



IFC – Bank Indonesia International Workshop and Seminar on *“Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data”*

Bali, Indonesia, 23-26 July 2018

Nowcasting New Zealand GDP using machine learning algorithms¹

Adam Richardson, Thomas van Florenstein Mulder, Tugrul Vehbi,
Reserve Bank of New Zealand

¹ This paper was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Nowcasting New Zealand GDP using machine learning algorithms

Adam Richardson¹, Thomas van Florenstein Mulder², Tuğrul Vehbi³

Abstract

We examine whether machine learning algorithms can improve nowcasts of real GDP growth in New Zealand. We use a large real-time dataset of around 550 New Zealand and international macroeconomic indicators. We train a range of popular machine learning algorithms over an expanding window and replicate an actual nowcasting situation starting from 2009 Q1 and moving forward a quarter at a time through to 2018 Q1. We compare the predictive accuracy of these nowcasts with that of other benchmarks such as a simple autoregressive model, a factor model, a large Bayesian VAR and a suite of statistical models used at the Reserve Bank of New Zealand. We find that the machine learning algorithms outperform the statistical benchmarks. Moreover, combining the nowcasts of the machine learning models leads to further improvements in performance. The results indicate that there are gains in nowcasting accuracy from using machine learning methods.

Keywords: Nowcasting, Machine learning, Forecast evaluation

JEL classification: C52, C53, C55,

¹ Manager, Modelling Team, Reserve Bank of New Zealand. Email: Adam.Richardson@rbnz.govt.nz

² Analyst, Modelling Team, Reserve Bank of New Zealand. Email: Thomas.vanFlorensteinMulder@rbnz.govt.nz

³ Adviser, Modelling Team, Reserve Bank of New Zealand. Email: Tugrul.Vehbi@rbnz.govt.nz

We would like to thank the seminar participants at the Reserve Bank of New Zealand 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.

Contents

Nowcasting New Zealand GDP using machine learning algorithms	1
1. Introduction	2
2. Empirical Application	3
2.1 Models	4
2.2 Forecast evaluation methodology	7
2.3 Data	8
3. Empirical Results	9
3.1 Forecast combination	11
4. Conclusion	14
References	15

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 2009-2018 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 small 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.

Our contributions in this paper are twofold. First, we provide new evidence on the *real-time* nowcast performance of commonly used ML algorithms by comparing them to other useful benchmark models commonly used at central banks. Second, our results have important repercussions for policy as the accuracy of GDP nowcasts play a big role in correctly assessing the overall health of the economy.

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 focusses on the real-time forecasting performance of the 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 and 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 are a promising alternative for predicting stock market variables.

To the extent, we use a large dataset for New Zealand for our analysis, our paper is also related to Eickmeier and Ng (2011) who use elastic net and ridge regression (amongst other shrinkage methods) to produce macroeconomic forecasts 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 macroeconomic forecasts from a large number of predictors. This model results in good forecast performance at longer-term horizons when compared to other statistical models and the RBNZ's own forecasts.

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.

2.1 Models

i. Autoregressive Model (AR)

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

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

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

ii. K-Nearest Neighbour Regression (KNN)

KNN is a widely used non-parametric method that uses a distance function, in our case the Euclidean distance, to measure the similarity between the data used for training and testing purposes. The algorithm finds which k most observations in the training set have features most similar to the features of the hold out set and predict the outcome of the new data to be the outcome of the average of the k observations previously defined. The prediction is based on the mean of the k –most similar instances. We choose k using the grid search method available in Python’s Scikit-Learn library.

iii. Boosted Trees (BT)

This is an ensemble method which allows converting a set of weak learners into one high-quality predictor. The general idea is to impute a sequence of simple trees, where each successive tree is built for the prediction residuals of the preceding tree. The algorithm works by dividing the predictor space into a set of possible values for $\{x_1, x_2, \dots, x_i\}$ where x_i corresponds to the feature set of observation i. Starting with the predictor whose splitting leads to the greatest reduction in the sum of squared residuals, the algorithm then selects the next feature whose split minimises the sum of squared residuals given the split of the initial feature. This process continues until the reduction in the sum of squared residuals falls below a certain threshold. We use gradient tree boosting (GBRT) which is a forward stagewise methodology to add regression tree models one after the other to increase predictive ability. The added model is chosen by minimising a given loss function. GBRT continues this additive process until the gains in predictability from adding new models drops below some threshold value.

