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.





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,





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.



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

Copyright © 2017 Hans Levenbach

Copyright © 2017 Hans Levenbach

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

ALL RIGHTS RESERVED. No part of this work may be used or reproduced, transmitted, stored or used in any form or by any means graphic, electronic, or mechanical, including but not limited to photocopying, recording, scanning, taping, Web distribution, information networks or information storage and retrieval systems, or in any manner whatsoever without prior written permission.

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

Agile Forecasting®, CPDF® are registered trademarks of Delphus, Inc.

Contents



Chapter 1 - Embracing Change & Chance

Inside the Crystal BallError! Bookmark not defined.

Determinants of Demand		
Demand Forecasting Defined		
Why Demand Forecasting?		
The Role of Demand Forecasting in a Cons	sumer-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 Pro	ocess 14	
The PEER Methodology: A Structured Der	nand Forecasting Process	14
Case Example: A Consumer Electronic	cs Company 15	
PEER Step 1: Identifying Factors Likely to	Affect Changes in Demand	16
The GLOBL Product Lines 17	5	
The Marketplace for GLOBL Products	18	
Step 2: Selecting a Forecasting Technique	19	
Step 3: Executing and Evaluating Forecast	ing Models22	
Step 4: Reconciling Final Forecasts	22	
Creating Predictive Visualizations	22	
-		

Takeaways 26



Chapter 2 - Demand Forecasting Is Mostly about Data:

29

Demand Forecasting Is Mostly about Data Exploring Data Patterns 29

Learning by Looking at Data Patterns30Judging the Quality of Data30

Data Visualization 35

Time Plots 35 Scatter Diagrams 36

Displaying Data Distributions 37

Overall Behavior of the Data 38

Stem-and-Leaf Displays 39 Box Plots 41 Quantile-Quantile Plots 43 Creating Data Summaries 44 **Typical Values** 44 The Trimmed Mean 45 Variability 45 Median Absolute Deviation from the Median 45 The Interguartile Difference 46 **Detecting Outliers with Resistant Measures** 47 The Need for Nonconventional Methods 48 **M-Estimators** 49 A Numerical Example 49 51 Why Is Normality So Important? Case Example: GLOBL Product Line B Sales in Region A 52 Takeaways 54 Chapter 3 - Predictive Analytics: Selecting Useful Forecasting All Models Are Wrong. Some Are Useful 56 Qualitative Methods 56 **Quantitative Approaches** 59 Self-Driven Forecasting Techniques 60 Combining Forecasts is a Useful Method 61 Informed Judgment and Modeling Expertise 62 A Multimethod Approach to Forecasting 64 Some Supplementary Approaches 64 Market Research 64 **New Product Introductions** 65 Promotions and Special Events 65 Sales Force Composites and Customer Collaboration 65 Neural Nets for Forecasting 66 A Product Life-Cycle Perspective 66 A Prototypical Forecasting Technique: Smoothing Historical Patterns 68 Forecasting with Moving Averages 69 Fit versus Forecast Errors 71 Weighting Based on the Most Current History 73 A Spreadsheet Example: How to Forecast with Weighted Averages 75 Choosing the Smoothing Weight 78 Forecasting with Limited Data 78 **Evaluating Forecasting Performance** 79 Takeaways 79

82 The Need to Measure Forecast Accuracy 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 95 The Myth of the MAPE . . . and How to Avoid It 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

22

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 135Step 2: Calculating a Centered Moving Average135Step 3: Trend-cycle and Seasonal Irregular Ratios136Step 4: Seasonally Adjusted Data 137

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

APPENDIX: A Two-Way ANOVA Table Analysis139

Percent Contribution of Trend and Seasonal Effects 140 Takeaways 140



Chapter 6 - Dealing with Seasonal Fluctuations 141

Seasonal Influences 141

Removing Seasonality by Differencing143Seasonal Decomposition145Uses of Sasonal Adjustment146Multiplicative and Additive Seasonal Decompositions146

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 Analysis180First Order utocorrelation182The Correlogram183

Trend-Variance Analysis 187 Using Pressures to Analyze Business Cycles 189

Mini Case: Business Cycle Impact on New Orders for Metalworking Machinery 191

1/12 Pressures1923/12 Pressures19312/12 Pressures193Turning Point Forecasting194

Ten-Step Procedure for a Turning-Point Forecast195Alternative Approaches to Turning-Point Forecasting195Takeaways196

Chapter 8 - Big Data: Baseline Forecasting With Exponential Smoothing

Models197

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 akegways 225

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 241 244 Seasonal ARIMA Models

A Multiplicative Seasonal ARIMA Model 244 Identifying Seasonal ARIMA Models 246 Diagnostic Checking: Validating Model Adequacy 247

