# Contents

Demand Forecasting: A Data-Driven Process What Is New? Scope Coverage Courseware Organization Dedication Acknowledgments



Chapter 1 - Embracing Change and 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

16

# Case Example: A Consumer Electronics Company15PEER Step 1: Identifying Factors Likely to Affect Changes in DemandThe GLOBL Product Lines17The Marketplace for GLOBL Products18Step 2: Selecting a Forecasting Technique19Step 3: Executing and Evaluating Forecasting Models22

Step 4: Reconciling Final Forecasts 22

# Creating Predictive Visualizations 22

# Takeaways 26



Improving Data Quality through Data Exploration and Visualization 28

30

# Smarter Forecasting Is Mostly about Data 29

Exploring Data Patterns 29 Learning by Looking at Data Patterns

# Judging the Quality of Data $\ 30$

Data Visualization 35 Time Plots 35 Scatter Diagrams 36

#### Displaying Data Distributions 37

Overall Behavior of the Data Stem-and-Leaf Displays 39 Box Plots 41 Quantile-Quantile Plots 43

#### Creating Data Summaries 44

Typical Values44 The Trimmed Mean 45 Variability 45 Median Absolute Deviation from the Median 45 The Interquartile Range 46 Detecting Outliers with Resistant Measures 47

#### The Need for Nonconventional Methods 48

M-Estimators 49 A Numerical Example 49

#### Why Is Normality So Important? 51

#### Case Example: GLOBL Product Line B Sales in Region A 52

38

Takeaways 54



Chapter 4 <u>-</u> Tama Accuracy 79	ing Uncertainty: W	That You	Need to I	Know about Me	asuring Forecast
<b>The Need to Measur</b> Analyzing Forecast Err Lack of Bias 8 <u>1</u> What Is an Acceptable	re Forecast Accur ors 81 Precision? 82	асу	80		
Ways to Evaluate A The Fit Period versus t Goodness of Fit versus Item Level versus Aggr Absolute Errors versus Measures of bias Measures of Precision Comparing with Naive Relative Error Measure	ccuracy 85 he Holdout Period Forecast Accuracy egate Performance Squared Errors 88 88 Techniques es94	85 86 87 87 93			
<b>The Myth of the MA</b> Are There More Reliab	<b>PE and How i</b> le Measures Than the	<b>to Avoid</b> e MAPE?	<b>It</b> 96	95	
<b>Predictive Visualiza</b> Ladder Charts 96 Prediction-Realization	t <b>ion Techniques</b> Diagram 97	96			
<b>Empirical Prediction</b> Prediction Interval as a Prediction Intervals as Trigg Tracking Signal	Intervals for Time Percentage Miss Early Warning Signals 103	<b>ne Series</b> 101 s	<b>Models</b>	100	
Spreadsheet Examp Quick and Dirty Contro The Tracking Signal in Adapting the Tracking	<b>le: How to Monit</b> II103 Action 103 Signal to Other Sprea	o <b>r Forec</b> adsheets	<b>asts</b> 107	103	
Mini Case: Accuracy	Measurements o	of Trans	portatio	n Forecasts	106
Takeaways 111					
Chapter Uncertainty Factor	5 - Characterizing • 113	g Deman	d Variab	ility: Seasonali	ty, Trend, and the
Visualizing Compon Trends and Cycles Seasonality 119 Irregular or Random F Weekly Patterns Trading-Day Patterns	ents in a Time Se 116 uctuations 122 122 124	ries	115		
Exploring Compone	nts of Variation	126			

Contribution of Trend and Seasonal Effects 126 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	
Step 1: Trading-Day Adjustment 134	
Step 2: Calculating a Centered Moving Average	134
Step 3: Trend-cycle and Seasonal Irregular Ratios	135
Step 4: Seasonally Adjusted Data 136	

#### GLOBL Case Example: Is the Decomposition Additive or Not? 136

#### APPENDIX: A Two-Way ANOVA Table Analysis 138

Percent Contribution of Trend and Seasonal Effects 139

#### Takeaways 139





Chapter 7 - Trend-Cycle Forecasting with Turning Points 169

170

#### Demand Forecasting with Economic Indicators

Origin of Leading Indicators173Use of Leading Indicators173Composite Indicators175Reverse Trend Adjustment of the Leading Indicators175Sources of Indicators177Selecting Indicators177

