Evaluating Point Forecasts

Building upon Forecasting Note #1, this note delves more deeply into point forecasts and their validation. While predicting the future implies inherent uncertainty, a point forecast can be very useful to quickly describe the general expected path with less complexity.

Forecast evaluation is a crucial part of forecasting because it allows the forecast user to assess how much trust to place in a given forecast. A perfect forecast correctly predicts, ahead of time, the subsequent outcome. However, absent the ability to see the future, forecasts are necessarily imperfect. Thus, it is important to be able to distinguish good forecasts from bad ones. This requires measuring the accuracy of the forecast by comparing it to the actual or "realized" future value. The bigger the "distance" between the forecast and the actual outcome, the worse the forecast. In this note, we discuss how this distance is measured, i.e. how the forecast is evaluated, in the case of a point forecast. Point forecasts are the most common and simple type of forecasts, and take the form of a single number that conveys the forecaster's expectation as to the most likely outcome.

Measuring Predictive Accuracy

Consider Apple's quarterly sales for the first quarter of 2016, which was $50.6bn, down by 13% relative to the 2015Q1 value. A perfect forecast would have predicted this exact revenue value some time prior to the release of the revenue information by Apple. Naturally, in practice perfect forecasts are extremely rare. In the more realistic scenario where the outcome and the forecast are not identical, a measure is needed of what constitutes a good, let alone "optimal", forecast.

Forecast error

A natural starting point for measuring the precision of a forecast is to note that a good forecast is, in some sense, "close" to the outcome. Closeness is often measured as the size of the forecast error, i.e., the difference between the outcome and the forecast:

Forecast error = outcome – forecast

The forecast error is commonly used to measure the precision of forecasts of variables whose levels fluctuate around a constant mean, such as growth rates of revenues, percentage stock returns, the unemployment rate, or the inflation rate.

For many other economic variables, the level as well as the size of the unpredictable component trend up over time. One example is the level of the Dow Jones index over long periods of time. Measuring the simple forecast error for the Dow Jones would not be sufficiently informative for comparisons that are made over long spans of time. For example, a one-day change of 100 points in the Dow Jones index is a common occurrence when the index is close to a level of 17,000, but much rarer when the index is at 1,000. In the former case, a 100-point change in the index would correspond to a 0.6% daily change; in the latter case, it would correspond to a 10% change.

Other examples of economic variables that are trending upward over time include the number of unemployed people in the work force, as opposed to the unemployment rate, the level of the consumer price index, as opposed to the inflation rate

and, in many cases, the level of company revenues. As an illustration, **Figure 1** below shows the time-series evolution in Apple's quarterly revenues along with the average one-quarter-ahead analyst forecast of these revenues. Both are increasing sharply over time.

