4

Taming Uncertainty: What You Need to Know about Measuring Forecast Accuracy

Television won't be able to hold on to any market it captures after the first six months. People will soon get tired of staring at a plywood box every night.

A WELL-KNOWN MOVIE MOGUL IN THE EARLY DAYS OF TV

This chapter describes

- why it is necessary to start first with an unambiguous definition of forecast error
- what bias and precision mean for accuracy measurement
- how, when and why to make accuracy measurements
- the systematic steps in a forecast evaluation process

` After reading this chapter, you should be able to

- distinguish between fit and forecast errors
- recognize that there is not one best measure of accuracy for products, locations and summary hierarchies
- understand why simple averaging is one of the worst best practices for summarizing accuracy measurements
- engage with forecast users and demand managers to identify appropriate accuracy standards for the business

The Need to Measure Forecast Accuracy

The Institutional Broker's Estimate System (I/B/E/S), a service that tracks financial analysts' estimates, reported that forecasts of corporate yearly earnings that are made early in the year are persistently optimistic, that is to say, upwardly biased. In all but two of the 12 years studied (1979–1991), analysts revised their earnings forecasts downward by the end of the year. In a *New York Times* article (January 10, 1992), Jonathan Fuerbringer wrote that this pattern of revisions was so clear that the I/B/E/S urged stock market investors to take early-in-the year earnings forecasts with a grain of salt.

It is generally recognized in most organizations that accurate forecasts are essential for achieving measurable improvements in business operations. Developers of forecasting models use accuracy measures, and accuracy at all levels is relevant to the users of forecasts in a company. Will the model prove reliable for forecasting units or revenues over a planned forecast horizon, such as item-level product demand for the next 12 weeks or aggregate revenue demand for the next four quarters? Will it have a significant impact on marketing, sales, budgeting, logistics and production activities? Will it have an effect on inventory investment or customer service? In a nutshell, inaccurate forecasts can have a direct effect on setting inadequate safety stocks, ongoing capacity problems, massive rescheduling of manufacturing plans, chronic late shipments to customers, and adding expensive manufacturing flexibility resources (Figure 4.1). In addition, the Income Statement and Balance Sheet are also impacted by poor demand forecasts for financial planning (Figure 4.2).

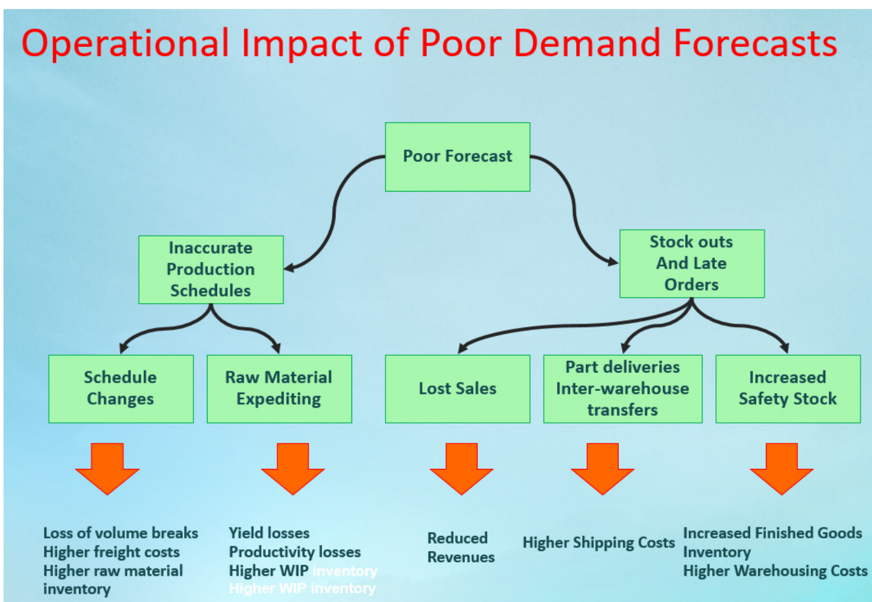


Figure 4.1 Operational impact of poor forecasts. (*Source*: Simon Conradie, Noetic Business Consulting)

Analyzing Forecast Errors

Whether a method that has provided a good fit to historical data will also yield accurate forecasts for future time periods is an unsettled issue. Intuition suggests that this may not necessarily be the case. There is no guarantee that past patterns will persist in future periods. For a forecasting technique to be useful, we must demonstrate that it can forecast reliably on ongoing basis and with consistent accuracy. It is not credible to simply produce a model that performs best only in a historical (within sample) fit period.