# Characterizing Trending Data Patterns 179 Autocorrelation Analysis 179

First Order utocorrelation179The Correlogram182

Trend-Variance Analysis 186

#### Using Pressures to Analyze Business Cycles 188

#### Mini Case: Business Cycle Impact on New Orders for Metalworking Machinery 190

 1/12 Pressures
 191

 3/12 Pressures
 192

 12/12 Pressures
 192

#### Turning Point Forecasting 193

Ten-Step Procedure for a Turning-Point Forecast194Alternative Approaches to Turning-Point Forecasting194

#### Takeaways 195



Chapter 8 - Big Data: Baseline Forecasting With Exponential Smoothing Models 196

What is Exponential Smoothing?197Smoothing Weights198The Simple Exponential Smoothing Method 200

#### Forecast Profiles for Exponential Smoothing Methods 201

Smoothing Levels and Constant Change203Damped and Exponential Trends207Some Spreadsheet Examples209Trend-Seasonal Models with Prediction Limits215The Pegels Classification for Trend-Seasonal Models218Outlier Adjustment with Prediction Limits220Predictive Visualization of Change and Chance – Hotel/motel roomroom demand220

Takeaways 224



#### Why Use ARIMA Models for Forecasting? 226

The Linear Filter Model as a Black Box

#### A Model-Building Strategy 229

Identification: Interpreting Autocorrelation and Partial Autocorrelation Functions230Autocorrelation and Partial Autocorrelation Functions231An Important Duality Property233Seasonal ARMA Process 234

227

#### 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 243

A Multiplicative Seasonal ARIMA Model 243

Identifying Seasonal ARIMA Models 245

#### Diagnostic Checking: Validating Model Adequacy 246

#### Implementing a Trend/Seasonal ARIMA Model for Tourism Demand 247

Preliminary Data Analysis248Step 1: Identification248Step 2: Estimation248Step 3: Diagnostic Checking249

ARIMA Modeling Checklist 253

Takeaways 254

Postcript 255



Chapter 10 - Demand Forecasting with Regression Models 256

What Are Regression Models? 257 The Regression Curve 258 A Simple Linear Model 258 The Least-Squares Assumption 258 CASE: Sales and Advertising of a Weight Control Product 260 Creating Multiple Linear Regression Models 261 Some Examples 262 CASE: Linear Regression with Two Explanatory Variables 264 Assessing Model Adequacy 266 Transformations and Data Visualization 266 Achieving Linearity 267 Some Perils in Regression Modeling 268 Indicators for Qualitative Variables 271 Use of Indicator Variables 271 Qualitative Factors 272 Dummy Variables for Different Slopes and Intercepts 273 Measuring Discontinuities 273 Adjusting for Seasonal Effects 274 Eliminating the Effects of Outliers 274 How to Forecast with Qualitative Variables 275 Modeling with a Single Qualitative Variable 276 Modeling with Two Qualitative Variables 277 Modeling with Three Qualitative Variables 277 279 A Multiple Linear Regression Checklist Takeaways 280



# L Chapter 11 - Gaining Credibility Through Root-Cause

Analysis and Exception Handling 281

The Diagnostic Checking Process in Forecasting 282			
The Role of Correlation Analy	sis in Regression Modeling	282	
Linear Association and Correlati The Scatter Plot Matrix 284 The Need for Outlier Resistance	ion 283 in Correlation Analysis 285		
Using Elasticities 286			
Price Elasticity and Revenue Der Cross-Elasticity 289 Other Demand Elasticities Estimating Elasticities 290	mand Forecasting 288 290		
Validatina Modelina Assum	nptions: A Root-Cause Analysis	292	
A Run Test for Randomness Nonrandom Patterns 295 Graphical Aids 297	294		
Identifying Unusual Patterns	297		
<b>Exception Handling: The Ne</b> Why Robust Regression? M-Estimators 2991 Calculating M-Estimates	2991 300	Modeling 299	
Using Rolling Forecast Simu Choosing the Holdout Period Rolling Origins 303 Measuring Forecast Errors over	Ilations 302 302 Lead Time 304		
<i>Mini Case: Estimating Elast</i> Procedure 306 Taming Uncertainty 308	icities and Promotion Effects	304	
Takeaways 311			
Chapter 12 - The Final I	Forecast Numbers: Reconciling Cha	unge & Chance 313	

