Economic Modelling 122 (2023) 106204

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.journals.elsevier.com/economic-modelling

Nowcasting Chinese GDP in a data-rich environment: Lessons from machine learning algorithms

Qin Zhang [a,b], He Ni [a,*], Hao Xu [c]

[a] Tailong Finance School, Zhejiang Gongshang University, China
[b] Collaborative Innovation Center of Statistical Data Engineering Technology & Application, Zhejiang Gongshang University, China
[c] Fintech Research Centre, Zhejiang Lab, China

ARTICLE INFO

JEL classification:
C32
C53
E37

Keywords:
Nowcasting
China's macroeconomy
Machine learning algorithm
Dynamic factor model
Real GDP

ABSTRACT

The ability to estimate current GDP growth before official data are released, known as "nowcasting", is crucial for the Chinese government to effectively implement economic policy and manage economic uncertainties; however, there is limited research on nowcasting China's GDP in a data-rich environment. We evaluate the performance of various machine learning algorithms, dynamic factor models, static factor models, and MIDAS regressions in nowcasting the Chinese annualised real GDP growth rate in pseudo out-of-sample exercise, using 89 macroeconomic variables from years 1995 to 2020. We find that some machine learning methods outperform the benchmark dynamic factor model. The machine learning method that deserves more attention is ridge regression, which dominates all other models not only in terms of nowcast error but also in effective recognition of the impacts of the Global Financial Crisis and Covid-19 shocks. Policy-wise, our study guides practitioners in selecting appropriate nowcasting models for China's macroeconomy.

1. Introduction

The implementation of macroeconomic policy in real-time heavily requires assessments of precise information regarding the current state of the economy; however, obtaining accurate and timely estimates of key macroeconomic indicators such as Gross Domestic Product (GDP), in particular during times of economic uncertainty, is well known to be challenging because macroeconomic variables are released with substantial delays. Nowcasts, which refer to assessments of the current or most recent aggregate state of an economy based on a range of partially available indicators before the official economic figures are released, have been widely used by many central banks and other international institutions to mitigate some of these uncertainties (Hendry et al., 2016). Nowcasts can be undertaken at the stage of initial estimation of macroeconomic variables, possibly with incomplete data, in order to better understand current economic conditions, with the ability to update estimates as more information becomes available (Bańbura et al., 2013). They can also help identify turning points or significant shifts in an economy's momentum (Giannone et al., 2008; Zhang et al., 2018).

The task of nowcasting GDP is even more challenging in emerging markets due to the relatively short time span of macroeconomic data available, missing observations at the start and during the sample periods, variances in data quality, and the swiftness of continuous structural change. China, which was largely closed before the 1970s, is fully integrated into the global economy and is now the second-largest economy in the world. The story of China's rapid economic development began with a boom in investment and imports and exports since the 1980s. Recently, China's GDP growth has been slowing down, though, as the country's domestic overcapacity has become serious and the external demand has continued to fall. The government has implemented a series of reforms to restructure and upgrade its economy since reform and opening up. Domestic demand, particularly consumption, slowly began to assume responsibility for sustaining GDP growth. Given how China's macroeconomic condition has changed over the past few years, we argue that choosing dominating predictors according to economic theories is insufficient. Instead, selecting predictors with dynamic properties is a must to nowcast China's GDP growth. The nowcasting model should be able to handle a large number of predictors, deal with missing observations, and be robust to data quality in rigorous and systematic matters (Yiu and Chow, 2010; Higgins et al., 2016; Sun et al., 2018).

Given the growing popularity of models that address big data issues and the urgent need for systematic research in the evaluation of out-of-sample nowcasts of China's macroeconomy, a comprehensive study that compares the performance of various nowcasting models is required. This paper contributes to filling this gap by undertaking a challenging

* Correspondence to: Tailong Finance School, Zhejiang Gongshang University, No. 18, Xuezheng Str., Xiasha University Town, Hangzhou, 310018, China.
E-mail addresses: zhangqinwenya@mail.zjgsu.edu.cn (Q. Zhang), nihe@zjgsu.edu.cn (H. Ni), xuhao.econ@outlook.com (H. Xu).

https://doi.org/10.1016/j.econmod.2023.106204
Received 14 July 2022; Received in revised form 15 January 2023; Accepted 22 January 2023
Available online 4 February 2023

yet difficult task: nowcasting annualised quarterly real GDP growth rate for mainland China in a data-rich environment. Our work is inspired by the following question: do the nowcasting methods that are well-established for Western macroeconomics also work well for China? This is particularly important as only limited studies address the problem of nowcasting China's GDP in the existing literature, and none are sufficient to answer the question that we pose above: see Yiu and Chow (2010),[1] Barnett and Tang (2016),[2] and Zhang et al. (2018).[3]

We consider three classes of nowcasting models, each of which has proved itself useful for nowcasting GDP growth rate in developed countries (Mullainathan and Spiess, 2017; Bok et al., 2018). The first class consists of various specifications of machine learning (ML) algorithms, namely ridge regression, least absolute shrinkage and selection operator (LASSO), elastic net, feed-forward neural network, K-nearest neighbour regression, regression tree, the macroeconomic random forest of Coulombe (2020a), and two-stage ridge regression of Coulombe (2020b). The second class of model contains factor models that assume the dynamics of the state of an economy are governed by a small number of common components. We consider two broad streams of factor models: the dynamic factor model (DFM) of Giannone et al. (2008), and the static factor models with factors extracted by principal component analysis and partial least squares. In the third class of models, we have mixed-data sampling (MIDAS) of Ghysels et al. (2004), and AR-MIDAS of Clements and Galvão (2009).

Naturally constrained by data availability and relatively short samples for many series, we implement nowcasting algorithms with a relatively small-scale dataset (compared to Western economics) of monthly indicators, which are manually selected to best represent the Chinese economy. Our dataset contains 89 monthly series over the period from January 1995 to December 2020. The sample period covers both the sharp downturn of economic activity during the Global Financial Crisis and the Covid-19 period, allowing us further examine which nowcasting models are effective during times of economic uncertainty. Furthermore, to examine the effects of various data dimensions on the effectiveness of nowcasting, we divide our dataset into two subsets: the soft dataset and the hard dataset. The soft dataset contains survey data, financial market variables and various price indexes, whereas the hard dataset contains variables used for GDP calculations, such as the industrial production index and international trade data.

We set up our investigation on the fixed window scheme, which facilitates comparing nowcasts from nested models over a fixed estimation sample size.[4] We train each nowcasting algorithm once on the 1995Q1 to 2008Q1 in-sample period, and then generate nowcasts over the 2008Q1 to 2020Q4 out-of-sample period. Given our total sample size, this creates a rough 50:50 per cent split[5] between training and testing samples. Comparison of the nowcasting models is based on their pseudo out-of-sample performance along three metrics: the root mean square nowcasting error, the conditional predictive test of Giacomini and White (2006), and the model confidence set (MCS) of Hansen et al. (2011).[6]

We find two intriguing and interesting empirical results. First, ridge, LASSO, elastic net, regression tree and two-stage ridge are the best-performing models as they produce nowcasts that are superior to the nowcast generated by the benchmark DFM model and are covered by the MCS at the 25% level of significance. The model that stands out is ridge regression, due to not only it reduces overall nowcast errors by 77% but also demonstrating an extraordinary ability to recognise the impacts of the Global Financial Crisis and Covid-19 shock. Second, the total dataset possesses more useful information to differentiate the nowcast than the hard and soft dataset; therefore, policymakers should incorporate as much information as possible to make a nowcast. To open the black box of ML methods, we compare the predictors selected by some of the best-performing ML algorithms and investigate the relative importance of individual predictors using the variable importance measures. We find price indices and trade data have the most predictive power and variable importance measure provides a highly consistent summary of which variables are most influential in improving nowcast accuracy, which can help uncover the main predictors for nowcasting and forecasting future economic conditions, possibly shedding light on the drivers of price dynamics.

This paper joins the growing body of empirical application in evaluating the relative success of nowcasting models for China. We contribute to the literature in a number of ways. First, to the best of our knowledge, our paper is a rare attempt to put different classes of models together and compare their predictive performance in a pseudo out-of-sample nowcasting experiment for China. In contrast, the existing studies such as Yiu and Chow (2010), Jiang et al. (2017) and Zhang et al. (2018) typically consider one model (i.e. the dynamic factor model and MIDAS) at a time. Among others, we note the work by Jiang et al. (2017), who use a MIDAS regression to forecast China's GDP. Our work differs from Jiang et al. (2017) by comparing a large number of nowcasting algorithms that have been widely used by Western economies, including MIDAS models, dynamic and static factor models, and various specifications of ML algorithms. The selected nowcasting algorithms have the ability to capture nonlinear and non-parametric relationships, deal with issues raised by mixed frequency and missing observations, and are robust to the structural change that China has undergone. Second, we draw attention to the primary set of variables that account for these nowcast enhancements. Our findings indicate that this set of variables is not dense, which contradicts the conclusions of Giannone et al. (2008). In fact, we discover that the most effective ML models are those that do enforce sparsity and regularisation. Dynamic factor models, which have been among the most widely used models for macroeconomic nowcasting, have a high level of aggregation, yet this level is insufficient for China.

The remainder of the paper is organised as follows. Section 2 reviews the literature related to nowcasting in a data-rich environment. Section 3 describes our data. Section 4 outlines our nowcasting methodology and algorithms, while Section 5 presents model estimation and nowcasting evaluation. The empirical results are discussed in Section 6, and Section 7 provides concluding remarks.

[1] Yiu and Chow (2010) employ the dynamic factor model to nowcast the GDP growth rate for the most recent quarter and discover that it can outperform the benchmark random walk model.

[2] Barnett and Tang (2016) find that monetary aggregate variables are more illuminating than other variables, aiding the dynamic factor model in producing the best nowcasting outcomes currently possible.

[3] Zhang et al. (2018) find that the Bayesian approach may be a viable option for nowcasting China's GDP.

[4] The one major benefit of using fixed window is that it deals with model instability, which is the issue facing China. With the fixed window scheme, the estimation process excludes "old" data, which prevents bias in the parameter estimates. In circumstances where the parameters do not change considerably over time, expanding window estimates, which make better use of all data, tend to be less volatile than fixed window estimates. To compare the nowcasting results generated by the fixed window and expanding window methods, we present rRMSE and MCS p-value results for the expanding window method in Table 7 of Appendix A. We find that the nowcasting results generated by the fixed window and expanding window are similar.

[5] To further examine if nowcasting results are robust for different percentages of sample splitting, we simulate our nowcasting exercise using 75% and 25% in-sample and out-of-sample split in Table 6 of Appendix A. We show that regularised regressions, especially ridge regression, dominate other nowcasting models in terms of relative RMSE; therefore, our results are robust for a 75:25 percent training and testing sample split.

[6] The question of model selection and how it affects macroeconomic forecasting has long been a focus of applied macroeconometrics. This was due to the fact that empirical work typically lacks knowledge of the most parsimonious model incorporating the data-generating process. The best nowcasting model is first identified, and the subset of models that are not statistically different from the best model at the specified confidence level are then chosen. The best set of nowcasting models in the literature have frequently been chosen using the MCS; for instance, Kotchoni et al. (2019) and Coulombe et al. (2022).

