Analysis of the Treasury’s Macroeconomic and Tax Forecast Accuracy

26 May 2016

Executive Summary

• The Treasury has updated the analysis of its GDP growth, CPI inflation and tax revenue forecast accuracy, last reported in 2013.

• The accuracy of the Treasury’s GDP growth and tax revenue forecasts has improved over the last three years, whereas the accuracy of the CPI inflation forecasts has declined a little.

• The Treasury’s average forecast errors for real GDP growth and CPI inflation are towards the lower end of the error range comprising forecasts from 15 other forecasters of the New Zealand economy. This indicates that, at the time they are prepared, the Treasury’s forecasts are amongst the best available to Government Ministers for their decision making.

• In each of the last six years, the Treasury has achieved its target of having one-year-ahead Budget tax revenue forecast errors of less than ±3%.
Why does the Treasury produce forecasts?

The Public Finance Act 1989 requires the Treasury to produce two economic and fiscal updates each year using its best professional judgements. The *Budget Economic and Fiscal Update*, which informs the Government’s decision-making for its annual Budget, is usually released in May. The *Half-Year Economic and Fiscal Update* is typically released in December each year. A third official forecast, the *Pre-Election Economic and Fiscal Update* must be produced in years in which there is a general election, usually 20-30 working days before polling day. These updates provide the Government with five-year forecasts of the New Zealand economy and the Crown’s financial situation.

The economic forecasts include forecasts of over 400 macroeconomic variables relevant to New Zealand’s economic well-being. The fiscal forecasts are a full set of financial statements for the Crown covering the current and following four fiscal years and include all the things that are typically found in financial statements, e.g. statements of financial performance and position, statements of cash flows, etc.

The Government uses these forecasts to inform its decision making. For example, when considering its Budget package each year, the Government has been briefed on the Treasury’s latest economic and fiscal outlook in the form of a preliminary forecast. These forecasts form the backdrop for any policy measures that the Government is considering.

It is therefore desirable that these economic and fiscal forecasts be as accurate as possible. They should be neither overly-optimistic nor overly-pessimistic, so that the Government is able to make decisions that are both beneficial and affordable. Over time, forecast errors should remain within some small range, so that Government Ministers can have confidence in the forecasts, thereby giving them some confidence that their decisions will have the desired economic and/or fiscal effects.

When the forecasts are released, they are usually accompanied by one or two scenarios. These scenarios present alternative views on how the economy could evolve and the likely effect that these alternative growth paths would have on the Government’s finances. The scenarios are not intended to cover all possible outcomes nor define upper and/or lower bounds for possible outcomes. Rather, they demonstrate that there is uncertainty around the central forecast and that reality is almost certainly going to differ from forecast.

How can we analyse forecast performance?

Essentially, there are two desirable properties that we would like the forecasts to have:

1. We would like the variance of the forecast errors to be as small as possible, which would indicate that, on average, the forecasts stay within some relatively small distance of the actual outcomes.

2. We would also like the average forecast errors, i.e. the average difference between the actual outcomes and the forecasts, to be zero, which would indicate that, on average, the forecasts are neither too low nor too high.

To test the first property, we examine the root mean square error (RMSE), which is a measure of the average magnitude of the errors disregarding the signs (positive or negative) of the errors. We will look at whether or not these RMSEs are reducing over time, i.e. is forecast accuracy improving over time?, and whether or not the RMSEs reduce as the forecast horizon shortens, i.e. do the forecasts become more accurate as the time between the forecast and the actual result reduces?
To test the second property, we calculate the average of the forecast errors over time. Ideally, this mean error should be zero, indicating that the forecasts are neither persistently too high nor too low. In practice, the mean error is unlikely to be exactly zero, so we test it to see if it is ‘significantly different’ from zero. If so, we conclude that a particular forecast is ‘biased’, tending to be either too high or too low, on average. If the mean error is not significantly different from zero, then we do not have enough evidence to conclude that the forecast is biased.

In the following sections, we analyse the forecast performance of the real GDP growth and CPI inflation forecasts, and compare the performance of those forecasts with the performance of similar forecasts published by 15 other agencies. We also analyse the forecast performance of nominal GDP growth and tax revenue, the latter of which is one of the key aggregates in the Crown’s financial statements.
**Real GDP growth forecasts**

Real GDP is a key macroeconomic parameter forecast by the Treasury. It measures the size of economic activity in New Zealand in constant-dollar terms, i.e. the volume of economic activity ignoring the effects of price inflation. Growth in real GDP is thus the chief indicator of aggregate growth in the New Zealand economy. It is forecast by many other forecasters, which yields a sizeable sample against which the Treasury can compare its forecasting performance, as shown in the comparison section later in this report.