#### Establishing Credibility 314

Setting Down Basic Facts: Forecast Data Analysis and Review314Establishing Factors Affecting Future Demand315Determining Causes of Change and Chance315Preparing Forecast Scenarios315

#### Analyzing Forecast Errors 316

Taming Uncertainty: A Critical Role for Informed Judgment317Forecast Adjustments: Reconciling Sales Force and Management Overrides318

#### CHANGE&CHANCE EMBRACED:

Combining Forecasts and Methods	
Verifying Reasonableness 320	
Selecting 'Final Forecast' Numbers	321

# Gaining Acceptance from Management 322

The Forecast Package322Forecast Presentations323

Case: Creating a Final Forecast for the GLOBL Company	325	
Step 1: Developing Factors 326		
Impact Change Matrix for the Factors Influencing Product Demand	327	
The Impact Association Matrix for the Chosen Factors 328		
Exploratory Data Analysis of the Product Line and Factors Influencing De	mand	329
Step 2: Creating Univariate and Multivariable Models for Product Lines	331	
Handling Exceptions and Forecast Error Analysis 332		
Combining Forecasts from Most Useful Models 334		
An Unconstrained Baseline Forecast for GLOBL Product Line B, Region A	335	
Step 3: Evaluating Model Performance Summaries 338		
Step 4: Reconciling Model Projections with Informed Judgment 339		

# Takeaways 340



# Demand Management in the Supply Chain 342

Data-Driven Demand Management Initiatives 344 Demand Information Flows 345	
<b>Creating Planning Hierarchies for Demand Forecasting</b> 346 What Are Planning Hierarchies? 347 Operating Lead Times 348 Distribution Resource Planning (DRP)—A Time-Phased Planned Order Forecast Spreadsheet Example: How to Create a Time-Phased Replenishment Plan 350	347
A Framework for Agility in Forecast Decision Support Functions The Need for Agile Demand Forecasting 352 Dimensions of Demand 353 A Database-Driven Forecast Decision Support Architecture 353 Dealing with Cross-Functional Forecast Data Requirements 356 Specifying Customer/Location Segments and Product Hierarchies 356	351
Automated Statistical Models for Baseline Demand Forecasting Automated Model Selection 359 Selecting Useful Models Visually 360 Searching for Optimal Smoothing Procedures 364 Error-Minimization Criteria 365 Searching for Optimal Smoothing Weights 365 Starting Values 366	357
Computational Support for Management Overrides 367	

#### Takeaways 369

Chapter 14 - Blending Agile Forecasting with an Integrated Business Planning Process 370

#### PEERing into the Future: Creating a Framework for Agile Forecasting in Demand Management 371

The Elephant and the Rider Metaphor371Prepare371Execute373Evaluate373Reconcile378

#### Creating an Agile Forecasting Implementation Checklist 382

383		
384		
384		
385		
ce	385	
Presenting Models to Management		
	383 384 384 385 385 ce nt	

#### **Engaging Agile Forecasting Decision Support** 386

Economic/Demographic Data and Forecasting Services386Data and Database Management 387387Modeling Assistance387Training Workshops387

#### The Forecast Manager's Checklists 388

Forecast Implementation Checklist388Software Selection Checklist389Automated Demand Forecasting Checklist390

#### Takeaways 391



# 2

# **Smarter Forecasting Is Mostly about Data**

Improving Data Quality through Data Exploration and Visualization



Time series analysis consists of all the techniques that, when applied to time series data, yield, at least sometimes, either insight or knowledge, AND everything that helps us choose or understand these procedures.

