Australian National University

Crawford School of Public Policy

CAMA

Centre for Applied Macroeconomic Analysis

Nowcasting New Zealand GDP Using Machine Learning Algorithms

CAMA Working Paper 47/2018
September 2018

Adam Richardson
Reserve Bank of New Zealand

Thomas van Florenstein Mulder
Reserve Bank of New Zealand

Tuğrul Vehbi
Reserve Bank of New Zealand
Centre for Applied Macroeconomic Analysis, ANU

Abstract

This paper analyses the real-time nowcasting performance of machine learning algorithms estimated on New Zealand data. Using a large set of real-time quarterly macroeconomic indicators, we train a range of popular machine learning algorithms and nowcast real GDP growth for each quarter over the 2009Q1-2018Q1 period. We compare the predictive accuracy of these nowcasts with that of other traditional univariate and multivariate statistical models. We find that the machine learning algorithms outperform the traditional statistical models. Moreover, combining the individual machine learning nowcasts further improves the performance than in the case of the individual nowcasts alone.

Keywords

Nowcasting, Machine learning, Forecast evaluation

JEL Classification

C52, C53, C55

Address for correspondence:

(E) cama.admin@anu.edu.au

ISSN 2206-0332

Nowcasting New Zealand GDP Using Machine Learning Algorithms*

Adam Richardson†, Thomas van Florenstein Mulder‡, Tuğrul Vehbi§

September 27, 2018

Abstract

This paper analyses the real-time nowcasting performance of machine learning algorithms estimated on New Zealand data. Using a large set of real-time quarterly macroeconomic indicators, we train a range of popular machine learning algorithms and nowcast real GDP growth for each quarter over the 2009Q1-2018Q1 period. We compare the predictive accuracy of these nowcasts with that of other traditional univariate and multivariate statistical models. We find that the machine learning algorithms outperform the traditional statistical models. Moreover, combining the individual machine learning nowcasts further improves the performance than in the case of the individual nowcasts alone.

Keywords: Nowcasting, Machine learning, Forecast evaluation

JEL classification: C52, C53, C55

*We thank Christopher Ball, Punnoose Jacob and the seminar participants at the Reserve Bank of New Zealand and the BIS/Bank Indonesia "2018 International Seminar on Big Data for Policy Making" for their valuable comments. The views expressed here are the views of the authors and do not necessarily reflect the views of the Reserve Bank of New Zealand.

†Reserve Bank of New Zealand. Email: Adam.Richardson@rbnz.govt.nz

‡Reserve Bank of New Zealand. Email: Thomas.vanFlorensteinMulder@rbnz.govt.nz

§Reserve Bank of New Zealand and CAMA (ANU). Email: Tugrul.Vehbi@rbnz.govt.nz

1 Introduction

Policy makers typically make decisions in real time using incomplete information on current economic conditions. Many key statistics are released with lags and are subject to frequent revisions. Nowcasting models have been increasingly popular tools developed to mitigate some of these uncertainties and they have been widely used by forecasters at many central banks and other institutions (Giannone et al. 2008, Banbura et al. 2013, Jansen et al. 2016, Bloor 2009).

Prompted by advances in computing power, machine learning (ML hereafter) methods have recently been proposed as alternatives to time-series regression models typically used by central banks for forecasting key macroeconomic variables. The ML models are particularly suited for handling large datasets when the number of potential regressors is larger than that of available observations. In this paper, we investigate the performance of different ML algorithms in obtaining accurate nowcasts of the current quarter real gross domestic product (GDP) growth for New Zealand. We use multiple vintages of historical GDP data and multiple vintages of a large features set - comprising approximately 550 domestic and international variables - to evaluate the real-time performance of these algorithms over the 2009Q1-2018Q1 period. We then compare the forecasts obtained from these algorithms with the forecasting accuracy of a naive autoregressive benchmark as well as other data-rich methods such as a factor model, a Bayesian VAR (BVAR) and a suite of statistical models used at the RBNZ. To our knowledge, our study is the first to evaluate the relative nowcast performance of alternative ML methods using real-time data.