iv. 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 squared 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.⁴

$$\beta^{Lasso} = \underset{\beta}{\operatorname{argmin}} \left[\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p (\alpha \beta_j^2 + (1 - \alpha) |\beta_j|) \right]$$

The equation above shows the ridge case when $\alpha = 1$, the lasso case when $\alpha = 0$, and the elastic net case for any values in between. α and λ are hyperparameters that determine the relative impact of the penalty terms. When $\lambda = 0$, we have the standard ordinary least squares (OLS) approach.

v. Support Vector Machine Regression (SVM)

Support vector machine regression model is a non-parametric method first proposed by Vapnik (1995). The idea of SVM is to find a linear function of the form

$$f(x) = \langle w, x \rangle + b$$

where w is the weight vector, b is the bias and x is the input or feature vector while ensuring that the function is as flat as possible meaning one seeks a small w . One way to implement this is to minimize the norm, i.e. $\|w\|^2$. This can be formulated as a convex optimization problem:

$$0.5\|w\|^2 + C \sum_{i=1}^l |y_i - f(x_i)|_\epsilon$$

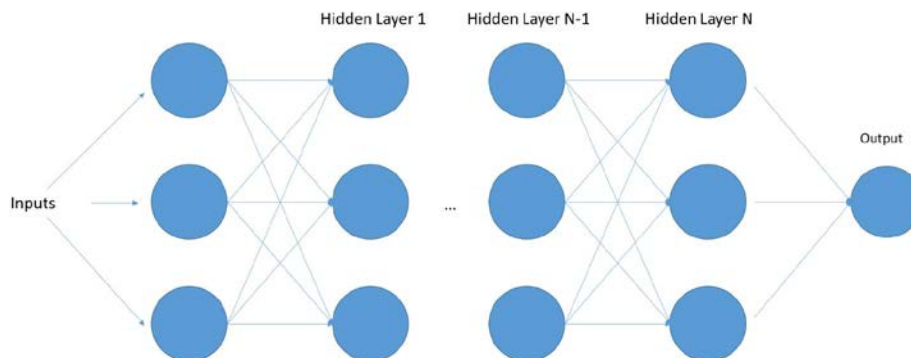
⁴ For a detailed explanation of the elastic net refer to (Zou & Hastie 2005)

where $C > 0$ is the regularisation parameter. The first term in the error function is a penalty term that increases as the model becomes more complex. The second term is the ϵ -insensitive loss function that penalises errors that are greater than ϵ , allowing flexibility to the model. For estimating the model, we use Matlab's Statistics and Machine Learning Toolbox which implements linear epsilon-insensitive SVM (ϵ -SVM) regression, which is also known as L1 loss.

vi. Neural Network (NN)

Neural network model 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 as shown in figure 1. 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 reweighted 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.

Figure 1: Multi Layer Perceptron Process



The weights are initially set with random values and are updated on each iteration using an algorithm called backpropagation. With backpropagation, the input data is repeatedly presented to the neural network where the output of the neural network and the desired output is produced.⁵

⁵ For more details on backpropagation see Hecht-Nielsen 1992

vii. 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

$$y_t = \alpha_0 + \alpha_1 f_1 + \dots + \alpha_k f_k + u_t$$

where the $f_j, j = 1, \dots, k$, 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 (8) using the Bai and Ng (2002) two-step procedure and use the Bayesian Information Criteria (BIC) as the benchmark for selection.

viii. Bayesian VAR (BVAR)

This model is part of the statistical models suite at the RBNZ and utilises 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. We use the quarterly GDP growth rate forecasts from this model as the basis for our comparisons.

ix. 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. The suite also includes several models that are aimed towards picking up shorter-run fluctuations.

2.2 Forecast evaluation methodology

We evaluate the performance of the models by means of an out-of-sample forecast exercise. We train each algorithm over an expanding window thereby replicating an actual forecasting situation starting from 2009 Q1 and moving forward a quarter at a time through to 2018 Q1. 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 2009 Q1 growth rate of real GDP. Overall, we generate 37 real-time nowcasts of quarterly GDP growth. We choose the default 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:

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

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

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.