2. A brief literature review

Nowcasting has a long tradition in empirical macroeconomics and has become the full-time job of dedicated economists at central banks and government agencies. In the past, policy-makers typically relied on the combination of the "bridge equation" and qualitative judgement to make a nowcast.[7] However, these methods might not be effective in a data-rich environment. Today, data is released to the public on a regular basis: Almost every week, new information is made available to analyse, comment on and interpret. This motivates macroeconomists to embrace the big data challenge, pushing the frontier of statistical methods and refined measurement in macroeconomic forecasting and nowcasting. Over the past 20 years, developments in time series econometrics have made it possible to include arbitrarily many series in nowcasting systems and to incorporate data release in real-time (Stock and Watson, 2017). This section reviews the related literature from the following three aspects: the dynamic factor model, ML algorithms, and the MIDAS regressions.

The initial attempts to produce nowcasts within a formalised and coherent statistical framework were due to Giannone et al. (2008) and Evans (2005) under the name of "dynamic factor models". The model has three appealing features to deal with typical problems arising from the different releases of the economic indicators. First, the model is a simple algorithm that can be automatically updated, which allows researchers to interpret and comment on various data released at different frequencies with different publication lags in terms of the signals they provide on current economic conditions. This is possible because the Kalman filter generates projections of missing observations for all variables and updates the nowcast as new data are released. Second, the model is robust for potential economic instabilities by reducing the weight and loading factor of any variable that loses predictive power over time. Third, the model is dynamically complete in the sense that it takes into account the dynamics of all indicators used in the analysis. Subsequent research using the dynamic factor model to nowcast current GDP include (Lahiri and Monokroussos, 2013) for the United States; Bańbura et al. (2013), Bańbura and Rünstler (2011) and Camacho and Perez-Quiros (2010) for the aggregate Euro area; Barhoumi et al. (2010) for France; and Matheson (2010) for New Zealand. To date, the dynamic factor model has been widely used by a number of international institutions and central banks as a standard tool for up-to-date GDP nowcasting, some example users being the Federal Reserve Bank of Atlanta (Higgins, 2014),[8] the Federal Reserve Bank of New York (Aarons et al., 2016),[9] the Bank of England (Hendry et al., 2016),[10] the European Central Bank (Bańbura et al., 2013)[11] and the International Monetary Fund (Matheson, 2011).

The MIDAS framework was proposed by Ghysels et al. (2004, 2007) to flexibly deal with data sampled at different frequencies. It incorporates high-frequency data into the regression without transforming predictors into latent factors, avoiding possible loss of potentially useful information and misspecification; therefore, MIDAS is a powerful tool for direct forecasting tasks, linking low-frequency data with current and lagged high-frequency indicators at multi-steps forecast horizon. In literature, MIDAS models have been used to forecast quarterly GDP using monthly economic variables. For example, Jiang et al. (2017) combine the dynamic factor model and MIDAS model to forecast China's quarterly GDP growth rate. They find the MIDAS models based on latent factors extracted from the dynamic factor model have better forecasting performance compared to traditional forecasting methods.

A similar conclusion is obtained by Kuzin et al. (2011) for nowcasting GDP in the euro area. Foroni et al. (2020) consider using MIDAS and Unrestricted MIDAS models with a variety of indicators to improve the growth nowcasts and forecasts during the Covid-19 crisis and recovery period. Financial variables that are forward-looking and contain information about the future state of the economy are later considered to be highly relevant to macroeconomic nowcasting (Andreou et al., 2013). Compared to the dynamic factor model, the MIDAS has the advantage of using high-frequency financial data that are observed without measurement error and are not subjected to revisions as opposed to macroeconomic variables.

Despite the effectiveness of dynamic factor models and MIDAS in producing timely and accurate nowcasts, there are arguments as to the parametric restrictions that make these models work discard potentially important information (Carrasco and Rossi, 2016; Stock and Watson, 2017). Instead, there is an exploitable nonlinear and non-parametric structure that could be revealed by modern ML algorithms (Stock and Watson, 2017; Mullainathan and Spiess, 2017). ML algorithms have recently been proposed as new tools by which empirical economists could solve the prediction problem rather than the parameters estimation problem (Mullainathan and Spiess, 2017; Athey, 2018). They offer a solution for *the curse of dimensionality* and *regularisation* to standard regression analysis by using a more robust and complex architecture, and without having the need to transform variables to latent factors (Athey, 2018; Athey and Imbens, 2019).

The current literature has shown large improvements in favour of ML algorithms in macroeconomic forecasting and nowcasting tasks. For instance, Coulombe et al. (2021) assess the forecasting performance of various standard and ML forecasting methods for key UK economic variables. They find ML methods are capable of providing substantial gains for forecasting several indicators of the UK macroeconomy in the short term, and methods with non-linearity features are particularly suitable for the Covid-19 period. Medeiros et al. (2021) show that, among various ML methods, the random forest and LASSO model dominate all other models, perhaps their superior performance are due to non-linearities and variable selection mechanism. Richardson et al. (2021) commence the work of utilising 41 real-time data vintages to nowcast New Zealand GDP. Their findings show that ML techniques can outperform both the dynamic factor model and the autoregressive. With an actual application to nowcasting US GDP growth, Babii et al. (2021) present oracle inequalities for the sparse-group LASSO estimators in recognition of the possibility that the financial and macroeconomic data may have heavier than exponential tails. They demonstrate that the estimator beats competing solutions and that text data can be an advantageous addition to more conventional macroeconomic data. The most comprehensive study to date is conducted by Coulombe et al. (2022). They move beyond the question of, *Is machine learning useful for macroeconomic forecasting?* by adding the question, *how* and they find that machine learning is useful for macroeconomic forecasting because it mostly captures important non-linearities that arise in the context of uncertainty and financial frictions.

3. Data

We collect a panel of 89 monthly predictors, spanning from January 1995 to December 2020. Our sample period covers the transition of

[7] The idea of bridge equation comes from using small-scale models to "bridge" the information contained in one or a few key monthly data with the quarterly GDP, which is released after the monthly data.

[8] The Quantitative Economic Research Department in the Federal Reserve Bank of Atlanta frequently produces nowcasts of United States GDP before United States Bureau official releases. It nowcasts 13 separate expenditure components of the GDP to mimic the expenditure approach to calculating GDP using dynamic factor modelling.

[9] The Federal Reserve Bank of New York uses a dynamic factor approach to provide a model-based nowcast of GDP. However, it does not mimic a particular approach to calculating GDP.

[10] The Bank of England's Monetary Policy Committee uses a compilation of nowcasts from three different models (the MIDAS model, the dynamic factor model and judgements based on different industries) to form its initial view on the current state of the economy.

[11] The European Central Bank also uses dynamic factor-based nowcasting models to inform its policy decisions. Its staff have released a number of working papers, which are cornerstones of the nowcasting literature.

the Chinese economy since the reform and opening up, the Asian financial crisis (from July 1997 to December 1998),[12] the Global Financial Crisis period (from December 2007 to June 2009),[13] and China's rapid subsequent recovery (from July 2009 to December 2019), and the Covid-19 period (from March 2020 to December 2020). The set of predictors covers all important sectors of China's macroeconomy, including the government sector, international trade, the industrial sector, the service sector, the monetary sector, and the financial sector. We source our Chinese data from the CEIC data manager. The variables used to compute nowcasts are shown in Table 8 of Appendix B.

Since proposed ML algorithms require predictors to be quarterly to make a nowcast, we apply the temporal aggregation of collected monthly variables X_M to obtain the set of quarterly predictors X_Q. For stock variables such as interest rates and price indices, we treat X_Q as the monthly average of X_M in a quarter. For flow variables such as investment and industrial production, we treat X_Q as the monthly aggregation of X_M in a quarter. This temporal aggregation process avoids the issue of conflating information sets with different frequencies and ensures an apples-to-apples comparison between the ML algorithms, factor models and MIDAS. Finally, we induce all predictors to stationarity following McCracken and Ng (2016)'s approach.

All predictors were subjected to five preliminary steps: possible deseasonalisation, temporal aggregation, possible transformation by taking either the difference or the percentage change, standardisation, and screening for possible outliers. Seasonal patterns are common in China's macroeconomic data, and we use *X-13ARIMA-SEATS* in *R* platform to address them.[14] A judgement decision was made over whether to use the difference or the percentage change, primarily based on an examination of the time series plots and unit root tests. Predictors that are already expressed in index or percentage terms typically have the difference taken into account, while predictors that have actual quantities are typically given the percentage change. Additionally, we clean the dataset by replacing outliers with missing data and deleting outliers for which the first difference deviates from the median by more than five times the interquartile range. Our last concern is to balance the resulting panel since some predictors have missing observations. Following Stock and Watson (2002) and McCracken and Ng (2016), we apply an expectation–maximisation algorithm which uses factor models to fill missing observations.[15]

The time lag of the selected economic and financial variables is a must to consider. We do not include any predictors that have a time lag longer than one month because the release of the first estimate of Chinese GDP data is quite timely, with a time lag of about three weeks after the end of the corresponding quarter. Our monthly variables are all released during the reference month (i.e., the calendar date of the observation). Variables with a publication lag greater than 3 weeks are treated as missing observations since Chinese GDP is usually released with 3 weeks lag, and we estimate them using the expectation–maximisation algorithm to obtain the balanced set of predictors. Concerning the timing of data releases, we detailed release time and publishing lag in Table 8 of Appendix B.

The final note is that the data we examine are not real-time because we do not examine a sequence of revisions for each calendar-dated observation. In this sense, our experiments are only pseudo real-time. We defer to future studies since real-time datasets for China are still being developed, which would enable us to undertake true real-time prediction if they were available in the future.

[12] Sources: Federal Reserve History.

[13] Sources: NBER Business Cycle Reference Dates.

[14] Developed by the United States Census Bureau, the *X-13* supports user-defined holiday variables such as Chinese New Year. It returns deseasonalised predictors if seasonality is detected and returns original predictors if seasonality is not detected.

[15] Predictors for the construction of factor models are further standardised to mean zero and unit variance.

4. Nowcasting framework

We consider algorithms that form predictions about the current growth rate of Chinese real GDP on the basis of the available historical and contemporaneous information. Let N be the number of variables and T be the number of time series observations. For $t = 1, \dots, T$ and $n = 1, \dots, N$, we have $X_t = [X_{1t}, \dots, X_{nt}]$ be the vector of variables found in macroeconomic dataset and y_t is our target variable at time t. With reference to Coulombe et al. (2021), we consider the general nowcasting framework:

$$\min_{g \in \mathcal{G}} \left\{ \sum_{t=1}^{T} (y_t - g(f(X_t))^2) + \text{pen}(g; \tau) \right\} \tag{1}$$

The function g, chosen in the functional space $\mathcal{G}$, maps the transformed inputs into our target variable. $f()$ represents the function of data pre-processing (or feature engineering in ML literature) effects on nowcasting performance, and $pen()$ is the regularisation function whose strength depends on hyperparameter(s) τ. We now discuss our nowcasting algorithms.

4.1. Machine learning algorithms

ML algorithms offer ways to approximate unknown and potentially complicated functional forms to minimise the expected loss of a nowcast. For each ML algorithm considered, myriad varieties are proposed in the literature, so it would be challenging to employ all possible varieties.[16] Since this study is an early attempt to examine the predictive performance of ML algorithms for China's macroeconomy, we focus on investigating off-the-shelf ML algorithms susceptible to improve macroeconomic nowcasting. These ML algorithms are some of the most commonly used models in the literature on macroeconomic forecasting. Below are the considered algorithms.