Our results show that the majority of the ML models produce point nowcasts that are superior to the simple AR benchmark. The top-performing models such as the support vector machines, Lasso (Least Absolute Shrinkage and Selection Operator) and neural networks are able to reduce the average nowcast errors by approximately 16-18 per cent relative to the AR benchmark. Moreover, combining the nowcasts of the ML models using various weighting schemes leads to further improvements in performance. The majority of the ML algorithms also outperform the other two commonly used statistical benchmarks, namely the factor model and the small Bayesian VAR model.

This paper joins a growing literature that evaluates the relative success of the ML models in forecasting over the more traditional time-series techniques. However, to our knowledge, none of these papers focuses on the real-time forecasting performance of these models.

Makridakis et al. 2018 compares the forecast accuracy of various popular ML algorithms with eight types of traditional statistical benchmarks and finds that the out-of-sample forecasting accuracy of ML models is lower than that of more traditional statistical methods.

Chakraborty & Joseph 2017, on the other hand, conduct an out-of-sample forecasting exercise using UK data and argue that ML models generally outperform traditional modelling approaches in prediction tasks. Similarly, Kim 2003 finds that support vector machines in particular are a promising alternative for stock market prediction.

Given we use a large dataset for New Zealand for our analysis, our paper is also related to Eickmeier & Ng 2011 who use elastic net and ridge regression (amongst other shrinkage methods) to forecast New Zealand's GDP from a large number of domestic and international predictors. They find data-rich methods result in gains in forecast accuracy over common statistical methods using small data sets. Also, Matheson 2006 uses a factor model to produce real-time forecasts of New Zealand's GDP, inflation, interest rate and exchange rate from a large number of predictors which has good forecast performance at longer-term horizons when compared to other statistical models.

The remainder of this paper is as follows. Section 2 explains the models and the data used. Section 3 presents the results and Section 4 concludes.

2 Empirical Application

In this section, we provide a brief description of the various ML and benchmark models we considered for nowcasting GDP. We also discuss how we made key hyperparameter choices.[1]

2.1 Models

2.1.1 Autoregressive Model (AR)

As a benchmark, we use an univariate AR model of order 1 for quarterly GDP growth (y_t):

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + u_t \tag{1}$$

where α_0 and α_1 are parameters and u_t is the residual term.

2.1.2 *K* Nearest Neighbour Regression (KNN)

KNN is a non-parametric method that works by storing all available training data and predicting the outcome of new data based on the k-most similar observations (neighbours) in the training set to the new data. The similarity is typically measured using the Euclidean

[1] We initially investigated using hyperparameter optimization through k-fold cross validation. However, this methodology proved computationally intensive, and resulted in poor out-of-sample forecast performance compared to adopting standard or default values for hyperparameters.

distance and the prediction is computed as the mean of the k most similar instances. We choose k as 4 which produces the lowest average RMSE calculated over all iterations.

2.1.3 Least-squares boosting (LSBoost)

LSBoost is an ensemble aggregation methodology. In this approach, the algorithm builds up a higher quality predictor from a number of weak individual predictors. The individual predictors (called learners) in this case are regression trees. At each step in the algorithm, a new tree is fitted to the difference between the observed response and the aggregated prediction of all trees grown previously. This ensemble minimises the mean-squared error at each step.

We have chosen hyperparameters for this methodology with a focus on improving the out-of-sample generalization of the model. The learning rate, which governs how the information from a new tree is incorporated into the model, is set to 0.1, as discussed in Hastie et al. (2001). This means that, for every new learner, corrections made to previous residuals are weighted less than 1, to avoid over-fitting the in-sample data. Our initial investigations into hyperparameter optimization pointed to the need for only a small number of learners - so our number of learning cycles is set to 15. We also need to make choices about the structure of the regression trees. Specifically, we restrict the maximum number of splits to 10. We also opt for a minimum leaf size of 8, to avoid arbitrary splits with too few predictors.