The forecasts have been grouped together by forecast horizon, i.e. the length of time between when the forecast was made and when the actual event occurred. Since the Treasury usually produces two forecasts per year, the forecasts have been grouped together into six-monthly forecast-horizon bands. The 0-6 month-ahead forecasts are typically *Half-Year Update* forecasts of the current March year’s GDP outturn, 6-12 month-ahead forecasts are *Budget Update* forecasts of the current March year’s GDP outturn, 12-18 month-ahead forecasts are typically *Half-Year Update* forecasts of the following March year’s GDP outturn, and so on. The forecasts have been analysed in these forecast-horizon groups as it is possible that forecasts of differing horizons display different characteristics.

![Figure 1](https://example.com/f1.png)

**Figure 1** – Root-mean-square errors of the Treasury’s forecasts of annual average percentage change in March year GDP in constant prices (i.e., real GDP growth), 1990 to 2015\(^1\).\(^2\)

Source: The Treasury

In Figure 1, the solid blue line shows the RMSEs of forecasts made at the various forecast horizons. For instance, the 42-48 month data point shows the RMSE of all the forecasts over the 1990 to 2015 period that were made 42-48 months in advance of the actual event. The forecasts are ordered in decreasing forecast horizon order as, reading left to right, time passes and the interval between the date the forecast was made and the date of the GDP outcome reduces.

---

\(^1\) Forecast errors are presented as actual percentage change less forecast percentage change. Thus, the forecast errors may be interpreted as percentage point errors rather than relative errors.

\(^2\) The latest actual result used in this analysis was the March 2015 outturn for GDP growth.
All of the RMSEs have decreased since the previous report (the solid blue line is below the broken lighter-blue line in Figure 1), indicating that the accuracy of the real GDP growth forecasts has increased over the past three years. There were some large forecast errors for the March 2009 & 2010 years, as forecasts prior to then did not adequately predict the size of the recession or, in the case of the longer-horizon forecasts, did not predict a recession at all. Forecast errors through the post-recession period were smaller than through the recessionary period, which has brought the average forecast errors down a little from what was reported in the 2011 and 2013 reports.

This is particularly noticeable when we look at the average forecast errors over shorter time-frames: the ten most-recent observations and five most-recent observations.

**Figure 2** – Root-mean-square errors of the Treasury’s forecasts of real GDP growth, whole period vs more-recent forecasts

The average forecast errors over the last five years at all forecast horizons, and for the last 10 years for the shorter forecast horizons, are smaller than the average forecast errors over the whole sample period, which, like Figure 1, indicates that the accuracy of the real GDP growth forecasts has increased in recent times. The improvement in forecast accuracy over the past 10 years, particularly in the one- to two-year-ahead range, coincides with the introduction of the New Zealand Treasury Model (NZTM) as the principal macroeconomic forecasting tool.

All of the RMSE lines in Figure 1 and Figure 2 eventually reduce as the forecast horizon shortens, particularly once the forecast horizon is less than a year, indicating that forecast accuracy improves as we get closer to the outcome.
Figure 3 – Mean errors of the Treasury’s forecasts of real GDP growth

If the mean forecast error is zero, then we say that the forecast is unbiased, i.e. neither persistently too low nor too high. Since the mean error is unlikely to ever be exactly zero, we test it to see if it is ‘significantly different’ from zero. Figure 3 does this by building a 95% confidence interval around the mean errors, based on the distribution of the individual forecast errors at each forecast horizon. If the point corresponding to a zero mean error lies outside the confidence band, then we can be 95% certain that the mean error at that forecast horizon is significantly different from zero and we can conclude that that forecast is biased, i.e. either persistently too high or too low, depending on whether the zero line is above or below the confidence interval.

At all of the forecast horizons shown in Figure 3, the zero forecast error lies within the green band. This means that we cannot state with 95% certainty that the forecasts were persistently too high or too low. In other words, we do not have enough evidence to conclude that the real GDP growth forecasts were biased at any forecast horizon.

CPI inflation forecasts

The Consumers Price Index (CPI) is the headline indicator of prices of consumption goods and services in the New Zealand economy. It is important in the context of the forecasts as the Reserve Bank of New Zealand (RBNZ) adjusts the Official Cash Rate (OCR\(^3\)) to keep CPI inflation, i.e. the growth rate of CPI, within the target range (currently 1% - 3%) set out in its Policy Targets Agreement (PTA). CPI inflation is also important for its influence on nominal, i.e. current-price, GDP, which is a key influence on the Crown’s tax revenue.

