Nowcasting: Towards Real-time GDP Prediction

Teo Susnjak*

School of Natural and Computational Sciences, College of Sciences, Massey University

Christoph Schumacher*

Massey Business School, School of Economics and Finance, Massey University

Abstract

The large amount of rich and relevant data that are available now and the technological advances in machine learning make it possible to create a real-time GDP prediction system. This research presents a world-first, live GDP nowcasting system (GDPLive[1]) which estimates the economic activity at the national, regional and industry-sector levels on a daily basis for the New Zealand context. We use consumption, import/export and cargo shipments data on a daily basis and currently available macroeconomic indicators and apply state-of-the-art machine learning algorithms to derive daily estimates. GDPLive makes predictions for the current quarterly percent GDP change at the national level, as well as forecasted current quarter-end and next quarter-end GDP rates. At the regional level and industry-sector level, current estimates of GDP rates are reported at an annual and quarterly percent change respectively.

This research presents experimental results of the accuracy of the first-phase GDPLive models on historic predictions. Overall, the results are encouraging. The models have indicated their ability to learn from historic patterns and broadly generalise. A degree of uncertainty is present in the predictions and this will remain, though it is anticipated that with future iterations of GDPLive, this will decrease.

Keywords: GDP, Nowcasting, Forecasting, Machine-learning, Macroeconomics, Analytics, GDPLive

1. Introduction

Gross national product (GDP) shows the market value of all finished goods and services produced nationally within that year. It is the universal standard measure for determining a nations economic growth and for comparing economies internationally. The underlying assumption is that the value of production, expenditure or income of an economy adequately reflects performance and the state of society. A high GDP suggests a healthy economy, which in turn tends to signal that unemployment is low and wages are on the rise as labour and production demands also increase. Investors also are more likely to invest when the economy is on the rise, leading to higher levels of prosperity.

Calculating GDP is a complex task and it is not surprising that governments release national and industry figures with a two to three month delay and regional figures take up to a year before being announced. In addition to the delay, GDP values are frequently revised afterwards.

To speed up the release time, the Office for National Statistics in the UK has recently started to produce monthly GDP estimates which are released with a five to six week delay. The Federal Reserve Bank of Atlanta is providing estimated values of real GDP based on available data for the running quarter and releases figures every week. Both of these initiatives provide GDP figures that are more up-to-date, and represent the desire and need for a real-time GDP forecasting tool.

Given the availability of high frequency data and recent technological advances, it seems surprising that when GDP figures are released we still do not know how the economy is performing right now, but only how it was performing several months ago. This frustration has sparked increased research in the field of real-time forecasting of macroeconomic indicators. Our research contributes to this literature by presenting a world-first, live GDP nowcasting system (GDPLive[2]) which estimates the economic activity at the national, regional and industry-sector levels on a daily basis for New Zealand. To achieve this we use consumption, import/export and cargo shipments data on a daily basis as well as currently available macroeconomic indicators and apply state-of-the-art machine learning algorithms. Overall, the accuracy of our estimates is encouraging. While our research is still in an experimental phase, our models have shown an ability to learn from historic patterns and broadly generalise.

This paper is organised as follows. We first provide some background information on predictions using proxy variables and briefly review relevant literature. Chapter 3 then outlines

*Corresponding author

Email addresses: `t.susnjak@massey.ac.nz` (Teo Susnjak), `c.schumacher@massey.ac.nz` (Christoph Schumacher)

[2]www.gdplive.net

the challenge and process of building a real-time GDP nowcasting tool and introduces our solution architecture. Chapter 4 describes our methodology, the data set and the algorithms used. In chapter 5, we present our results and review national, regional and industry forecasts. After briefly discussing our findings in chapter 6, we outline limitations (chapter 7), discuss future work (chapter 8) and make concluding remarks (chapter 9).

2. Background

2.1. Predictions using proxy variables

In 2012, Vernon, Storeygard and Weil published a paper in the American Economic Review[1] linking night lights to income growth. The authors introduced a statistical framework that uses satellite data of lights that were turned on at night to augment official growth figures. The underlying premise is that increasing income has a positive effect on light consumption at night as people try to extend daylight hours.

Using consumption data as a proxy for economic growth is not new. According to the DG Environment News Alert Service[2], a positive and causal link exists between the amount of solid waste produced by a country and its GDP. There is also anecdotal evidence that Li Keqiang, the former Party Committee Secretary of Liaoning, China, used the railway cargo volume, electricity consumption and bank loan disbursements as a measure of GDP[3].

More recently, ANZ Bank New Zealand has created the Truckometer, a set of economic indicators based on traffic volume[4]. The underlying assumption is that traffic flows are a good proxy for economic activity. The Truckometer splits traffic flows into heavy (commercial) and light (private) traffic to map quarterly GDP growth. Nowcast, a Japanese financial research and technology company is currently using satellite images of night lights to estimate inflation data and is in the process of using the same data to forecast GDP of economies including the US, China, India and Taiwan[5].

Our research is in line with the outlined work as we use consumption data as a proxy for economic activity.

2.2. Existing Literature

There has been an increasing amount of research on using large data sets to measure or predict macroeconomic indicators. The theoretical literature is adapting existing or developing new statistical and econometric methods for the analysis of large data sets with a large number of explanatory variables. Researchers are working on techniques that resolve issues such as high fluctuations commonly present in high frequency data or seasonality effects of time series data. Further research in the field addresses issues of variable selection and dimensional reduction as well as investigating nowcasting techniques. Nowcasting is the production of timely short-term analysis and forecasting. Kapetanios and Papailias (2018) provide a detailed and current methodological review of the outlined research areas[6].

The empirical literature has focused on using easily available large datasets (mainly Google Trends) and simple econometric techniques to derive macroeconomic forecasts[6]. The key challenge for researchers is finding available relevant data sets. Credit card and financial transaction data, direct consumption information or communication data is often difficult to obtain due to privacy issues. Again, for a review of current literature see Kapetanios and Papailias (2018)[6].

3. GDP Real-time Prediction

3.1. Challenges

Building a real-time GDP nowcasting tool is non-trivial. The primary reason why such a tool does not yet exist is due to numerous challenges as well as difficult design decisions which need to be made. Key challenges are listed below. Some are general challenges encountered in any real-world machine learning problem, while others are specific to, or are amplified by, the intrinsic nature of this project.

- **Input Data Acquisition**: The utility of any GDP nowcasting technology is predicated on the quality of data inputs. GDP is a complex phenomenon. Estimating its value at any given point in time requires data that is ideally causally linked with it and, in the absence of this proxy data, that shares some correlation with it. In order to satisfy this, it becomes important to seek input data from very diverse sources which presents challenges. We do currently live in a world that is saturated by ever increasing volumes and variety of data. Some of the data is held by public institutions while others are in the corporate sector. Rich and useful data is often in the possession of companies who rightly regard it as a valuable and sensitive resource to protect. This ensures that there are many corporate hurdles to overcome, assurances to be made and compliances to adhere to from the perspective of governing privacy laws before access to data can be granted.

- **Input Data Utility**: Having achieved access to data from a third party source does not guarantee that effective models for predictive purposes can be generated using it. Not all data is of equal value and not all data is conducive to producing effective prediction models. Data acquisition is followed by a lengthy experimental process in order to ascertain the viability of its usefulness for predictive purposes. The utility of particular data cannot be ascertained *a priori* and this process represents large time investments which may lead to dead-ends.

- **Input Data History**: Being in possession of valuable data which is correlated and predictive of a target concept is not sufficient for building machine learning models unless the data covers a sufficiently large span of history. Machine learning algorithms rely on learning patterns from historic data in order to make predictions on current and future events. Generally, the larger the time-span that the data covers, the more accurate the predictive models will be. It is not straightforward to satisfy both the constraints of attaining the right data, as well as data that has been captured for a long enough period to facilitate the generation of accurate models.