2.1.4 Lasso, Ridge and Elastic Net (ENET)

These three methods are very similar to ordinary least squares (OLS) but incorporate different types of shrinkage for creating parsimonious models in presence of a large number of features. Lasso performs $L1$ regularisation which involves adding a penalty equivalent to the absolute value of the magnitude of the coefficients and shrinking some of them to zero. Ridge is a similar method to lasso but it performs $L2$ regularisation where the penalty is on the absolute value of the magnitude of the coefficients. Therefore the coefficients estimated by ridge are never reduced to exactly zero. The elastic net regression is a hybrid of the ridge and lasso regressions and as such, it can shrink the coefficients of the features as well as removing them completely (a coefficient of zero). The regression penalty is, therefore, a convex sum of the ridge and lasso penalties.[2]

$$\beta = argmin[\sum_{i=1}^{l}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j) + \lambda \sum_{j=1}^{p}(\alpha\beta_j^2 + (1-\alpha)|\beta_j|)] \tag{2}$$

[2]For a detailed explanation of the elastic net refer to Zou & Hastie 2005

For the Lasso and ridge regression, we use a λ value of 0.1. For the Elastic Net, we set both α and λ to 0.1.

2.1.5 Support Vector Machine Regression (SVM)

SVM is a non-parametric approach which aims to identify the hyperplane that maximizes the margin between classes while also making sure that the perpendicular distance between the two closes points from either of these two classes is maximized.[3]For our modelling, we use linear epsilon-insensitive SVM (ε-SVM) regression with a polynomial kernel function of order 2. As is commonly used in the literature, we use an epsilon value of a tenth of the interquartile range of the target variable; a proxy for the noise level in the training sample. The hyperparameter C (ie the box constraint) is set in a similar way using the interquartile range of the target variable.

2.1.6 Feed Forward Neural Network (NN)

A feed forward neural network is a model that is able to capture and represent complex relationships. The model works by taking input data and using weights and an activation function to pass them through to N hidden layers of a perceptron. Each input is weighted and passed through the activation function to determine the value of a given node within the first layer of perceptrons and this is repeated for each node in the first layer. Each node of the first layer then becomes the new input variable for layer 2 and gets re-weighted and passed through the activation function to determine the value of a given node in layer 3. The process is repeated until the N^{th} layer is created. The nodes in the N^{th} layer are weighted and passed through an activation function to give the output value. The weights are initially set with random values and are updated on each iteration using the backpropagation algorithm.[4]

The key parameter choices for this model are the number of nodes and the number of hidden layers. Our initial modelling used 10 nodes at each layer, with 2 hidden layers. We choose 10 nodes as this captures the overall variation of the feature set (for comparison, 10 factors from a principle components model covers about 70 percent of the variation of the feature set), and 2 layers are typically enough to capture any potential non-linear relationships. However, practically, we found a 10 node, 1 hidden layer model produced very similar results, and substantially reduced computational time. After our initial investigation on fitting neural networks, we used 50 neural networks at each vintage and average the nowcasts from these 50 models. [5]

[3]See Vapnik 1995 for detailed information

[4]For more details on backpropagation see Hecht-Nielsen 1992

[5]As a sensitivity check, we also tested models with 5 nodes, and models with an early stopping criteria

2.1.7 Factor Model (FM)

Factor models use a large number of time series to produce forecasts, and therefore do not require the model builder to make strong assumptions about what particular series are important for forecasting the variable of interest. We estimate a linear factor model which assumes that the quarterly growth rate of GDP is given by

$$y_t = \alpha_0 + \alpha_1 f_1 + ... + \alpha_k f_k + u_t \tag{3}$$

where the $f_j, j = 1, ..., k$, are the common factors obtained using the principal components technique. These factors are the linear combinations of all the data in the model that explain the highest proportion of the variance of the data. They can, therefore, be thought of as picking up the underlying movements in the economy that influence a large number of variables. The estimated factors are used in linear regressions of the variables of interest. We choose the optimal number of factors to incorporate in (3) using the Bai & Ng 2001 two-step procedure and use the Bayesian Information Criteria (BIC) as the benchmark for selection.