Implementing a Trend/Seasonal ARIMA Model for Tourism Demand 249

Preliminary Data Analysis 249 Step 1: Identification 250 Step 2: Estimation 250 Step 3: Diagnostic Checking 251 ARIMA Modeling Checklist 254

Takeaways 255

Postcript 256



Chapter 10 - Demand Forecasting with Regression Models 258

What Are Regression Models? 259

The Regression Curve 260 A Simple Linear Model 260 The Least-Squares Assumption 260 CASE: Sales and Advertising of a Weight Control Product 262 Creating Multiple Linear Regression Models 263 Some Examples 264 CASE: Linear Regression with Two Explanatory Variables 266 Assessing Model Adequacy 268 Transformations and Data Visualization 268 Achieving Linearity 269 Some Perils in Regression Modeling 270 Indicators for Qualitative Variables 273 Use of Indicator Variables 273 Qualitative Factors 274 Dummy Variables for Different Slopes and Intercepts 275 **Measuring Discontinuities** 275 Adjusting for Seasonal Effects 276 Eliminating the Effects of Outliers 276 How to Forecast with Qualitative Variables 277

Modeling with a Single Qualitative Variable278Modeling with Two Qualitative Variables279Modeling with Three Qualitative Variables279A Multiple Linear Regression Checklist281

Takeaways 282



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 292 293 Validating Modeling Assumptions: A Root-Cause Analysis 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 301 301 Why Robust Regression? **M-Estimators** 301 Calculating M-Estimates 302 304 Using Rolling Forecast Simulations Choosing the Holdout Period 304 **Rolling Origins** 305 Measuring Forecast Errors over Lead Time 306 306 Mini Case: Estimating Elasticities and Promotion Effects 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 Review317Establishing Factors Affecting Future Demand318Determining Causes of Change and Chance318Preparing Forecast Scenarios318

Analyzing Forecast Errors 319

Taming Uncertainty: A Critical Role for Informed Judgment320Forecast Adjustments: Reconciling Sales Force and Management Overrides32Combining Forecasts and Methods322Verifying Reasonableness323Selecting 'Final Forecast' Numbers324	1
Gaining Acceptance from Management 325	
The Forecast Package 325 Forecast Presentations326	
Case: Creating a Final Forecast for the GLOBL Company 328	
Step 1: Developing Factors 329	
Impact Change Matrix for the Factors Influencing Product Demand 330	
The Impact Association Matrix for the Chosen Factors 331	
Exploratory Data Analysis of the Product Line and Factors Influencing Demand 33	2
Step 2: Creating Univariate and Multivariable Models for Product Lines 334	
Handling Exceptions and Forecast Error Analysis 335	
Combining Forecasts from Most Useful Models 337	
An Unconstrained Baseline Forecast for GLOBL Product Line B, Region A 338	
Step 3: Evaluating Model Performance Summaries 341	
Step 4: Reconciling Model Projections with Informed Judgment 342	
Takeaways 343	



Demand Management in the Supply Chain 345	
Data-Driven Demand Management Initiatives 346 Demand Information Flows 347 Creating Planning Hierarchies for Demand Forecasting 349	
What Are Planning Hierarchies? 349 Operating Lead Times 350 Distribution Resource Planning (DRP)—A Time-Phased Planned Order Forecast	350
A Framework for Agility in Forecast Decision Support Functions	
The Need for Agile Demand Forecasting354Dimensions of Demand354A Data-Driven Forecast Decision Support Architecture355Dealing with Cross-Functional Forecasting Data Requirements358Specifying Customer/Location Segments and Product Hierarchies358Automated Statistical Models for Baseline Demand Forecasting	
Selecting Useful Models Visually 363 Searching for Optimal Smoothing Procedures 367 Error-Minimization Criteria 368 Searching for Optimal Smoothing Weights 368 Starting Values 368	
Computational Support for Management Overrides 369 Takeaways 372	



Chapter 14 - Blending Agile Forecasting with an Integrated Business Planning Process 373

PEERing into the Future: A Framework for Agile Forecasting in Demand Management 374

The Elephant and the Rider Metaphor 374 Prepare 374 Execute 376 Evaluate 376 Reconcile 381 385 Creating an Agile Forecasting Implementation Checklist Selecting Overall Goals 385 Obtaining Adequate Resources 386 Defining Data 386 Forecast Data Management 387 Selecting Forecasting Software 387 Forecaster Training 388 Coordinating Modeling Efforts 388 Documenting for Future Reference 388 Presenting Models to Management 389 Engaging Agile Forecasting Decision Support 389 Economic/Demographic Data and Forecasting Services 389 Data and Database Management 390 Modeling Assistance 390 Training Workshops 390 391 The Forecast Manager's Checklists Forecast Implementation Checklist 391 Software Selection Checklist 392 Large-Volume Demand Forecasting Checklist 393 Takeaways 394