- **Input Data Timeliness**: Once input data is acquired, its efficacy established, and enough of it collected, it is not certain that the given data can be supplied in a timely fashion for making the actual predictions on an ongoing basis. For example, it can be established that immigration rates affect GDP. This can be shown for instance by matching first quarter immigration rates with the first quarter GDP and concluding that immigration at the reported time is to some degree predictive of GDP. However, immigration rates for a given period of time are only known 2-3 months after the fact. Therefore, immigration for a given period cannot be used as a predictor of GDP for that point in time. If immigration data were to be used, the time lag in its reporting must be taken into account and a correlation between immigration rates at a given point in time and its effect on GDP 2-3 months after the fact must be established. This is not a given and may invalidate a number of useful data sources due to the inability to capture and provide them in a timely fashion.

- **Data Granularity**: The term 'real-time' is imprecise and carries different connotations across various application domains. For some, this might mean a time span of seconds, while in others days. In the context of GDP, once data from various sources is collected, a design decision must be made as to what real-time means and therefore upon which level of granularity ongoing predictions will be made. Will GDP be predicted on a monthly, weekly or daily basis? Standard machine learning algorithms require that all input data is in the same level of granularity. The technical challenge arises when the input data does not meet this requirement. The larger the variety of data sources, the smaller the probability that the data will be available at identical levels of granularity.

- **Data Interpolation**: The data granularity problem essentially means that the dataset has large amounts of missing data. For example, if CPI were reported quarterly then only one data point would exist per quarter. If a design decision is made to operate at a daily level of granularity, approximately 90 new input data points would need to be created between each end-of-quarter CPI figure. Technically, the missing data problem can be addressed by creating new data through interpolation methods. The difficulty in this process lies in the fact that there are multiple ways of generating the missing data and each method is associated with a set of assumptions which may or may not hold, thus possibly introducing bias into the models and compromising their eventual accuracy. Furthermore, for any level of granularity higher than quarterly, the GDP data itself will also require interpolation since the granularity of input data must match that of the GDP data, and GDP is only reported quarterly. The interpolation of the 'dependent variable' represents additional risks of compromise to the accuracy of the final models.

- **Variable Input Data Time-span**: Machine learning algorithms also require various data inputs to cover the same lengths of time, otherwise, missing data will most likely negatively affect the results. For instance, if historic immigration input data is captured which goes back to 1995 and consumer spending input data is collected which goes back to 2010, then both data sources can be used in combination only as far back as 2010. Immigration data prior to 2010 will need to be disregarded. This means that if all data sources are used in unison to train machine learning models, the utility of the entire dataset will only be as strong as the weakest link from the perspective of historic data coverage.

- **Variable Input Data Quality**: An important factor to consider is how to utilise input data whose quality degrades the further back in time it is recorded. This issue can for instance be encountered if using traffic volume data. Some geographic regions have much more comprehensive data capturing now than they did 10 years ago. The pertinent question arises if all the data stretching back 10 years should be used in machine learning. Indeed, if input data is sourced from private companies which have seen significant growth in their business, the same scenario will be experienced in respect to the phenomenon they measure. Careful investigation of input data is in these cases necessary and trade-off decisions need to be made as to how much of the data is utilised.

- **Uncertainty of Value of Dependent Variable**: The calculation of GDP is an immensely complicated undertaking. There are numerous assumptions and subjective decisions that factor in its calculation. This means that the official GDP values are not strictly speaking absolutely accurate at all times and are also subject to revision. The uncertainty surrounding the actual values of GDP presents difficulties for machine learning. In order to extract a pattern that generalises into the future, a mapping function between the data inputs and the dependent variable must be induced. If the dependent variable is incorrect, unreliable models will result.

- **Input Data Feature Engineering**: Input data in its raw form is seldom suitable for the purposes of machine learning and the very success of machine learning often relies more on the quality of the extracted features than it does on the algorithms themselves. The task of engineering meaningful and descriptive features from raw data requires domain knowledge and manual intervention. An additional complication with the process of feature engineering is the fact that it is possible to generate large numbers of features. This presents serious problems (curse of dimensionality) if the dataset used for machine learning has a high ratio of features with respect to the total number of data points. The higher this ratio becomes, the greater the probability that the machine learning algorithms will uncover spurious patterns which have no relation to reality and will thus exhibit low accuracy at deployment. Strategies to mitigate this will need to be employed and only useful features selected.

- **Machine Learning Algorithm Selection**: There is no shortage of algorithms to choose from and this is both a positive as well as potentially a bewildering challenge. There is no 'best' algorithm that is guaranteed to outperform others across different problem domains. The danger in selecting one algorithm on a complex time-series problem like this lies in the very real possibility that it might work well for a given time range, but subsequently begin to underperform as unforeseen changes in data or underlying economic conditions begin to occur. What is more, sophisticated algorithms are accompanied by numerous tunable parameters. Small changes in these as well as perturbations in the training data can result in vastly different models.

- **Forecast Horizon**: A time-series problem domain such as GDP prediction embodies some additional challenges compared to other machine learning domains. A prediction that is made for one day needs to depend on the value of the prediction that was made on the previous day. A model can be tuned so that it depends on more than one previous prediction. Whichever way the model is configured, the fact remains that any future prediction is influenced by previous predictions. If each prediction is a small way off, that means that the errors accumulate. The further out into the horizon the predictions stretch out beyond the last known 'true value', the greater the probability that the forecast will fall into significant error. The nature of the timing of official GDP figures means that the predictions stretch almost six months into the unknown. For instance, in mid-December, the last known GDP rates cover the period up to June of that year. If predictions are made for economic activity in December, the prediction horizon therefore is almost six months from the last reported GDP rates. The challenge is even greater for predicting regional economic activity since the reporting delay for these figures is in excess of 12 months.

The above challenges present many potential sources of bias, and the difficulties in capturing all the variables that affect GDP. Together, they all make the task of producing an accurate GDP nowcasting tool a formidable undertaking.

3.2. Solution Architecture

We describe here the strategies and design decisions we made in order to overcome the challenges highlighted in Section 3.1.

Data Solutions

In order to guarantee that the input data for our models is rich and representative of economic activity, we have resorted to complement up-to-date industry data with key macroeconomic data. In the expenditure approach of calculating GDP, the largest component is household consumption. This is captured by credit card and financial transactions. Thus, we have secured data provision from Paymark Ltd., a company that captures around 75 percent of all credit and debit/eftpost card transactions made from around 140,000 terminals in New Zealand.

Export and import data is another key component of the GDP calculation. In order to capture a portion of this component, we secured data from Portconnect Ltd., New Zealand's comprehensive online cargo management system. Portconnect tracks all shipping container movements from Ports of Auckland and Port of Tauranga, the two largest ports in New Zealand and covers some two thirds of all shipments in the country. The data also includes figures from the Auckland satellite ports at Wiri and Metroport.

Movement of goods across New Zealand can also be used as a proxy of economic activity. Higher volumes suggest more economic activity. To capture this effect, we use data from KiwiRail, New Zealand's largest rail transport operator, incorporating business units of KiwiRail Freight, The Great Journeys of New Zealand and Interislander. In addition, heavy and light traffic volume data from across the country, collected by the New Zealand Transport Agency (NZTA) also forms a part of the overall solution.

Apart from NZTA, the above data sources constitute dynamic inputs into our system. The data from our industry partners is collected from their servers on a daily basis and satisfy the need for data timeliness, and a high level of granularity. These data sources are also highly correlated with economic activity and satisfy the utility constraints. The historic coverage of the data sources does vary, and the manner in which this is handled is described below in the machine learning strategies.