2.1.8 Bayesian VAR (BVAR)

This model uses 95 macroeconomic time series in a Bayesian VAR framework. The model produces forecasts for key macroeconomic variables such as GDP, consumption, inflation, the exchange rate and interest rates. The BVAR has 4 lags and a Minnesota prior (Doan et al. 1984) as the method of shrinkage. We use the quarterly GDP growth rate forecasts from this model as the basis for our comparisons.

2.1.9 RBNZ Statistical Suite (RBNZ SS)

The suite contains a range of different models that vary across size and complexity (Bloor 2009). The models in the suite are particularly designed to forecast medium-term movements in the economy and are used as a cross check for the central forecasts produced by the main policy model and sectoral experts. The suite also includes several models that are aimed towards picking up shorter-run fluctuations.

2.2 Data

The data consist of a number of continuous real-time vintages of a range of macroeconomic and financial market statistics. These include: New Zealand business surveys; consumer

linked to a minimum MSE and found similar out-of-sample results.

and producer prices; general domestic activity indicators (e.g. concrete production, milk-solids production, spending on electronic cards etc.); domestic trade statistics; international macroeconomic variables and international and domestic financial market variables. The data range from daily to quarterly - with the mean used to aggregate higher frequency data to quarterly for model estimation. The series are then seasonally adjusted. Each series is individually assessed, and either left in a level form or transformed to a stationary form depending on which form is likely to be more predictive of GDP growth. All data are standardized.

The storage of historical model runs by the RBNZ has allowed us to create 37 real-time vintages of these data. Every 3 months, the RBNZ publishes a *Monetary Policy Statement*. In working towards this publication, the RBNZ's staff put together an initial set of macroeconomic projections. The data banks containing the data described above were saved down along with these projections each quarter. This process, therefore, gives us a quarterly real-time snapshot of a range of macroeconomic and financial market series. Conveniently, these snapshots were generally taken about four weeks before the release of the preceding quarters GDP estimate. For example, the initial projections for the March 2015 *Monetary Policy Statement* would have been finalized on about the 20th of February. At this point, almost all of the key macroeconomic and financial market indicators for the December 2014 quarter would have been released - and it is at this point the snapshot of these macroeconomic statistics has been saved. The December quarter GDP estimate was then released 19th March. New Zealand does not produce flash estimates of GDP, and so there is a significant lag between the end of the quarter and the publication of the GDP estimate.

From this point, the data storage methodology was changed. The 'global database' containing 668 series (the detail of which we described above) was routinely saved in estimating the Bank's suite of statistical models. A version of this data set is saved at the end of the month following each *Monetary Policy Statement*. For example, for the May 2018 Monetary Policy Statement, a version of the data set was saved on the last working day of May. This data set contains most of the indicators up to the end of 2018Q1. The 2018Q1 GDP figures were then released on the 21 June 2018 - around 3 weeks after the data snapshot was taken. This process gives us a set of data vintages from 2015Q3 to 2018Q1.

The data available with each vintage differ somewhat, as data were added and removed from the RBNZ's data banks through time. After making the modifications described above, from a candidate of 668 series, we are left with between 532 to 634 series at each vintage.

In total, we have a 37 real-time vintages of this dataset, covering the period 2009Q1 to 2018Q1. The data in each vintage begin in 1995Q1. These data vintages enable us to test how the forecast performance of these models compares under the conditions which the practitioner would use them - capturing the revision properties of the predictors.

