A Details of the Dataset

Table A.1: Panel of survey balances

Code	Variable Name	Source	Transf
1	CBI Distributive trades total-SALES	CBI	L
2	CBI Distributive trades total-ORDERS	CBI	L
3	CBI Distributive trades total-SALES FOR TIME OF YEAR	CBI	L
4	CBI Distributive trades total-STOCKS	CBI	L
5	CBI Distributive trades Retailing-SALES	CBI	L
6	CBI Distributive trades Retailing-ORDERS	CBI	L
7	CBI Distributive trades Retailing-SALES FOR TIME OF YEAR	CBI	L
8	CBI Distributive trades Retailing-STOCKS	CBI	L
9	CBI Distributive tradesWholesaling-SALES	CBI	L
10	CBI Distributive tradesWholesaling-ORDERS	CBI	L
11	CBI Distributive tradesWholesaling-SALES FOR TIME OF YEAR	CBI	L
12	CBI Distributive tradesWholesaling-STOCKS	CBI	L
13	CBI Distributive trades Motor Trades-SALES	CBI	L
14	CBI Distributive trades Motor Trades-ORDERS	CBI	L
15	CBI Distributive trades Motor Trades-SALES FOR TIME OF YEAR	CBI	L
16	CBI Distributive trades Motor Trades-STOCKS	CBI	L
17	CBI Monthly Trends-Export Orders	CBI	L
18	CBI Monthly Trends-FG Stocks	CBI	L
19	CBI Monthly Trends-Expected Output	CBI	L
20	CBI Monthly Trends-Average Prices	CBI	L
21	CBI Monthly Trends-Reported Output	CBI	L
22	CIPSManufacturing-PMI	HIS/MARKIT CIPS	L
23	CIPSManufacturing- new orders	HIS/MARKIT CIPS	L
24	CIPSManufacturing-New export orders	HIS/MARKIT CIPS	L
25	CIPSManufacturing-Output	HIS/MARKIT CIPS	L
26	CIPSManufacturing-Employment	HIS/MARKIT CIPS	L
27	CIPSManufacturing-Suppliers' deliver times	HIS/MARKIT CIPS	L
28	CIPSManufacturing-Stock of purchases	HIS/MARKIT CIPS	L
29	CIPSManufacturing-Input prices	HIS/MARKIT CIPS	L
30	CIPSManufacturing-Quantity of purchases	HIS/MARKIT CIPS	L
31	CIPSManufacturing-Stock of finished goods	HIS/MARKIT CIPS	L
32	CIPSManufacturing-Output prices	HIS/MARKIT CIPS	L
33	CIPSManufacturing-Work backlogs	HIS/MARKIT CIPS	L
34	CIPSConstruction-Total activity	HIS/MARKIT CIPS	L
35	CIPSConstruction-Commercial activity	HIS/MARKIT CIPS	L
36	CIPSConstruction-Civil engineering activity	HIS/MARKIT CIPS	L
37	CIPSConstruction-Suppliers delivery times	HIS/MARKIT CIPS	L
38	CIPSConstruction-Employment	HIS/MARKIT CIPS	L
39	CIPSConstruction-Future business activity	HIS/MARKIT CIPS	L
40	CIPSConstruction-Housing activity	HIS/MARKIT CIPS	L