We also use several macroeconomic indicators that correlate with economic growth, namely household disposable income, employment numbers, immigration figures, housing stock and CPI. Higher levels of disposable income encourage either spending or saving which has a positive effect on GDP. The same can be said about employment levels and immigration figures. High levels of employment and immigration increase overall household consumption and therefore positively impact on GDP. Increasing housing stock values does not directly impact on GDP as only the value of new houses is captured by the annual GDP figure. However, the total value of a country's housing stock can be seen as a proxy for wealth which is linked to GDP. Finally, we use the CPI value not only to adjust our industry values to real prices but also as a variable of our model. Although the relationship between CPI and economic growth can, under certain circumstance, be ambiguous, a comprehensive empirical study of around 100 countries concludes that a negative relationship exists (Barro, 2013)[7].

Design Decisions

The ability to source high-quality input data at regular intervals enabled us to define real-time GDP estimation for our system to be at a daily level of granularity. The consequence of this design decision was that all macroeconomic data as well as the official GDP figures needed to be converted into a daily level of granularity. In order to do this, missing data situated between known data points were interpolated.

The decision to carry out predictions at a daily level was made in parallel with the decision on what should actually be predicted. There were two options. One option was to predict what the end-of-current-quarter GDP value is likely to be, which is what most financial institutions opt to forecast. This could be updated on a daily basis. The other option, and genuinely unique in the world to the best knowledge of the authors, was to predict what the GDP rate is right now. We decided to do both and in addition, we decided build to models that predict the end-of-next-quarter GDP for national rates too. For national-level forecasts three models are trained. The first is trained to predict what the quarterly percent GDP change is for each day after the last known official GDP release. The next model is trained using the outputs of the first model in order to continue the predictions and use the latest input data to predict the end-of-current-quarter GDP. Lastly, a model is trained that takes the previous prediction outputs and is trained with the most recent input data in order to continue the predictions through to the end-of-next-quarter GDP. These models are retrained with new input data each day and the long-term forecasts updated accordingly.

Lastly, given the richness of the data, the decision was made to make predictions for all regions in NZ as well as for all industry sectors. Both regional and industry-sector specific economic activity would be predicted based on the annual percent GDP change for a given day. At this stage, end-of-current-quarter and end-of-next-quarter GDP predictions are not implemented, but are a possibility for subsequent updates of GDPLive.

Machine Learning Strategies

Ensemble-based machine learning in combination with stochastic processes at the level of algorithm selection, algorithms parameter selection, feature and feature size selection, date range of the training data as well as the number of lag periods, were all part of an overarching strategy to mitigate the problems of algorithm bias resulting from overfitting.

Each machine learning algorithm has its own sets of assumptions and underlying biases. It is not always clear which algorithm is necessarily the best for a given problem at hand. The literature is clear that no algorithm is superior to all others across a full spectrum of possible combinations of problems on a given domain. Many algorithms are also quite unstable, in that they produce considerably different results as a response to small variations in their tunable parameters and the underlying data.

Ensemble-based machine learning solutions generate multiple models and combine them together to achieve a final prediction. A key component of this approach is diversity of models. A number of ensemble-based algorithms have a built-in mechanism for ensuring and maximising diversity. Here, we induce diversity between models manually through a stochastic process by randomising the key components of the training process and thus addressing many of the pitfalls highlighted in Section 3.1.

4. Methodology

The key attributes of the data sources used in current models are described here, as well as the details surrounding the types of individual machine learning algorithms used and the ensemble design.

4.1. Data Sources

Paymark:

The structure of the Paymark data is in aggregate form. For each day, region and industry, the total sum of transactions, the total number of transactions and the number of terminals in operation is provided. The transactions were adjusted for inflation based on most recent CPI values provided by the Reserve Bank. The CPI is provided quarterly and the most recent transactions occurring after the release of CPI were thus not adjusted. The data was normalised based on the number of terminals in operation and smoothed using rolling means with different sizes of sliding windows. The Paymark data coverage reaches back to January 2009 and given its daily granularity and the ability to subdivide into multiple categories, it was used for predicting all national, industry and regional forecasts.

Non-Real-time Data:

CPI, Labour Market, Household Debt Levels, Housing Stock and Immigration data are acquired on a quarterly basis. The reporting time of each of these data varies for a given period. Labour Market data is made available approximately one month following the period it describes. CPI, Household Debt Levels and Immigration have a lag of about two months, while Housing Stock has a delay of up to four months.

Once the data is interpolated to a daily basis, each of the these data sources is shifted in proportion to their reporting delay as well as the span of their coverage. For example, this means that Labour Market data is shifted one month forward as well as an additional three months forward to account for its reporting period. The net result is that the machine learning algorithm is required to discover the possible effect of Labour Market data from four months on the GDP response in the present day.

Coverage and Interpolation

The coverage of Paymark data at present stretches as far back as January 2009. Even though data points from non-real-time data sources stretch further back in time, these data sources were truncated and lined up with Paymark's coverage. Data points between each quarterly 'true data point' were interpolated (estimated) using spline interpolation. Spline interpolation was also used to generate dependent variable values for quarterly percent change and annual percent change of GDP values.

4.2. Feature Engineering

Approximately 60 features were engineered and extracted from Paymark data. These features were derived from consumption spending for each national, regional and industry sector datasets. Once a smoothing operator was applied to minimize the effect of noise, features were created based on different permutations of rates of change in the values from one period to the next.

An estimated annual as well as quarterly percent change in CPI and unemployment rates were created for each point in time. Likewise, the estimated unemployment rate at each day was selected as a feature. Two features describing the levels of household debt levels were created. These were, the daily estimated debt servicing as a percent of disposable income as well as the estimated quarterly percent change in debt servicing on any given day. Similarly, the estimated percent change in the annual housing stock value on any given day, as well as the estimated quarterly percent change on the resident population of NZ on a daily basis were generated and integrated into the machine learning.

Correlation analysis was performed against features and their respective dependent variables across national, regional and industry data sets. Based on the outcome of this analysis, highest ranked features were selected as constituents of a total pool of features for random selection during the learning process.

4.3. Machine Learning Algorithms

Three machine learning algorithms are used. These being XGBoost, Light GBM and Kernel Ridge Regression. All three algorithms represent some of the more recent advances in machine learning for continuous valued output data. Both XGBoost and Light GBM belong to the family of gradient boosting algorithms. XGBoost represents 'state-of-the-art' in its category of algorithms with it being cited as one of the most effective and successful algorithms on machine learning competition platforms like Kaggle. Light GBM, while sharing similarities with XGBoost, differs in the manner it calculates the best split and comes with some additional advantages associated with a faster training speed and higher efficiency. Kernel Ridge Regression is situated in a different family of algorithms, being a non-parametric form of Ridge Regression combined with the kernel trick.

All three algorithms possess a range of tunable parameters. For XGBoost and Light GBM, the number of estimators ranged from 100 to 500 and the learning rate from 0.1 to 0.7. For Kernel Ridge Regressor, a liner kernel was used as well as an alpha value which ranged from 0.001 to 2.0.

4.4. Ensemble-based Meta Learning Strategy

Given the bias of each individual algorithm with its associated sets of parameters, and the resulting uncertainty in their generalisability, an ensemble-based strategy was devised in the attempt to mitigate the risks. The strategy relied on combining the predictions of multiple models. The total number of models trained for each component of GDPLive ranged from 15 to 50 models. Uniform weights were assigned to all models; thus, the mean across all model predictions constituted the final prediction. In instances where algorithms with given sets of parameters produced predictions in stark disagreement from others, those were removed from the ensemble. The threshold for removing estimates ranged from 3.5 to 4 standard deviations. The remaining predictions were then used for constructing confidence intervals.