---

\(^3\) The OCR is effectively the wholesale price of borrowing or lending money in New Zealand (Reserve Bank of New Zealand Fact Sheet: What is the Official Cash Rate?)
RMSEs of the CPI inflation forecasts at all forecast horizons have increased to at least some degree since the last report, indicating that the accuracy of these forecasts has decreased over the last few years, although the increase in the average forecast errors is not large. This has occurred mainly because Treasury has continued to forecast CPI inflation coming back towards the midpoint of the RBNZ’s 1% to 3% target range, but actual inflation has remained persistently low.

This effect can be seen in Figure 5, which shows the average CPI inflation forecast errors over shorter timeframes. The average errors of the more-recent forecasts are generally higher than the average over the whole 25-year period, which has caused the whole-period RMSEs to creep upwards in recent years.
Previous forecast performance reports revealed evidence of bias in the longer-horizon CPI inflation forecasts; forecasts of horizons of more than 24 months tended to under-estimate actual CPI inflation. Figure 6 shows that this is no longer the case, as the zero forecast error line is now within the 95% confidence interval at these forecast horizons.

**Figure 6** – Mean errors of the Treasury’s forecasts of annual CPI inflation

With the recent period of relatively low inflation, the longer-dated forecasts that were produced before the low-inflation period started were too high. This has brought the mean errors closer to zero than at the last report and the standard statistical test now shows that, over the 25-year period as a whole, we do not have enough evidence to conclude that these longer-dated CPI inflation forecasts were persistently too high or too low. For the very shortest time period under examination, i.e. forecasts of up to six months ahead, the zero forecast error is above the 95% confidence interval, indicating that these forecasts tended to over-estimate actual CPI inflation. This was also the case in previous reports.

**Nominal GDP growth forecasts**

Nominal GDP measures the size of economic activity in New Zealand in current-dollar terms, i.e. the amount of economic activity including the effects of price inflation. Nominal GDP is important for the Crown’s financial management because of its close relationship with tax revenue. It also affects government expenditure. For example, in periods of slow economic growth, government spending via social assistance benefits may increase. The Treasury’s forecasts of tax revenue are based on forecasts of nominal GDP and its components.
Figure 7 – Root-mean-square errors of the Treasury’s forecasts of average annual percentage change in March year GDP in current prices (ie, nominal GDP growth), 1991 to 2015

Source: The Treasury

RMSEs for all forecast horizons have decreased since the last report for the same reason as for the real GDP growth forecasts, i.e. large forecast errors through the 2009/10 recession were followed by relatively smaller forecast errors. RMSEs for forecast horizons from 4 years ahead down to one year ahead were all quite similar to each other, indicating that the one-year-ahead Half-Year Update forecasts were no more accurate, on average, than the three-year-ahead Budget forecasts. However, for forecast horizons of less than one year, the RMSEs decreased dramatically.

None of the mean errors were found to be significantly different from zero, indicating that there was no strong evidence of bias in the forecasts at any forecast horizon.

Figure 8 – Mean errors of the Treasury’s forecasts of nominal GDP growth
Comparisons with other forecasters

We have compared the performance of the Treasury’s real GDP growth and CPI inflation forecasts with those produced by other forecasters as published in Consensus Economics Inc.’s monthly report “Asia Pacific Consensus Forecasts”. Only those forecasters that had a time series of forecasts of a similar length to that of the Treasury were used. This yielded a sample of 15 comparator forecasts.

For this exercise, we have used the forecasts published in Asia Pacific Consensus Forecasts in the same month in which the Treasury’s forecast was finalised. In this way, we are determining how the Treasury’s forecasts rated against all forecasts that were available to Ministers at the time they used those forecasts to inform their decision making.

However, it also means that most, if not all, of the comparator forecasts were prepared earlier than the Treasury’s forecasts, which means that the Treasury’s forecasts will have had the benefit of more recent data than the comparator forecasts. This will bias the results of the comparison in the Treasury’s favour and readers should bear this in mind when interpreting the results.

Figure 9 – Comparison of forecasters’ average real GDP growth forecast errors, December years 2000 to 2015

Each diamond in Figure 9 represents the average (RMSE) forecast error for a particular forecaster for that forecast horizon. Each set of diamonds therefore represents the range of forecast errors at each forecast horizon for the 15 forecasters, plus Treasury and consensus. The Treasury’s average forecast errors (red diamonds) are consistently towards the lower end of each range, indicating that the Treasury’s forecasts of real GDP growth were amongst the most accurate of these forecasters. For forecast horizons of up to 18 months, the Treasury’s forecast performance was similar to that of the consensus forecast (green diamonds), and better than the consensus forecast for the 18-24 month forecast horizon.
In a similar fashion to the real GDP growth forecast result, the Treasury’s CPI inflation average forecast errors were consistently towards the lower ends of the error ranges. The Treasury’s forecast performance was similar to that of consensus in three out of the four forecast horizons examined, and removing the outliers from the calculations for the 12-18 month and 6-12 month forecast horizons moves the consensus results closer to the Treasury results.