2.3 Data

The data consist of a number of real-time vintages of a range of macroeconomic and financial market statistics. These include: New Zealand business surveys; consumer 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 - and are aggregated to quarterly data for model estimation. The series are 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. We also remove the series that contain missing values.

The storage of historical model runs 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. This process gives us a set of data vintages from 2009 Q1 to 2018 Q1.

From this point, the data storage methodology was changed. The 'global database' containing 668 series (the detail of which I 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 2018 Q1. The 2018 Q1 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 2015 Q3 to 2018 Q1.

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 532-634 series at each vintage.

In total, we have a 37 real-time vintages of this dataset, covering the period 2009 Q1 to 2018 Q1. The data in each vintage begin in 1995 Q1. 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.

3. Empirical Results

In this section, we describe the main results of our analysis. Table 1 presents data on the nowcast performances of the models for the sample period 2009 Q1-2018 Q1. 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 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. The most computationally intensive model is the NN model, taking approximately 31 minutes to solve. All the other models are relatively less intensive taking only several seconds to solve.⁶

Figure 2 presents the quarterly GDP growth and its nowcasts obtained from each model over the 2009 Q1-2018 Q1 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.

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

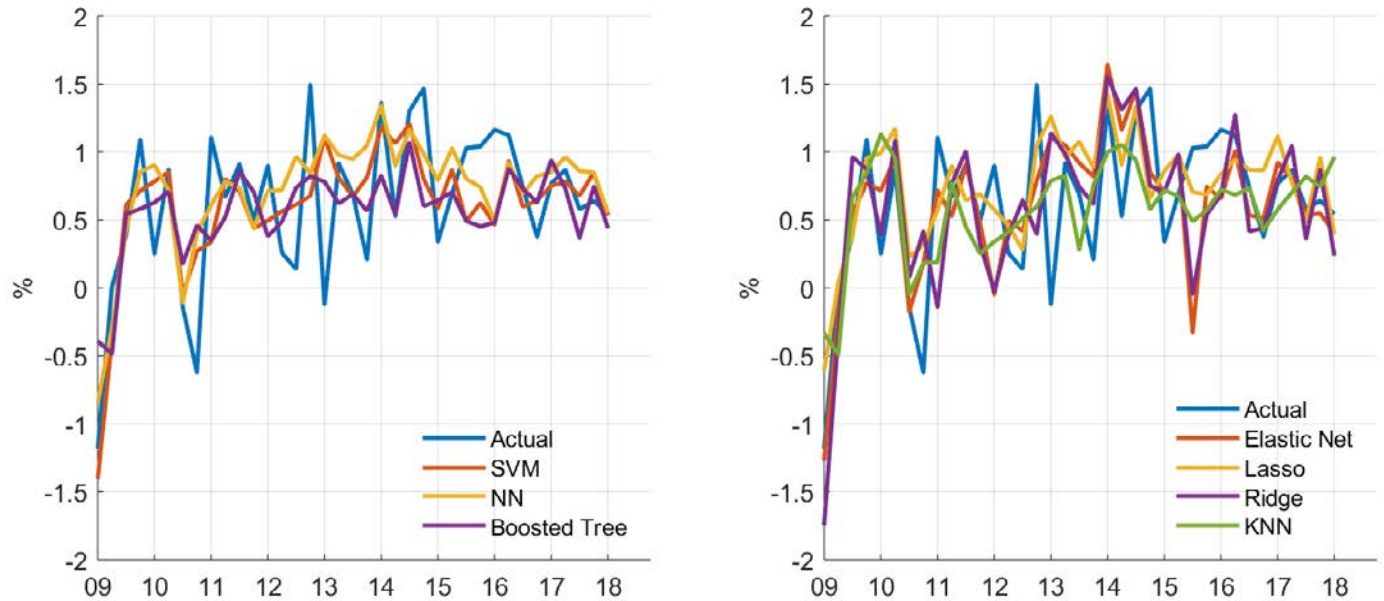
Models	RMSE	RMSE (Rel. to AR)	p-value	MAD	MAD (Rel. to AR)	p-value
SVM	0.445	0.820	0.166	0.338	0.770	0.037
NN	0.446	0.821	0.182	0.336	0.766	0.041
Lasso	0.454	0.836	0.206	0.348	0.793	0.061
BT	0.469	0.865	0.153	0.386	0.880	0.107
ENET	0.488	0.899	0.526	0.349	0.796	0.140
KNN	0.491	0.905	0.349	0.405	0.923	0.402
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