A stochastic process was employed in order to ensure diversity amongst the different models. Diversity is a fundamental element of all ensemble-based solutions. To achieve this, each algorithm was randomly selected as well as its tunable parameters. The feature sets were also randomly selected. The number of features ranged from 3 to a possible set of 25. The number of lag features were also randomly selected. These ranged from 5 to 10 previous predictions. And lastly, the length of the size of the training data was also randomly determined for each model and at every different round of training.

4.5. Model Execution

The full suite of models is re-trained each day as new data is made available[3]. Models from previous training rounds are not persisted. The consequence is that due to the nature of randomness inherent in the devised strategy, a degree of fluctuation in the predicted forecasts can be expected to occur from one day to the next. The fluctuation can be expected to be larger on certain datasets where generic tunable parameters are insufficient and further model optimisation is required.

5. Results

A selection of experimental results are presented here, showing the accuracy of the models on national, regional and industry sector datasets. The models are trained using the parameters and procedures described in the methodology, and represent the same settings of the models in the GDPLive production system as of 17 December 2018.

The purpose of this section is to document the generalisability of the models from the initial release phase, and to serve as a point of reference and benchmarking in the ongoing evolution of the GDPLive models and their improvements in accuracy.

5.1. National GDP Predictions

Figures 1 to 7 depict a sequence of predictions made between quarters from December 2016 through to December 2018. The figures simulate the current capability of the GDPLive models to predict the final value of the quarterly percent change in GDP from one quarter to the next over the specified time range. The shaded areas on the graphs represent the 95% and the 99% range of the confidence intervals. The dashed line represents the predicted quarterly GDP percent change. The simulation continues through to the time of writing (14 December 2018) on which day the most recent input data was available. Also, at the time of writing, the most recently known GDP figures were known for the second quarter at the end of June 2018. Thus, the final set of predictions in Figure 7 almost spans half a year beyond the last known GDP value.

The figures portray a mixture of outcomes. Figures 1, 3 and 4 demonstrate the ability of the models to detect the signal within the data and produce predictions that are within the ideal ballpark of the actual values.

Meanwhile, there are examples of the models detecting elements of the underlying trends in the movements of the GDP; however, they have in these instances either overshot (Figure 2)

[3]The data from our industry partners that covers their previous day's trading is made available around 11am each morning. The models execute soon after that. The aim is to make refreshed forecasts available on the website by 1pm each day.

Figure 1: National quarterly percent GDP change showing predictions from end of December 2016

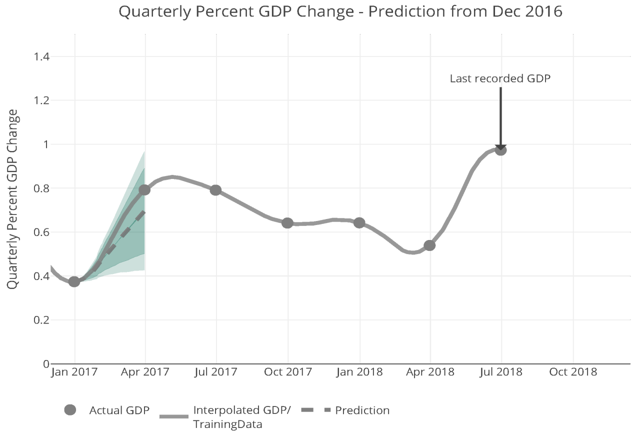


Figure 2: National quarterly percent GDP change showing predictions from end of March 2017

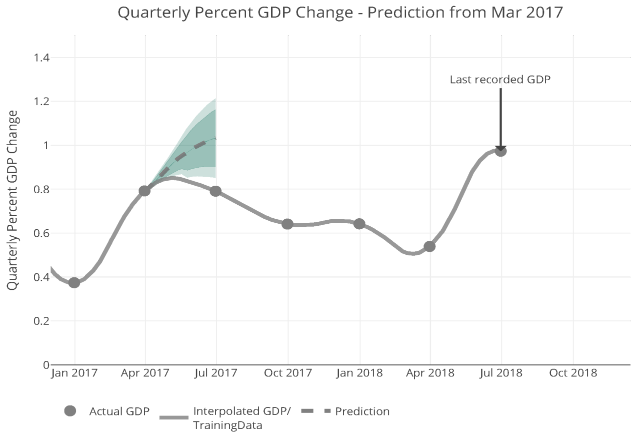


Figure 3: National quarterly percent GDP change showing predictions from end of June 2017

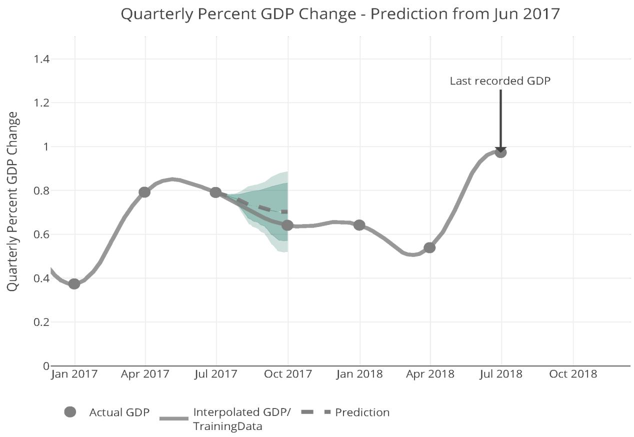


Figure 4: National quarterly percent GDP change showing predictions from end of September 2017

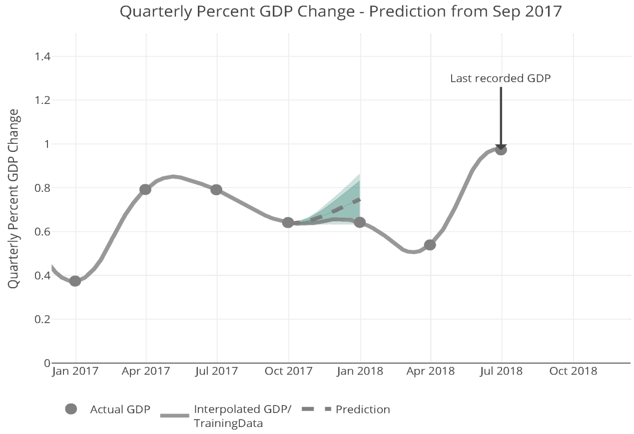
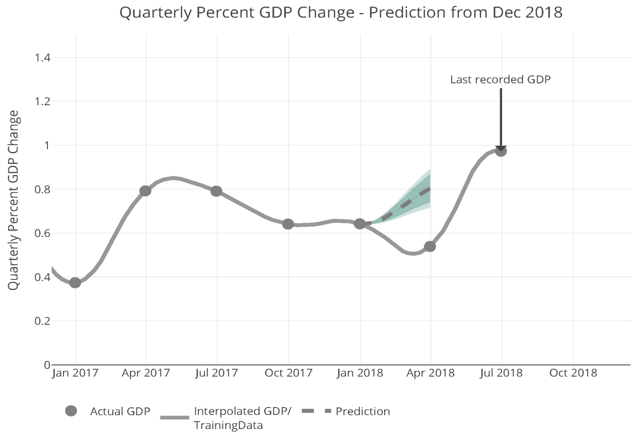


Figure 5: National quarterly percent GDP change showing predictions from end of December 2017



or undershot (Figure 6) the true patterns. Finally, an example of a significant failure to detect the true movement of GDP is depicted in Figure 5.