Code	Variable Name	Source	Transf
41	CIPSConstruction-Input Prices	HIS/MARKIT CIPS	L
42	CIPSConstruction-New orders	HIS/MARKIT CIPS	L
43	CIPSConstruction-Quantity of purchases	HIS/MARKIT CIPS	L
44	CIPSServices-Business activity	HIS/MARKIT CIPS	L
45	CIPSServices-Incoming new business	HIS/MARKIT CIPS	L
46	CIPSServices-Outstanding business	HIS/MARKIT CIPS	L
47	CIPSServices-Employment	HIS/MARKIT CIPS	L
48	CIPSServices-Average prices charged	HIS/MARKIT CIPS	L
49	CIPSServices-Average input prices	HIS/MARKIT CIPS	L
50	CIPSServices-Business expectations	HIS/MARKIT CIPS	L
51	GfK CC-Financial situation of households over last 12 months	European Commission	L
52	GfK CC-Financial situation of households over next 12 months	European Commission	L
53	GfK CC-General economic situation over last 12 months	European Commission	L
54	GfK CC-General economic situation over next 12 months	European Commission	L
55	GfK CC-Price trends over last 12 months (LHS)	European Commission	L
56	GfK CC-Price trends over next 12 months	European Commission	L
57	GfK CC-Unemployment over next 12 months	European Commission	L
58	GfK CC-Major purchases at present	European Commission	L
59	GfK CC-Major purchases over next 12 months	European Commission	L
60	GfK CC-Savings at present	European Commission	L
61	GfK CC-Savings over next 12 months	European Commission	L
62	GfK CC-Current financial situation of households (saving a lot v running into debt)	European Commission	L
63	CBI SSS-BusPro-Present level of Business	CBI	L
64	CBI SSS-BusPro-Volume business past 3 months	CBI	L
65	CBI SSS-BusPro-Volume business next 3 months	CBI	L
66	CBI SSS-BusPro-Number of employed past 3 months	CBI	L
67	CBI SSS-BusPro-Number of employed next 3 months	CBI	L
68	CBI SSS-BusPro-Average selling prices next 3 months	CBI	L
69	CBI SSS-Consumer-Present level of Business	CBI	L
70	CBI SSS-Consumer-Volume business past 3 months	CBI	L
71	CBI SSS-Consumer-Volume business next 3 months	CBI	L
72	CBI SSS-Consumer-Number of employed past 3 months	CBI	L
73	CBI SSS-Consumer-Number of employed next 3 months	CBI	L
74	CBI SSS-Consumer-Average selling prices next 3 months	CBI	L
75	CBI SSS-Business-Present level of Business	CBI	L
76	CBI SSS-Business-Volume business past 3 months	CBI	L
77	CBI SSS-Business-Volume business next 3 months	CBI	L
78	CBI SSS-Business-Number of employed past 3 months	CBI	L
79	CBI SSS-Business-Number of employed next 3 months	CBI	L
80	CBI SSS-Business-Average selling prices next 3 months	CBI	L
81	CBI SSS-Professional-Present level of Business	CBI	L
82	CBI SSS-Professional-Volume business past 3 months	CBI	L
83	CBI SSS-Professional-Volume business next 3 months	CBI	L
84	CBI SSS-Professional-Number of employed past 3 months	CBI	L
85	CBI SSS-Professional-Number of employed next 3 months	CBI	L
86	CBI SSS-Professional-Average selling prices next 3 months	CBI	L
87	CBI SSS-Total Sector-Present level of Business	CBI	L
88	CBI SSS-Total Sector-Volume business past 3 months	CBI	L
89	CBI SSS-Total Sector-Volume business next 3 months	CBI	L
90	CBI SSS-Total Sector-Number of employed past 3 months	CBI	L
91	CBI SSS-Total Sector-Number of employed next 3 months	CBI	L
92	CBI SSS-Total Sector-Average selling prices next 3 months	CBI	L
93	NIESR-Monthly GDP	NIESR	LD
94	Lloyds Business Barometer-Overall Business Confidence	Lloyds Bank	L
95	Lloyds Business Barometer-Business Activity next year	Lloyds Bank	L
96	Lloyds Business Barometer-Economic Optimism	Lloyds Bank	L
97	Monthly GDP	ONS	LD

Note: Sources are the Office for National Statistics (ONS), the Bank of England database (BOE), IHS Markit/CIPS, the Confederation of British Industries (CBI), LLoyds Bank, the European Commission. Transformation codes: LDD = log double difference, LD = log difference, L = levels, D = first difference.

Table A.3: Dataset augmented with Macroeconomic Series

Code	Variable Name	Source	Transf
1	IoS: Services, Index	ONS	LD
2	PNDS: Private Non-Distribution Services: Index	ONS	LD
3	IoS: G: Wholesales, Retail and Motor Trade: Index	ONS	LD
4	IoS: 47: Retail trade except of motor vehicles and motorcycles: Index	ONS	LD
5	IoS: 46: Wholesale trade except of motor vehicles and motorcycles: Index	ONS	LD
6	IoS: 45: Wholesale And Retail Trade And Repair Of Motor Vehicles And Motorcycles: Index	ONS	LD
7	IoS: O-Q: PAD, Education and Health Index	ONS	LD
8	IoP:Production	ONS	LD
9	IoP:Manufacturing	ONS	LD
10	Energy output (utilities plus extraction) Pound Sterling (Index	ONS	LD
11	IoP: SIC07 Output Index D-E: Utilities: Electricity, Gas, Water Supply, Waste Management.	ONS	LD
12	IOP: B:MINING AND QUARRYING:	ONS	LD
13	RSI:VolumeAll Retailers inc fuel:All Business Index	ONS	LD
14	Construction Output: Seasonally Adjusted: Volume: All Work	ONS	LD
15	BOP Total Exports (Goods)	ONS	LD
16	BOP:EX:volume index:SA:Total Trade in Goods	ONS	LD
17	BOP Total Imports (Goods)	ONS	LD
18	BOP:IM:volume index:SA:Total Trade in Goods	ONS	LD
19	CPI all items	ONS	LDD
20	RPI all items	ONS	LDD
21	RPI ex Mortgages Interest Payments (RPIX)	ONS	LDD
22	PPI Output	ONS	LDD
23	PPI Input	ONS	LDD
24	Nationwide House Price MoM	BoE database	D
25	RICS House Price Balance	BoE database	D
26	M4 Money Supply	BoE database	LD
27	New Mortgage Approvals	BoE database	LD
28	Bank of England UK Mortgage Approvals	BoE database	LD
29	Average Weekly Earnings	ONS	LD
30	LFS Unemployment Rate	ONS	D
31	LFS Number of Employees (Total)	ONS	LD
32	Claimant Count Rate	ONS	D
33	New Cars Registrations	BoE database	LD
34	Oil Brent	BoE database	LD
35	UK mortgage base rate	BoE database	L
36	3m LIBOR	BoE database	L
37	FTSE all share	BoE database	LD
38	Sterling exchange rate index	BoE database	LD
39	FTSE volatility	BoE database	LD
40	GBP EUR spot	BoE database	LD
41	GBP USD spot	BoE database	LD
42	FTSE 250 INDEX	BoE database	LD
43	FTSE All Share	BoE database	LD
44	UK focused	BoE database	LD
45	S&P 500	BoE database	LD
46	Euro Stoxx	BoE database	LD
47	Sterling ERI	BoE database	LD
48	VIX	BoE database	LD
49	UK VIX - FTSE 100 VOLATILITY INDEX - PRICE INDEX	BoE database	LD