⁶ These models were estimated using a system with 64 GB of RAM, a Xeon E5-1660, 8 core, 3.20 GHz CPU, with a Windows 10 operating system. It is important to note that increasing the number of training iterations for optimising the hyperparameters would increase these computation times significantly. The computation time of these models can be further addressed by making use of cloud computing services.

model using the methodology outlined in subsection 2.2. 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).

Figure 2: Real-time nowcasts of quarterly GDP growth



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’ve 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 weighting	0.437	0.805	0.324	0.738*
Least-squares weighting	0.427	0.786	0.317	0.722*
Inverted MSE weighting	0.434	0.799	0.324	0.738*
Inverted MSE ranks weighting	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 weighting: assigning equally weighted forecasts; Least-squares weighting: 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 weighting: 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 Romer and Romer (2008) and estimate the following regression equation:

$$y_t = c + \alpha_1 FC + \alpha_2 SS + e_t$$

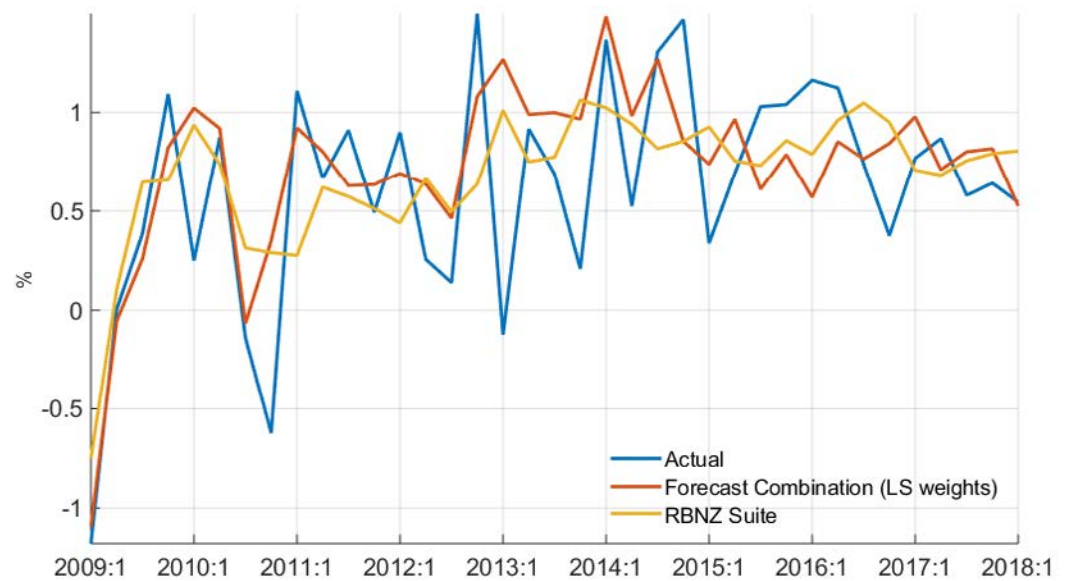
where y_t is the real-time GDP nowcasts, FC is the combined forecasts using the least-squares weights method outlined above, SS is the real-time GDP nowcasts obtained from the combination of models in the RBNZ statistical suite and e_t is the residual term. The results presented in Table 3 suggest that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

suggest that ML models would be a useful addition to the RBNZ's current suite of forecasting models.