5.2. *Regional GDP Predictions*

This section considers the prediction of GDP across NZ regions. There are in total 16 regions. The predictions for eight of these are presented here and cover the majority of main centres. Regional GDP predictions differ from national predictions in the previous section in that the annual percent change is estimated instead of quarterly percent change. This is due to the fact that the regional estimates are only reported once a year, at the end of each financial year. The necessity of having to estimate GDP 12 months or more in the future presents additional difficulties due to the accumulation of errors and thus regional estimates entail considerable uncertainty. The range of prediction for each of the graphs covers 1 April 2016 through to the last known regional GDP rates at the time of writing, 31 March 2017.

Despite the challenge, there is evidence that the models possess a promising degree of capability to generalise across mul-

Figure 6: National quarterly percent GDP change showing predictions from end of March 2018

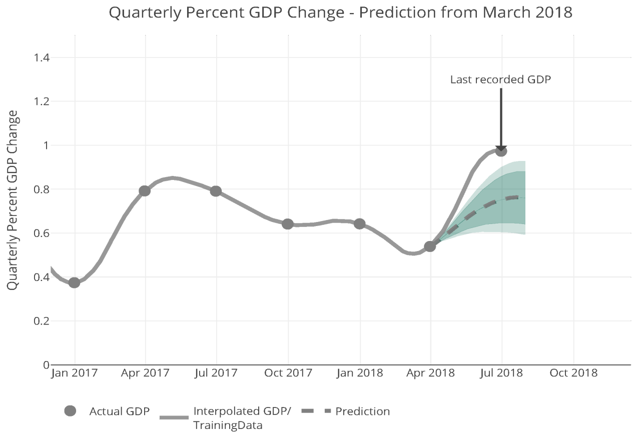


Figure 7: National quarterly percent GDP change showing predictions from end of June 2018

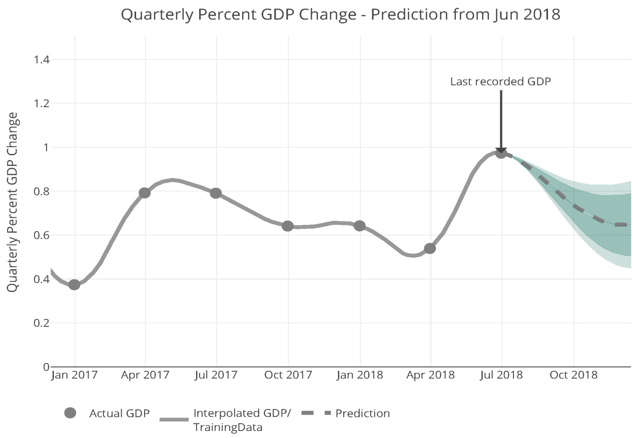


Figure 8: Annual percent GDP change prediction for Canterbury from end of March 2016.

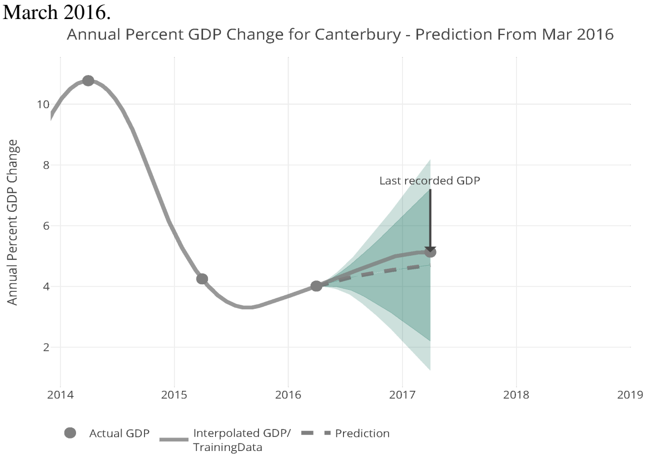


Figure 9: Annual percent GDP change prediction for Wellington from end of March 2016.

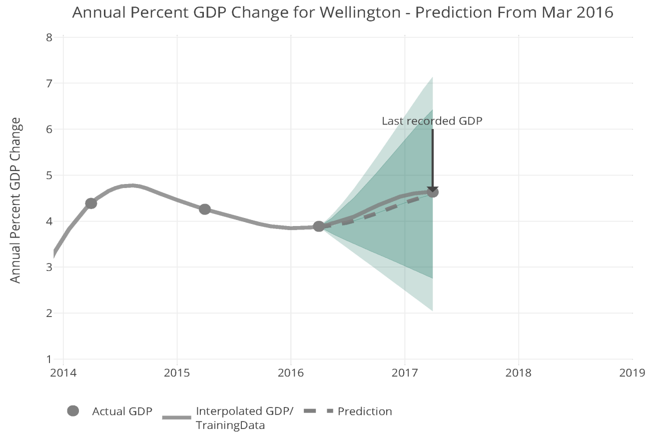


Figure 10: Annual percent GDP change prediction for Waikato from end of March 2016.

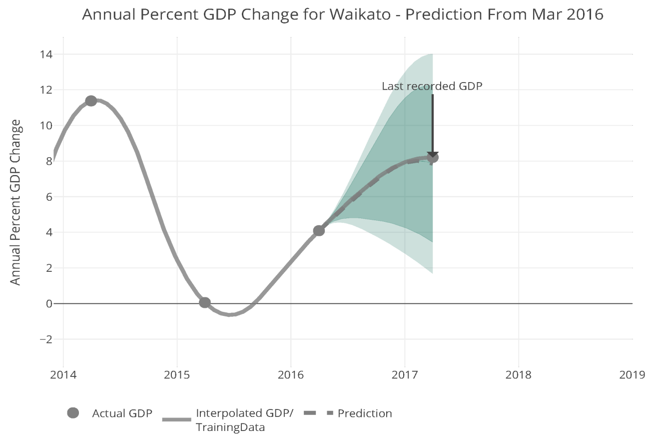


Figure 11: Annual percent GDP change prediction for Taranaki from end of March 2016.

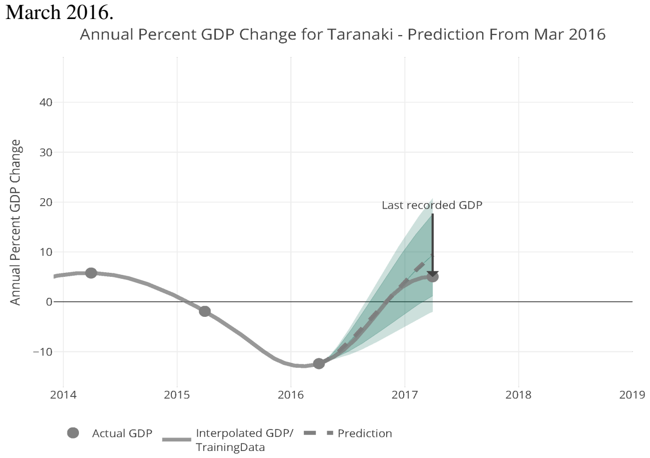


Figure 12: Annual percent GDP change prediction for Manawatu-Wanganui from end of March 2016.

