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).

![Operational Impact of Poor Demand 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.

![Financial Impact of Poor Demand Forecasts](image)

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.

![Bias and Precision diagram](image)

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).

![Precision in forecasts](image)

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:

\[ \text{Forecast error } (E) = \text{Actual } (A) - \text{Forecast } (F) \]

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 overforecast, 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.

![Figure 4.6](image)

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 \( PE = 100\% \times \frac{A - F}{A} \).

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
Contents

Chapter 1 - Embracing Change & Chance ........................................

Inside the Crystal Ball

Determinants of Demand
Demand Forecasting Defined
Why Demand Forecasting?
The Role of Demand Forecasting in a Consumer-Driven Supply Chain 4
Is Demand Forecasting Worthwhile? 7
Who Are the End Users of Demand Forecasts in the Supply Chain? 8
Learning from Industry Examples 9
Examples of Product Demand 10
Is a Demand Forecast Just a Number? 11

Creating a Structured Forecasting Process 14

The PEER Methodology: A Structured Demand Forecasting Process 14

Case Example: A Consumer Electronics Company 15

PEER Step 1: Identifying Factors Likely to Affect Changes in Demand 16
The GLOBL Product Lines 17
The Marketplace for GLOBL Products 18
Step 2: Selecting a Forecasting Technique 19
Step 3: Executing and Evaluating Forecasting Models 22
Step 4: Reconciling Final Forecasts 22

Creating Predictive Visualizations 22

Takeaways 26

Chapter 2 - Demand Forecasting Is Mostly about Data:

Demand Forecasting Is Mostly about Data 29

Exploring Data Patterns 29
Learning by Looking at Data Patterns 30

Judging the Quality of Data 30

Data Visualization 35

Time Plots 35
Scatter Diagrams 36

Displaying Data Distributions 37

Overall Behavior of the Data 38
Creating Data Summaries

- Typical Values
- The Trimmed Mean
- Variability
- Median Absolute Deviation from the Median
- The Interquartile Difference
- Detecting Outliers with Resistant Measures

The Need for Nonconventional Methods

- M-Estimators
- A Numerical Example

Why Is Normality So Important?

Case Example: GLOBL Product Line B Sales in Region A

Takeaways
Chapter 4 - Taming Uncertainty: What You Need to Know about Measuring Forecast Accuracy

The Need to Measure Forecast Accuracy 82
- Analyzing Forecast Errors 82
- Lack of Bias 82
- What Is an Acceptable Precision? 83

Ways to Evaluate Accuracy 86
- The Fit Period versus the Holdout Period 86
- Goodness of Fit versus Forecast Accuracy 87
- Item Level versus Aggregate Performance 88
- Absolute Errors versus Squared Errors 88
- Measures of bias 89
- Measures of Precision 90
- Comparing with Naive Techniques 93
- Relative Error Measures 94

The Myth of the MAPE . . . and How to Avoid It 95
- Are There More Reliable Measures Than the MAPE? 96

Predictive Visualization Techniques 96
- Ladder Charts 96
- Prediction-Realization Diagram 97

Empirical Prediction Intervals for Time Series Models 100
- Prediction Interval as a Percentage Miss 101
- Prediction Intervals as Early Warning Signals 101
- Trigg Tracking Signal 103

Spreadsheet Example: How to Monitor Forecasts 104

Mini Case: Accuracy Measurements of Transportation Forecasts 107

Takeaways 112

Chapter 5 - Characterizing Demand Variability: Seasonality, Trend, and the Uncertainty Factor 114

Visualizing Components in a Time Series 115
- Trends and Cycles 116
- Seasonality 119
- Irregular or Random Fluctuations 122
- Weekly Patterns 124
- Trading-Day Patterns 124

Exploring Components of Variation 126
- Contribution of Trend and Seasonal Effects 127
- A Diagnostic Plot and Test for Additivity 130

Unusual Values Need Not Look Big or Be Far Out 132

The Ratio-to-Moving-Average Method 134
Step 1: Trading-Day Adjustment 135
Step 2: Calculating a Centered Moving Average 135
Step 3: Trend-cycle and Seasonal Irregular Ratios 136
Step 4: Seasonally Adjusted Data 137

GLOBL Case Example: Is the Decomposition Additive or Not? 137

APPENDIX: A Two-Way ANOVA Table Analysis 139

Percent Contribution of Trend and Seasonal Effects 140

Takeaways 140

Chapter 6 - Dealing with Seasonal Fluctuations ..................... 141

Seasonal Influences 141
Removing Seasonality by Differencing 143
Seasonal Decomposition 145
Uses of Seasonal Adjustment 146

Multiplicative and Additive Seasonal Decompositions 146
Decomposition of Monthly Data 146
Decomposition of Quarterly Data 151
Seasonal Decomposition of Weekly Point-of-Sale Data 153