JOHN W. TUKEY, pioneer data scientist (1980)

This chapter deals with the statistical basis for data analysis in demand forecasting. As in the modeling process presented in Chapter 1, you will find that

- data analysis is open-ended and iterative in nature
- the steps may not always be clearly defined
- the nature of the process depends on what information is revealed at various stages. At any given stage, various possibilities may arise, some of which will need to be explored separately
- As a statistical methodology, much data analysis in demand forecasting is informal and exploratory.

However, it is important to realize that

- an understanding of historical data in forecasting demand will be enhanced when we can identify key patterns in a time series
- analyzing data is part of the smarter forecasting process. For example, when data contain trends, contain seasonal patterns, or have outliers, the use of some commonly used projection techniques, such as moving averages is inappropriate
- data visualizations are beneficial in describing the shape or distribution of data patterns, model residuals and forecast errors
- assuming unrealistic (non-robust) distributions for forecast errors can lead to misleading results when assessing forecasting performance. Most testing procedures implicitly assume that the underlying data follow a normal

*distribution,* and this may not often be the case in demand forecasting applications

 most demand planners and forecasters use accuracy measurements to make oversimplified use of the arithmetic mean (as in the case of the MAPE, MAD, MASE) for (*non-outlier-resistant*) forecast performance measurement, not realizing that such data rarely follow conventional statistical assumptions

# Smarter Forecasting Is Mostly about Data



When a multidisciplinary research study group at Princeton University undertook a study of the paired uses of electricity and gas in townhouses, it contacted the residents of Twin Rivers, a nearby planned community in New Jersey. Over a five-year study period, it learned how to eliminate three-quarters of the energy used by the furnace in quite ordinary, reasonably well-built townhouses, as chronicled in *Saving Energy in the Home: Princeton's Experiments at Twin Rivers*, edited by Robert H. Socolow (Cambridge, MA: Ballinger, 1977).

The purpose of the Princeton study, during a winter in the mid-1970s, was to examine differences in energy use and make comparisons with structural aspects of the 152 individual townhouses and the behavioral aspects of their inhabitants. As an applied statistician (a.k.a. data scientist), I took great delight in being a participant and was intrigued by later looking at the results and the data from the study. I was a resident at Twin Rivers at the time, not realizing that some new analysis techniques used on the data would eventually be published in 1977 in the ground-breaking book *Exploratory Data Analysis* by data science pioneer John W. Tukey (1915–2000).

The data were gathered automatically through a special device that was hooked up to the landline telephones and the energy sources in the home. There were questions to be answered periodically about our lifestyle, the details of which have long escaped my memory. Nevertheless, some novel uses of graphing techniques with schematic data plots (data visualization) can be found throughout this book. These techniques, new at the time, have now become a familiar part of many business statistics books.

#### Exploring Data Patterns

Studying the patterns in the data improves the forecaster's chances of successfully modeling data for forecasting applications. Through **exploratory data analysis** (EDA), a demand forecaster can start the important task of finding **factors** (drivers of demand) that are generally quantitative in nature.

Tukey likens EDA to detective work: "A detective investigating a crime needs both tools and understanding. If he/she has no fingerprint powder, the detective will fail to find fingerprints on most surfaces. If detectives do not understand where criminals are likely to have put their fingers, they will not look in the right places." A planned forecasting and modeling effort that does not include provisions for exploratory data analysis often misses the most interesting and important results; but it is only a first step, not the whole story.

Exploratory data analysis means looking at data, absorbing what the data are suggesting, and using various summaries and display methods to gain insight into the process generating the data.



Many business forecasting books describe a variety of classical ways to summarize data. For the practitioner, an entertaining yet informative cartoon guide covering these is Gonick and Smith's *A Cartoon Guide to Statistics,* published in 1993. For example, the familiar histogram is widely used in practice. In addition, there are a number of lesser-known techniques that are specifically useful for analyzing large quantities of data that have become accessible as a result of the increased flexibility in data management, computer processing, and predictive analytics. Because of their potential value to demand forecasting, we describe them in some detail.