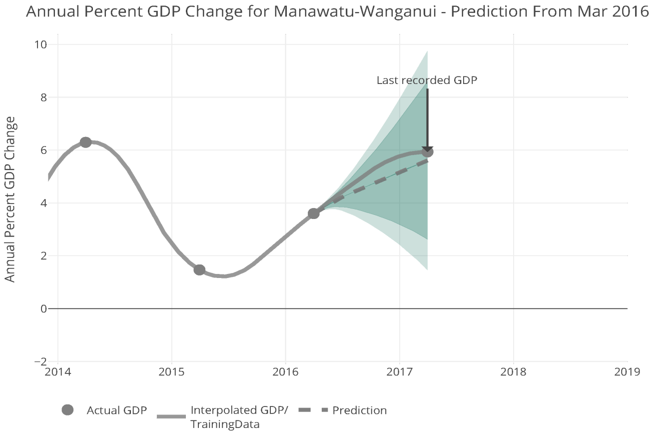
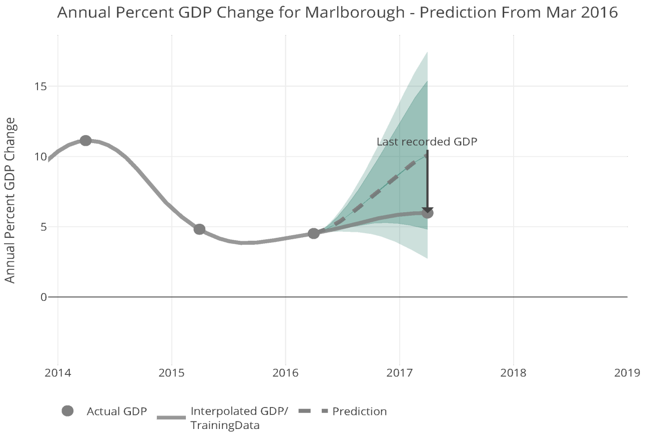


Figure 13: Annual percent GDP change prediction for Marlborough from end of March 2016.



tiple regions. Figures 8 to 12 cover the regions of Canterbury, Wellington, Waikato, Taranaki and Manawatu-Wanganui. The predictions for these regions were in line with the actual reported GDP values in March 2017.

Poorer generalisations can be observed on Marlborough (Figure 13) and Auckland (Figure 15), with a sizable overestimate in GDP growth.

The opposite is depicted in the Figure 14 where a non-existent slowdown in economic growth is detected in the underlying data, with the final prediction ending in negative territory for this region.

5.3. *Industry GDP Predictions*

The Ministry of Business Innovation and Enterprise categorises business into 16 sectors in three broad categories: Primary Industries, Goods-Producing Industries and Service Industries. Due to the nature of our current input data, Mining and Agriculture, forestry & fishing, which constitute the Primary Industries Sector are not tracked by GDPlive. Out of the 14 remaining industry sectors, this section highlights the predictive accuracy on ten sectors.

Figure 14: Annual percent GDP change prediction for West Coast from end of March 2016.

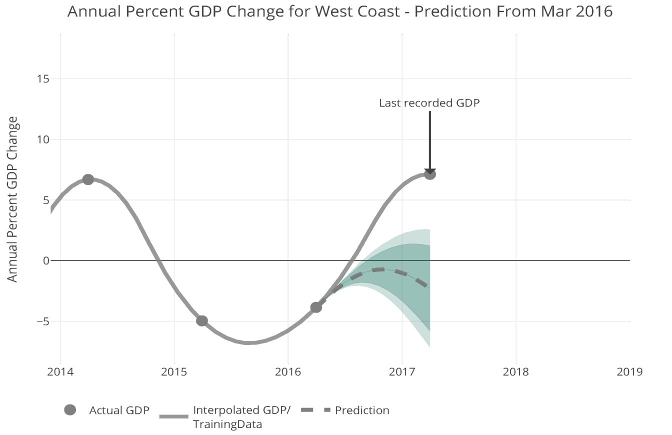


Figure 15: Annual percent GDP change prediction for Auckland from end of March 2016.

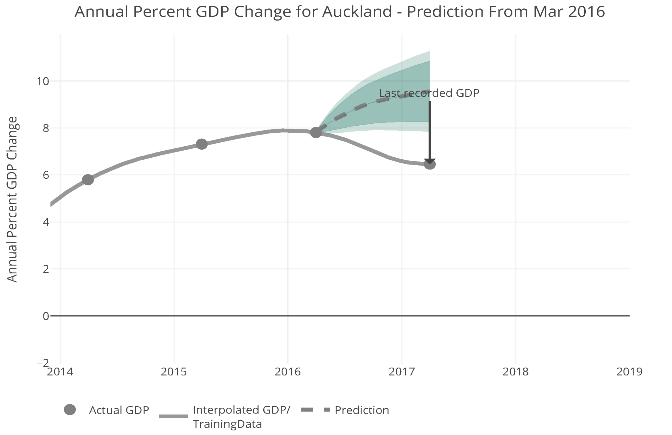


Figure 16: Annual percent change prediction for the Rental hiring and real estate services, from end of March 2018.

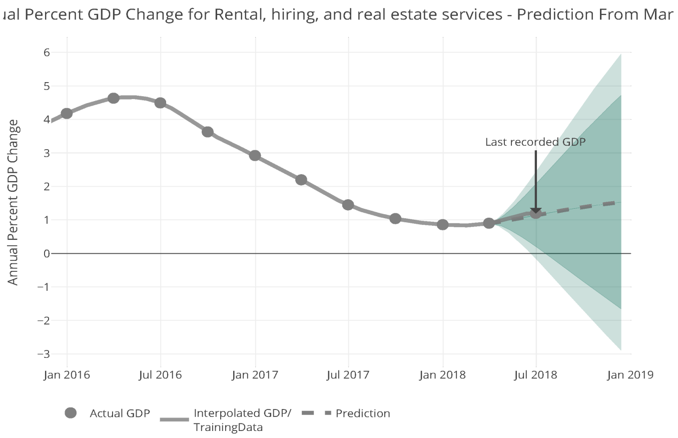


Figure 17: Annual percent change prediction for Construction, from end of March 2018.

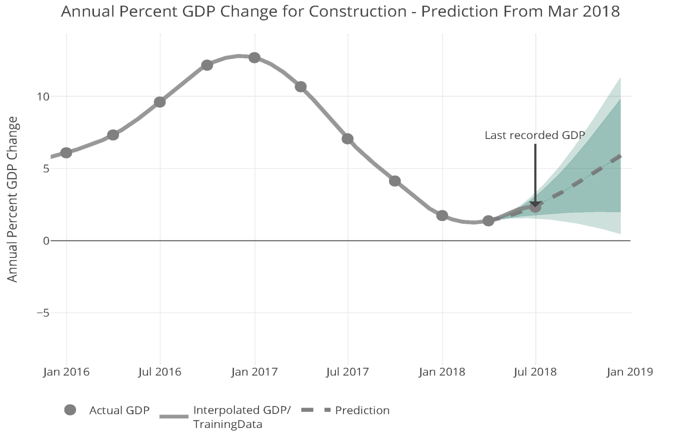
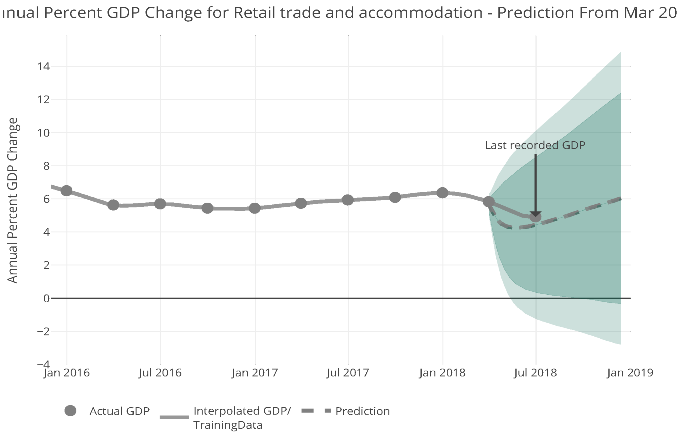


Figure 18: Annual percent change prediction for Retail, trade and accommodation, from end of March 2018.



The quarterly percent GDP change is reported in the industry experiments. The generalisability of the models is evaluated on most recent data covering the March 31 2018 to June 30 2018 period. However, a small variation is performed and presented in these experiments to the prior examples in the national and regional predictions. Namely, the predictions are allowed to run right through to the most recently available input data at the time of the experiment (14 December 2018). This constitutes a nearly nine month prediction period into the 'future' with the data being unknown on GDP in the last six months on that period. As errors accumulate the further out into the future the models are permitted to forecast, this becomes an interesting exercise in evaluating the robustness of the models.