2.3 Forecast evaluation methodology

We evaluate the performance of the models using an out-of-sample forecast exercise. We train each algorithm over an expanding window thereby replicating an actual forecasting situation starting from 2009Q1 and moving forward a quarter at a time through to 2018Q1. For example, for the first vintage of the data, the models are estimated over the period 1995Q1 to 2008Q4 using real-time data for both the predictor and response variables. The resultant fitted models are used to nowcast the 2009Q1 growth rate of real GDP. Overall, we generate 37 real time nowcasts of quarterly GDP growth. As discussed above, we choose fairly standard parameter settings for each algorithm. Next, we measure the forecast accuracy of each model by calculating the Root Mean Square Error (RMSE) and the Mean Absolute Deviation (MAD) defined as:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}{T}} \tag{4}$$

$$MAD = \frac{1}{T}\sum_{t=1}^{T}|y_t - \hat{y}_t| \tag{5}$$

where y_t and $\hat{y}_t$ are the actual and forecast values of GDP growth and T is the total number of forecasts. The forecasts of a univariate (i.e. AR(1)) model provide the main benchmark for our comparisons. We use the Diebold-Mariano (Diebold & Mariano 1995) test to determine whether the forecasts obtained from each ML model are significantly different than those from the AR model.

3 Empirical Results

In this section, we describe the main results of our analysis. Table 1 documents the nowcast performance of the models for the sample period 2009Q1-2018Q1. In addition to the models outlined in Section 2.1, we also present the results obtained by combining the forecasts from all the ML models using alternative weighting schemes.

The results indicate that the large majority of the ML models produce forecasts that have RMSEs and MADs lower than the AR benchmark. The top three models are the SVM, Lasso and NN models which are able to reduce the average forecast errors by approximately 16-18 per cent relative to the AR benchmark. The relative success of the neural network and support vector machine models are not surprising and is in line with previous findings in the literature (Teräsvirta et al. 2005, Ahmed et al. 2010). The majority of the ML models are also able to produce RMSEs and MADs lower than the BVAR, factor model and the combination forecasts obtained from RBNZ's statistical suite. However, the DM test results based on the RMSE loss function indicate that in most cases, the null hypothesis that the forecast errors are equal cannot be rejected. The null hypothesis of equal forecast errors based on the mean absolute deviations, on the other hand, is rejected for the case of SVM, NN and the Lasso at the conventional levels of statistical significance. The DM test results, however, should be treated with caution given our small sample size of 37 observations.

It is important to note that some of these models come with the added costs of increased computational time, and a lack of tractability when it comes to understanding the drivers of certain results. This could be significant in practice in two regards. First, it may limit the practical use of such models in situations that require a quick turnaround. Second, if the forecast accuracy of a model started to deteriorate, it may be difficult to pick apart the factors leading to such a deterioration.

Figure 1 presents the quarterly GDP growth and its nowcasts obtained from each model over the sample period. It can be seen that all ML models have successfully predicted the sharp downturn in activity occurred in the first quarter of 2009 and also predicted the other major upturns and downturns in the GDP data successfully.