#### Learning by Looking at Data Patterns

Because most forecasting techniques require data, a forecaster analyzes the availability of data from both external (outside the company) and internal (within the company or its industry) sources. For example, one potential source of internal data is a corporate data warehouse or **Enterprise Resource Planning** (ERP) system, which normally contains a rich history of product sales, shipments, prices, revenues, expenses, capital expenditures, and marketing programs.

The availability of **external data** is improving rapidly. Most of the required demographic factors (age, race, sex, households, and so forth), forecasts of **economic indicators**, and related variables can be readily obtained from computerized data sources and from industry and government publications on the Internet.

With the explosion of Internet websites, potential sources of valuable data are becoming limitless. With unstructured data, the need for data mining tools has become a necessity for exploring potential sources of data for consumer analyses and predictive modeling purposes.

# Judging the Quality of Data

The analysis of data for forecasting purposes requires a careful consideration of the quality of data sources. **Data quality** is important for modeling, because a model based on historical data will be no better than the quality of its data source.

Definitions may vary because of changes in the structure of an organization, accounting procedures, or product and service definitions. As part of the forecasting process, throughout this book various kinds of demand data will be related to economic/demographic indicators, survey data, and other external data in a number of examples and spreadsheet exhibits.

There are several criteria that can be applied to data to determine their appropriateness for modeling:

Accuracy. Proper care must be taken that the required data are collected from a reliable source with proper attention given to accuracy. Survey data exemplify the need to ensure accurate data: Survey data are collected by government agencies and research firms from questionnaires and interviews to determine future plans of consumers and businesses. The consumer confidence index in Figure 2.1 is the result of a survey made by the Conference Board (https://www.conference-board.org/data/consumerconfidence.cfm). These data have certain limitations because

they reflect only the respondents' anticipation (what they expect others to do) or expectations (what they themselves plan to do), not firm commitments. Nevertheless, such information may be regarded as a valuable aid to forecasting demand directly or as an indication of the state of consumer confidence concerning the economic outlook.



Figure 2.1 Time plot of the Conference Board's Consumer Confidence Index - US (1977 - 2014 (Units: 1985 = 100 - *Source*: https://www. Conference-board.org/); the variation in the Consumer Confidence Index shows the dominant 89% trend-cycle effect.

Based on an ANOVA decomposition (cf. MS Excel Data Analysis Add-in > ANOVA: Two-Factor without Replication), we determine that the total variability in the data is made up of 1% seasonality (Month effect), 89% trend-cycle (Year effect) and 10% unknown. When modeling we can ignore models with seasonal forecast profiles, like the Holt-Winters exponential smoothing family, but when using these data as an explanatory factor or driver of demand in causal models, we recognize its value in explaining primarily economic cycle and trending behavior.

• Conformity. The data must adequately represent the phenomenon for which it is being used. If the data purport to represent economic activity, the data should show upswings and downswings in accordance with past historical business cycle fluctuations. Data that are too smooth or too erratic may not adequately reflect the patterns desired for modeling. The Federal Reserve Board Index of Industrial Production (Figure 2.2) is an example of a cyclical indicator of the economy. It is evident that the data are consistent with economic *expansions* and *contractions* (to be discussed in Chapter 7). The index of industrial production measures changes in the physical volume of the output of manufacturers, mineral suppliers, and electric and gas utilities. The index does not cover production on farms, in the construction industry, in transportation, or in various trade and service sectors. Since the U.S. Federal Reserve Board (FRB) first introduced the index in 1920, it has been revised from time to time to take account of the growing complexity of the US economy, the availability of more data, improvements in statistical processing techniques and refinements in methods of analyses. Such indices are now widely available in many other national economies.

The ANOVA decomposition of the original (unadjusted) FRB Index of Industrial Production suggests that the total variability in the data is made up of 10% seasonality (Month effect), 89% trend-cycle (Year effect) and 1% irregular (Not identified). When using these data as an explanatory factor or driver of demand in causal models, we recognize its value in explaining primarily trend and economic cycles.