For this time-range, strong generalisability is noted in Figures 16 to 19 covering some of the largest industry sectors: Rental hiring and real estate services, Construction, Retail trade and accommodation and Wholesale trade.

Examples of closely underestimating and overestimating economic output can be seen in Figure 20, Arts, recreation and other services, and Figure 21 for Transport, postal and warehousing respectively.

Figure 19: Annual percent change prediction for Wholesale trade, from end of March 2018.

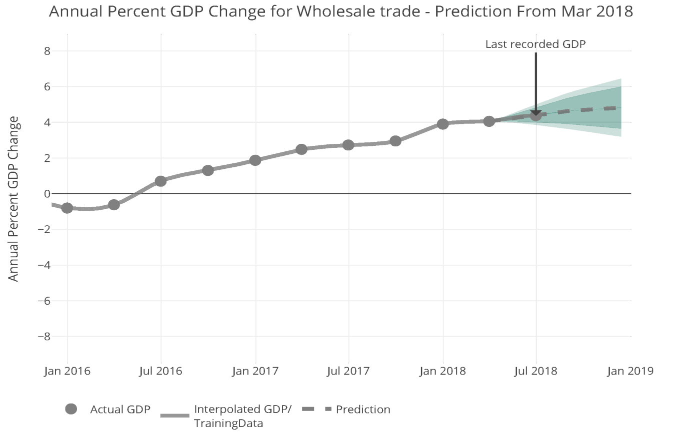


Figure 20: Annual percent change prediction for the Arts, recreation and other services, from end of March 2018.

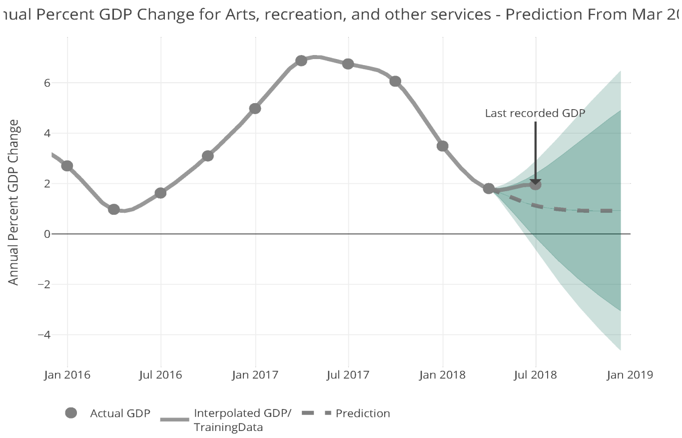


Figure 21: Annual percent change prediction for Transport postal and warehousing, from end of March 2018.

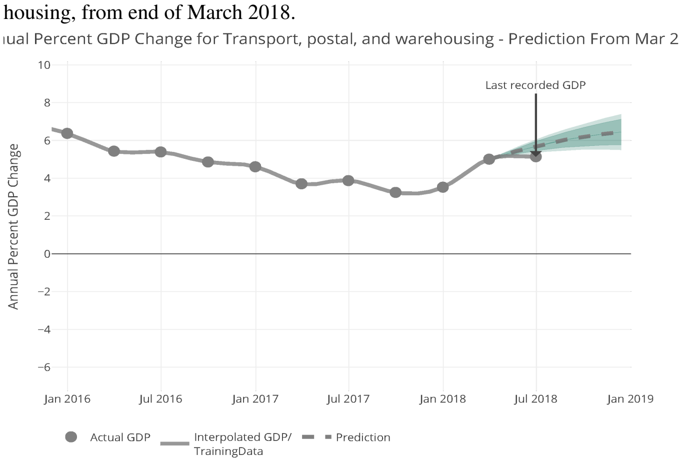


Figure 22: Annual percent change prediction for the Financial and insurance services, from end of March 2018.

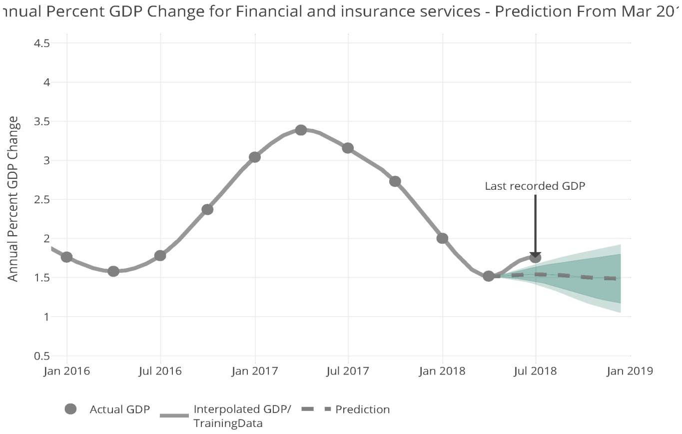


Figure 23: Annual percent change prediction for Education and training, from end of March 2018.

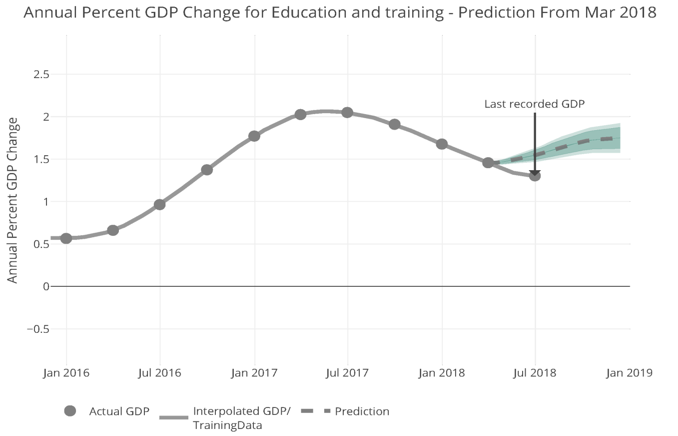


Figure 24: Annual percent change prediction for Manufacturing, from end of March 2018.

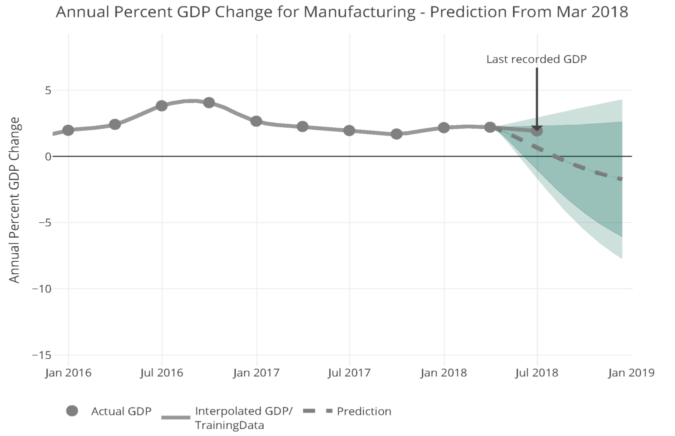
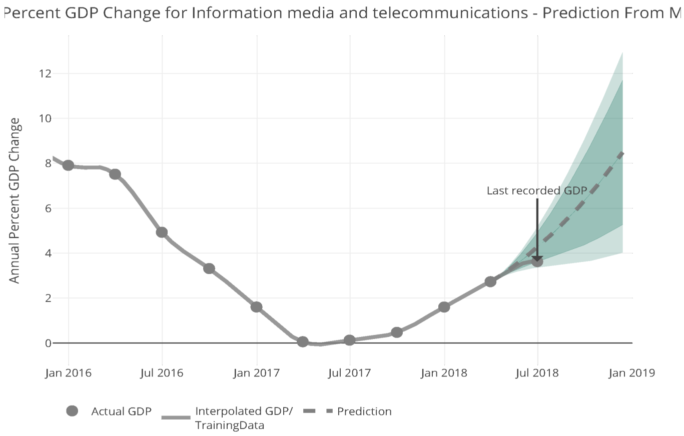


Figure 25: Annual percent change prediction for Information, media and telecommunications, from end of March 2018.