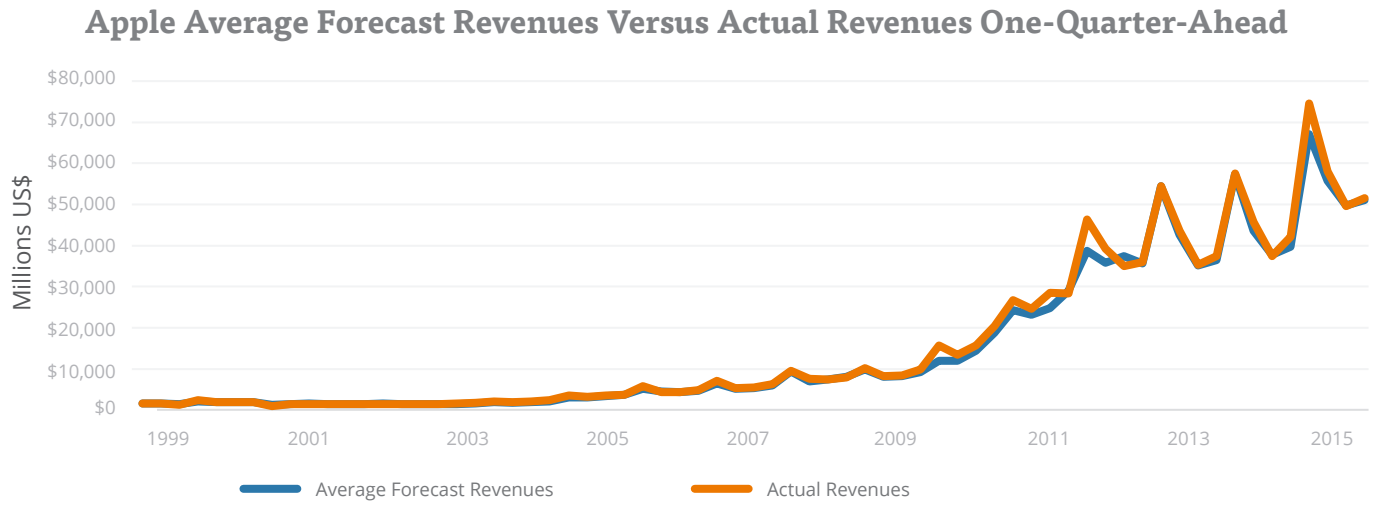

Figure 1 compares Apple's average forecast revenues to actual revenues between 1999 and 2015 for one-quarter-ahead.[1]

Upward trends in economic variables generally translate into larger forecast errors. For example, the average analyst forecast error for Apple's first-quarter revenues in 2001 was $61.2 million, compared to $2.17 billion in 2015Q1. This difference need not indicate that analysts have gotten systematically worse at predicting Apple's revenues over the preceding 15-year period. Rather, it simply reflects that Apple's revenues have grown systematically over time.

Figure 2 shows the time-series evolution in the dollar forecast error of Apple's quarterly revenues, measured as the actual revenue number minus the average analyst forecast. The magnitude of the dollar level of the forecast errors is sharply increasing over time and thus cannot naturally be compared at the beginning and end of the sample.

Figure 2 shows Apple's average forecast error for one-quarter-ahead.

1 Apple, Inc. quarterly revenues via Thomson Reuters QA Direct (accessed May 4, 2016).
Analyst revenue forecasts via Thomson Reuters QA Direct (accessed May 4, 2016).

Percentage forecast error

To account for such trends in revenues, and to be able to compare the precision of forecasts across long spans of time, it is common to report the percentage forecast error, i.e. the forecast error relative to the most recent outcome.

Percentage Error = Forecast Error / Previous Outcome

The forecast error and the percentage forecast error can be calculated for any period for which both a forecast and historical data are observed. Returning to the earlier figures for Apple's first-quarter revenues, the $61.2 million forecast error in 2001 amounted to a 6.1 percentage error, whereas the much bigger forecast error of $2.17 billion in 2015Q1 amounted to a smaller percentage error of 2.9%.

Further illustrating this point, **Figure 3** below shows the percentage forecast errors for Apple, obtained by dividing the dollar forecast errors in Figure 2 by the previous quarter's revenue figure. In contrast with the forecast error measured in levels, there is no systematic tendency for the percentage error to increase in magnitude over time.