LASSO, Ridge Regression (RR) and Elastic Net (E-Net): LASSO, RR and E-Net are similar methods to avoid over-fitting the regression model by incorporating penalties on the vector of the regression coefficients. The difference between LASSO, RR and E-Net is the type of regularisation they implement. They estimate the parameters by minimising the sum of least squares residuals subject to different penalties, which lead to the following objective functions:

$$\begin{aligned} &\arg\min_{\beta} \\ &\text{1. RR (L1 norm)}: \quad \sum_{t=1}^{T} \left(y_t - X_t\beta\right)^2 + \lambda \sum_{i=1}^{N} \left(\beta_i^2\right) \\ &\text{2. LASSO (L2 norm)}: \quad \sum_{t=1}^{T} \left(y_t - X_t\beta\right)^2 + \lambda \sum_{i=1}^{N} \left(|\beta_i|\right) \\ &\text{3. E-Net (L1+L2 norm)}: \quad \sum_{t=1}^{T} \left(y_t - X_t\beta\right)^2 + \lambda \sum_{i=1}^{N} \left(\alpha |\beta_i| + (1-\alpha)\beta_i^2\right) \end{aligned} \tag{2}$$

In RR, parameters of least relevant variables are shrunk uniformly towards zero, whereas LASSO performs the variable selection procedure to eliminate irrelevant variables. E-Net can be interpreted as a hybrid version of RR and LASSO regressions. It can either shrink the coefficients close to zero or completely remove them. Hyperparameters λ_1 and λ_2 control the strength of the penalty, and α assigns the weight of L1- and L2-norm penalties. A larger β leads to a greater penalty and a stronger effect of shrinkage and regularisation of regression

[16] For example, common varieties of artificial neural networks include feed-forward neural networks, radial basis function neural networks, multilayer perceptrons, convolutional neural networks, recurrent neural networks, modular neural networks, and sequence-to-sequence models.

coefficients. The values of λ_1 and λ_2 are tuned by the K-Fold cross-validation of the training dataset, whereas α is selected via minimum within the sample MSE.

Feed-Forward Neural Network (FFNN): Inspired by how the human brain works, the neural network is an information-processing algorithm that resembles biological nervous systems. While there are several variations of the neural network, we consider the FFNN, which is perhaps the most general type of artificial neural network.[17] One distinct feature of FFNN is that connections between nodes do not form a cycle. The model works forward only by feeding datasets into the network through input neurons, which trigger the layers of hidden neurons and finally turn into output neurons. More precisely, each input is weighted and passed through an activation function to determine the value of nodes in the first layer of perceptrons.[18] Then, nodes from the first layer become new inputs (re-weighted) for the second layer and are passed through a new activation function to determine the value of nodes in the third layer. This process is repeated until the N_{th} layer is created. Finally, nodes in the last layer are re-weighted and passed through the final activation function to generate the output value.

The key hyperparameters of FFNN are the number of nodes on each layer and the number of hidden layers. We use a single hidden layer because the universal approximation theorem states that FFNN with a single hidden layer and a finite number of neurons can approximate any continuous function for inputs within a specific range (Hornik et al., 1989). The number of nodes is selected via K-fold cross-validation.

K-Nearest Neighbour (KNN): KNN is a non-parametric method that works by predicting the outcome of new data based on K most similar observations (neighbours) in the training dataset. More specifically, given a data point (in this case, the GDP growth rate), we compute the Euclidean distance between that point and all points in the training dataset. Then, the nowcast of China's GDP growth rate is the average of the target output value by the closest K training data points. The hyperparameter is the value of K and is determined by K-Fold cross-validation.

Decision Tree (DT): DT for regression is based on a hierarchical tree-like partition of the input space wherein data are partitioned into smaller and smaller subsets that contain similar values. The tree consists of internal decision nodes and terminal leaves. The flow begins at the root node for any given data point and proceeds through a series of tests and decision nodes until it reaches a leaf node. The tree uses within-node variance in the regression process to divide the data into the paths that lead to the greatest variance reduction. Up until the termination criterion is reached, the data splitting is carried out iteratively. Usually, the within-node sample mean of each leaf node serves as the final prediction. The hyperparameters include the tree's maximum depth, the minimum number of samples needed to split an internal node, and the number of features to take into account when looking for the best split.

Macroeconomic Random Forests (MRF): MRF is proposed by Coulombe (2020a) as a new form of the random forest better suited for macroeconomic data by considering generalised time-varying parameters. It reads as:

$$y_t = \tilde{X}_t \beta_t + \varepsilon_t \tag{3}$$

$$\beta_t = \mathcal{F}(S_t) \tag{4}$$

where S_t (usually smaller than X_t) is the state variable determining the linear model that we want to be time-varying, X_t is a high-dimensional macroeconomic data set, and $\mathcal{F}$ is a forest. The problem is then finding an optimal variable (j) out of the random subset of predictors $(\mathcal{J}^-)$ to split the sample with and at which value (c) we should split the variable. Coulombe (2020a) develops a greedy algorithm to recursively construct trees as:

$$\min_{j\in\mathcal{J}^-,c\in\mathbb{R}}\left[\min_{\beta_1}\sum_{t\in I_1^{RW}(j,c)} w(t;\zeta)\left(y_t-\tilde{X}_t\beta_1\right)^2+\lambda\left\|\beta_1\right\|_2 + \min_{\beta_2}\sum_{t\in I_2^{RW}(j,c)} w(t;\zeta)\left(y_t-\tilde{X}_t\beta_2\right)^2+\lambda\left\|\beta_2\right\|_2\right] \tag{5}$$

where the Kernel trick $w(t;\zeta)$ that assigns a weight of 1 on observation t, a weight of $\zeta < 1$ for observations $t-1$ and $t+1$ and a weight of ζ for observations $t-2$ and $t+2$.

Coulombe (2020a) and Coulombe et al. (2021) demonstrate that the newly proposed algorithm forecasts better than off-the-shelf ML algorithms and traditional econometric approaches. The intuition is rather straightforward: it nests a number of significant applied macroeconomic parameters (structural breaks/changes, threshold effects, regime-switching, etc.) and lets the data determine which combination of them is best appropriate.

Two-stag Ridge Regression (TS-RR): Coulombe (2020b) shows that time-varying parameter models can be approximated as ridge regressions, making macroeconomic policy implementation much easier than in the state-space paradigm. The critical issue in the time-varying parameter models –the "amount of time variation" can be then easily solved by tunning cross-validation. He proposes TS-RR to deal with evolving volatility and let the algorithm select which parameters vary and which do not. The first step involves interacting with each individual regressor with a scatter of moving indicators to accomplish a plain ridge regression (potentially in high-dimensional space). Then, by restricting to incremental location shifts caused by heavy ridge shrinkage, all of them are maintained in the model rather than only a small number of them (sparsity) (smooth time variation). Coulombe (2020b) demonstrates how well the suggested TS-RR performs when compared to various factor model and VAR model specifications.

4.2. Factor models

Dynamic Factor Model (DFM): DFM is well suited to a context in which parameter proliferation of a parsimonious model should be avoided, but at the same time, the salient features of high-dimensional economic data should be captured. In DFM framework, each economic variable can be divided by the sum of two orthogonal components: a handful of latent and unobserved factors that capture the joint dynamics and an idiosyncratic residual. Similarly to Giannone et al. (2008)'s approach, we specify that the high-frequency variables (in this case, monthly variables), X_t, have a factor structure following a VAR process:

$$X_{m|t} = \mu + \Lambda F_t + E_t, \quad E_t \sim i.i.d.N(0, \Sigma_E), \tag{6a}$$

$$F_t = AF_{t-1} + Bu_t, \quad u_t \sim WN(0, I_q), \tag{6b}$$

where $X_{m|t}$ is $N\times 1$ vector of monthly series, Λ is $N\times r$ of the factor loading, F_t is $r\times 1$ vector of factors, and E_t is an $N\times 1$ residual vector.[19] B is a $r\times q$ matrix of full rank q,[20] A is a $r\times r$ matrix, all roots of $\det(I_r - A_Z)$ lie outside the unit circle, and u_t is a q dimensional vector of the macroeconomic stochastic shocks to the common factors

[17] The identity function is used as the activation function. The robustness check was performed using two and three hidden layers and a different number of neurons. The results indicate that the single hidden feed-forward neural network is the best nowcasting model among various feed-forward neural nets.

[18] A perceptron is an algorithm used for supervised learning of binary classifiers.

[19] r is the number of common factors.

[20] q is the number of macroeconomic shocks.

Giannone et al. (2008) estimate DFM framework of (6a)–(6b) using a two-step procedure. In the first step, the factors and factor loadings are estimated using the principal component analysis (PCA) to a balanced panel of X_t by considering only the series for which all observations are observed. In the second step, factors are re-estimated by applying the Kalman filter to the entire information set. Unlike in the work of Giannone et al. (2008), which uses United States macroeconomic data, we raise particular concerns about issues related to the availability of relevant Chinese macroeconomic variables, most of which started being published only recently and encounter issues related to missing observations during the Chinese New Year.[21] These two issues could potentially eliminate the majority of variables in the first estimation procedure step. We adopt the first step used by Giannone et al. (2008), using an expectation–maximisation (EM) algorithm to estimate the factors and the factor loadings. Stock and Watson (2002) and McCracken and Ng (2016) demonstrate that the EM algorithm is an iterative algorithm for the efficient estimation of factors and factor loadings when data irregularities are present. Then, we update the estimated factors and factor loadings using the Kalman filter.[22] Finally, the nowcasts are obtained through a regression of GDP on temporally aggregated factors:

$$y_{t,1}^{k1} = \hat{\alpha} + \hat{\beta} F_{t|\Omega_v}^{k1} + e_t^{k1}, \quad t = k_1, 2k_1, \ldots .. \tag{7}$$

The data-processing function $f()$ in (1) involves PCA and Kalman filter, whereas the key hyperparameters in the dynamic factor model are the number of common factors (r), and the number of shocks (q). We assume that the macroeconomic dynamics are affected by the real shock and monetary shock ($q = 2$) and select the number of common factors using (Bai and Ng, 2002)'s information criteria.

Static Factor Models: A DFM can be expressed as a static factor model with r contemporary common factors. It reads as:

$$y_t = \beta' F_t + \gamma(L) y_t + \epsilon_{t+1}, \tag{8a}$$

$$X_t = \Lambda F_t + e_t, \tag{8b}$$

where in (8a) and (8b), $\gamma(L)$ is a lag operator for y_t and $e_t = (e_{1t}, \ldots, e_{Nt})'$ is the $N \times 1$ vector of idiosyncratic disturbances. Static factor models belong to the class of dense modelling where data-processing function $f()$ in (1) utilises dimensional reduction techniques, and hyperparameters in $pen()$ are the number of common factors and lag of y_t. From a practical perspective, the primary advantage of the static framework is that it is easily estimated using time domain methods and involves few choices of auxiliary parameters. Therefore, the construction of the nowcast can be easily achieved by a linear combination of contemporary F_t and y_t (with or without its lags).