Census Seasonal Adjustment Method 156
The Evolution of the X-13ARIMA-SEATS Program 157
Why Use the X-13ARIMA-SEATS Seasonal Adjustment Program? 157
A Forecast Using X-13ARIMA-SEATS 158

Resistant Smoothing 158

Mini Case: A PEER Demand Forecasting Process for Turkey Dinner Cost 162

Takeaways 168

Chapter 7 - Trend-Cycle Forecasting with Turning Points ........ 171

Demand Forecasting with Economic Indicators 171
Origin of Leading Indicators 174
Use of Leading Indicators 174
Composite Indicators 176
Reverse Trend Adjustment of the Leading Indicators 176
Sources of Indicators 178
Selecting Indicators 178

Characterizing Trending Data Patterns 180
Autocorrelation Analysis 180
First Order autocorrelation 182
The Correlogram 183
Chapter 8 - Big Data: Baseline Forecasting With Exponential Smoothing Models ................................................................................................................................. 197

What is Exponential Smoothing? 198

Smoothing Weights 199
The Simple Exponential Smoothing Method 201

Forecast Profiles for Exponential Smoothing Methods 202

Smoothing Levels and Constant Change 204
Damped and Exponential Trends 208
Some Spreadsheet Examples 210
Trend-Seasonal Models with Prediction Limits 216
The Pegels Classification for Trend-Seasonal Models 219
Outlier Adjustment with Prediction Limits 221
Predictive Visualization of Change and Chance – Hotel/Motel Demand 221

Takeaways 225

Chapter 9 - Short-Term Forecasting with ARIMA Models ..226

Why Use ARIMA Models for Forecasting? 226

The Linear Filter Model as a Black Box 227
A Model-Building Strategy 229

Identification: Interpreting Autocorrelation and Partial Autocorrelation Functions 230
Autocorrelation and Partial Autocorrelation Functions 231
An Important Duality Property 233
Seasonal ARMA Process 234

Identifying Nonseasonal ARIMA Models 236

Identification Steps 236
Models for Forecasting Stationary Time Series 236
White Noise and the Autoregressive Moving Average Model 237
One-Period Ahead Forecasts 239
L-Step-Ahead Forecasts 239
Three Kinds of Short-Term Trend Models 241
A Comparison of an ARIMA (0, 1, 0) Model and a Straight-Line Model 241

Seasonal ARIMA Models 244
Chapter 11 - Gaining Credibility Through Root-Cause Analysis and Exception Handling

The Diagnostic Checking Process in Forecasting ................................................................. 284

The Role of Correlation Analysis in Regression Modeling ............................................... 284

Linear Association and Correlation 285
The Scatter Plot Matrix 286
The Need for Outlier Resistance in Correlation Analysis 287

Using Elasticities ................................................................. 288

Price Elasticity and Revenue Demand Forecasting 290
Cross-Elasticity 291
Other Demand Elasticities 292
Estimating Elasticities 292

Validating Modeling Assumptions: A Root-Cause Analysis ........................................... 293

A Run Test for Randomness 296
Nonrandom Patterns 297
Graphical Aids 299
Identifying Unusual Patterns 299

Exception Handling: The Need for Robustness in Regression Modeling 301

Why Robust Regression? 301
M-Estimators 301
Calculating M-Estimates 302

Using Rolling Forecast Simulations ................................................................. 304

Choosing the Holdout Period 304
Rolling Origins 305
Measuring Forecast Errors over Lead Time 306

Mini Case: Estimating Elasticities and Promotion Effects 306

Procedure 308
Taming Uncertainty 310

Multiple Regression Checklist 311

Takeaways 313

Chapter 12 - The Final Forecast Numbers: Reconciling Change & Chance

Establishing Credibility ................................................................. 317

Setting Down Basic Facts: Forecast Data Analysis and Review 317
Establishing Factors Affecting Future Demand 318
Determining Causes of Change and Chance 318
Preparing Forecast Scenarios 318
Analyzing Forecast Errors

Gaining Acceptance from Management

Case: Creating a Final Forecast for the GLOBL Company

Chapter 13 - Creating a Data Framework for Agile Forecasting and Demand Management
Chapter 14 - Blending Agile Forecasting with an Integrated Business Planning Process

PEERing into the Future: A Framework for Agile Forecasting in Demand Management

The Elephant and the Rider Metaphor
Prepare
Execute
Evaluate
Reconcile

Creating an Agile Forecasting Implementation Checklist

Selecting Overall Goals
Obtaining Adequate Resources
Defining Data
Forecast Data Management
Selecting Forecasting Software
Forecaster Training
Coordinating Modeling Efforts
Documenting for Future Reference
Presenting Models to Management

Engaging Agile Forecasting Decision Support

Economic/Demographic Data and Forecasting Services
Data and Database Management
Modeling Assistance
Training Workshops

The Forecast Manager’s Checklists

Forecast Implementation Checklist
Software Selection Checklist
Large-Volume Demand Forecasting Checklist

Takeaways