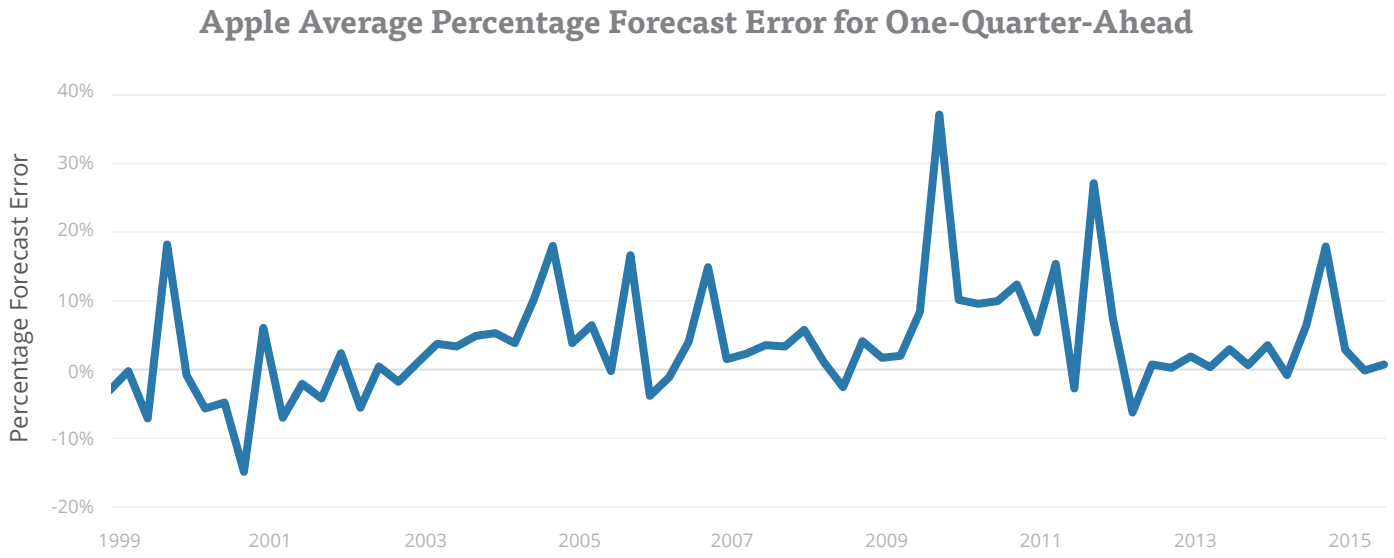

Figure 3 shows Apple's average percentage forecast error for one-quarter-ahead.

Average forecast error

Outcomes differ from their predicted values for random and unpredictable reasons. Thus, a forecast could be accurate in a particular period due to luck, or it could be poor because of a large and unforeseen event, such as a company's introduction of a new product. One example of such an event in history is the introduction of the iPhone by Apple. Another example is the terrorist attacks on September 11, 2001, which could not have been foreseen by any economic forecaster.

Forecasts in any given period can be accurate either because of skill or luck. Distinguishing between these two possibilities is possible when a long time-series record of forecasting performance is considered. One way to do this is to calculate the average forecast error, or percentage forecast error, over this long period of time. However, the average forecast error conveys only limited information. For example, consider a forecaster whose percentage errors alternate between -10% or +10%, with each value occurring about one-half of the time. On average, this forecaster's percentage error is equal to zero.

This is the same average forecast error as that of a forecaster whose percentage errors alternate between -1% or +1%, each occurring half of the time. The use of average forecast error as the measure of analyst performance obscures the fact that the second forecaster produces far more accurate predictions.

Mean absolute errors and mean squared errors

To correct for this problem, rather than averaging over forecast errors, it is common to compute averages over either the absolute values or the squares of the forecast errors. Both methods consider the size of the error, but not its direction, so that the average error cannot be made artificially small unless the forecaster is accurate. In the example above, the mean absolute percentage forecast error of the first forecaster is 10%, while that of the second forecaster is 1%, thus clearly capturing the superiority of the second forecaster. Similarly, the mean squared percentage error of the first forecaster is 100, while that of the second forecaster is 1.

In addition, in order to preserve the same scale as the original forecast, a common approach is to report the square root of the mean squared error. This measure is called the root mean squared error, or RMSE. The root mean squared error is equal to 10 and 1, respectively, for the two forecasters in the example above. Note that the root mean squared error again correctly reveals that the second forecaster performs best, but without distorting the size of the error of the first forecaster to be an order of magnitude larger than the error itself.

Forecaster	Percentage Error	Average Error	Mean Absolute % Error	Mean Squared % Error	RMSE
Forecaster 1	-10% to 10%	0%	10%	100	10
Forecaster 2	-1% to 1%	0%	1%	1	1

Correlation

Another measure of forecasting performance that can be useful is the correlation between the forecasts and actual values, subject to certain conditions as mentioned below. This is a number that always lies between -1 and +1. Negative values of the correlation indicate that the outcomes and the forecasts tend to move in opposite directions so that the outcome tends to rise when a decline is forecasted, and vice versa. Positive values of the correlation indicate that the outcomes and forecasts move in the same direction most of the time; said another way, when the forecast predicts an increase in the variable, the outcome of the variable also tends to increase. As the correlation gets closer to +1, movements in the forecast and the subsequent outcomes become closer, indicating more accurate forecasts. Because the correlation is constructed in such a way as to be independent of the scale of the underlying variable, it is easy to interpret and compare across different economic variables.

However, correlations can be misused to suggest a relationship exists when there actually is no such reliable relationship. As often noted, correlation does not imply causation. One distinction is that evaluation of correlations can be performed in or out of sample. Please see Forecasting Note #4 for the difference between in-sample and out-of-sample evaluation.