Figure 9 and Figure 10 demonstrate that, at the time they were compiled, the Treasury’s forecasts of real GDP growth and CPI inflation were amongst the best available to Ministers. Although forecasts of nominal GDP growth are not available for comparison, the combination of the real GDP and CPI inflation forecast performance analyses suggests that the same might be true for nominal GDP growth. Furthermore, no single forecaster was consistently the best forecaster at every forecast horizon for either real GDP growth or CPI inflation.
**Tax revenue forecasts**

Tax revenue is a measure of the tax that is due for a particular period, regardless of whether or not the tax has actually been paid. Tax revenue is an important component of the Crown’s operating balance, accounting for the majority of the Crown’s total revenue.

There are several different measures of tax revenue, depending on which entities are included in, or excluded from, the calculation. In this analysis, we used total unconsolidated tax revenue, as this is the measure that is forecast by the Treasury’s tax forecasters. Other components of tax revenue that are used in the calculation of core Crown and total Crown tax, e.g. GST paid by the Crown and income tax paid by SOEs, have been excluded as they are not forecast by the Treasury but are instead forecast by the relevant departments, SOEs, etc. Total unconsolidated tax revenue is also the measure that the OECD uses in its monitoring of government revenues.

The sample period for the analysis of tax revenue forecasts starts with the forecast prepared for the 1994 Budget. 1994 was the year in which the Treasury first started publishing financial results on an accrual basis.

Results are presented in percentage terms (percentage of average outturn) to provide an indication of the relative size, rather than the nominal size, of the average forecast error. The forecasts were adjusted, where appropriate, for changes in tax legislation and/or accounting method that were made after the forecasts were compiled. Performance statistics have been calculated for tax receipts, the cash measure of tax, but are not presented here as they are similar to the revenue results.

**Figure 11** – Root-mean-square errors of the Treasury’s tax revenue forecasts, 1994 to 2015

All of the RMSEs have decreased since the last report, indicating that the accuracy of the tax revenue forecasts has increased over the last few years. As the forecast horizon shortens, the

---

4 Measured as the percentage by which the actual result deviated from the forecast.
RMSEs decline, indicating that the shorter-horizon forecasts were more accurate than the longer-horizon forecasts, as would be expected.

**Figure 12 – Mean errors of tax revenue forecasts**

There is not enough evidence to conclude that the tax revenue forecasts are biased, as the mean errors at all forecast horizons were quite small and the zero forecast error sits comfortably inside the confidence interval around each forecast horizon’s mean error.

**Serial correlation in tax forecasts**

If forecast errors are serially correlated, then an under/over-forecast in one year increases the probability of an under/over-forecast in the following year. This will be manifested by long sequences of either solely positive or solely negative forecast errors. This is not a desirable characteristic to have in forecast errors as, ideally, we would like forecast errors to be randomly distributed about zero. Furthermore, if forecast errors are serially correlated, then this will affect the credibility of the forecasts and reduce their utility for policymakers.

We used a Durbin-Watson test to determine if there was any serial correlation in the tax forecasts. The test did not produce any evidence of serial correlation at forecast horizons of under 12 months. However, it did return positive results for forecast horizons of 12 months or more, i.e. there is evidence of serial correlation in the tax forecast errors at these forecast horizons.

Figure 13 demonstrates the serial correlation in the forecast errors for the one-year-ahead forecasts. These forecasts are typically Budget Update forecasts for the next June year’s result, e.g. the first bar in the chart represents the forecast error for the 1994 Budget’s forecast of the 1995 June year’s tax revenue outcome. Note the sequences of solely positive and solely negative forecast errors, in particular, the sequence of nine positive forecast errors in a row in the 2000s.
The RMSE for one-year-ahead Budget tax revenue forecasts over the whole period under examination is 3.2%. In 2009, when the average one-year-ahead tax forecasting error was over 4%, the Treasury set itself a target of achieving tax forecasting errors of less than ±3%. Figure 13 shows that:

- six of the seven forecast errors from 2003 to 2009 were outside this range, and
- the six forecast errors from 2010 to 2015 were all within the target range.

Although it is difficult to test for serial correlation with a small number of observations, a visual inspection of the forecast error charts shows that serial correlation between forecast errors may have reduced over the past six years. Further investigation is required to understand the cause of remaining such serial correlations in order to eliminate them altogether. This work is ongoing as Treasury strives to refine and improve its economic and tax forecasting models.