PCA has been widely used in the literature to construct factors (Stock and Watson, 2002; Coulombe et al., 2021). It uses weighted cross-sectional averaging to remove the influence of the idiosyncratic disturbances in (8b). Such weights are the resulting factor estimators explaining as much data variance as possible. Partial least square (PLS) is a serious competitor for PCA. Unlike PCA factors that depend only on the covariance of X_t while ignoring the correlation between y_t and X_t, PLS factors are capable of capturing the variation between y_t and X_t while ignoring the covariance of X_t. We consider four variants of static factor models. The first and the second, denoted as FAAR-PCA and FAAR-PLS, include contemporaneous $\hat{F}_t$ estimated by PCA and PLS, respectively, and a lag of y_t.[23] The lag order of y_t is chosen by BIC. The third and the fourth, denote as DI-PCA and DI-PLS, include only contemporaneous $\hat{F}_t$.[24]

4.3. Mixed Data Sampling (MIDAS) and Autoregressive (AR)

MIDAS: MIDAS regressions were developed to cope with time series data sampled at various frequencies without needing to transform predictors into latent factors. We concentrate on stationary single-equation MIDAS regression models where the dependent variable is observed every three months and the explanatory variables are observed every month. MIDAS reads as follows:

$$y_t = \alpha + \sum_{i=1}^{p} \beta_i L^i y_t + \gamma \sum_{k=1}^{m} \Phi(k;\theta) L_{HF}^k X_t + \varepsilon_t \tag{9}$$

where L is the lag operator, m is the frequency alignment indicator so that the high-frequency variable is observed m times in the low-frequency variables,[25] $\Phi(k;\theta)$ is a polynomial that determines the weights for temporal aggregation of X_t.[26]

The MIDAS leads to sparse models ex-post as it uses functional forms to constrain the coefficients. The key hyperparameters are which out of several functional constraints is better suited to the MIDAS regression model and the number of lags for y_t and X_t. Since the selection of lags can differ with different functional constraints, Ghysels et al. (2016) suggest using an information criterion to select the best model in terms of the functional constraint and the lag order simultaneously in the out-of-sample precision measure. We opt for Bayesian information criteria (BIC) in the hyperparameter selection.

Following Andreou et al. (2013), we consider the direct nowcast of GDP growth rate with factors estimated by PCA, and we propose two variants of factor-based MIDAS. First, the AR term $\sum_{i=1}^{p} \beta_i L^i y_t$ is ignored and we obtain the MIDAS specification. Second, the AR term $\sum_{i=1}^{p} \beta_i L^i y_t$ in (9) is included and we denote the model as MIDAS-AR.

AR: The AR model is a commonly used univariate model in economic forecasting literature (Stock and Watson, 2002; Gupta and Kabundi, 2011; Lin and Wang, 2013; Kotchoni et al., 2019). The key hyperparameter in the AR model is the lag order p, which can be determined using information criteria. The AR nowcasting model is essentially a one-step-ahead forecast, as the National Bureau of Statistics of China usually releases GDP figures with a two- to three-week delay. We choose the lag order based on the minimum BIC of the training dataset.

4.4. Computer codes

We implement the computation of nowcasting algorithm in the *R* platform. All nowcasting methods are estimated in *R* using standard and well-established packages. For the linear ML methods, we use *glmnet* package by Friedman et al. (2010). For single layer feed-forward neural network, we use *nnet* package by Ripley et al. (2016). The K-nearest neighbour for regression is implemented using *caret* package of Kuhn (2015). Decision tree is estimated using *rpart* package of Breiman et al. (2017). Macroeconomic random forest and two-stage ridge regression are implemented using *MacroRF* package and *TVPRidge* package, respectively. For a dynamic factor model and factor-augmented AR with factors estimated by PLS, we use *nowcasting* package by de Valk et al. (2019). MIDAS regressions are implements using *midasr* package by Ghysels et al. (2016).

[21] In the replication file of the research by Giannone et al. (2008), the data only contain a few missing observations at the end of the sample period.

[22] The Kalman filter exploits the dynamics of the common factors and the cross-sectional heteroscedasticity of the idiosyncratic components, thereby providing efficiency improvements over simple PCA.

[23] FAAR refers to factor augmented autoregression.

[24] DI refers to diffusion index by Stock and Watson (2002) where they propose a variant of factor models that only contains contemporaneous factors and name it diffusion index.

[25] For example, if low-frequency GDP is quarterly and high-frequency CPI is monthly, $m = 3$.

[26] The weighting function $\Phi(k;\theta)$ can have any number of functional constraints, for example, Exponential Almon lag polynomial, Beta polynomial, Gompertz polynomial, Log-Cauchy polynomial and Nakagami polynomial.

Table 1
List of nowcasting models.

Nowcasting models	Acronym	Input (Z_t)	R packages
Machine learning algorithms			
Ridge Regression	**RR**	X_Q	glmnet of Friedman et al. (2010)
Least Absolute Shrinkage and Selection Operator	**LASSO**	X_Q	glmnet of Friedman et al. (2010)
Elastic Net	**E-NET**	X_Q	glmnet of Friedman et al. (2010)
Single Layer Feed-forward Neutral Network	**SL-FFNN**	X_Q	nnet of Ripley et al. (2016)
K-nearest Neighbour	**KNN**	X_Q	caret of Kuhn (2015)
Decision Tree	**DT**	X_Q	rpart (Breiman et al., 2017)
Macroeconomic Random Forest	**MRF**	X_Q	MacroRF of Coulombe (2020a)
Two-stage Ridge Regression	**TSRR**	X_Q	TVPRidge of Coulombe (2020b)
Factor models			
Dynamic Factor Model	**DFM**	F by X_M	Nowcasting of de Valk et al. (2019)
Diffusion Index with factors estimating by PCA	**DI-PCA**	F by X_Q	
Factor-augmented AR with factors estimating by PCA	**FAAR-PCA**	F by X_Q	
Diffusion Index with factors estimating by PLS	**DI-PLS**	F by X_Q	pls of Mevik and Wehrens (2007)
Factor-augmented AR with factors estimating by PLS	**FAAR-PLS**	F by X_Q	pls of Mevik and Wehrens (2007)
MIDAS and AR			
Mixed Data Sampling Regression	**MIDAS**	X_M	midasr of Ghysels et al. (2016)
Mixed Data Sampling Regression with AR term	MIDAS-AR	X_M	midasr of Ghysels et al. (2016)
Autoregression (with p chosen by BIC)	AR(p)	y	

Notes: Superscript Q and M indicates the frequency of input data so X_Q is quarterly whereas X_M is monthly. The set of predictors X_Q is obtained by applying the temporal aggregation to X_M. For stock variables such as interest rates and price indices, we treat X_Q as the monthly average of X_M in a quarter. For flow variables such as investment and industrial production, we treat X_Q as the monthly aggregation of X_M in a quarter. Regarding the acronym, PCA refers to the principal component analysis, while PLS refers to the partial least square estimator. F refers to estimated factors.

Table 2
Hyperparameter, selection methods and suggested range in previous literature.

Model	Hyperparameters	Selection methods	Suggested range
Machine learning algorithms			
RR	λ	CV	0.01–0.99
LASSO	λ	CV	0.01–0.99
EN	α; λ	Minimum within-sample MSE; CV	0–1; 0.01–0.99
SL-FFNN	No. of neurons	CV	1–10
KNN	No. of neighbours	CV	2–10
DT	Max depth; Min split; No. features	CV	1–30; 5–60; 5–60
MRF	λ; Kernel weight $w(t;\zeta)$	CV	0.01–0.99; Symmetric 5-step Olympic podium
TS-RR	λ; amount of time variation	CV	0.01–0.99
Factor models			
DFM	No. factor; No. shock	Bai and Ng (2002)'s IC; Set to be 2	1–6; 1–4
DI-PCA	No. factors	Bai and Ng (2002)'s IC	1–6
FAAR-PCA	No. factors ; Lag of y_t	Bai and Ng (2002)'s IC; BIC	1–6; 1–12
DI-PLS	No. component	CV method with least MSE	1–4
FAAR-PLS	No. component; Lag of y_t	CV method with least MSE; BIC	1–4; 1–12
MIDAS & AR			
MIDAS	Polynomial weighting function; Lag of X	Out-of-sample precision; BIC	Exponential Almon and Beta; 1–4
MIDAS-AR	Polynomial weighting function; Lag of X_t and y_t	Out-of-sample precision; BIC	Exponential Almon and Beta; 1–4
AR(p)	Lag of y_t	BIC	1–12

The model acronyms used are explained in Table 1. RMSE refers to the root mean squared error. CV is cross-validation, while BIC refers to Bayesian information criteria.

Table 1 lists all the models implemented in the nowcasting exercise, together with their respective acronym, input matrices Z_t and R packages used for computation process.

5. Model estimation and nowcast evaluation

5.1. Model estimation and hyperparameter optimisation

We partition the target data into training and validation sets in order to estimate nowcasting models and select the best hyperparameters. In this setup, the model is first estimated using the training set's data. The parameters with the lowest nowcast error in the validation set are then chosen as the optimal hyperparameters. We limit the hyperparameter search to a range around values established in the prior literature to prevent overfitting the model to the training data set, see for example Coulombe et al. (2022) for hyperparameter tuning of ML algorithms, Stock and Watson (2016) for factor models, and Ghysels et al. (2016) for MIDAS models. This helps the models generalise to unseen data. Table 2 presents the hyperparameters we are optimising for each model, the range we are exploring, and the criteria we used for hyperparameters selection.

5.2. Nowcast evaluation

We evaluate the nowcasting performance using three measures: (i) rRMSE, (ii) the test of conditional predictive of Giacomini and White (2006), and (iii) MCS of Hansen et al. (2011). The first measure, rRMSE, is commonly used in forecast evaluation metrics to compare model's performance against the chosen benchmark, see for example Ahmed et al. (2010) and Kotchoni et al. (2019) and other macroeconomic forecasting literature.

The second measure tests null hypotheses of RMSE equality of competing models against the benchmark model. According to Giacomini and White (2006), the conditional test has the ability to capture the effect of estimation uncertainty on relative forecast performance when models may be misspecified. The test employs a null of equal expected loss evaluated at the current parameter estimates as in (10), so it is dependent on the information set available at time t.

$$H_0 : E[L(f_{1t+1|t}(\hat{\beta}_{1t}), y_{t+1})] = E[L(f_{2t+1|t}(\hat{\beta}_{2t}, y_{t+1}))]. \tag{10}$$

The third measure, MCS, produces a set of the best model(s) from all candidate models with a given level of significance, wherein the

definition of 'best model(s)' can be user specified. It is constructed from sample information about the relative performances of all competing models; therefore, it is a random data-dependent set of models that contains the best nowcasting model(s), just like a standard confidence interval for the population parameter. Let $\mathcal{M}_0 = M_1 \ldots M_m$ denote initial set competing models, the expected loss (in our case RMSE) from model m and n at time t are compared using:

$$d_{mn,t+1} = L_{m,t} - L_{n,t} \quad m, n \in \mathcal{M} \tag{11}$$

At the start of the procedure, $\mathcal{M}_0$ includes all the candidate models. The test of superiority is performed by testing the null that $H_{0,\mathcal{M}} : E\left[d_{mn,t+1}\right] = 0$ for m and n at a critical level α. If the null hypothesis is accepted at the critical level, then the set $\mathcal{M}^* = \mathcal{M}$. The weakest model in the set gets eliminated if the null hypothesis is rejected. The process is repeated until the null is no longer rejected, in which the MCS procedure produces models with the same level of nowcasting performance.

The elimination of the nowcast models in each iteration can be done according to two test statistics:

$$t_{mn} = \frac{\bar{d}_{mn}}{\sigma_{\bar{d}_{mn}}} \tag{12a}$$

$$t_m = \frac{\bar{d}_m}{\sigma_{\bar{d}_m}} \tag{12b}$$

where $\bar{d}_{mn} = T^{-1} \sum_{t=1}^{T} d_{mn,t}$ and $\bar{d}_m = (m-1)^{-1} \sum_{m \in \mathcal{M}} d_{mn,t}$, $\sigma_{\bar{d}_{mn}}$ and $\sigma_{\bar{d}_m}$ are bootstrapped estimates of standard deviation. In order to allow for comparisons of the model sets, the proposed test statistics (12a) and (12b) map naturally into:

$$T_{R,\mathcal{M}} = \max_{m,n \in \mathcal{M}} |t_{mn}| \tag{13a}$$

$$T_{\max,\mathcal{M}} = \max_{m \in \mathcal{M}} t_m \tag{13b}$$

The first statistic measures the relative sample loss of models m and n, while the second measures the relative sample loss of model m against the average across the M models.

5.3. Pseudo out-of-sample experimental design

We evaluate the performance of the nowcasting algorithms using an out-of-sample simulation. We adopt a fixed estimation window method[27] in which we use the first R observations to estimate parameters of each model where R refers the number of within-sample periods, and $T - R$ refers to the number of out-of-sample periods. Then, we generate the first nowcast at the time period R. To compute the next nowcast at the period $R + 1$, we drop the first observation from the previous estimation sample and add an extra observation to the end of the previous estimation sample. Hyperparameters of each model are re-estimated, and the nowcast is computed based on estimated parameters. This process–updating the estimation sample, re-estimating model parameters and computing a nowcast for the next time period—is continued until $T - R$ time period out-of-sample nowcasts have been computed.

We consider the total sample size in a 50:50 per cent split between in-sample and out-of-sample periods to ensure a sufficient number of the in-sample period for model selection and evaluation.[28] Therefore, the within-sample period covers the period from 1995Q1 to 2008Q1. The first out-of-sample nowcast of China's real GDP growth rate is made for 2008Q1. To compute this nowcast, we deseasonalise all series, screen for potential outliers, standardise all series and estimate models using data only available from 1995Q1 to 2008Q1. To compute the next nowcast, we re-determine the value of hyperparameters and re-estimate models using data from 1995Q2 to 2008Q2. This process repeats until the final simulated out-of-sample nowcast is made for 2020Q4.

6. Empirical results

6.1. Out-of-sample experiment

In this section, we describe the main results of our analysis. Each MCS is constructed using the quadratic loss function and 5000 bootstrap replications. Following Kotchoni et al. (2019), we compute MCS across all the competing models and set the level of significance $\alpha = 25\%$. Thus, models with p-values greater than 25% are covered by MCS, while models with p-values less than 25% are excluded from MCS. We use $T_{\max,\mathcal{M}}$ criteria to reach our set of superior nowcasting models. We also calculate the rRMSE of each model relative to the benchmark DFM model as it is a standard nowcasting model for Western economies. In addition, we use a test of conditional predictive procedure to compare the nowcast accuracy of each model against the benchmark.

We now examine the performance of the various models at nowcasting China's real GDP growth. Table 3 presents rRMSE, results of the test of conditional predictive, and the MCS p-value of each nowcasting algorithm for the simulated out-of-sample period from 2008Q1 to 2020Q4. We first inspect the nowcasting performance of the total dataset. Overall, RR, LASSO, E-Net, DT, TS-RR, FAAR-PCA, FAAR-PLS, MIDAS-AR and AR models are covered by MCS at $\widehat{M}^*_{75\%}$, constituting the set of best models for the total data. In terms of rRMSE, the best nowcast is generated from RR, which employs the shrinkage technique to discipline the behaviour of the parameters and cross-validates hyperparameters with the K-fold method. The improvement in nowcast accuracy is extraordinarily substantial, which is about 77% better than the benchmark DFM model. LASSO and E-Net, which incorporate the L2-norm of the penalty and a mixture of L1- and L2-norm penalties to constrain the least relevant predictors, are included in MCS at 25% significance level. By looking at rRMSE, LASSO and E-Net are able to statistically reduce nowcast error by 37% and 49%, respectively. Moreover, these gains are also statistically significant at the 1% level of significance according to Giacomini and White (2006)'s tests of conditional predictive ability.

DT, which splits data into paths according to the largest variance reduction, and TS-RR, which uses cross-validation to determine the optimal amount of time variation parameters, also perform well in terms of rRMSE. For example, TS-RR, which is the second best performing model, generates a nowcast error variance that is only 33% that of the benchmark DFM model, and the test of conditional predictiveness is statistically significant at 1% level of significance. Nowcasts generated by FAAR-PCA and FAAR-PLS are also superior to those produced by DFM, outperforming the benchmark at the 5% level of Giacomini and White (2006) test.

An intriguing empirical finding is observed to the rejection of DFM as it fails to enter into $\widehat{M}^*_{75\%}$ of three datasets and is the second worst performing model among all competing models. MCS p-values for DFM are relatively small: 0.22 for the total dataset, 0.25 for the soft dataset

[27] One significant benefit of using the fixed window is that it deals with model instability, which is an issue facing China. With the fixed window scheme, "old" data get excluded in the estimation and so do not cause biases in the parameter estimates. The fixed window estimates tend to be more stable than expanding window estimates, making more efficient use of all data in situations in which the parameters do not change much over time (Elliott and Timmermann, 2016).

[28] Hansen and Timmermann (2012) conclude that short assessment samples are most likely to result in erroneous rejections, whereas long out-of-sample intervals increase the power of prediction evaluation tests.; therefore, we opt for the long out-of-sample period. We perform the robustness analysis using a 70:30 split and find similar nowcasting results.

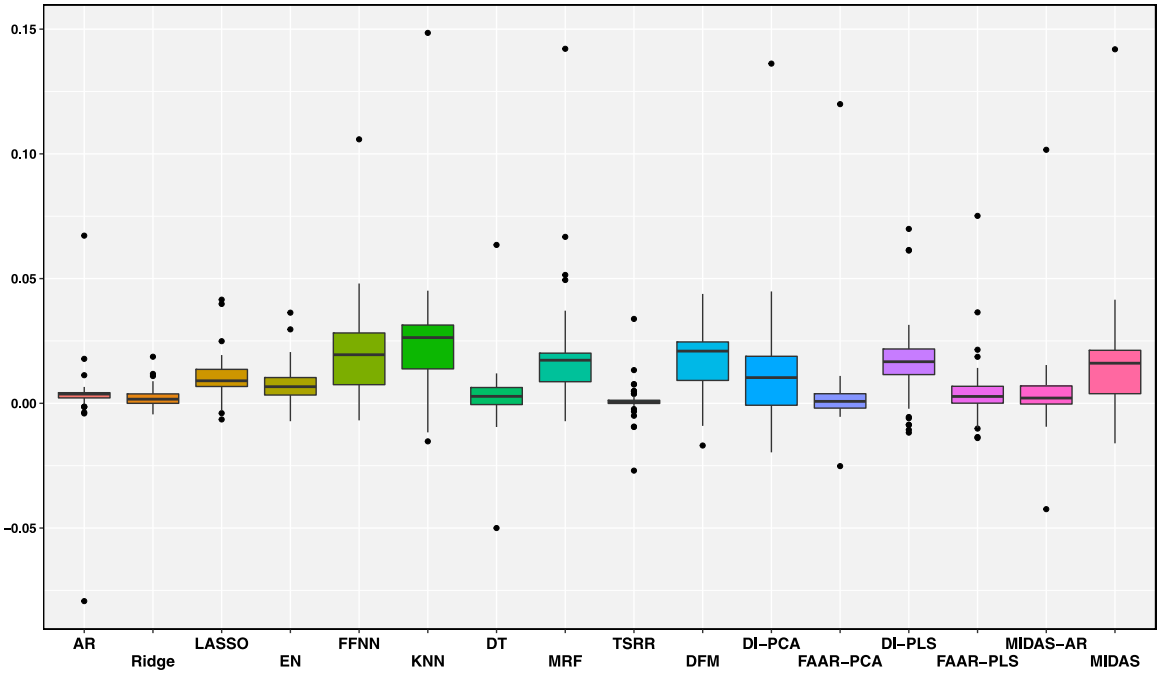

Fig. 1. Boxplot of nowcast errors of each model. Note: the boxplot represents the nowcast error for each model, supplying information regarding the minimum, first quartile (Q1), median, third quartile (Q3), and maximum. The black dots at the end of the boxplot represent outliers, which are calculated as less than Q1 minus 1.5 times the interquartile range.

and 0.16 for the hard dataset. This is surprising because our results are contrary to the actions of most central banks and International Monetary Fund policies, which use dynamic factor-based nowcasting models to make nowcasts and inform their policy decisions. We find conclusive evidence that DFM is inferior to some simple ML algorithms for nowcasting China's GDP, such as RR and LASSO. By contrast, static factor models augmented with an AR term that use only contemporaneous smoothing to estimate factors perform relatively well in terms of the rRMSE and the test of conditional predictiveness. They are able to reduce the averaged nowcast error by 25.5% compared with the benchmark DFM model.

Fig. 1 represents a box plot of each measure drawn by using the nowcast error obtained from each model of the total dataset, which supplies information on the variability of the nowcast error for the out-of-sample period. We observe that there is a greater variation and higher error values for SL-FFNN, DFM, DI-PLS, FAAR-PLS, and MIDAS than for other models. RR, LASSO, E-Net, DT and TS-RR perform relatively well, consistent with our rRMSE and MCS results. However, TS-RR experiences a hard time detecting extreme outlier observations that violate the overall performance. DFM and SL-FFNN produce the worst median nowcast error and the largest interquartile range, indicating great dispersion of the nowcast performance. Figure 9 is provided in Appendix B for additional information regarding the nowcast error variability for the soft dataset and hard dataset. We also consider the differences in nowcasting performance among the total dataset, soft dataset and hard dataset. The comparison of the rRMSE of each nowcasting algorithm is reported in Fig. 2. Although the number of nowcasting algorithms that enter into $\widehat{M}^{*}_{75\%}$ based on the total dataset is similar to those numbers based on the hard and soft dataset,[29] we find that the use of the total dataset leads to reductions in rRMSE for RR, LASSO and E-Net. This finding is further supported by Fig. 3, which shows that, for E-Net, LASSO and RR, the averaged R^2 based on the total dataset is large than the corresponding R^2 based on the soft and hard dataset. The improvements are substantial, with averaged R^2 boosted by more than 10%. Therefore, the total dataset possesses more useful information to differentiate nowcasts from the hard and soft datasets, and policymakers should rather incorporate as much information as possible to make a nowcast.