Forecast Accuracy

Measures of forecast accuracy such as the mean absolute error and the root mean squared error are informative about the extent to which a variable is predictable, and are sometimes called absolute measures of forecasting performance. They can be used to capture the fact that some variables are intrinsically very difficult to predict. For example, stock market returns, changes in commodity futures prices, and exchange rates are notoriously difficult to predict because of their high and changing variation over time. For these variables, we would expect to find low correlation values and high mean squared errors. Other variables, such as the unemployment rate or interest rates, are relatively persistent and, therefore, easier to predict, at least at relatively short time horizons. **See Figure 4.** Absolute measures of forecasting performance convey which of these cases the variable being forecasted and the relevant forecasting model most closely resemble.

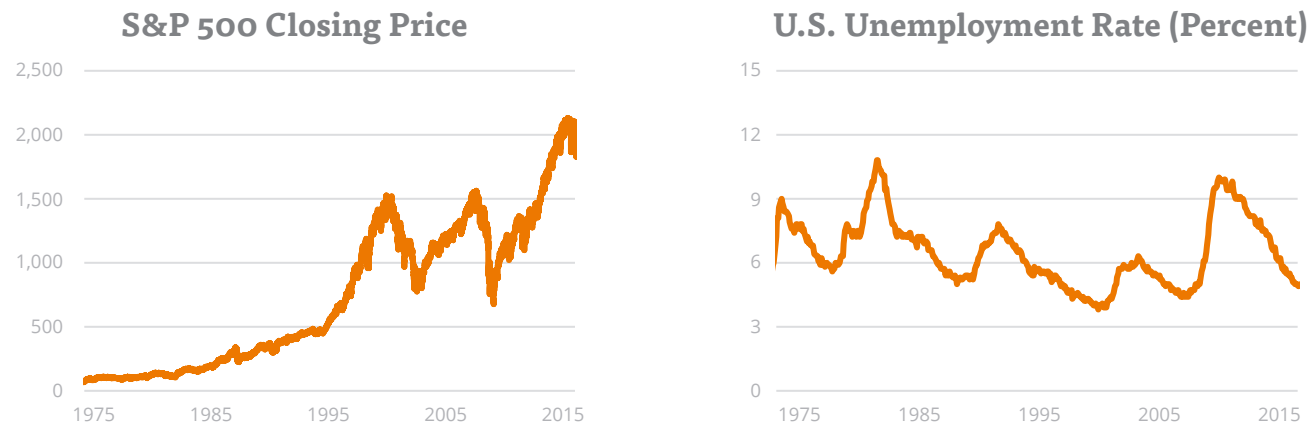

Figure 4 shows historical S&P 500 closing prices and U.S. unemployment rates. [1] Relative to the stock market, unemployment rates are less volatile across time, and do not feature a trend.[2]

Absolute measures of forecasting performance do not, however, convey information about whether a particular forecasting model is good relative to other possible forecasts, let alone "best practice". This is the goal of forecast comparisons.

Properties of an Optimal Forecast

Optimal forecasts possess certain properties that can be tested by comparing predicted values to the actual outcome.

Unbiasedness

First, a good point forecast should be unbiased. This means that even if at a given point in time, the forecast and the outcome may be very different, on average over a reasonably long period of time, the two values should be very similar. Suppose, for example, that forecasts for a given variable tend to be 10% above the historical outcomes over some time period. This would suggest that these forecasts could be improved simply by lowering them by the amount of the bias, i.e. 10%.

[1] Thomson Reuters, S&P 500 Index (accessed May 18, 2016).

[2] Federal Reserve Bank of St. Louis, Civilian Unemployment Rate (accessed May 5, 2016).

Unpredictable forecast errors

A second property is that forecast errors, i.e. the difference between the outcome and the forecast, should not themselves be predictable. In other words, if one could systematically predict, ahead of time, the likely error one will make, then that error can be corrected as soon as it is predicted. Most notably in terms of unpredictability, forecast errors should not be serially correlated. This means that it should not be possible to find sustained periods with forecasts differing from outcomes in the same direction, either consistently over-predicting or under-predicting the outcome. If the direction of the forecast error is seen repeating, so that one observes a string of overpredictions, this would suggest reducing next period's forecast.

PROPERTIES OF AN
OPTIMAL FORECAST

1. Forecasts and outcome can be very different at a given point in time, but should be same on average
2. Forecast errors should not be predictable
3. Short-term outcomes should be easier to predict than long-term outcomes.