Figure 2.2 Time plot of the monthly seasonally adjusted index of industrial production from the Federal Reserve Board, July 2010–August 2014 (Units: 2007 = 100). (*Source*: Board of Governors of the Federal Reserve System - <u>http://www.federalreserve.gov/</u>)

• **Timeliness**. It takes time to collect data. Data collected, summarized, and published on a timely basis are of greatest value to the forecaster. Often preliminary data are available first, so that the time delay before the data are declared official may become a significant factor. Demographic data may fall into this category for many users. The monthly housing starts data shown in Figure 2.3 are demographic data reported by contractors and builders for use by government and private industry. Such external data are of course, subject to adjustment because of data collection delays and reporting inaccuracies.



Figure 2.3 Time plot of monthly U.S. new housing starts (seasonally adjusted annual rates). (*Source: www.census.gov*)

Once a factor has been identified as a key driver of demand, we should not assume that its usefulness is good for all times. Data need to be analyzed on an ongoing basis for quality assurance. For example, a preliminary ANOVA decomposition, while useful to assess the strength of a seasonal or trend-cycle effect, can point to a change in the composition of components of variability. For the ten-year period 1960–1969 (Figure 2.4 *left frame*), the total variability in the data is made up of 0.1 % seasonality, 64.5% trend and 35.4% irregular (unknown). On the other hand, for the ten-year period 2004–2013 (Figure 2.4 *right frame*),

the total variability in the data is made up of 0.3 % seasonality, 96.5% trend and 3.2% irregular (unknown). Because these are seasonally adjusted data, no significant seasonal effect can be present. The non-trending variability in the early period, masked by the deep dip, is suggested by the high irregular contribution. When using these data as an explanatory factor or driver of demand in causal models, we recognize its value in explaining primarily economic and demographic consumer behavior.

• **Consistency**. Data must be consistent throughout the period of their use. When definitions change, adjustments need to be made in order to retain logical consistency in historical patterns. The monthly demand for refrigerator sales, shown in Figure 2.5, is an example of internal data that would be made available to a demand forecaster of a consumer goods manufacturer. It shows a consistent (seasonal with business cycle?) pattern. If the data pattern shows abrupt level changes or unusual variation, the forecaster should check into how the data are defined.



Figure 2.4 Time plots of monthly new housing units started in the United States (seasonally adjusted annual rates for periods 1960–1969 and 2004–2013). (*Source*: www.census.gov)

In Figure 2.5 (*left frame*), the total variability in the refrigerator data is made up of 81% seasonality, 7% trend and 12% Other (unknown). It can be seen that the second seasonal peak appears unusually low (domain expert's explanation: lack of inventory resulting in lost sales). A simple interpolation between the seasonal peaks makes the data more representative of the demand.





It is important to recognize that modeling with unexamined data can distort the expected forecast profile (*change*) as well as the uncertainty (width of the prediction limits: *chance*). With the adjustment, the total variability in the data is now made up of 87%

#### CHANGE&CHANCE EMBRACED:

seasonality, 4% trend and 9% Other (unknown). The seasonality is still dominant but the unknown component, comprised of the uncertain variation is significantly reduced—the impact of just a single data point!

When using these data as demand in causal models, we recognize the need for explanatory factors explaining primarily seasonality or habit in consumer behavior, along with a measure of the nature of the uncertainty (Other).

The consequences of jumping into the modeling phase before thoroughly investigating the quality of the data and checking for anomalies can be quite severe. For instance, if we blindly applied a credible trend-seasonal model (e.g. Holt-Winters exponential smoothing or an ARIMA (011)(011)<sub>12</sub> "airline" model) to the original data, we would see that the seasonal peak profile becomes level and the uncertainty range is wide; on the other hand with a single point adjustment based on informed judgment and the same trend-seasonal model, the results become much more credible and practical (Figure 2.6). Not only is the forecast profile representative of what may lie in the future, but the prediction limits have become much narrower, indicating a tamed (reduced) uncertainty. The variability in the Other category of the exploratory ANOVA decomposition is reduced and our forecasts become more precise.





To see the rest of the Chapter, you can buy the book at Amazon.com