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

Coefficients	Estimated value	t-Statistic
α_1	0.94	4.11
α_2	-0.11	-0.31
C	0.02	0.13
R^2	0.46	

Figure 3: 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 multiple vintages of 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 various ML forecasts. Our results thus recommend the use of ML algorithms as a useful 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.
- Banbura, M., Giannone, D., Modugno, M. & Reichlin, L. (2013), 'Nowcasting 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.
- Diebold, F. X. & Mariano, R. S. (1995), 'Comparing Predictive Accuracy', *Journal of Business & Economic Statistics* 13(3), 253–263.
- Eickmeier, Sandra & Ng, Tim. (2011), 'Forecasting national activity using lots of international predictors: An application to New Zealand,' *International Journal of Forecasting*, Elsevier, vol. 27(2), pages 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.
- 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*, *International Journal of Central Banking*, vol. 2(2), May.
- 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.

IFC – Bank Indonesia International Workshop and Seminar on *“Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data”*

Bali, Indonesia, 23-26 July 2018

Nowcasting New Zealand GDP using machine learning algorithms¹

Adam Richardson, Thomas van Florenstein Mulder, Tugrul Vehbi,
Reserve Bank of New Zealand

¹ This presentation was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Nowcasting New Zealand GDP using Machine Learning algorithms

*Adam Richardson, Thomas Mulder, Tugrul
Vehbi*

Motivation



- Applications using machine learning are transforming our daily life
- ML methods have recently been proposed as alternatives to time-series regression models typically used by central banks
- We want to know how useful are these models for nowcasting New Zealand GDP?

The evidence is mixed...



- Makridakis et al. (2018)
 - out-of-sample forecasting accuracy of ML models is lower than that of more traditional statistical methods
- Chakraborty and Joseph (2017)
 - ML models generally outperform traditional modelling approaches in prediction tasks
- Matheson (2006)
 - Factor Model Forecasts for New Zealand.
- Eickmeier, & Ng (2011)
 - Forecasting national activity using lots of international predictors: An application to New Zealand.

Our approach



- We train a range of popular machine learning algorithms over an expanding window
- Using real-time data, we replicate an actual nowcasting situation starting from 2009 Q1 through to 2018 Q1
- We compare the predictive accuracy of these nowcasts with other benchmarks
 - AR(1), a factor model, a large Bayesian VAR and a suite of statistical models used at the Reserve Bank of New Zealand.

The models



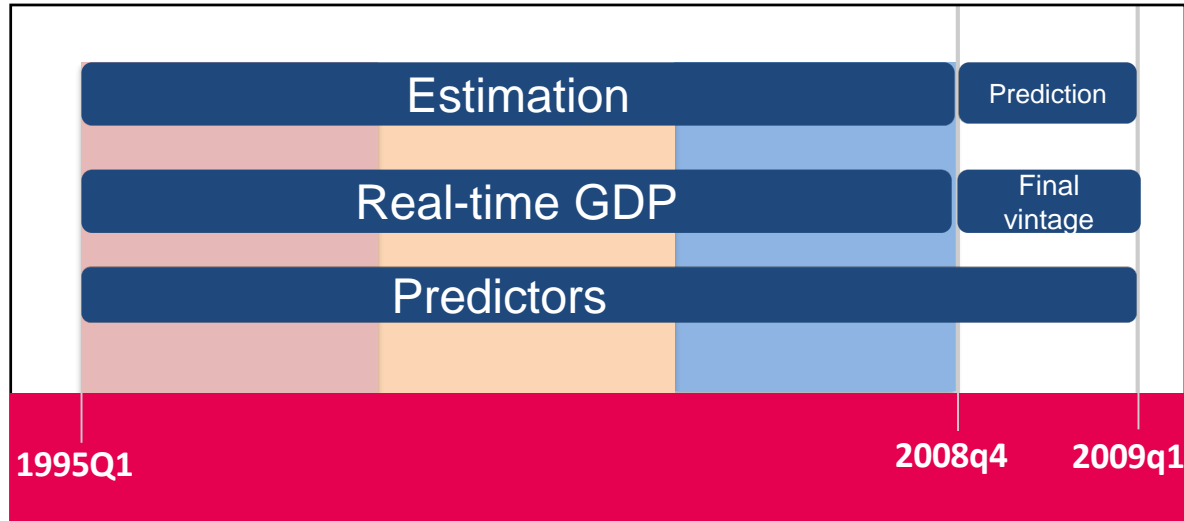
- **Machine Learning models:**
 - K-Nearest Neighbour (KNN)
 - Boosted Trees (BT)
 - Lasso, Ridge and Elastic Net (ENET)
 - Support Vector Machines (SVM)
 - Neural Network (NN)
- **Statistical benchmarks**
 - AR model
 - Factor model (FM)
 - Large BVAR
 - RBNZ statistical suite