The implication of these measures of forecast quality and optimality is that if a forecaster has skillfully used all available information to produce a forecast, then this same information should not be useful for predicting the forecast error. Conversely, if this is not the case, then the forecaster did not optimally exploit all the available information. For example, suppose that forecasters underreact when presented with a particular piece of news and only adjust their expectations slowly. Such a pattern of underreaction would give rise to predictability in the forecast error. For example, if the news were positive and suggested a beneficial revision of the forecast by 10%, but the initial revision to their forecast is only 5%, then we would be able to predict that the future forecast error will be higher by an amount (5%) reflecting the underreaction.

These properties can be tested on a sample of forecasts and outcomes. For example, the basic optimality property that the forecast error should not itself be predictable can be tested by regressing forecast errors on a variety of predictors, and checking if the slope coefficients (as well as the intercept) all obtain coefficients of zero.

Another commonly used strategy for evaluating forecasts is by regressing the outcome on a constant and the forecast:

$$\textbf{Outcome} = a + b \times \textbf{Forecast} + \textbf{error}$$

If the realized and forecasted values vary one-to-one, then the constant coefficient in this regression (α) should equal zero, and the slope (β) should equal one. Any other finding would suggest that the forecast may be biased, and imply a simple strategy for correcting the forecast for this bias. Suppose, for example, that this regression shows that the outcome is equal to one-half plus 0.8 times the forecast (α= 0.5, β=0.8) plus some error that has zero mean. Then the bias-adjusted forecast would be equal to one-half plus 0.8 times the forecast:

$$\textbf{Bias-adjusted forecast} = 0.5 + 0.8 \times \textbf{forecast}$$

This bias-adjusted forecast would, on average, be closer to the actual outcome than the raw forecast itself.

Better forecast performance at shorter horizons

A third property of a good forecast is that its accuracy can be expected to decline the longer the forecast horizon. This property reflects that the short term is, on average, easier to predict than the long term. In other words, forecasts of one-year-ahead outcomes should be more accurate than forecasts of two, three, or five-year-ahead outcomes. More unpredictable events can happen between now and the distant future (say between April 2016 and April 2021, a five-year forecast horizon) than between now and the near future (say between April 2016 and May 2016, a one-month forecast horizon). The expected decline in the predictive accuracy, the longer the forecast horizon, can be tested by comparing mean squared forecast errors across different forecast horizons. For an optimal forecast we would expect to find that the mean squared errors would increase, the longer the forecast horizon.

To illustrate this for analysts' forecasts of Apple's revenues, the figures below compare the time-series evolution of the one-quarter-ahead forecast errors against the four-quarters-ahead forecast errors, both measured in percentage terms. Clearly the four-quarters-ahead percentage forecast errors are notably bigger than the one-quarter-ahead percentage forecast errors, illustrating that analysts find it harder to accurately predict the relatively distant future than the near future.

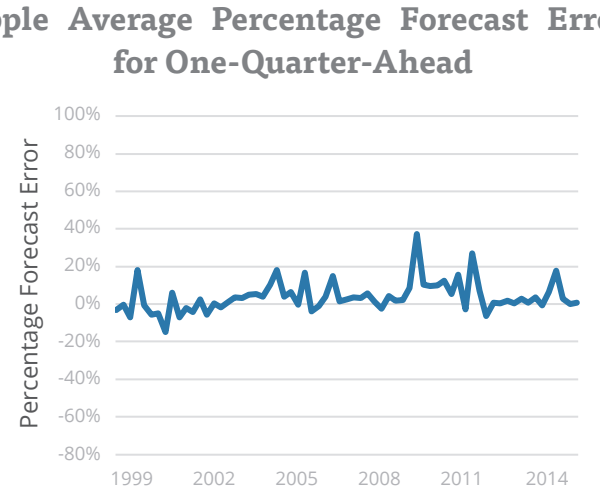

Figure 5 compares Apple average percentage forecast errors for one-quarter-ahead to four-quarters-ahead. Percentage forecast error is equal to actual revenues minus forecasted revenues, divided by actual revenues in the previous quarter.

All figures described herein are for illustrative purposes only

INTENSITY

About Intensity

Intensity generates excellence through powerful research, analysis, and expertise to solve the most complex challenges in the marketplace and courtroom. We consistently deliver reliable results that are built upon meticulous research, intense scrutiny, and scientific analysis.

Please contact us any time!

12730 High Bluff Drive Suite 300
San Diego, CA 92130
858.876.9101

intensity.com