Figures 22 to 24 arguably represent somewhat stronger deviations from the general trend in the true GDP values. This is particularly accentuated for Manufacturing. Estimates for Information, media and telecommunications in Figure 25 represent an interesting case in which the predictions are generally in line with the true GDP values; however, when allowed to run for a further six months, they appear to have detected a much stronger trend in economic activity than might appear to be the case. Though, this cannot be ascertained until actual figures for current forecasts are released in February of 2019.

6. Discussion

The GDPLive models demonstrate strong evidence that real-time GDP estimation at the level of daily granularity is within reach. This has been shown to be the case for predicting national-level economic activity as well as regional and industry sector outputs. Based on the results of a limited set of experiments presented here, arguably it can be claimed that marginally better accuracies have been realised on predicting regional and industry predictions than on overall national economic activity.

It is also self-evident that there remains at this stage a considerable degree of uncertainty in the accuracy of predictions. This is not surprising. The GDPLive models represent the first generation of models and their accuracy is expected to improve with time as additional data sources and features are integrated, machine learning algorithms are optimised in their parameters, and training strategies are refined. However, no matter how much fine tuning is implemented in successive iterations of GDPLive models, the fact will always remain that making predictions about a complex phenomenon such as GDP is non-trivial and will be accompanied with uncertainly and a potential for errors.

7. Limitations

There are several limitations of the first-generation GDPLive models. Firstly, the models conflate the economic output of both Auckland and Northland regions and are only able to produce predictions for the two regions in combination. This is unfortunate for a couple of reasons. Firstly, the two regions represent both ends of the wealth/deprivation spectrum, with Northland representing the region of greater deprivation. The current inability to track Northland's attempts to increase its economic output is therefore unfortunate. Also, Auckland represents approximately 40% of New Zealand's economic activity. Accurately tracking Auckland's economic output is vital for the economy of the entire country. However, and the fact that the data from Northland is included in its models raises some concerns as to whether or not the ability to achieve predictive accuracy is compromised.

In addition, the current GDPLive models are not able to make predictions on economic activity for the Primary Industries Sector covering Mining and Agriculture, forestry & fishing. This is also regrettable given that Primary Industries Sector is a key focus area of growth.

Both of the above limitations are reflections of the nature of the underlying data and are seen as being only temporary.

A further limitation is the fact that we were unable in the initial release to include data from KiwiRail, Portconnect and NZTA into the first-generation GDPLive models. The inclusion of import/export data is expected to improve the overall accuracies, especially for the industry sector predictions. Rail cargo shipments and heavy and light road traffic is also expected to make a contribution as proxies of economic activity. The inclusion of these data sources is also expected to help remedy current limitations surrounding the Northland/Auckland data entanglement as well as the primary industry forecasts.

The GDP figures are released by the NZ Treasury. The machine learning models depend on these to be accurate since the input data is trained against them. However, GDP figures are often revised at later dates and mostly centre on most recent data points. This represents a point of vulnerability for the GDPLive models, since any future revisions are likely to have detrimental effects on the accuracy of the models which are in the production system.

An additional source of current limitations of the models are in the manner in which the dependent variable is being interpolated. The use of spline interpolation injects bias into the final models and potentially has a more detrimental effect on the function curve as it approaches the final known GDP point, from which new predictions set off. This means that the lag periods in the initial predictions could adversely affect the direction of the prediction curve.

8. Future Work

Efforts are already underway for planning the update and release of the second-generation GDPLive models. The next phase primarily centres around addressing a number of limitations highlighted in the previous section. Foremost of these will be the integration of KiwiRail, Portconnect and NZTA data sources. Disentanglement of Auckland/Northland models and a creation of new models for the Primary Industries Sector. In addition, model parameter adjustments will be further fine-tuned and a stochastic process will be introduced into the generation of multiple versions of the dependent variable.

The secondary focus will be on integrating new machine learning algorithms into the ensemble and investigating the utility of using a more informed ensemble-weighting scheme for each model based on their competence in prior prediction performance.

9. Conclusion

Despite its many detractors, GDP is still the primary measure of economic activity that informs Government policy making, corporate decision making and public discourse on the health and wellbeing of an economy. The delay in the reporting and release of GDP data can range from 2-3 months for national and industry-specific figures, to over a year for regional figures. The release delays are understandable given the magnitude of the task of calculating GDP for any country. However, in this day and age there is an increasing appetite and need to know with more certainty how the economy is tracking right now, rather than how it was months or a year ago.

The enormous amount of rich and descriptive volumes of data that are available now and the technological advances in machine learning and computational power make it possible for the first time to attempt to create a real-time GDP prediction system. This research presents a world-first, live GDP nowcasting system (GDPLive[4]) which estimates the economic activity at the national, regional and industry-sector levels on a daily basis for the New Zealand context.

GDPLive ingests consumption, import/export and cargo shipments data on a daily basis as well as currently available macroeconomic indicators and applies state-of-the-art machine learning algorithms and strategies in order to derive estimates for each day and nowcasts them online. GDPLive makes predictions for the current quarterly percent GDP change at the national level, as well as forecasted current quarter-end and next quarter-end GDP rates. At the regional level and industry-sector level, current estimates of GDP rates are reported at an annual and quarterly percent change respectively.

This research presents experimental results of the accuracy of the first-phase GDPLive models on historic predictions. Overall, the results are encouraging. The models have indicated their ability to learn from historic patterns and broadly generalise. A degree of uncertainty is present in the predictions and this will remain, though it is anticipated that with future iterations of GDPLive, this will decrease.

10. Acknowledgements

We would like to thank our industry partners Paymark, KiwiRail and Portconnect for their participation in the project. We appreciate their belief and trust in us. We thank also NZTA for their assistance in granting us access to their traffic data. We would also like to acknowledge the financial support of the Knowledge Exchange Hub (KEH) at Massey University (New

[4]www.gdplive.net

Zealand) who funded and project-managed GDPLive. Particular thanks to Alison Brook, the project manager from the KEH. Several Massey University students helped with the project and we would like to thank them for their involvement. Special thanks go to Erdem Alpkaya, William Wrigley and Ryan Weyers for their dedication, help and hard work with the development of the website.

Appendix A. Bibliography

[1] J. V. Henderson, A. Storeygard, D. N. Weil, Measuring economic growth from outer space, American Economic Review 102 (2012) 994–1028.

[2] R. G. Alajmi, The relationship between economic growth and municipal solid waste testing the ekc hypothesis: Analysis for saudi arabia, Journal of International Business Research and Marketing 1 (2016) 20–25.

[3] Keqiang ker-ching, https://www.economist.com/asia/2010/12/09/keqiang-ker-ching, 2010.

[4] ANZ Truckometer, https://www.anz.co.nz/about-us/economic-markets-research/truckometer/, 2018.

[5] Start-up can forecast GDP by using satellite images of night lights, https://www.livemint.com/Companies/BM91ywDM94xlpIpLqKbf4O/Japanese-startup-can-forecast-Indias-GDP-by-using-satellit.html, 2014.

[6] K. G., F. Papailias, Big data and macroeconomic nowcasting: Methodological review, ESCoE Discussion Paper (2018).

[7] R. J. Barro, Inflation and economic growth, Annals of Economics and Finance 14 (2013).