Because of ever-present outliers and non-normality, the demand forecaster must also measure forecasting performance with multiple metrics, not just with a single metric, like the Mean Absolute Percentage Error (MAPE) or weighted MAPE. At the same time, the users of a forecast need forecasting results in a timely fashion. Using a forecasting technique and waiting one or more periods for history to unfold in future periods is not practical because our advice as forecasters will not be agile.

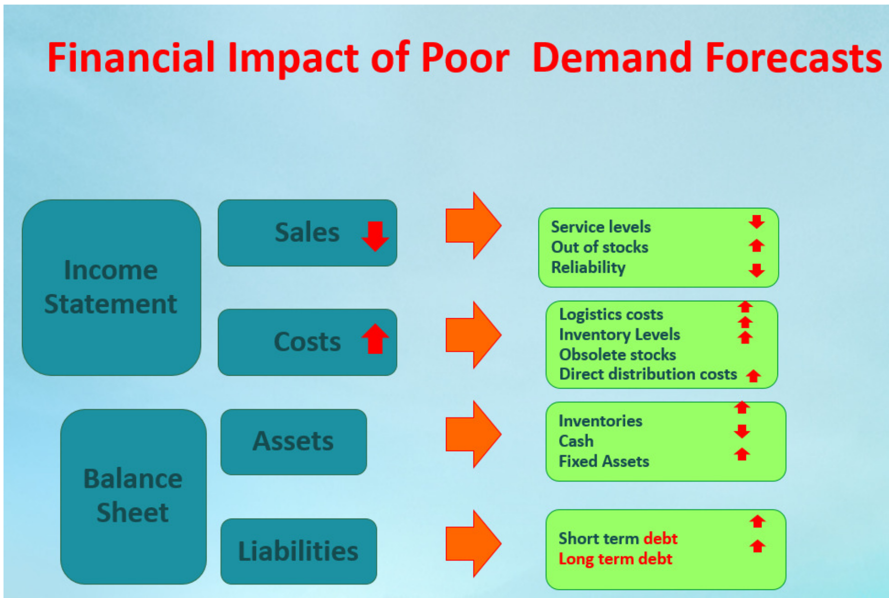


Figure 4.2 Financial impact of poor forecasts. (*Source*: Noetic Business Consulting)

Two important aspects of forecast accuracy measurement are bias and precision. **Bias** is a problem of direction: Forecasts are typically too low (downward bias) or typically too high (upward bias). **Precision** is an issue of magnitudes: Forecast errors can be too large (in either direction) using a particular forecasting technique. Consider first a simple situation - forecasting a single product or item. The attributes that should satisfy most forecasters include lack of serious bias, acceptable precision, and superiority over naive models.

Lack of Bias

If forecasts are typically too low, we say that they are downwardly biased; if too high, they are upwardly biased. If overforecasts and underforecasts tend to cancel one another out (i.e., if an average of the forecast errors is approximately zero), we say that the forecasts are unbiased.

Bias refers to the tendency of a forecast to be predominantly toward one side of the truth.

If bias is a problem of direction, we can think of forecasting as aiming darts at a target; then a bias implies that the aim is off-center. That is, the darts land repeatedly toward the same side of the target (Figure 4.3). In contrast, if forecasts are unbiased, they are evenly distributed around the target.

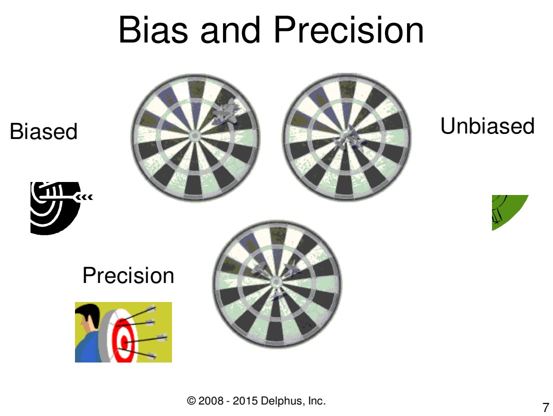


Figure 4.3 Biased, unbiased, and precise forecasts.