Figure 1: Real-time nowcasts of quarterly GDP growth

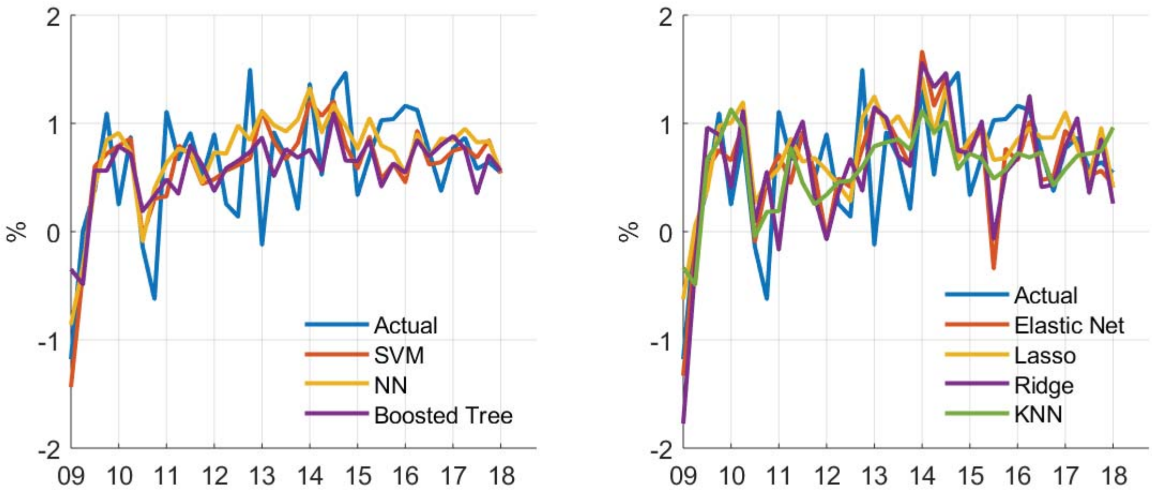


Table 1: Real-time nowcast performances of models (RMSE), 2009Q1-2018Q1

Models	RMSE	RMSE (Rel. to AR)	p-value	MAD	MAD (Rel. to AR)	p-value
SVM	0.445	0.820	0.166	0.338	0.770	0.037
NN	0.446	0.821	0.182	0.336	0.766	0.041
Lasso	0.454	0.836	0.206	0.348	0.793	0.061
BT	0.469	0.865	0.153	0.386	0.880	0.107
ENET	0.488	0.899	0.526	0.349	0.796	0.140
KNN	0.479	0.882	0.251	0.387	0.882	0.196
Ridge	0.565	1.040	0.770	0.427	0.972	0.823
FM	0.517	0.951	0.757	0.379	0.864	0.357
BVAR	0.608	1.119	0.286	0.450	1.025	0.961
RBNZ SS	0.471	0.867	0.262	0.384	0.874	0.160
AR	0.543	-	-	0.439	-	-

Notes: The first column refers to the models used to nowcast GDP. SVM: support vector machine; NN: neural network; ENET: elastic net; Lasso: lasso regression; BT: boosted tree; FM: factor model; Ridge: ridge regression; AR: autoregressive model; BVAR: bayesian VAR; KNN: k-nearest neighbours; RBNZ SS: the statistical models suite used by the RBNZ. The second column refers to the entries for the out-of-sample RMSEs obtained from each model using the methodology outlined in subsection 2.3. The third column refers to the RMSEs relative to the AR model. The fourth column refers to the p-values obtained from the Diebold-Mariano test for testing the significance of the forecast accuracy of each method versus that of the AR model. Columns 5-7 refer to the corresponding values in columns 2-4 when the loss function is the Mean Absolute Deviation (MAD).

3.1 Forecast combination

In the previous section, we compared the forecasts from individual models by ranking them individually according to their forecast accuracy. From a practical point of view, however, we may prefer to pick the "best" model amongst them to use for nowcasting. Therefore, an alternative approach is to combine forecasts from the set of all models under consideration to produce a single summary forecast. Forecast combinations have frequently been found in the literature to produce better forecasts than individual models. In this section, we implement this strategy using all the ML models we considered. More specifically, we use four types of forecast combination strategies for combining point forecasts: equal weighting, Least-squares weighting, inverted mean squared error (MSE) and MSE ranks weighting. Equal weighting is a particularly simple method which works by assigning equal weights to the forecasts from all individual models at each date in the forecast sample. The Least-squares weighting strategy, on the other hand, is implemented by regressing all the forecasts against the actual values and then using the coefficients from the resultant regression as weights. The final two strategies both small use the inverted MSEs computed over the forecast horizon for weighting the forecasts where the latter is based on the inverted MSE ranks rather than the actual MSE values.

The results summarised in Table 2 suggest that there are gains in combining forecasts. It can be seen that the combined forecasts produce lower RMSEs and MADs compared to the individual model results presented in Table 1. Amongst the different weighting schemes we've considered, the Least-squares weighting scheme generates the best gains in predictive accuracy.

Table 2: Forecast combination results

	RMSE	RMSE (Rel. to AR)	MAD	MAD (Rel. to AR)
Equal	0.437	0.805	0.324	0.738*
Least-squares	0.427	0.786	0.317	0.722*
Inverted MSE	0.434	0.799	0.324	0.738*
Inverted MSE ranks	0.429	0.790	0.323	0.736*

Notes: The first column refers to the weighting methods used to combine the individual ML forecasts. Equal: assigning equally weighted forecasts; Least-squares: weights are assigned by regressing all the forecasts against the actual values and then using the coefficients from the resultant regression as weights neural network; Inverted MSE: weights are assigned using the inverted MSEs as weights; Inverted MSE ranks: weights are assigned using the inverted MSE ranks as weights. See Table 2 for column definitions. The * indicates statistical significance of the Diebold-Mariano test at the 5% significance level .

Furthermore, we investigate whether the optimal combination of the ML model nowcasts (i.e. the Least-squares weighting) adds value to the nowcasts generated by the combination of models in the RBNZ's statistical model suite. To test this formally, we follow the approach by D. Romer & H. Romer 2008 and estimate the following regression equation:

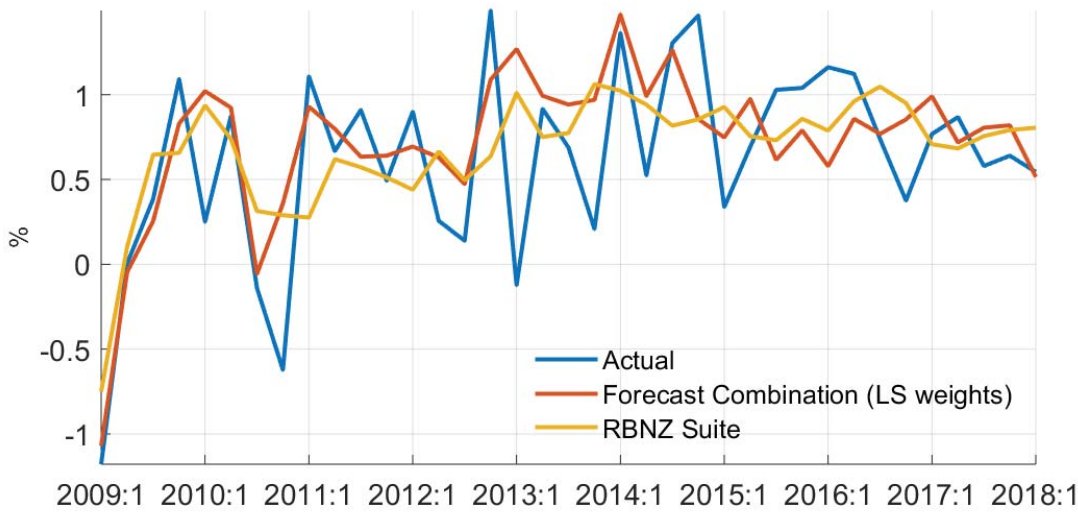
$$y_t = c + \alpha_1 FC + \alpha_2 SS + e_t \tag{6}$$

where y_t is the actual real GDP, FC is the combined nowcasts from the least-squares weights method outlined above, SS is the real-time GDP nowcasts obtained from the combination of models in the RBNZ statistical suite and e_t is the residual term. The results presented in Table 3 suggest that forecast combination adds significant value to the combined statistical-suite nowcasts as implied by the large, positive and significant α_1 coefficient. Figure 2 presents the quarterly GDP growth together with the two sets of nowcasts over the period 2009Q1-2018Q1.

Table 3: Does forecast combination add value to the RBNZ's statistical-suite nowcasts?

Coefficients	Estimated value	t-Statistic
α_1	0.94	4.14
α_2	-0.10	-0.30
C	0.004	0.04
R^2	0.46	

Figure 2: Quarterly GDP growth and its nowcasts (2009Q1-2018Q1)



4 Conclusion

In this paper, we evaluate the real-time performance of popular ML algorithms in obtaining accurate nowcasts of the real gross domestic product (GDP) growth for New Zealand. We estimate several ML models over the 2009-2018 period using multiple vintages of historical GDP data and a large features set comprising approximately 550 domestic and international variables. We then compare the forecasts obtained from these models with the forecasting accuracy of a naive autoregressive benchmark as well as other data-rich methods such as a factor model, a large Bayesian VAR and the combined GDP nowcasts obtained from the suite of statistical models used at the RBNZ. We find that the majority of the ML models are able to produce more accurate forecasts than those of the AR and other statistical benchmarks. The results also suggest that there are some gains in combining individual ML forecasts. Our results thus recommend the use of ML algorithms as an addition to a forecaster's suite of GDP nowcasting models.

References

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 Reviews* **29**(5-6), 594–621.

Bai, J. & Ng, S. (2001), 'Determining the number of factors in approximate factor models', **70**.

Banbura, M., Giannone, D., Modugno, M. & Reichlin, L. (2013), 'Now-casting and the real-time data flow', *Handbook of economic forecasting* **2**(Part A), 195–237.

Bloor, C. (2009), 'The use of statistical forecasting models at the Reserve Bank of New Zealand', *Reserve Bank of New Zealand Bulletin* **72**, 21–26.

Chakraborty, C. & Joseph, A. (2017), Machine learning at central banks, Bank of England working papers 674, Bank of England.

D. Romer, C. & H. Romer, D. (2008), 'The fomc versus the staff: Where can monetary policymakers add value?', **98**, 230–35.

Diebold, F. X. & Mariano, R. S. (1995), 'Comparing Predictive Accuracy', *Journal of Business & Economic Statistics* **13**(3), 253–263.

Doan, T., Litterman, R. & Sims, C. (1984), 'Forecasting and conditional projection using realistic prior distributions', *Econometric Reviews* **3**(1), 1–100.

Eickmeier, S. & Ng, T. (2011), 'Forecasting national activity using lots of international predictors: An application to New Zealand', *International Journal of Forecasting* **27**(2), 496–511.

Giannone, D., Reichlin, L. & Small, D. (2008), 'Nowcasting: The real-time informational content of macroeconomic data', *Journal of Monetary Economics* **55**(4), 665 – 676.

Hastie, T., Tibshirani, R. & Friedman, J. (2001), *The Elements of Statistical Learning*, Springer Series in Statistics, Springer New York Inc., New York, NY, USA.

Hecht-Nielsen, R. (1992), Theory of the backpropagation neural network, *in* 'Neural networks for perception', Elsevier, pp. 65–93.

Jansen, W. J., Jin, X. & de Winter, J. M. (2016), 'Forecasting and nowcasting real gdp: Comparing statistical models and subjective forecasts', *International Journal of Forecasting* **32**(2), 411 – 436.

Kim, K. (2003), 'Financial time series forecasting using support vector machines', *Neurocomputing* **55**(1), 307 – 319. Support Vector Machines.

Makridakis, S., Spiliotis, E. & Assimakopoulos, V. (2018), 'Statistical and machine learning forecasting methods: Concerns and ways forward', *PloS one* **13**(3).

Matheson, T. D. (2006), 'Factor Model Forecasts for New Zealand', *International Journal of Central Banking* **2**(2).

Teräsvirta, T., van Dijk, D. & Medeiros, M. C. (2005), 'Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination', *International Journal of Forecasting* **21**(4), 755 – 774. Nonlinearities, Business Cycles and Forecasting.

Vapnik, V. N. (1995), *The Nature of Statistical Learning Theory*, Springer-Verlag, Berlin, Heidelberg.

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