To further investigate the performance of each nowcasting algorithm in a specific time period, we plot the annualised real GDP growth rate and its nowcasts obtained from each algorithm over the out-of-sample period in Fig. 4. We divide the out-of-sample period into three sub-periods: (i) the Global Financial Crisis and its recovery period that stretches from 2008Q3 to 209Q2, (ii) the quiet period during the period from 2009Q3 to 2019Q4, and (iii) the Covid-19 period in the year 2020. Panel (a) illustrates that RR, LASSO, and E-Net can successfully predict the sharp downturn and subsequent recovery in activity from 2008Q1 to 2009Q2, most fluctuations during the quiet period, and the effects of Covid-19 in the year 2020. Revealed in panel (b), SL-FFNN and KNN largely over-nowcast the actual growth rate. DT presented in panel (c) succeed in nowcasting the overall trend well. MRF and TS-RR in panel (d), which allow time-varying parameter components, reveal a similar historical movement to the DT, but the level of over-prediction is more substantial for MRF.

In panels (e) and (f), we observe that DI-PCA and FAAR-PCA perform less well to recognise the sharp downturn in economic activity due to the Covid-19 crisis. In contrast, DI-PLS and FAAR-PLS are successful in recognising the impacts of the Global Financial Crisis and Covid-19 crisis. During the quiet period, they all slightly over-nowcast the GDP growth, resulting in a relatively weak overall performance compared to RR and LASSO. Two variants of MIDAS models in panel (g) respond mildly to the Global Financial Crisis; however, they are having a difficult time reducing nowcast errors during the quiet period and the Covid-19 period. In panel (h), DFM produces nowcasts that are greater than the actual growth rate during the quiet period, but it performs exceptionally during the Covid-19 period.

[29] MCS of the total dataset include 9 nowcasting models, whereas MCS of the soft data and the hard dataset contain 7 and 11 models at the 25% level of significance, respectively.

Table 3
rRMSE and MCS P-Value results.

	Total dataset		Soft dataset		Hard dataset	
	rRMSE	MCS P-value	rRMSE	MCS P-value	rRMSE	MCS P-value
ML algorithms						
RR	0.23***	**1.00**	0.69**	**0.87**	0.36***	**0.81**
LASSO	0.63***	**0.79**	1.04	0.14	0.53***	**0.81**
E-NET	0.51***	**0.79**	0.92	0.23	0.46***	**0.81**
SL-FFNN	1.25	0.00	1.16	0.01	1.30	0.08
KNN	1.56	0.03	1.46	0.04	1.19	0.12
DT	0.60***	**0.79**	0.50***	**1.00**	0.53***	**0.81**
MRF	1.39	0.15	1.30	0.09	1.18	0.16
TS-RR	0.33***	**0.79**	0.78*	**0.71**	0.45***	**0.81**
Factor models						
DFM	1.00	0.22	1.00	0.25	1.00	0.16
DI-PCA	1.17	0.22	1.31	0.23	0.95	0.23
FAAR-PCA	0.83*	**0.79**	0.78*	**0.87**	0.73**	**0.81**
DI-PLS	1.10	0.15	1.03	0.23	0.50***	**0.81**
FAAR-PLS	0.66*	**0.67**	0.62**	**0.87**	0.49***	**0.81**
MIDAS						
MIDAS-AR	0.79	**0.77**	0.93	**0.59**	0.68	**0.78**
MIDAS	1.24	0.15	1.13	0.14	1.06	0.11
AR	0.72	**0.79**	0.72	**0.87**	0.72	**0.81**

Noted: The reported rRMSE of each model is relative to the benchmark DFM model for the full out-of-sample period (2008Q1 to 2020Q4). The grey shading corresponds to the benchmark. The best models (with minimum rRMSE) are underlined. The p-values that are covered in $\widehat{M}^{*}_{75\%}$ are marked in bold. According to the Federal Reserve Bank of St. Louis, the soft dataset contains survey data, financial market variables and various price indexes, whereas the hard dataset contains variables used for GDP calculations, such as the industrial production index and international trade data.

*Indicate 10% significance level for the Giacomini and White (2006) conditional predictive test.

**Indicate 5% significance level for the Giacomini and White (2006) conditional predictive test.

***Indicate 1% significance level for the Giacomini and White (2006) conditional predictive test.

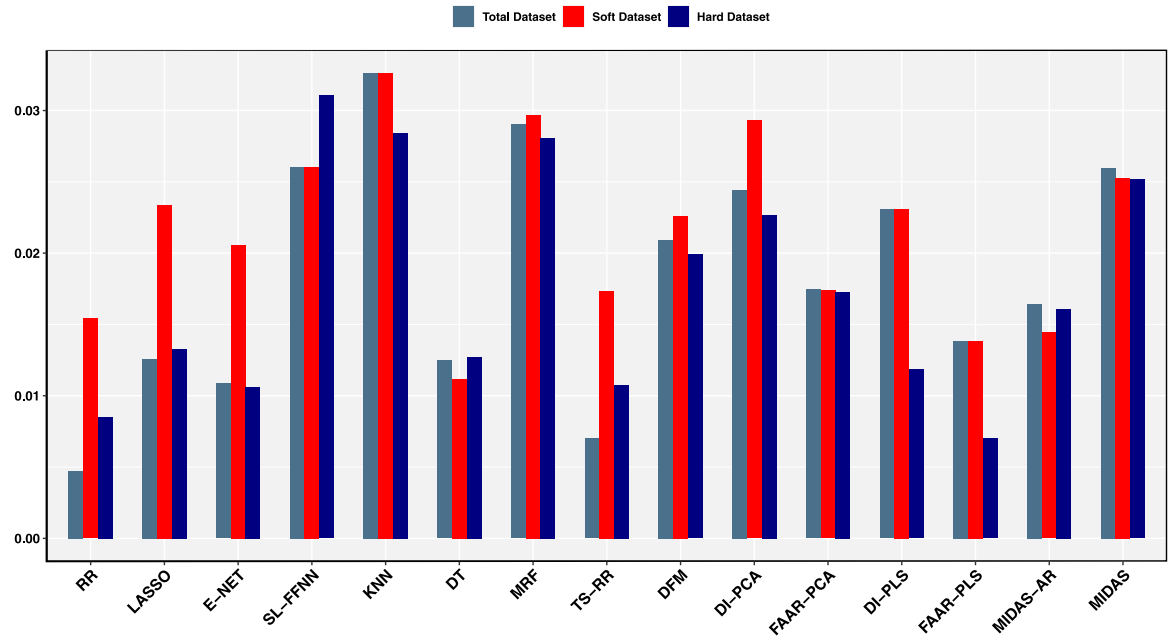

Fig. 2. RMSE of each nowcasting model for total, soft and hard dataset. Note: the RMSE computed from each nowcasting model using total, soft, and hard datasets over the out-of-sample period 2008Q1 to 2020Q4.

6.2. Robustness checks

Our empirical analyses have thus far been based on annualised real GDP growth rate. To further reaffirm the effectiveness of ML learning algorithms in nowcasting China's annualised real GDP growth rate, we provide two types of robustness checks by comparing the performance of the various models at forecasting annualised real GDP growth rate one-quarter-ahead and nowcasting the China's quarter-on-quarter nominal GDP growth rate. Table 4, Table 5 and Figure 7 in Appendix A demonstrate that the comparative patterns in performance across different models for forecasting one-quarter-ahead and nowcasting China's nominal GDP growth rate are similar to those of nowcasting annualised real GDP growth rate.

6.3. Opening the black box of machine learning algorithms

We now compare the predictors selected by some ML algorithms and investigate the relative importance of individual predictors for the performance of the selected nowcasting model using the variable importance (VI) measures. The aim of VI is to identify covariates that

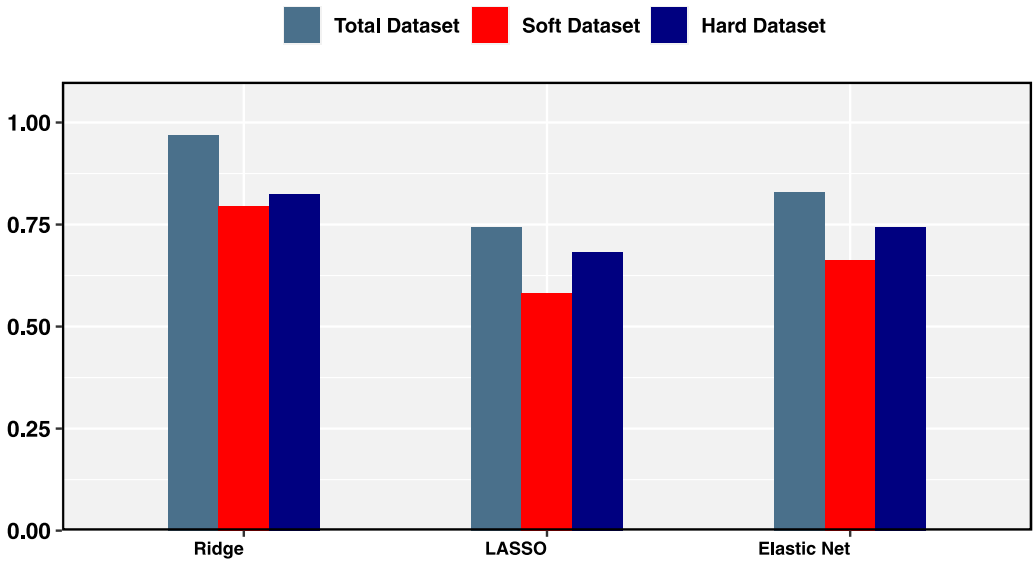

Fig. 3. Average R^2 for hard, soft and total dataset. Note: the averaged R^2 on the nowcasting model of RR, LASSO and E-Net computed by total, soft and hard datasets over the out-of-sample period 2008Q1 to 2020Q4.

have an important influence on the cross-section of nowcasting results while simultaneously controlling for the many other predictors in the nowcasting system. We select four models for two reasons: (i) they perform exceedingly in rRMSE and are covered by $\widehat{M}^{*}_{75\%}$, and (ii) they have different characteristics. LASSO is a sparsity-inducing model, while RR, E-Net and DT are approximately sparse. LASSO and RR are modifications of linear regression, whereas DT can be highly nonlinear specification. For RR, LASSO, and E-Net, we compute VI measures as the average coefficient size and standardised regression coefficients so that variable importance can be measured and compared directly. For DT, we use the out-of-bag (OOB) samples, which compute VI as the decrease in accuracy due to the permutation being averaged over all trees.

Regarding which individual predictor has the most predictive power, we report the resulting importance of the 10 most important predictors in Fig. 6.[30] Predominantly, the most influential predictor for RR, LASSO and E-Net is nominal consumption, which accounts for more than 50% of predictive power among all 89 monthly predictors. Government revenue, 3 months term deposit rate, RMB required reserve ratio, and electricity production also constitute the 10 most important predictors for RR, LASSO and E-net. DT is more democratic, drawing predictive information from a broader set of predictors, including 1-year deposit rate, nominal exchange rate, RPI, median and long-term lending rates, and macroeconomic climate index. Fig. 5 demonstrate the importance of each variable group for RR, LASSO, E-Net and DT methods, with the values in the plots are rescaled to the sum of one. Universally, VIs suggest the predominance of certain categories. More precisely, these four nowcasting models are generally in close agreement regarding the most influential groups of predictors. The set of the most relevant groups for RR, LASSO and E-Net include predictors from the government sector, industrial sector, international trade sector and interest rate sector. The least influential group for RR, LASSO, and E-Net is predictors from the exchange rate sector. For DT, the most influential group comes from the price indexes group; together, they contribute to approximately 35% of predictive power, whereas the least influential group of predictors is the energy sector. Overall, VI measures provide a highly consistent summary of which variables are most influential for nowcast accuracy.

[30] To better compare the measure across different models, we scale VIs of the 10 predictors and sum them to one.

7. Conclusion

Nowcasting models have become an increasingly important tool for mitigating uncertainties regarding the current state of the economy and have been widely used by policymakers at many central banks to nowcast the current quarter's GDP growth rate. The recent "big data" boom has deeply motivated scholars to use ML algorithms in nowcasting macroeconomics, which has led to extensive nowcasting gains. However, the literature with regard to nowcasting China's GDP is relatively limited compared to that of Western economies, and the performance of a wide range of nowcasting models is not well examined. In this paper, we evaluate the performance of various ML algorithms in obtaining accurate nowcasts of the real GDP growth rate for China. ML algorithms consist of regularisation methods such as LASSO, RR, and E-Net; non-parametric methods such as KNN, SL-FFNN and DT; time-varying parameter methods such as MRF and TS-RR; and ensemble ML algorithms such as RF. We then compare the nowcasts obtained from these ML algorithms with the nowcasts generated by the AR model, dynamic and static factor models, and three specifications of MIDAS models.

We find three notable results. First, based on the total dataset, RR, LASSO, E-Net, DT, and TS-RR generate nowcasts that are significantly superior to the nowcast generated by the benchmark DFM model. Second, compared to the hard dataset, the total dataset contains more useful information and is more informative for nowcasting the Chinese GDP. Third, we reject the superiority of DFM in nowcasting China's GDP growth rate, which is contrary to the literature for Western economies. Overall, our paper joins the growing body of literature that addresses the relative success of ML algorithms in the field of macroeconomic nowcasting and forecasting. To the best of our knowledge, our study is one of the first few papers to consider using ML algorithms to nowcast the current state of economic growth for China. Our main takeaway result is that regularisation techniques yield the set of 'best' models according to the MCS. This is an interesting result, as it provides a direction for future research investigating whether other types of penalties can generate better nowcasts than RR and LASSO.

Arguably, there may exist other variants of ML algorithms that are not included in this paper but are potentially helpful for macroeconomic analysis. This is unavoidable, as there is a wide range of ML algorithms that could be useful to improve nowcasting accuracy. Another research opportunity is in non-parametric factor models that require fewer restrictive assumptions than parametric factor models

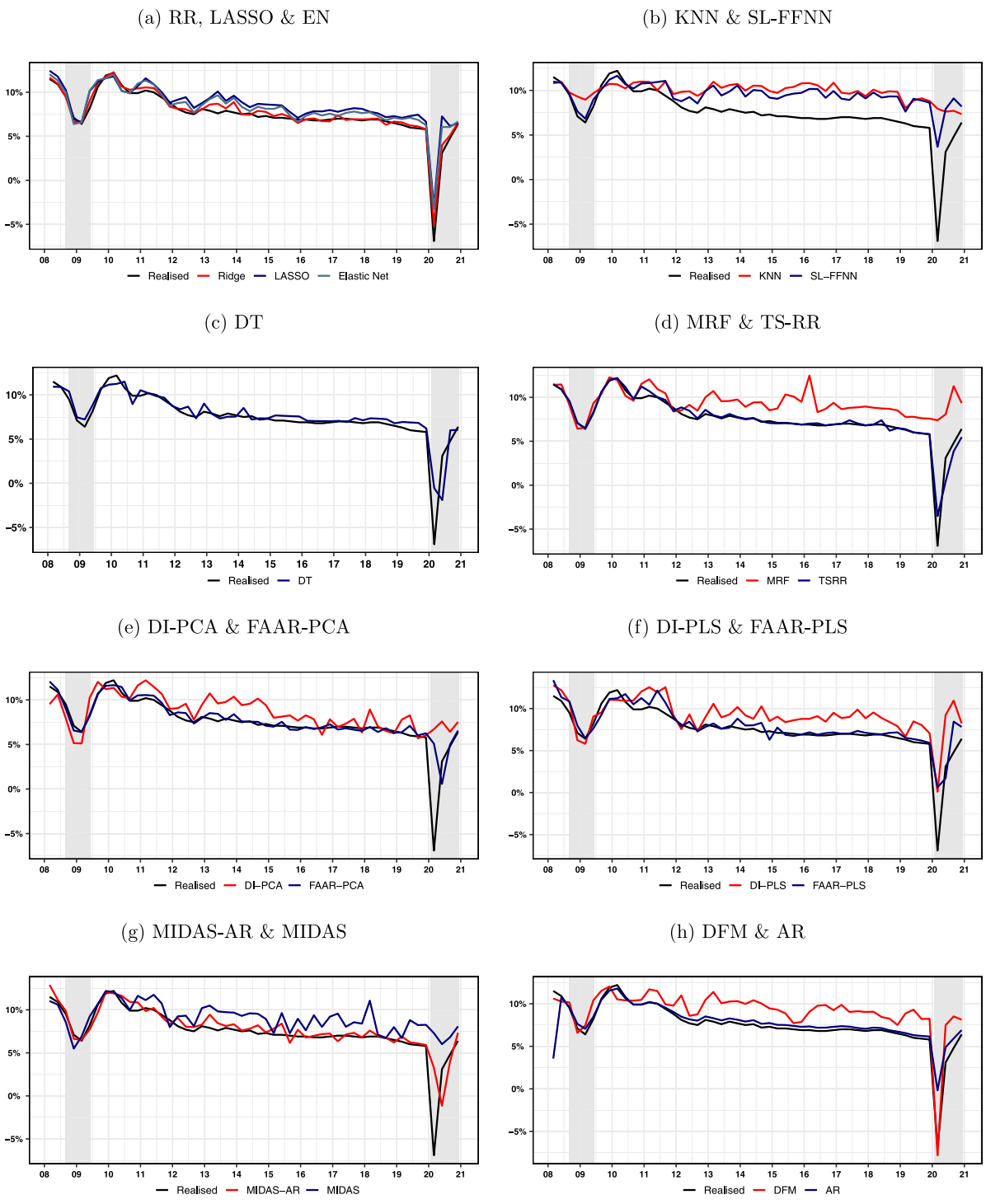

Fig. 4. Nowcasts vs. Actual annualised real GDP growth rate. Note: Annualised real GDP growth rate in percent change. Light grey shading corresponds to recessions due to the 2008 GFC and 2020 Covid crisis.

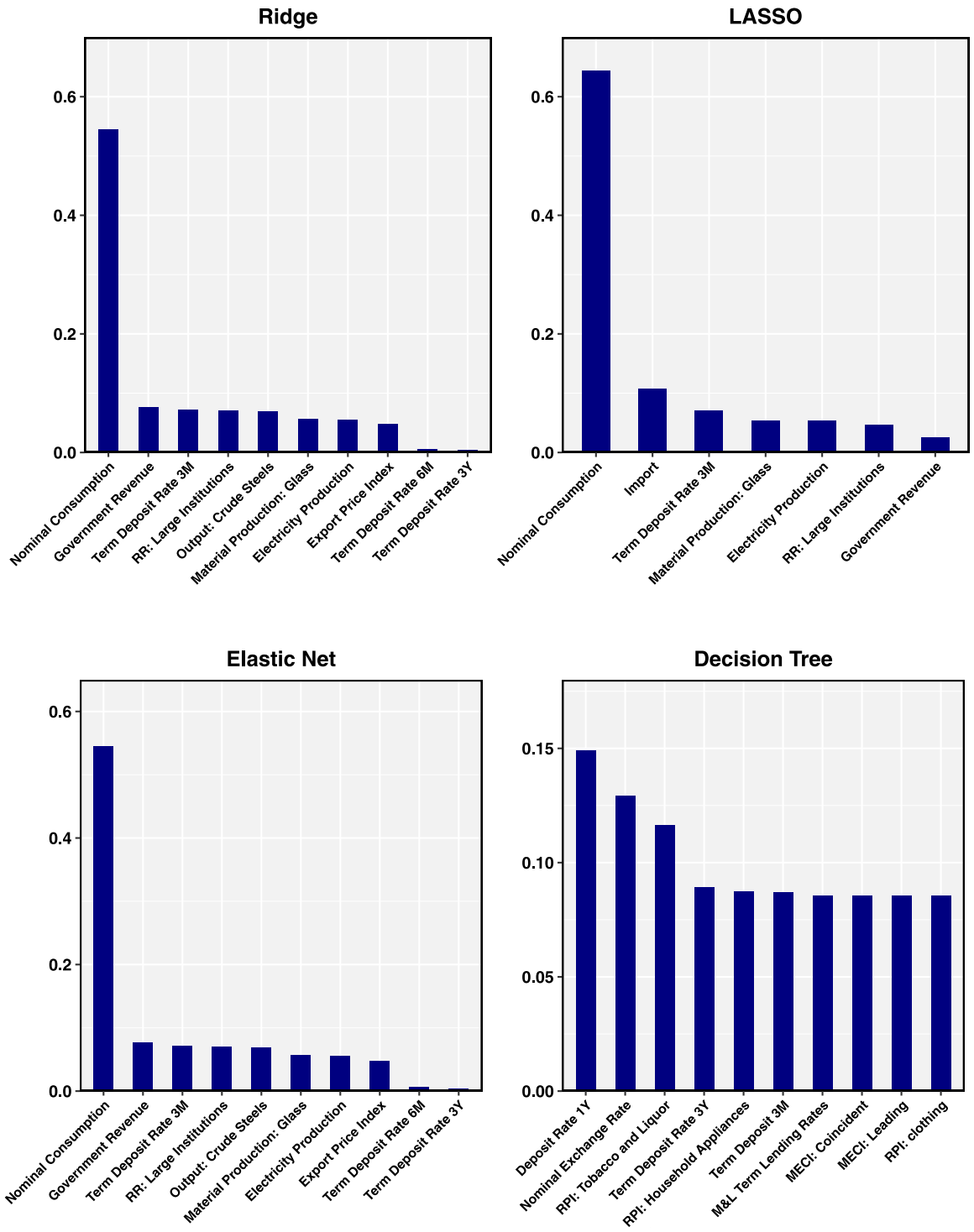

Fig. 5. Variable importance for RR, LASSO, E-Net and DT. Note: 10 most important series according to the variable importance (VI) criteria. The values in the plots are rescaled to sum one. For ridge regression, LASSO and elastic net, the relative importance measure is computed as the average coefficient size (multiplied by the respective standard deviations of the regressors). To measure the importance of each variable for the decision tree, we use out-of-bag samples. RPI is the retail price index. MECI is the macroeconomic climate index. M&L term lending rate refers to medium and long-term bank lending rates. RR: Large Institutions is RMB deposit reserve ratio: Large depository financial institutions.

and do not impose a prior structure on the underlying data-generating process. Therefore, non-parametric factor models that allow for data-driven flexibility are robust alternatives to parametric factor models and could be fruitful in improving nowcasting performance when the structure of the data-generating process is unknown (Stock and Watson, 2016).