Note: Sources are the Office for National Statistics (ONS), the Bank of England database (BOE), IHS Markit/CIPS, the Confederation of British Industries (CBI), LLoyds Bank, the European Commission. Transformation codes: LDD = log double difference, LD = log difference, L = levels, D = first difference.

B Additional Empirical Results

Table B.1: Relative RMSFE Across Specifications

relative to PCR	(1)	(3)	(6)	(9)	(12)	(24)
Lasso	0.964***	0.958	0.941	0.931	0.944	0.996
Ridge	0.702***	0.860	0.857	0.904	0.953	1.115
Elastic	0.940***	0.939	0.926	0.920	0.937	1.009
PLS	0.837***	0.841	0.827	0.844	0.869	1.045
Random Forest	0.722***	0.874	0.877	0.890	0.927	1.051
SVM	0.702***	0.866	0.856	0.884	0.923	1.111
NN	0.698	0.886	0.850	0.898	0.948	1.116
relative to AR(1)	**(1)**	**(3)**	**(6)**	**(9)**	**(12)**	**(24)**
Lasso	0.957***	0.955	0.940	0.931	0.945	0.997
Ridge	0.696***	0.858	0.856	0.904	0.953	1.115
Elastic	0.933***	0.937	0.925	0.920	0.938	1.010
PLS	0.831***	0.839	0.826	0.844	0.869	1.045
Random Forest	0.809***	0.884	0.879	0.891	0.923	1.078
SVM	0.697***	0.864	0.855	0.884	0.923	1.112
NN	0.779***	0.873	0.887	0.897	0.942	1.143***

Note: Top panel: Relative RMSFEs across different specifications using combined *Survey*, *Text* and *Macro* data against a **PCR** model. Mid panel: RSMFEs relative to an **AR(1)**. *,** & *** denote rejection at 10%, 5% and 1% level of the null hypothesis of equal forecasting method accuracy of the Diebold and Mariano (1995) test against the benchmark models

Table B.2: RMSFE Across Different Models

relative to PCR	(1)	(3)	(6)	(9)	(12)	(24)
Lasso	0.900***	0.913	0.896	0.898	0.918	1.033
Ridge	0.902***	0.914	0.897	0.898	0.918	1.033
Elastic	0.984	0.974	0.959	0.940	0.950	0.990
PLS	0.935***	0.931	0.924	0.916	0.932	1.012
Random Forest	0.930***	0.941	0.935	0.926	0.941	1.013
SVM	0.821***	0.881*	0.905	0.912	0.939	1.040
NN	0.870***	0.960	0.887	0.904	0.945	1.018
	(1)	(3)	(6)	(9)	(12)	(24)
Lasso	0.957***	0.955	0.940	0.931	0.945	0.997
Ridge	0.854***	0.883	0.862	0.882	0.896	1.065
Elastic	0.944***	0.946	0.932	0.925	0.941	1.003
PLS	0.883***	0.899	0.876	0.888	0.899	1.045
Random Forest	0.931***	0.946	0.920	0.922	0.939	1.020
SVM	0.720***	0.885	0.872	0.896	0.929	1.118
NN	0.858***	0.888	0.901	0.895	0.935	1.126
	(1)	(3)	(6)	(9)	(12)	(24)
Lasso	0.984***	0.974	0.959	0.940	0.950	0.990
Ridge	0.692***	0.828*	0.854	0.901	0.946	1.097
Elastic	0.798***	0.835**	0.843	0.873	0.914	1.062
PLS	0.811***	0.831*	0.839	0.868	0.909	1.058
Random Forest	0.787***	0.880	0.872	0.882	0.917	1.044
SVM	0.657*	0.871	0.864	0.895	0.936	1.125
NN	0.661*	0.849	0.877	0.903	0.952	1.148

Note: Top panel: Relative RMSFEs across different specifications using only *Text* data. Middle panel: Relative RMSFEs across different specifications using only *Macro* data. Bottom panel: Relative RMSFEs across different specifications combining *Survey* and *Text* data. All results are reported against an **AR(1)** model as a benchmark. *,** & *** denote rejection at 10%, 5% and 1% level of the null hypothesis of equal forecasting method accuracy of the Diebold and Mariano (1995) test against the benchmark specification.