What Is an Acceptable Precision?

Imprecision is a problem if the forecast errors tend to be too large. Figure 4.3 shows three patterns that differ in terms of precision. The upper two forecasts are the less precise - as a group, they are farther from the target. If bias is thought of as bad aim, then imprecision is a lack of constancy or steadiness. The precise forecast is generally right around the target (Figure 4.4).

Figure 4.4 Precision in forecasts.

Precision refers to the distance between the forecasts as a result of using a particular forecasting technique and the corresponding actual values.

Figure 4.5 illustrates the measurement of bias for three hypothetical forecasting techniques. In each case, the fit period is periods 1 - 20. Shown in the top row of Figure 4.5b are actual values for the last four periods (21 – 24). The other three rows contain forecasts using forecasting models X, Y, and Z. These are shown in a bar chart in Figure 4.5a. What can we say about how good these forecast models are? On the left graph, the three forecasts do not look all that different,

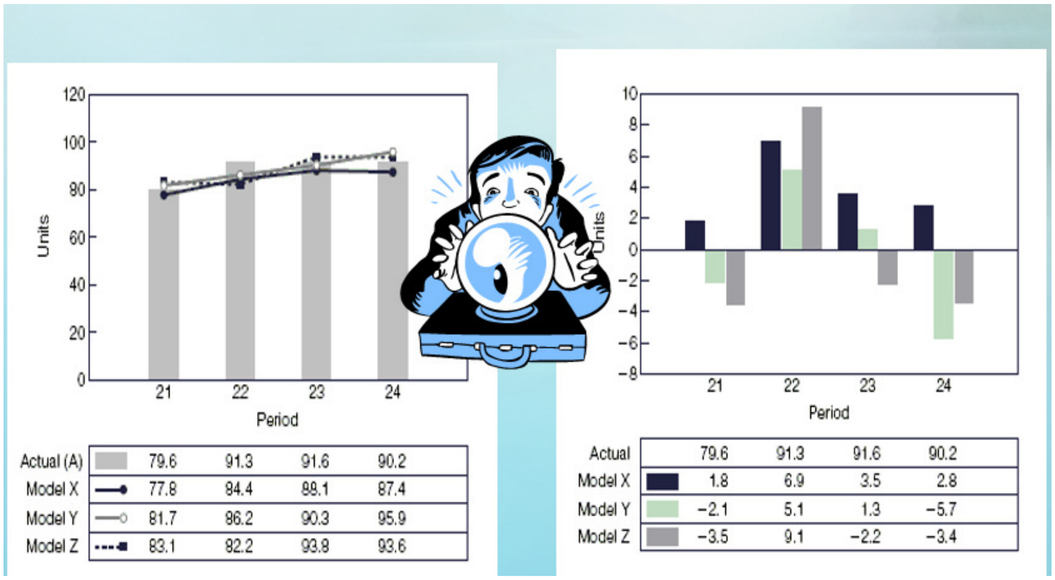


Figure 4.5 Bar charts and tables showing actuals (A), forecasts, and forecast errors for three forecasting techniques.

But, what is a forecast error? In the CPDF® professional development training workshops I conduct, it was not unusual to hear inconsistent definitions and interpretations among practitioners, even within the same company.

Figure 4.5 records the deviations between the actuals and their forecasts. Each deviation represents a **forecast error** (or forecast miss) for the associated period:

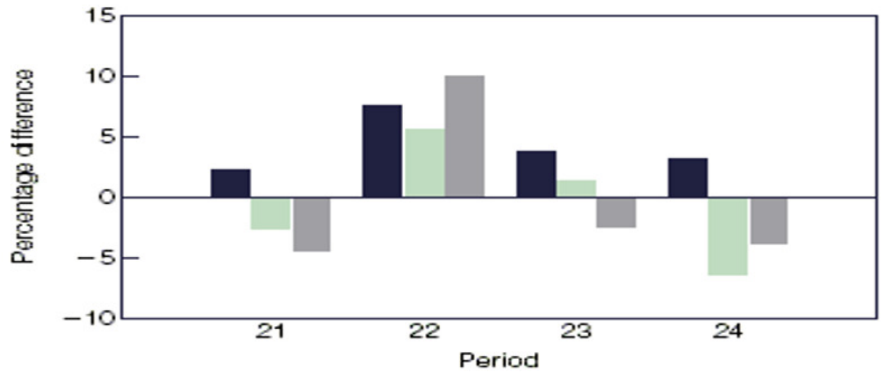
Forecast error (***E***) = Actual (***A***)–Forecast **(*F*)**