Policy-wise, we provide an empirical model for China's macroeconomy that can be replicated and evaluated independently. Based on our results, the ML algorithm should serve as a benchmark for the comparison of policy projections and macroeconomic nowcast accuracy. Overall, we believe that our study not only guides practitioners in selecting appropriate nowcasting models for China's economy but also

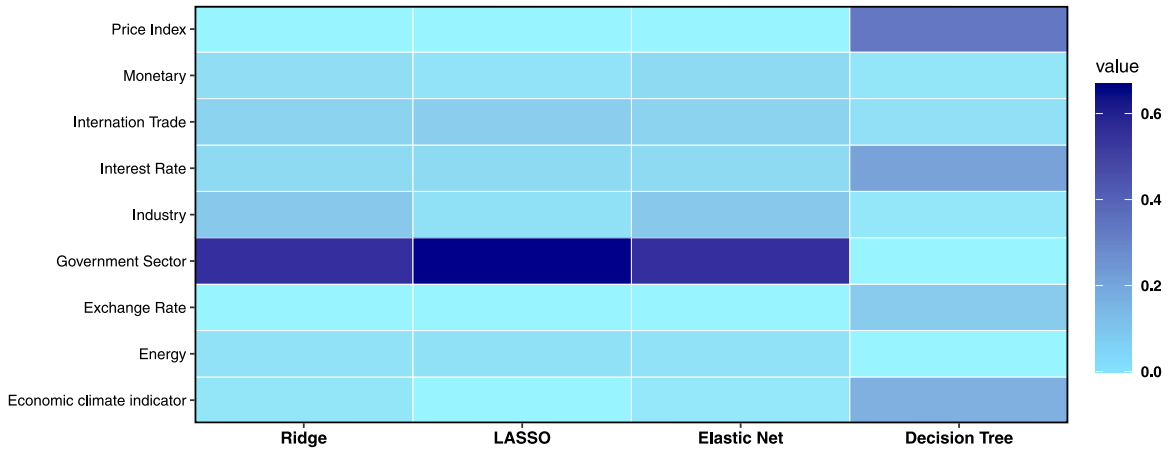

Fig. 6. Variable importance by model. Note: the importance of each variable group for the decision tree, elastic net, LASSO and ridge regression. The values in the plots are rescaled to sum one. Methods used to calculate variable importance can be seen in Fig. 5.

provides the research community with more feasible directions for ML algorithms in macroeconomic nowcasting and forecasting.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This research was supported by the characteristic & preponderant discipline of key construction universities in Zhejiang Province (Zhejiang Gongshang University - Statistics), China, the Tailong Finance School of Zhejiang Gongshang University. Dr He Ni acknowledges financial support from the Natural Science Foundation of Zhejiang Province, China, China (Grant Nos. LY20G010002). Dr Hao Xu acknowledges financial support from the Natural Science Foundation of Zhejiang Province, China (Grant Nos. LQ22G030021). Meanwhile, the authors would like to thank the data collection services provided by Hangzhou Woma Info. Tech. Co. Ltd.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.econmod.2023.106204.

Robustness Test, Data Description and Additional Tables and Figures.

References

Aarons, G., Caratelli, D., Cocci, M., Giannone, D., Sbordone, A., Tambalotti, A., 2016. Just released: Introducing the FRBNY nowcast. Lib. Street Econ..

Ahmed, N.K., Atiya, A.F., Gayar, N.E., El-Shishiny, H., 2010. An empirical comparison of machine learning models for time series forecasting. Econometric Rev. 29 (5–6), 594–621.

Andreou, E., Ghysels, E., Kourtellos, A., 2013. Should macroeconomic forecasters use daily financial data and how? J. Bus. Econom. Statist. 31 (2), 240–251.

Athey, S., 2018. The impact of machine learning on economics. In: The Economics of Artificial Intelligence: An Agenda. University of Chicago Press.

Athey, S., Imbens, G.W., 2019. Machine learning methods that economists should know about. Annu. Rev. Econ. 11, 685–725.

Babii, A., Ghysels, E., Striaukas, J., 2021. Machine learning time series regressions with an application to nowcasting. J. Bus. Econom. Statist. 1–23.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70 (1), 191–221.

Bańbura, M., Giannone, D., Modugno, M., Reichlin, L., 2013. Now-casting and the real-time data flow. In: Handbook of Economic Forecasting, Vol. 2. Elsevier, pp. 195–237.

Bańbura, M., Rünstler, G., 2011. A look into the factor model black box: publication lags and the role of hard and soft data in forecasting GDP. Int. J. Forecast. 27 (2), 333–346.

Barhoumi, K., Darné, O., Ferrara, L., 2010. Are disaggregate data useful for factor analysis in forecasting French GDP? J. Forecast. 29 (1–2), 132–144.

Barnett, W.A., Tang, B., 2016. Chinese divisia monetary index and GDP nowcasting. Open Econ. Rev. 27 (5), 825–849.

Bok, B., Caratelli, D., Giannone, D., Sbordone, A.M., Tambalotti, A., 2018. Macroeconomic nowcasting and forecasting with big data. Annu. Rev. Econ. 10, 615–643.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 2017. Classification and Regression Trees. Routledge.

Camacho, M., Perez-Quiros, G., 2010. Introducing the Euro-sting: Short-term indicator of Euro area growth. J. Appl. Econometrics 25 (4), 663–694.

Carrasco, M., Rossi, B., 2016. In-sample inference and forecasting in misspecified factor models. J. Bus. Econom. Statist. 34 (3), 313–338.

Clements, M.P., Galvão, A.B., 2009. Forecasting US output growth using leading indicators: An appraisal using MIDAS models. J. Appl. Econometrics 24 (7), 1187–1206.

Coulombe, P.G., 2020a. The macroeconomy as a random forest. Available at SSRN 3633110.

Coulombe, P.G., 2020b. Time-varying parameters as ridge regressions. arXiv e-prints arXiv–2009.

Coulombe, P.G., Leroux, M., Stevanovic, D., Surprenant, S., 2022. How is machine learning useful for macroeconomic forecasting? J. Appl. Econometrics 37 (5), 920–964.

Coulombe, P.G., Marcellino, M., Stevanović, D., 2021. Can machine learning catch the Covid-19 recession? Natl. Inst. Econ. Rev. 256, 71–109.

de Valk, S., de Mattos, D., Ferreira, P., 2019. Nowcasting: An R package for predicting economic variables using dynamic factor models. R J. 11 (1), 230–244.

Elliott, G., Timmermann, A., 2016. Forecasting in economics and finance. Annu. Rev. Econ. 8, 81–110.

Evans, M.D., 2005. Where are We Now? Real-Time Estimates of the Macro Economy. Tech. Rep., Bank of International Settlement.

Foroni, C., Marcellino, M., Stevanovic, D., 2020. Forecasting the Covid-19 recession and recovery: Lessons from the financial crisis. Int. J. Forecast..

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33 (1), 1.

Ghysels, E., Kvedaras, V., Zemlys, V., 2016. Mixed frequency data sampling regression models: the R package midasr. J. Stat. Softw. 72 (1), 1–35.

Ghysels, E., Santa-Clara, P., Valkanov, R., 2004. The MIDAS touch: Mixed data sampling regression models.

Ghysels, E., Sinko, A., Valkanov, R., 2007. MIDAS regressions: Further results and new directions. Econometric Rev. 26 (1), 53–90.

Giacomini, R., White, H., 2006. Tests of conditional predictive ability. Econometrica 74 (6), 1545–1578.

Giannone, D., Reichlin, L., Small, D., 2008. Nowcasting: The real-time informational content of macroeconomic data. J. Monetary Econ. 55 (4), 665–676.

Gupta, R., Kabundi, A., 2011. A large factor model for forecasting macroeconomic variables in South Africa. Int. J. Forecast. 27 (4), 1076–1088.

Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79 (2), 453–497.
Hansen, P.R., Timmermann, A., 2012. Choice of sample split in out-of-sample forecast evaluation.
Hendry, D., Castle, J., Kitov, O., 2016. Forecasting and nowcasting macroeconomic variables: a methodological overview. Handb. Rapid Estim..
Higgins, P.C., 2014. GDPNow: A model for GDP nowcasting.
Higgins, P., Zha, T., Zhong, W., 2016. Forecasting China's economic growth and inflation. China Econ. Rev. 41, 46–61.
Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Netw. 2 (5), 359–366.
Jiang, Y., Guo, Y., Zhang, Y., 2017. Forecasting China's GDP growth using dynamic factors and mixed-frequency data. Econ. Model. 66, 132–138.
Kotchoni, R., Leroux, M., Stevanovic, D., 2019. Macroeconomic forecast accuracy in a data-rich environment. J. Appl. Econometrics 34 (7), 1050–1072.
Kuhn, M., 2015. A short introduction to the caret package. R Found. Stat. Comput. 1, 1–10.
Kuzin, V., Marcellino, M., Schumacher, C., 2011. MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the Euro area. Int. J. Forecast. 27 (2), 529–542.
Lahiri, K., Monokroussos, G., 2013. Nowcasting US GDP: The role of ISM business surveys. Int. J. Forecast. 29 (4), 644–658.
Lin, C.-Y., Wang, C., 2013. Forecasting China's inflation in a data-rich environment. Appl. Econ. 45 (21), 3049–3057.
Matheson, T.D., 2010. An analysis of the informational content of New Zealand data releases: the importance of business opinion surveys. Econ. Model. 27 (1), 304–314.
Matheson, T., 2011. New Indicators for Tracking Growth in Real Time. International Monetary Fund, pp. 11–43.
McCracken, M.W., Ng, S., 2016. FRED-MD: A monthly database for macroeconomic research. J. Bus. Econom. Statist. 34 (4), 574–589.
Medeiros, M.C., Vasconcelos, G.F., Veiga, Á., Zilberman, E., 2021. Forecasting inflation in a data-rich environment: the benefits of machine learning methods. J. Bus. Econom. Statist. 39 (1), 98–119.
Mevik, B.-H., Wehrens, R., 2007. The pls package: principal component and partial least squares regression in R. J. Stat. Softw. 18, 1–23.
Mullainathan, S., Spiess, J., 2017. Machine learning: an applied econometric approach. J. Econ. Perspect. 31 (2), 87–106.
Richardson, A., van Florenstein Mulder, T., Vehbi, T., 2021. Nowcasting GDP using machine-learning algorithms: A real-time assessment. Int. J. Forecast. 37 (2), 941–948.
Ripley, B., Venables, W., Ripley, M.B., 2016. Package 'nnet'. R Pack. Ver. 7 (3–12), 700.
Stock, J.H., Watson, M.W., 2002. Macroeconomic forecasting using diffusion indexes. J. Bus. Econom. Statist. 20 (2), 147–162.
Stock, J.H., Watson, M.W., 2016. Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In: Handbook of Macroeconomics, Vol. 2. Elsevier, pp. 415–525.
Stock, J.H., Watson, M.W., 2017. Twenty years of time series econometrics in ten pictures. J. Econ. Perspect. 31 (2), 59–86.
Sun, Y., Wang, S., Zhang, X., 2018. How efficient are China's macroeconomic forecasts? Evidences from a new forecasting evaluation approach. Econ. Model. 68, 506–513.
Yiu, M.S., Chow, K.K., 2010. Nowcasting Chinese GDP: information content of economic and financial data. China Econ. J. 3 (3), 223–240.
Zhang, Y., Cindy, L.Y., Li, H., Hong, Y., 2018. Nowcasting China's GDP using a Bayesian approach. J. Manage. Sci. Eng. 3 (4), 232–258.