The data



- Real-time data
- 2009q1 to 2018q1 (37 quarters)
- 1995q1 – current
- Quarterly, stationary, non-NaN
- About 550 macro indicators

Estimation example (2009q2)



Forecast evaluation

Root Mean Squared Error

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

Mean Absolute Deviation

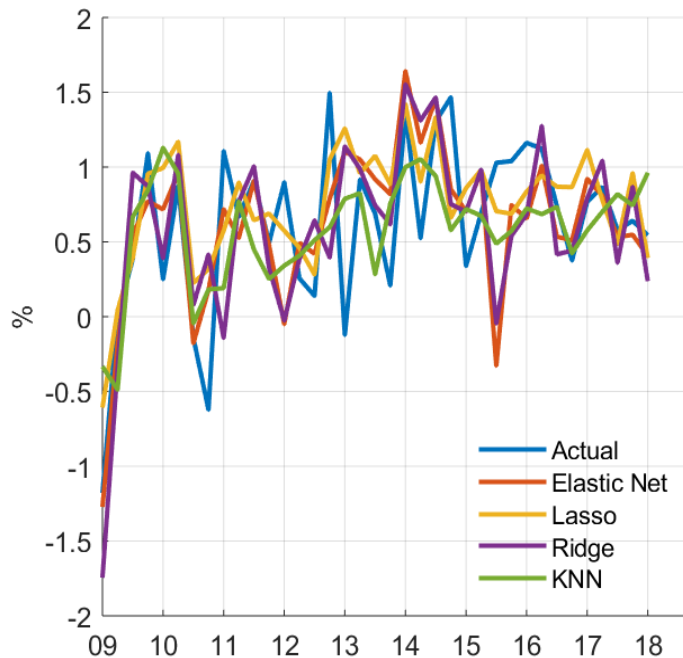
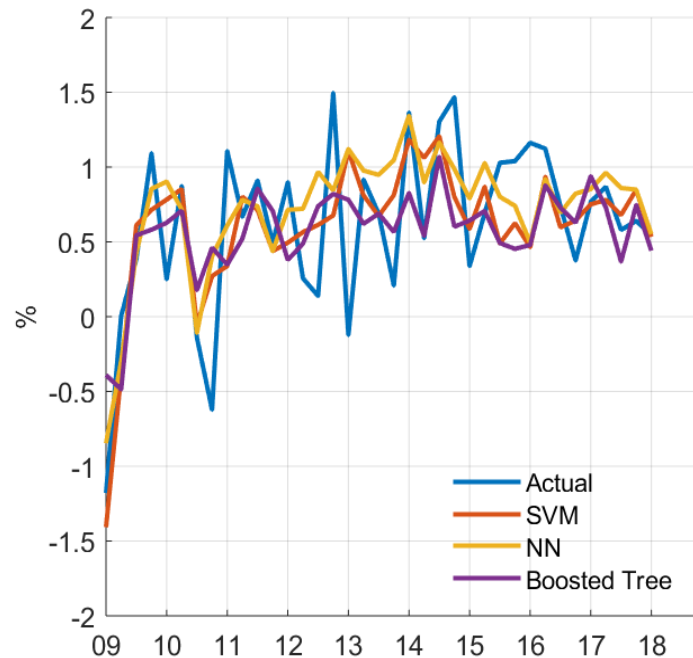
$$MAD = \frac{1}{T} \sum_{t=1}^T |y_t - \hat{y}_t|$$

Results



	RMSE	RMSE (rel to AR)	MAD	MAD (rel to AR)
SVM	0.445	0.820	0.338	0.770*
NN	0.446	0.821	0.336	0.766*
Lasso	0.454	0.836	0.348	0.793*
BT	0.469	0.865	0.386	0.880*
ENET	0.488	0.899	0.349	0.796
KNN	0.491	0.905	0.405	0.923
Ridge	0.565	1.040	0.427	0.972
FM	0.517	0.951	0.379	0.864
BVAR	0.608	1.119	0.450	1.025
RBNZ SS	0.471	0.867	0.384	0.874
AR	0.543	-	0.439	-

Real-time nowcasts of quarterly GDP growth (2009-2018)



Forecast combination



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

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



We follow the approach by Romer and Romer (2008) and estimate the following regression equation:

$$y_t = c + \alpha_1 FC + \alpha_2 SS + e_t$$

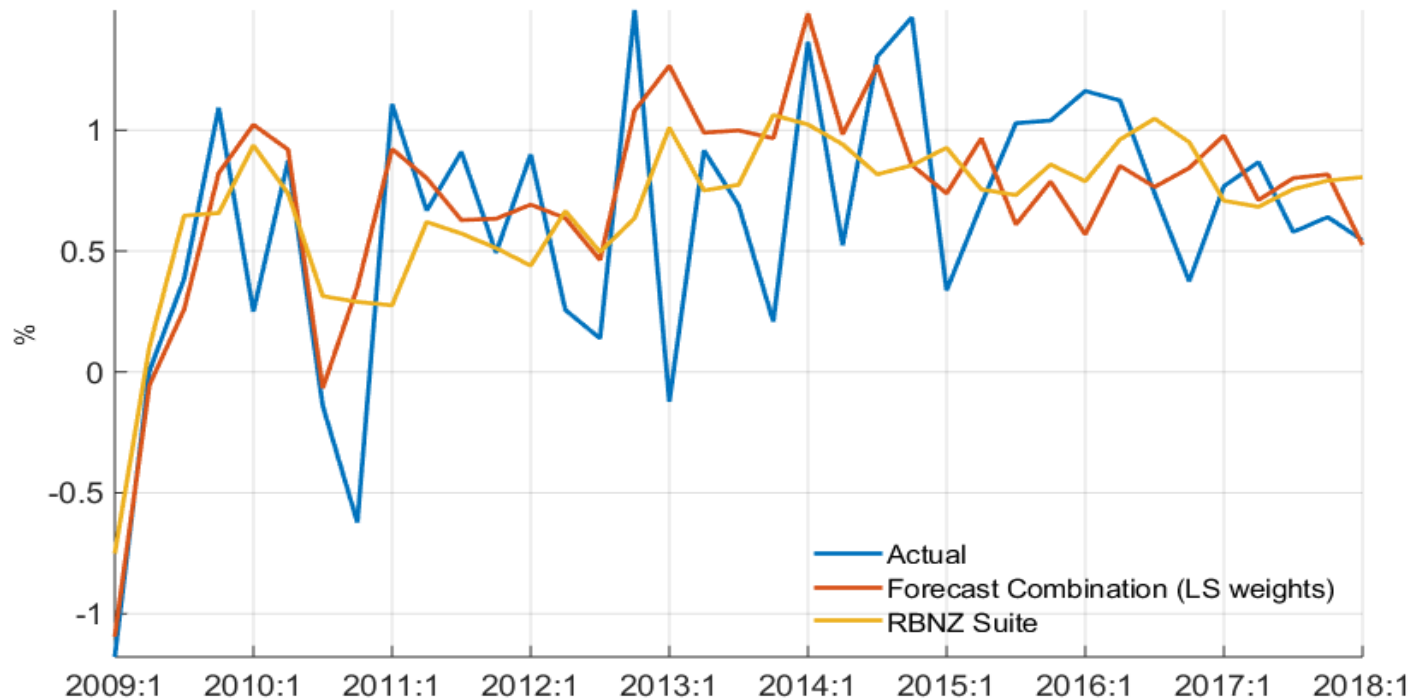
FC: combined nowcast using LS weights

SS: RBNZ statistical suite nowcast

ML forecast combination adds significant value

Coefficients	Estimate	t-Statistic
α_1	0.94	4.11
α_2	-0.11	-0.31
C	0.02	0.13
R^2	0.46	

Forecast combination vs RBNZ suite



Conclusion



- Machine learning algorithms outperform the statistical benchmarks
- Combining the nowcasts of the ML models leads to further improvements in performance.
- ML models would be a useful addition to the RBNZ's current suite of forecasting models.