If (F–A) is the preferred use in some organizations but not others, then demand forecasters and forecast users should name it something else, like **forecast variance**, a more conventional meaning among revenue-oriented planners. The distinction is important because of the interpretation of bias in under- and overforecasting situations. Contrast this with a fit error (or residual) of a model fit over a historical period, which is Fit error = Actual (***A***) - Model fit (***F***).

Forecast error is a measure of forecast accuracy. Fit error (or residual) is a measure of model adequacy.

In Figure 4.5, the forecast error shown for model X is 1.8 in period 21. This represents the deviation between the actual value in forecast period 21 (= 79.6) and the forecast using model X (= 77.8). In forecast period 22, the forecast using model X was lower than actual value for that period resulting in a forecast error of 6.9.

The period 24 forecast using model Z was higher than that period's actual value; hence, the forecast error is negative (- 3.4). When we ***over***forecast, we must make a *negative* adjustment to reach the actual value. Note that if the forecast is less than the actual value, the miss is a positive number; if the forecast is more than the actual value, the miss is a negative number.



	21	22	23	24
Actual	79.6	91.3	91.6	90.2
Model X	2.3	7.6	3.8	3.1
Model Y	−2.6	5.6	1.4	−6.3
Model Z	−4.4	10.0	−2.4	−3.8

Figure 4.6 Bar chart and table of the forecast error as percentage error (PE) between actuals and forecasts for three models.

To identify patterns of upward- and downward-biased forecasts, we start by comparing the number of positive and negative misses. As Figure 4.5 shows, Model X under-forecasts in all four periods, indicative of a persistent (downward) bias. Model Y underforecasts and overforecasts with equal frequency; therefore, it exhibits no evidence of bias in either direction. Model Z is biased slightly toward overforecasting. As one measure of forecast accuracy (Figure 4.6), we calculate a percentage error PE = 100% * (A – F)/A.

To reduce bias in a forecasting technique, we can either (1) reject any technique that projects with serious bias in favor of a less-biased alternative (after we have first compared the precision and complexity of the methods under consideration) or (2) investigate the pattern of bias in the hope of devising a bias adjustment; for example, we might take the forecasts from method X and adjust them upward to try to offset the tendency of this method to underforecasts certain periods. Also, forecasts for Models Y and Z could be averaged after placing Forecast X aside for the current forecasting cycle.

Ways to Evaluate Accuracy

A number of forecasting competitions have been held to assess the effectiveness of statistical forecasting techniques and determine which techniques are among the most useful. Starting with the original M competition in 1982, Spyros Makridakis (*left*) and his academic collaborators compared the accuracy of about 20 forecasting techniques across a sample of 111 time series—a very small dataset by today's standards. A subset of the methods was tested on 1001 time series. The last 12 months of each series were held out and the remaining data were used for model fitting. Using a range of measures on a holdout sample, the *International Journal of Forecasting* (IJF) conducted a competition in 1997 comparing a range of forecasting techniques across a sample of 3003 time series. Known as the M3 competition, these data and results can be found at the website *www.maths.monash.edu.au/~hyndman/forecasting/.* A number of IJF papers have been written summarizing the results of these competitions. These competitions have become the basis for how we should measure forecast accuracy in practice.

Forecast accuracy measurements are performed in order to assess the accuracy of a forecasting technique.

The complete chapter can be found in

Change & Chance Embraced

ACHIEVING AGILITY WITH DEMAND FORECASTING IN THE SUPPLY CHAIN

HANS LEVENBACH, PhD

ISBN-10:0692945989
ISBN-13:978-069294588

Cover art: Rose Tannenbaum, Berkshire TypeGraphic, www.tgo.com

Cartoon art: Don Engel, Isographic LLC

Copy editor: David Coen, Freelance copyeditor

Printed by Createspace, www.createspace.com

Professional development: Certified Professional Demand Forecaster (CPDF®), www.cpdftraining.org

Contents

: