A PUBLICATION OF THE INTERNATIONAL INSTITUTE OF FORECASTERS

FORESIGHT

The International Journal of Applied Forecasting

Issue 11 ■ Fall 2008

THE ESSENTIAL READ FOR THE PRACTICING FORECASTER

SPECIAL FEATURES

Benchmarking of Forecast Accuracy

Overcoming Challenges In Operational Forecasting

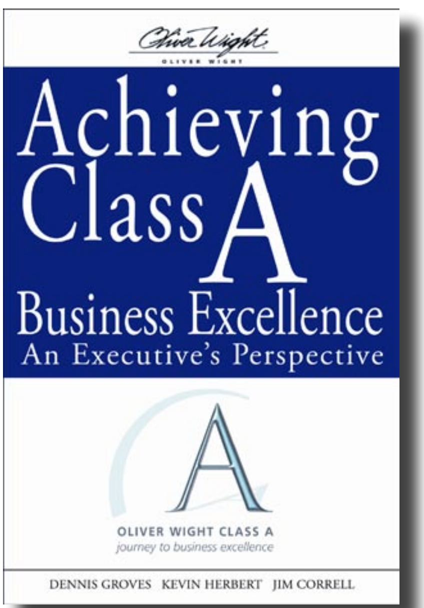
$40 per issue

Review of the International Best Seller *Super Crunchers*

The Presidential Election: Forecasts from the Statistical Models

Review of Off-the-Shelf Software for Large-Volume Forecasting

a new book from Oliver Wight – explaining the strategies and tactics of the innovative Proven Path methodology to achieve Business Excellence



- Revenue and profit falling short of projections?
- Customer service unacceptable with customers threatening to walk?
- New product introductions late with product deficiencies, disappointing the marketplace?
- Inventories overflowing the warehouse, while fill rate plummets?
- Functional managers operating in silos, protecting only their turf…

This realistic discussion shows you how to achieve a comprehensive state of excellence that will enable your company to deliver outstanding and sustainable business results.

ORDER TODAY TO RECEIVE YOUR AUTOGRAPHED COPY...WHILE SUPPLIES LAST

OLIVER WIGHT

oliverwight-americas.com
info@oliverwight.com
800-258-3862

CONTENTS

"Knowledge of truth is always more than theoretical and intellectual. It is the product of activity, as well as its cause. Scholarly reflection therefore must grow out of real problems, and not be the mere invention of professional scholars." *John Dewey, University of Vermont*

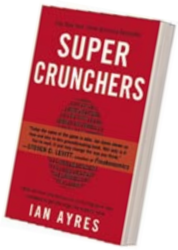
02 *Foresight* Editors and Advisory Boards

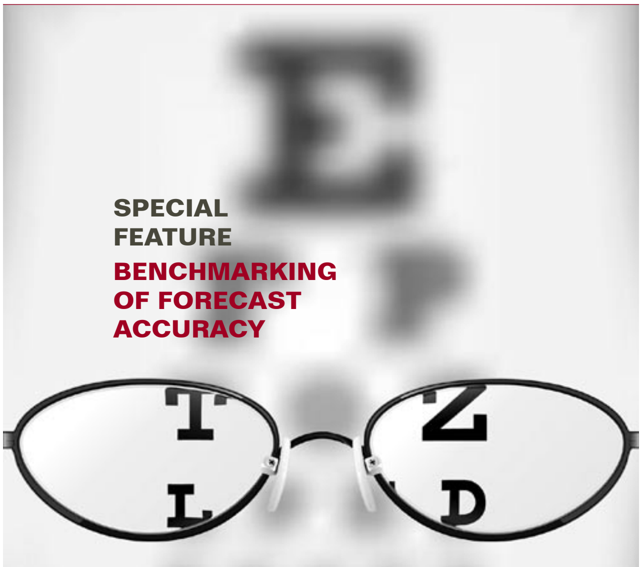
NOTE FROM THE EDITOR

03 The Fall 2008 Issue
LEN TASHMAN

BOOK REVIEW

04 *Super Crunchers* by Ian Ayres
ROY BATCHELOR





SPECIAL FEATURE

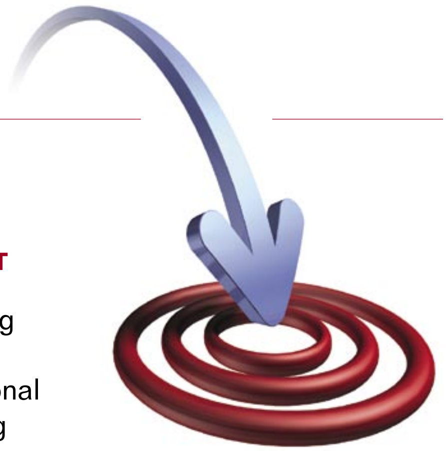
BENCHMARKING OF FORECAST ACCURACY

06 Can We Obtain Valid Benchmarks from Published Surveys of Forecast Accuracy?
STEPHAN KOLASSA

15 Measuring Improvement in Forecast Accuracy: A Case Study
ROBERT RIEG

21 Commentary on Benchmarking
TERESA MC CARTHY, DONNA DAVIS, SUSAN GOLICIC, AND JOHN MENTZER

24 Commentary on Benchmarking
JIM HOOVER



SPECIAL FEATURE

FORECAST PROCESS IMPROVEMENT

26 Overcoming Challenges in Operational Forecasting Projects
IAN WATSON-JONES

37 Commentaries on Overcoming Challenges in Operational Forecasting Projects
PATRICK WADER, MARK MOON

THE WORLD OF FORECASTING

41 Regression Model Forecasts of the U.S. Presidential Election
RANDALL JONES AND ALFRED CUZÁN

SOFTWARE REVIEW

43 *Forecast Pro Unlimited*: An Off-the-Shelf Solution for Large-Volume Forecasting
ULRICH KÜSTERS AND JANKO THYSON

FORECASTER IN THE FIELD

50 Mohsen Hamoudia

ISF2009 Hong Kong

Theme: Services Sector Forecasting

Kowloon, HONG KONG
Sheraton Hong Kong Hotel andTowers

THE HONG KONG POLYTECHNIC UNIVERSITY 香港理工大學

School of Hotel & Tourism Management 酒店及旅遊業管理學院

The International Institute of Forecasters
www.forecasters.org

School of Hotel and Tourism Management
The Hong Kong Polytechnic University
www.polyu.edu.hk/htm

General Chair

Professor Haiyan Song
The Hong Kong Polytechnic University, Hong Kong
hmsong@polyu.edu.hk

Confirmed Keynote & Featured Speakers

Professor Clive W.J. Granger
University of California, USA

Professor Rob J Hyndman
Monash University, Australia

Professor Stephen F. Witt
The Hong Kong Polytechnic University
Hong Kong

Professor J. Scott Armstrong
University of Pennsylvania

Professor Lindsay Turner
Victoria University, Australia

Symposium Theme and Papers

Time Series Models/Methods
Econometrics
Finance
Macro and Microeconomic applications
Data Mining
Judgmental Forecasting
Collaborative Forecasting
Environmental & Climate applications
Telecommunication forecasting
Business and Marketing applications
Supply Chain Forecasting
Non-parametric methods
Travel and Tourism forecasting
Demographic and Social applications
Technology and Innovation forecasting
Neural Network models

For General Enquiry Please Contact

The Symposium Secretariat
isf2009.hk@polyu.edu.hk

Pam Stroud
IIF Business Manager
isf@forecasters.org

Co-Sponsors:

PUBLIC POLICY RESEARCH INSTITUTE 公共政策研究所

HONG KONG TOURISM BOARD 香港旅遊發展局

www.TourismForecasting.Net

Editorial Statement

FORESIGHT, an official publication of the International Institute of Forecasters, seeks to advance the practice of forecasting. To this end, it will publish high-quality, peer-reviewed articles, and ensure that these are written in a concise, accessible style for forecasting analysts, managers, and students. Topics include:

- Design and Management of Forecasting Processes
- Forecast Model Building: The Practical Issues
- Forecasting Principles and Perspectives
- Integration of Forecasting Into Business Planning
- Forecasting Books, Software and Other Technology
- The World of Forecasting: Applications in Political, Climate and Media Forecasting
- Case Studies

Contributors of articles will include:

- Analysts and managers, examining the processes of forecasting within their organizations.
- Scholars, writing on the practical implications of their research.
- Consultants and vendors, reporting on forecasting challenges and potential solutions.

All invited and submitted papers will be subject to a blind editorial review. Accepted papers will be edited for clarity and style.

FORESIGHT welcomes advertising. Journal content, however, is the responsibility of, and solely at the discretion of, the editors. The journal will adhere to the highest standards of objectivity. Where an article describes the use of commercially available software or a licensed procedure, we will require the author to disclose any interest in the product, financial or otherwise. Moreover, we will discourage articles whose principal purpose is to promote a commercial product or service.

FORESIGHT is published by the International Institute of Forecasters
Business Office: 53 Tesla Avenue, Medford, MA 02155 USA

FORESIGHT
The International Journal of Applied Forecasting

EDITOR: Len Tashman
lentashman@forecasters.org

ASSOCIATE EDITOR: Nada Sanders
nadasanders@forecasters.org

FINANCIAL FORECASTING EDITOR: Roy Batchelor
R.A.Batchelor@city.ac.uk

FORECASTING INTELLIGENCE EDITOR: Roy Pearson
roy.pearson@mason.wm.edu

SUPPLY CHAIN FORECASTING EDITOR: John Boylan
jboyla01@bcuc.ac.uk

RESEARCH COLUMN EDITOR: Paul Goodwin
mnspg@management.bath.ac.uk

SOFTWARE EDITORS:
Jim Hoover
hooverjh@aol.com
Ulrich Küsters
ulrich.kuesters@t-online.de

MANAGING EDITOR: Bill Wicker
billwicker@forecasters.org

DESIGN EDITOR: Anne McLaughlin
Anne Matthew Design, amdvt@comcast.net

MANUSCRIPT EDITORS:
Mary Ellen Bridge
me.foresight@comcast.net
Ralph Culver
letter_perfect_rc@yahoo.com

SUBSCRIPTIONS: Pam Stroud, IIF Business Manager
forecasters@forecasters.org

(ISSN 1555-9068)

We are saddened to report that Steven Candiotti passed away on August 10, 2008 following a lengthy illness. Steve, of South Burlington, Vt., served as Manuscript Editor for Foresight on issues 3-6 (2006-7) before poor health forced him to step down. He also taught English at the University of Vermont in Burlington and St. Michael's College in Colchester, Vt., worked in computer programming, and had his own communications and editing business. He was 59.

Note from the Editor

The Fall 2008 Issue

The Fall 2008, Number 11 issue of *Foresight* includes two special sections: one on benchmarking of forecast accuracy and the second on the challenges of operational forecasting projects.

We lead off with Roy Batchelor's book review of *Super Crunchers* by Ian Ayres, another in the recent vintage of publications that beseech us to become more analytical in the ways we address our business challenges. If you haven't already read Roy's review of *Competing on Analytics* in our Spring 2008 issue, you'll want to check back on that one as well.

Benchmarking surveys are often cited by businesses as comparative indicators of an organization's forecasting performance. But do they provide useful performance standards? Stephan Kolassa's article analyzes the major surveys and finds their results untrustworthy as benchmarks. Stephan's discussion of the areas of noncomparability in surveys is an eye-opener. Internal benchmarking may be more reliable, and Robert Rieg presents a case study in which he tracks the forecasting performance of an automobile manufacturer over a 15-year period. The benchmarking section concludes with commentaries by Teresa McCarthy and colleagues, authors of one of the major benchmarking surveys, and Jim Hoover, who has studied the challenges of tracking forecast process improvement in military applications.

Overcoming Challenges in Operational Forecasting Projects is *Foresight*'s latest article on forecast process improvement. Author Ian Watson-Jones of IBM describes the key elements in successfully engineering a new forecasting process and gives us an invaluable checklist of process, system, and organizational hurdles. Commentaries by Mark Moon and Patrick Wader extend Ian's discussion and give particular emphasis to the overarching need for effective change management.

The U.S. presidential election is upon us, and Randall Jones and Alfred Cuzán summarize the predictions of some major **election-forecasting models**. According to most, it's a close call leaning toward the Democrats.

In the very competitive market for off-the-shelf forecasting software, Forecast Pro products have enjoyed remarkable longevity and praise from researchers for the accuracy of their automatic forecasting procedures. Software Editor Ulrich Küsters and coauthor Janko Thyson have examined **Forecast Pro Unlimited** from every angle and report on the strengths and weaknesses that they uncovered.

This issue's **Forecaster in the Field** is Mohsen Hamoudia of France Telecom, fresh from his notable success in spearheading the 2008 International Symposium on Forecasting in Nice, France. The ISF venue for June 2009 shifts to Hong Kong – see the announcement on a preceding page.

Finally, Foresight is pleased to welcome some new colleagues to its Editorial Staff and Advisory Boards.

- Roy Batchelor as Financial Forecasting Editor
- John Boylan as Supply-Chain Forecasting Editor
- Andreas Graefe, Stephan Kolassa, John Mello, and Peter Sephton to the Editorial Board
- Robert Stahl to the Practitioner Advisory Board

Len Tashman

lentashman@forecasters.org

SUPER CRUNCHERS by Ian Ayres

Subtitled in the UK: ANYTHING CAN BE PREDICTED

Subtitled in the US: WHY THINKING-BY-NUMBERS IS THE NEW WAY TO BE SMART

Reviewed by Roy Batchelor

Super Crunchers tells how simple statistical models applied to large data sets are increasingly being used to improve decision making in a wide range of business and professional applications. It highlights the fact that these models generally outperform the judgment of experienced and well-informed experts in tasks ranging from pricing airline seats, rating individuals' creditworthiness, diagnosing and treating disease, paroling prisoners, and finding a life partner, to writing a Hollywood blockbuster.

Following the spirit of the author's main premise (a premise with which I agree), one review in *Foresight* is not going to give the potential reader an optimal, unbiased evaluation of the merits of the book. So let me start by reporting that, out of 70 reviews posted on Amazon.com, 21 gave it the top 5-star rating, and its weighted score is 3.5 stars. This is pretty close to my own score. But while reading the book, I fell prey to many of the emotional reactions that make expert judgment fragile. I got angry. Then I warmed to the retelling of some familiar stories. Then I learned some cool new stuff. Then I got scared.

I read the front cover, and my heart sank. *Super Crunchers? Anything Can Be Predicted?* Are you kidding? Banks are collapsing all around us, economies careering into recession, but I don't remember all of last year's exotic option pricing formulae and clever econometric models telling us these were events we had to look forward to. Financial markets generate masses of data that are worked over by the smartest people and programs around, but I don't see any predictability emerging. Let's face it: a lot of things can't be predicted.

The title, we learn in Chapter 2, was chosen in preference to a more descriptive alternative because it would receive more hits in Google searches using the keywords "data mining" and "number crunching." This is a real shame, for two reasons. First, the book has little to do with data mining as properly understood – the search for unsuspected regularities in very large databases – and, of course, "number crunching" really doesn't mean much of anything. Second, the author clearly wanted the term *Super Crunchers* to be up there alongside other recent popular titles/buzzwords like *Blink*, *Freakonomics*, *Black Swans*, and *Tipping Point*. So while the book is otherwise written in an extremely elegant and lucid way, and I liked it a lot more than *Blink* etc., the author feels obliged to use the term "super crunching" about once per paragraph.

Roy Batchelor is HSBC Professor of Banking and Finance at Cass Business School, City University of London, and *Foresight*'s Financial Forecasting Editor. His previous contributions to *Foresight* include "A Primer on Forecasting with Neural Nets" (October 2005) and two book reviews: *Dow 36,000* (February 2006) and *Competing on Analytics* (Spring 2008). **Contact: R.A.Batchelor@city.ac.uk**

Having leaped this hurdle, I found the cases discussed are interesting and illustrate the author's argument well. Simple statistical models can help improve decision making under three circumstances. The first occurs when, by good fortune, a relevant cross-sectional database exists, and modern technology and freedom of information laws let us access and process the data. For example, one of Ayres's own published studies integrated credit company and demographic databases to demonstrate racial discrimination in the terms offered to auto purchasers, as described in Chapter 6.

A second scenario illustrating when simple statistical modeling can be helpful arises when data is generated from deliberate, controlled experiments. Evidence-based medicine (discussed in Chapter 4) has developed from pooling results of drug-company trials and patient outcomes. Evidence-based teaching (Chapter 7) in the U.S. has emerged from large, publicly funded studies of the achievements of children under sharply different learning regimes. The credit company CapOne (Chapter 2) generates business information by regularly sending offers of new products and new contract terms to test groups of customers, and then compares their responses with control groups that do not receive the offers.

ABOUT *SUPER CRUNCHERS* AUTHOR IAN AYRES

Ian Ayres is a lawyer and an economist. He is the William K. Townsend Professor at Yale Law School and a Professor at Yale's School of Management. He has published 9 books and over 100 articles on a wide range of topics.

The third situation emerges when, rather than building models directly, we model the way experts reach decisions, seeking the best predictors of expert decisions. Starting with Paul Meehl's famous monograph published back in 1954, all the evidence supports the proposition that simple models of expert decisions outperform the experts themselves. While experts appear to have more soft, nonquantitative information than computers, experts also forget more stuff, and can be cranky and inconsistent. On balance, the machines win. This powerful and counterintuitive idea is discussed in Chapter 5.

Ian Ayres (2007 in hardback, 2008 in paperback)
SUPER CRUNCHERS
John Murray Publishers, UK
Bantam Dell, A Division of Random House, NY
ISBN: Hardback: 9780719564635;
Paperback 978-0-553-38473-4

To his credit, Ian Ayres does not shy away from discussing the awkward implications of the power and increasing use of quantitative modeling in marketing and decision making. Data on individuals is now routinely collected and shared (and lost) by many government agencies and commercial firms. Respect for individual freedom and privacy is thereby diminished, and opportunities for identity theft and other cybercrime are multiplied. There has been little weighing of the costs and benefits of this process, and we seem powerless to stop it, or even rein it in.

It is a fact of life that the use of machine-generated rules for professional behavior diminishes the social status and *amor propre* of the expert. Such is human nature that doctors have proved reluctant to follow protocols created in this way, even when those protocols demonstrably improve outcomes for patients. Teachers, too, have shown themselves to be highly resistant to the outcomes of experiments on teaching methodologies, because the most successful technique in terms of student performance – direct instruction – allows teachers the least discretion and creativity. And so "super crunching" (damn!) can be expected to keep turning up more and more insights, forecasts, and decision rules from ever larger agglomerations of data. The challenge, for managers in both the public sector and in business, is, and will continue to be, to incentivize people to do what the machines tell them.

REFERENCE

Meehl, P. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press. Reprinted with new Preface, 1996, by Jason Aronson, Northvale, NJ. http://www.psych.umn.edu/faculty/meehlp/032ClinstixBook.pdf

CAN WE OBTAIN VALID BENCHMARKS FROM PUBLISHED SURVEYS OF FORECAST ACCURACY?

Stephan Kolassa

PREVIEW

Organizations often seek *benchmarks* to judge the success of their forecasts. Reliable benchmarks would allow the company or agency to see if it has improved upon industry standards and to evaluate whether investment of additional resources in forecasting would be money well spent. But can the existing benchmark surveys be trusted? "No," says Stephan Kolassa, who has analyzed the surveys and found them seriously deficient. In this article Stephan explains the many problems that plague benchmark surveys and advises that companies should redirect their search from external to internal benchmarks since the latter provide a better representation of the processes and targets the company has in place.

KEY POINTS

- In *benchmarking*, comparability is the key. Benchmarks can be trusted only if the underlying process to be benchmarked is assessed in similar circumstances.
- Published surveys of forecast accuracy are not suitable as benchmarks because of incomparability in product, process, time frame, granularity, and key performance indicators.
- It is doubtful that forecasting accuracy benchmarks can be compiled from cross-company surveys because the hurdles of establishing comparability are formidable.
- Quantitative targets themselves may be elusive. A better alternative for forecast improvement is a qualitative, process-oriented target. By focusing on process improvement, forecast accuracy and the use an organization makes of the forecasts will eventually be improved.

INTRODUCTION

Sales forecasters are frequently asked what a "good" forecast is; that is, what accuracy should be expected from the forecasting method or process?

This question is important for deciding how to allocate resources to the firm's forecasting function or forecast-improvement projects. If forecast accuracy is already as good as it can reasonably be expected to

Stephan Kolassa is Vice President of Corporate Research at SAF AG in Switzerland. He has worked extensively with some of Europe's largest retail chains in producing automatic forecasts for large batches of products. Stephan and his colleague Wolfgang Schütz coauthored "Advantages of the MAD/MEAN Ratio Over the MAPE" in *Foresight*'s Spring 2007 issue.

be, spending additional resources would be wasteful. Thus the company can benefit from true benchmarks of forecasting accuracy.

By true benchmarks, I mean reliable data on the forecast accuracy that can be achieved by applying best practices in forecasting algorithms and processes. Unfortunately, published reports on forecasting accuracy are rare, and those that exist suffer from shortcomings that sharply limit their validity in providing forecast-accuracy benchmarks. Consequently, I believe it is a mistake to use benchmark surveys.

PUBLISHED SURVEYS OF FORECAST ACCURACY

The McCarthy Survey

Teresa McCarthy and colleagues (McCarthy et al., 2006) studied the evolution of sales forecasting practices by conducting surveys of forecasting professionals in 1984, 1995, and 2006. Their results (see Table 1) provide some evidence on forecast accuracy both longitudinally and at various levels of granularity, from SKU-by-location to industry level. The forecast horizons shown are (a) up to 3 months, (b) 4-24 months, and (c) greater than 24 months. The number of survey responses is denoted by n. All percentage figures are Mean Absolute Percentage Errors (MAPEs).

One of the study's general conclusions is that the accuracy of short-term forecasts generally deteriorated over time, as shown by the weighted-average MAPEs in the bottom row. Considering the ongoing and vigorous research on forecasting, as well as vastly improved computing power since 1984, this finding is surprising. The McCarthy team conjectured that the deterioration could be due to decreasing familiarity with complex forecasting methods (as they found via interviews), product proliferation, and changes in the metrics used to measure forecast accuracy over the past 20 years.

Indeed, the survey results do suffer from problems of noncomparability. For one, the numbers of respondents in 1995 and especially in 2006 were much lower than those in 1984. In addition, I presume that the participants in 2006 differed from those in 1984 and 1995, so that lower forecast quality could simply reflect differences in respondents' companies or industries. For example, the meaning of "SKU-by-location" may have been interpreted differently by respondents in different companies and industries. Similarly, "Product Line" and "Corporate" forecasts may mean different things to different respondents.

So while the McCarthy survey provides some perspective on forecast accuracy at different times and levels, the usefulness of the figures as benchmarks is limited.

The IBF Surveys

The Institute of Business Forecasting regularly surveys participants at its conferences. The most recent survey results are reported in Jain and Malehorn (2006) and summarized in Table 2. Shown are MAPEs for forecast horizons of 1, 2, 3, and 12 months in different industries, together with the numbers of respondents. Jain (2007) reports on a similar survey taken at a 2007 IBF conference. The results are given in Table 3.

Table 1. MAPEs for Monthly Sales Forecast in 1984, 1995 and 2006 Surveys

Horizon	≤ 3 months			4 to 24 months			> 24 months		
Forecast Level	1984	1995	2006	1984	1995	2006	1984	1995	2006
Industry	8% n = 61	10% n = 1	15% n = 1	11% n = 61	12% n = 16	16% n = 10	15% n = 50	13% n = 36	7% n = 3
Corporate	7% n = 81	28% n = 2	29% n = 5	11% n = 89	14% n = 64	16% n = 31	18% n = 61	12% n = 42	11% n = 8
Product line	11% n = 92	10% n = 4	12% n = 6	16% n = 95	14% n = 83	21% n = 34	20% n = 60	12% n = 25	21% n = 5
SKU	16% n = 96	18% n = 14	21% n = 5	21% n = 88	21% n = 89	36% n = 36	26% n = 54	14% n = 10	21% n = 3
SKU by location		24% n = 17	34% n = 7		25% n = 58	40% n =22		13% n = 5	
Weighted average	15%	16%	24%						

Source: McCarthy et al. (2006)

Table 2. MAPEs for Monthly Sales Forecast Source: Jain & Malehorn (2006, Table 6.2)

Horizon	1 month			2 months			1 quarter			1 year		
Level	SKU	Category	Aggregate	SKU	Category	Aggregate	SKU	Category	Aggregate	SKU	Category	Aggregate
Automotive	25% n = 3	5% n = 1	36% n = 1	31% n = 3	33% n = 2	25% n = 2	42% n = 1			46% n = 1		10% n = 1
Computer/ Technology	19% n = 4	14% n = 4	12% n = 7	33% n = 2	11% n = 2	18% n = 4	30% n = 3	16% n = 4	25% n = 6	17% n = 2	30% n = 1	31% n = 4
Consumer Products	27% n = 35	20% n = 23	15% n = 21	29% n = 20	22% n = 14	15% n = 10	33% n = 11	23% n = 7	14% n = 6	48% n = 4	19% n = 4	8% n = 3
Food/ Beverages	26% n = 16	15% n = 10	18% n = 11	28% n = 10	22% n = 4	36% n = 5	26% n = 8	21% n = 3	40% n = 4	19% n = 4	14% n = 2	48% n = 3
Healthcare	25% n = 7	15% n = 6	9% n = 6	27% n = 5	19% n = 5	17% n = 5	41% n = 5	24% n = 5	25% n = 5	30% n = 2	20% n = 2	15% n = 2
Industrial Products	22% n = 4	15% n = 7	7% n = 8	16% n = 2	14% n = 5	8% n = 6	17% n = 3	15% n = 6	10% n = 7	40% n = 2	21% n = 5	15% n = 6
Pharma	26% n = 5	20% n = 4	23% n = 4	30% n = 3	35% n = 2	33% n = 2	31% n = 4	25% n = 4	25% n = 3	34% n = 4	35% n = 4	28% n = 3
Retail	24% n = 7	18% n = 4	7% n = 4	17% n = 5	17% n = 6	8% n = 4	24% n = 4	10% n = 3	9% n = 4	23% n = 4	6% n = 2	6% n = 3
Telco				30% n = 1	10% n = 1	30§ n = 1	40% n = 1	15% n = 1	35% n = 1			
Others	28% n = 13	21% n = 9	17% n = 16	23% n = 7	20% n = 5	11% n = 10	25% n = 6	15% n = 5	14% n = 9	15% n = 4	18% n = 4	12% n = 8
Overall	26% n = 94	18% n = 68	13% n = 80	27% n = 58	20% n = 46	15% n = 51	30% n = 46	19% n = 37	17% n = 45	29% n = 27	21% n = 24	16% n = 33

Tables 2 and 3 show large differences in forecasting accuracy among industries. For instance, the retail sector shows much lower errors than the more volatile computer/technology sector, especially for longer horizons. In general, the results show that forecast accuracy improves as sales are aggregated: forecasts are better on an aggregate level than on a category level and better on a category level than for SKUs. And, while we should expect forecast accuracy to worsen as the horizon lengthens, the findings here are not always supportive. For example, at the Category and Aggregate levels in Consumer Products (Table 2), the 1-year-ahead MAPEs are lower than those at shorter horizons.

Unfortunately, the validity of these results is again problematic. The sample sizes were very small in many categories (Table 2), reflecting a low response rate by the attendees. Jain (2007) does not even indicate the number of responses behind the results in Table 3. In addition, these tables are based on surveys done at IBF conferences – which, after all, are attended by companies that are sensitive enough to the strategic value of forecasting to attend conferences on forecasting! Thus the MAPEs may not reflect *average* performance, but instead may represent lower errors at better-performing companies. Finally, while the forecast errors are shown separately for different industries – and one clearly sees large differences across industries – the industry categories are broadly defined and encompass a range of types of companies and products.

The M-Competitions

Since 1979, Spyros Makridakis and Michèle Hibon have been coordinating periodic forecasting competitions, the so-called M-Competitions. Three major competitions have been organized so far, with forecasting experts analyzing 1001 time series in the M1-Competition, 29 in the M2-Competition, and 3003 in the M3-Competition.

Table 3. MAPEs for Monthly Sales Forecast Source: Jain (2007)

Horizon	1 month			2 months			1 quarter			1 year		
Level	SKU	Category	Aggregate	SKU	Category	Aggregate	SKU	Category	Aggregate	SKU	Category	Aggregate
Consumer Products	29%	19%	16%	31%	20%	16%	35%	23%	22%	35%	28%	21%
Food & Beverages	27%	24%	24%	22%	12%	11%	23%	14%	15%	29%	18%	18%
Industrial Products	19%	17%	16%	28%	24%	18%	29%	22%	18%	36%	30%	17%

Table 4. MAPEs for Monthly Sales Forecast

Source: Makridakis et al. (1993)

Company	Industry	Number of series	Forecast	1 month	2 months	1 quarter	1 year
Honeywell	Residential construction	6	Average	N/A	16.6%	15.9%	19.3%
			Best (Naive method including seasonality)	N/A	5.1%	6.7%	13.5%
Squibb	Pharma	7	Average	N/A	9.1%	10.6%	28.1%
			Best (Smoothing with dampened trend)	N/A	7.3%	7.2%	23,0%
Car company	Automotive	6	Average	10.1%	10.7%	14.6%	13.9%
			Best (Smoothing with dampened trend)	8.0%	9.5%	14.6%	14.2%
Aussedat-Rey	Paper	4	Average	3.7%	5.6%	6.8%	5.2%
			Best (Combination of smoothing methods)	2.8%	5.9%	6.7%	3.8%

I will restrict the analysis here to the M2-Competition (Makridakis et al., 1993), which featured 23 series of company sales data. It attempted to model closely the actual forecasting process used in firms: forecasters could include causal factors and judgmentally adjust statistical forecasts, and they were encouraged to contact the participating companies and obtain additional information which might influence sales. Table 4 shows the resulting MAPEs for monthly forecasts across different horizons, both for the average of 17 forecasting methods and for the "best" method (which I define here as the method that gave the best results, on average, across horizons up to 15 months ahead).

The table reveals that forecast accuracy varied considerably across the four companies on a 1-year horizon, the best method yielding a MAPE of 23% for the pharma data and 3.8% for the paper data. The authors attributed the variations to different seasonalities and noise levels in the data, with pharma sales fluctuating much more strongly than paper sales. Unsurprisingly, forecast accuracy generally deteriorated as forecast horizons increased. Finally, quite simple methods – a naïve forecast, exponential smoothing with a dampened trend, or a combination of smoothing methods – beat more complex methods, including human forecasters using market information and judgmental adjustments. In particular, the Honeywell dataset showed that a simple, seasonally adjusted naïve method could be more accurate than other methods that were more complex.

However, even the results of the M2-Competition are problematic candidates for forecasting benchmarks. These companies represent a very small sample of industries, and the sample contains only one company per industry. In addition, very few time series per company were considered; for example, the only Honeywell series included were channel sales of a safety device and fan control. The latter makes it problematic even to extrapolate, from the MAPEs on the series chosen, the accuracy achievable for other Honeywell products.

Another problem is that very different series are being averaged. For instance, the six series for the car manufacturer include not only sales of three individual models (without specification of whether sales were national or international), but also total company sales and the total of the entire car industry. Conceivably, a method may forecast well for the entire automobile industry but break down when forecasting sales of a single model – a situation where life cycles need to be taken into account, although they may be less important on the aggregate level.

Finally, even though forecasting experts were encouraged to contact the companies for additional explanation and data, some experts consciously decided not to. They doubted that a sufficient understanding of the companies' markets could be formed within a short period ("…it was hard to know what questions we should ask…."). Subsequently, they acknowledged that their forecast was "not comparable with the likely accuracy of a judgmental forecast prepared within a business organization" (Chatfield et al., 1993).

Makridakis and colleagues never intended the results of the M-Competitions to be used as benchmarks against which forecasting performance of companies should be measured. Instead, the M-Competitions aimed at comparing different forecasting algorithms on standardized datasets. Their failure to provide benchmarks does not mean the results are uninformative to practicing forecasters. On the contrary, they guide practitioners to consider relatively simple methods when seeking to improve their methodologies.

WHAT IS A BENCHMARK?

The concept of benchmarking is widely applied in business fields, from process benchmarking and financial benchmarking to IT performance benchmarking of new hardware. Common to any such endeavor is that measures of performance in similar and comparable fields are collected and analyzed in order to gain an understanding of what the best possible performance is.

In benchmarking, comparability is the key! Benchmarks can only be trusted if the underlying process to be benchmarked is assessed in similar circumstances. For instance, benchmarking profitability across "firms in general" fails the criterion of comparability; biotech and utility companies have widely different "normal" profitabilities, and using the best-in-class profitability of a biotech firm as a target for a utility is unrealistic.

Benchmarking is closely related to the search for *best practices*. Ideally, one would identify a performance benchmark and then investigate what factors enable achievement of the benchmark (Camp, 1989). For instance, an optimal sales forecast may be a result of very different factors: a good process for data collection, a sophisticated forecasting algorithm, or simply a clever choice of aggregating SKUs across stores and/or warehouses.

Any approach that leads to consistently superior forecasting performance would be a candidate for best practices. As forecasters, our search for benchmarks is really only part of our search for best practices. We try to optimize our forecasts and need to understand which part of our processes must be improved to reach this goal.

PROBLEMS WITH FORECAST ACCURACY SURVEYS

Can published figures on sales forecasting accuracy serve as benchmarks? My analysis indicates that the survey results suffer from multiple sources of incomparability in the data on which they are based. These include differences in industry and product, in spatial and temporal granularity, in forecast horizon, in metric, in the forecast process and in the business model.

Product Differences. Going across industries or even across companies, we have to forecast sales of wildly dissimilar products. Sales of canned soup and lawn mowers behave very differently; their forecasting challenges will be different, too. A manufacturer of canned soup may be faced with minor seasonality as well as sales that are driven by promotional activities whose timing is under the manufacturer's control. Lawn mower sales, however, will be highly seasonal, depending crucially on the weather in early summer. Thus, it's reasonable to expect lawn mower sales to be more difficult to forecast than canned soup sales and to expect that even "good" forecasts for lawn mowers will have higher errors than "good" forecasts for canned soup.

The comparability problem arises when both canned soup and lawn mowers are grouped together as *consumer products* or products sold by the *retail industry*. This is nicely illustrated by the differences between the company datasets in the M2-Competition (Table 4). In addition, as I noted above, separate products of a single company may vary in forecastability. A fast-moving staple may be easily forecastable, while a slow-moving, premium article may exhibit intermittency – and consequently be harder to forecast.

Forecasts, moreover, are not only calculated for products, but also for services and/or prices. For manpower planning, a business needs accurate forecasts for various kinds of services, from selecting products for a retailer's distribution center to producing software. And in industries where price fluctuation is strong, forecasting prices can be as important as forecasting quantities. Problems of comparability may apply to price forecasts as well as to quantity forecasts. Although most published surveys have focused on quantities of nonservice products, we can clearly see that benchmarking forecasts of services and prices face similar challenges.

Spatial Granularity. Published accuracy figures do not precisely specify the level of "spatial" granularity. When it comes to SKU-by-location forecasts, are we talking about a forecast for a single retail store, a regional distribution center (DC), or a national DC? Forecasting at all three locations may be important to the retailer. Forecasts at the national DC level will usually be of most interest to the manufacturer, as this is the demand from the retailer he normally faces – unless, of course, the manufacturer engages in direct store delivery (DSD), in which case he will certainly be interested in store-level sales and, it logically follows, store-level forecasts.

Aggregating sales from the retail stores serviced by a regional or national DC will usually result in more stable sales patterns. Consequently, forecasting at the retail store will usually be much harder than for the national DC. A given forecast error may be fine for a store forecast but unacceptably large for a DC forecast. Similarly, it will be easier to forecast car sales of General Motors in a mature and stable market, compared to car sales by a smaller company like Rolls-Royce, which builds limited runs of luxury cars for sale to aficionados.

Temporal Granularity. The time dimension of the forecasts reported in the surveys is often vague. Are the forecasts calculated for monthly, weekly, daily, or even intradaily sales? Forecasts for single days are important for retailers who need to replenish shelves on a daily basis, while weekly forecasts may be enough for supplying regional DCs. Manufacturers may only need to consider monthly orders from retailers' national DCs, but once again, in the case of DSD, they will need to forecast on a weekly or even daily level.

Just as aggregation of store sales to DC sales makes forecasting easier at the DC than in the store, it is usually easier to forecast monthly than weekly sales, easier to forecast weekly sales than daily sales, easier to forecast daily sales than intradaily sales. A given accuracy figure may be very good for a daily forecast but very bad for a monthly one.

Longer-term forecasting is harder than shorter-term, simply because the target time period is farther into the future. And long-range forecasts may differ in temporal granularity from short-range forecasts: often, a retailer forecasts in daily (or even intradaily) buckets for the immediate next few weeks, on a monthly basis for forecasts 2-12 months ahead, and in quarterly buckets for the long term. These forecasts correspond, respectively, to operational forecasts for store ordering and shelf replenishment, to tactical forecasts for distribution center orders, and to strategic forecasts for contract negotiations with the supplier.

This example clearly illustrates that forecasts with different horizons may have different purposes and different users and be calculated based on different processes and algorithms. It's important to note that errors on different time horizons may have different costs: an underforecast for store replenishment will lead to an out-of-stock of limited duration, but an underforecast in long-range planning may lead a retailer to delist an item that might have brought in an attractive margin.

Key Performance Indicators (KPIs). The published surveys employ the MAPE – or a close variation thereof – as the "standard" metric for forecast accuracy. In fact, there is little consensus on the "best" metric for sales forecast accuracy. While the MAPE is certainly the most common measure used in sales forecasting, it does have serious shortcomings: asymmetry, for one, and error inflation if sales are low. These shortcomings have been documented in earlier *Foresight* articles by Kolassa and Schütz (2007), Valentin (2007), and Pearson (2007), who proposed alternative forecast-accuracy metrics. Catt (2007) and Boylan (2007) go

Forecasting is an art which depends on good methods/algorithms *and* on sophisticated processes. Using results from purely scientific forecasting competitions will be difficult, as these competitions are often dissociated from the processes of the company that provided the data.

further, encouraging the use of cost-of-forecast-error (CFE) metrics in place of forecast-accuracy metrics.

Because of the proliferation of forecast-accuracy metrics, you can't be certain if survey respondents have actually correctly calculated the metric reported.

Then there's the asymmetry problem. Overforecasts (leading to excess inventory) and underforecasts (lost sales) of the same degree may have very different cost implications, depending on the industry and the product. Excess inventory may cost more than lost sales (as with short-life products like fresh produce, or high-tech items that quickly become obsolete), or it can be the other way around (e.g., for canned goods or raw materials). The MAPE and its variants, which treat an overforecast of 10% the same as an underforecast of 10%, may not adequately address the real business problem. KPIs that explicitly address over- and underforecasts may be more meaningful to forecast users.

Forecast Horizon. Most studies report the forecast horizon considered; I wish all of them did. Many different forecast horizons may be of interest for the user, from 1-day-ahead forecasts for the retailer to restock his shelves, to 18-months-ahead (and more) forecasts for the consumer-product manufacturer who needs to plan his future capacity and may need to enter into long-term contractual obligations.

Forecast Processes. Forecasting accuracy is intimately related to the *processes* used to generate forecasts, not only to the algorithmic *methods*. In the past 25 years, forecasters have tried a number of ways to improve accuracy within a company's forecasting process, from structured judgmental adjustments and statistical forecasts (Armstrong, 2001) to collaborative planning, forecasting and replenishment (CPFR) along the supply chain (Seifert, 2002). Yet the published surveys on forecast accuracy do not differentiate between respondents based on the maturity of their processes, whether a full-fledged CPFR effort or a part-time employee with a spreadsheet.

Benchmarking is deeply connected to process improvement (Camp, 1989). The two are, in a sense, inseparable. It follows that, as long as information on forecasting processes is not available, we really do not know whether reported MAPEs are "good" or "bad." Forecasting is an art which depends on good methods/algorithms *and* on sophisticated processes. Using results from purely scientific (what could be called *in vitro* or lab-based) forecasting competitions such as the M-Competitions or the recent competitions on Neural Network forecasting as benchmarks (Bunn & Taylor, 2001) will be difficult, as these competitions are often dissociated from the processes of the company that provided the data.

Business Model. The published surveys of forecast accuracy have examined business-to-consumer (B2C) sales in retail. In retail, we can only observe sales, not demand – if customers do not find the desired product on the shelf, they will simply shop elsewhere, and the store manager will usually be unaware of the lost sale. The information basis on which a forecast can be calculated is therefore reduced. We may want to forecast *demand* but only be able to observe historical *sales*.

This so-called *censoring* problem is especially serious for products where the supply cannot be altered in the short run, such as fresh strawberries. We may have a wonderful forecast for customer demand but miss sales by a large margin, simply because the stock was not high enough. Thus, comparing the accuracy of a strawberry sales forecast with a napkin sales forecast will be inappropriate: the censoring problems are more serious for strawberries than for napkins.

By contrast, in a business-to-business (B2B) environment, we often know the historical orders of our business clients, so even if the demand cannot be satisfied, we at least know how high it was. Therefore, B2B forecasts profit from much better historical data and should be more accurate than B2C forecasts. Any published benchmarks on forecasts for products that could be sold either B2B or B2C are consequently harder to interpret than forecasts for "pure" B2B or B2C products.

Moreover, in a build-to-order situation one may not even know the specific end-products that will be sold in the future. Here it makes sense to either forecast on a component level or to forecast sales volume in dollars rather than in units.

To summarize, none of the published sales forecasting studies can be used as a benchmark. All published indicators suffer from serious shortcomings regarding comparability of data and processes in which forecasts are embedded, as each industry and each company faces its own forecasting problems with its distinctive time granularity, product mix and forecasting processes. The issues of incomparability have been recognized for many years (Bunn & Taylor, 2001) but have not been solved.

All studies published to date have averaged sales forecasts calculated on widely varying bases, used poorly defined market categories, and ignored the underlying forecast processes at work. These shortcomings are so severe that, in my opinion, published indicators of forecast accuracy can only serve as a very rudimentary first approximation to real benchmarks. One cannot simply take industry-specific forecasting errors as benchmarks and targets.

EXTERNAL VS. INTERNAL BENCHMARKS

Are the survey problems of comparability resolvable? Could we, in principle, collect more or better data and create "real" benchmarks in forecasting?

The differences between companies and products are so large that useful comparisons among companies within the same market may be difficult to impossible. For instance, even in the relatively homogeneous field of grocery-store sales forecasting, I have seen "normal" errors for different companies varying between 20% and 60% (MAPE for 1-week-ahead weekly sales forecasts), depending on the number of fast sellers, the presence of promotional activities or price changes, the amount of fresh produce (always hard to forecast), data quality, etc. Thus comparability between different categories and different companies is a major stumbling block.

In addition, industries differ sharply on how much information they are willing to provide to outsiders. I have worked with retailers who threatened legal action if my company disclosed that they were considering implementing an automated replenishment system. These retailers considered their forecasting and replenishment processes as so much a part of their competitive edge that there was no possibility of publishing and comparing their processes, even anonymously. It simply was not to be done. This problem is endemic in the retail market and makes benchmarking very difficult. It may be less prevalent in other markets, but it is still a problem.

My conclusion is that the quest for external forecasting benchmarks is futile.

So what should a forecaster look at to assess forecasting performance and whether it can be improved? I believe that benchmarking should be driven not by external accuracy targets but by knowledge about what constitutes good forecasting practices, independent of the specific product to be forecast.

The article by Moon, Mentzer, and Smith (2003) on conducting a sales forecasting audit and the commentaries that follow it serve as a good starting point to critically assess a company's forecasting practices and managerial environment. It's important to note that no one – not the authors of the paper, not the commentators, and none of the other works made reference to – recommended that you rely upon or even utilize external forecast accuracy benchmarks. When discussing the "should-be" target state of an optimized forecasting process, they express the target in qualitative, process-oriented terms, not in terms of a MAPE to be achieved. Such a process-driven forecast improvement methodology also helps us focus our attention on the processes to be changed, instead of the possibly elusive goal of achieving a particular MAPE.

Forecast accuracy improvements due to process and organizational changes should be monitored over time. To support the monitoring task, one should carefully select KPIs that mirror the actual challenges faced by the organization. And historical forecasts as well as sales must be stored, so that you can answer the question, "How good were our forecasts for 2008 that were made in January of that year?" We can then evaluate whether, and by how much, forecasts improved as a result of an audit, a change in algorithms, the introduction of a dedicated forecasting team, or some other improvement project.

In summation, published reports of forecast accuracy are too unreliable to be used as benchmarks, and this situation is unlikely to change. Rather than look to external benchmarks, we should critically examine our internal forecast processes and organizational environment. If we focus on process improvement, forecast accuracy and the use an organization makes of the forecasts will eventually be improved.

CONTACT
Stephan Kolassa
SAF AG, Tägerwilen, Switzerland
stephan.kolassa@saf-ag.com

REFERENCES

Armstrong, J.S. (2001). *Principles of Forecasting: A Handbook for Researchers and Practitioners*, New York, NY: Springer.

Boylan, J. (2007). Key assumptions in calculating the cost of forecast error, *Foresight: The International Journal of Applied Forecasting*, Issue 8, 22-24.

Bunn, D.W. & Taylor, J.W. (2001). Setting accuracy targets for short-term judgemental sales forecasting, *International Journal of Forecasting*, 17, 159-169.

Camp, R.C. (1989). *Benchmarking: The Search for Industry Best Practices That Lead to Superior Performance*, Milwaukee, WI: ASQC Quality Press.

Catt, P. (2007). Assessing the cost of forecast error – a practical example, *Foresight: The International Journal of Applied Forecasting*, Issue 7, 5-10.

Chatfield, C., Hibon, M., Lawrence, M., Mills, T.C., Ord, J.K., Geriner, P.A., Reilly, D., Winkel, R. & Makridakis, S. (1993). A commentary on the M2-Competition, *International Journal of Forecasting*, 9, 23-29.

Jain, C.L. (2007). Benchmarking forecast errors, *Journal of Business Forecasting*, 26(4), Winter 2007/2008, 19-23.

Jain, C.L. & Malehorn, J. (2006). *Benchmarking Forecasting Practices: A Guide To Improving Forecasting Performance* (3rd ed.), Flushing, NY: Graceway.

Kolassa, S. & Schütz, W. (2007). Advantages of the MAD/MEAN ratio over the MAPE, *Foresight: The International Journal of Applied Forecasting*, Issue 6, 40-43.

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K. & Simmons, L. F. (1993). The M2-Competition: A real-time judgmentally based forecasting study, *International Journal of Forecasting*, 9, 5-22.

McCarthy, T.M., Davis, D.F., Golicic, S.L. & Mentzer, J.T. (2006). The evolution of sales forecasting management: A 20-year longitudinal study of forecasting practice, *Journal of Forecasting*, 25, 303-324.

Moon, M.A., Mentzer, J.T. & Smith, C.D. (2003). Conducting a sales forecasting audit (with commentaries), *International Journal of Forecasting*, 19, 5-42.

Pearson, R. (2007). An expanded prediction-realization diagram for assessing forecast errors, *Foresight: The International Journal of Applied Forecasting*, Issue 7, 11-16.

Seifert, D. (2002). *Collaborative Planning, Forecasting and Replenishment*, Bonn, Germany: Galileo.

Valentin, L. (2007). Use scaled errors instead of percentage errors in forecast evaluations, *Foresight: The International Journal of Applied Forecasting*, Issue 7, 17-22.

MEASURING IMPROVEMENT IN FORECAST ACCURACY

A CASE STUDY

Robert Rieg

PREVIEW

Over the past 15-20 years, improvements in forecasting methods, deepening practical experience, and increasing computing power should have allowed companies to significantly improve their forecasting accuracy. In this paper Robert Rieg examines the changes in forecasting accuracy of a large automobile manufacturer between 1991 and 2005. His analysis shows how a company can examine its track record over time and emphasizes the need to distinguish internal from external factors that impinge on forecasting accuracy.

IMPROVING FORECASTING ACCURACY OVER TIME

There are four basic means for improving forecast accuracy over time:

(1) Use better forecasting methods/algorithms.
(2) Acquire better software and hardware.
(3) Learn from past experience and mistakes.
(4) Reduce the uncertainty in the forecasting environment.

Deterioration in forecasting performance, or at least an absence of evidence of improvement, could occur despite reasons (1) and (2) because of an increasingly uncertain environment or loss of organizational knowledge. Such a result should prompt deeper analysis of the underlying factors. Are they internal factors, such as change of processes or use of inappropriate methods? If so, the problem is resolvable. However, external factors, such as an increasingly uncertain forecast environment, are considerably more difficult to resolve.

(1) Better forecasting methods/algorithms. The passage of time has seen a considerable enhancement of the toolbox of forecasting methods. But method upgrades do not automatically lead to better predictions. In the M3 forecasting competition (Makridakis & Hibon, 2000) which compared the performance of common forecasting methods on large, diverse data sets, newer, more sophisticated methods like Box-Jenkins and Artificial Neural Networks failed to outperform older and simpler methods such as Exponential Smoothing. Armstong (2006) cites an analysis that data mining, a very complex methodology, fails to improve even upon "random guessing." It is possible that the accuracy gained in upgrading forecast-method selection is initially very high, but the additional returns to increasing sophistication are negligible.

(2) Increased computing power and sophisticated software available at low costs. Companies can now process and store more data with ever-more-complex algorithms in a shorter period of time (Küsters et al., 2006). Some empirical studies show that the use of appropriate software can lay the groundwork for improved forecasts (Sanders & Manrodt, 2003), especially when the organization shifts from paper-based forecasts or spreadsheets to dedicated forecasting software.

(3) Improved learning, training, and knowledge sharing. New methods and software will prove beneficial only if organizations use them in a sensible way. Forecast quality can improve through training, as

Robert Rieg is a Professor of Management Accounting and Control and currently Dean of the Faculty of Business at Aalen University, Germany. He is interested in forecasting for planning, budgeting, and accounting purposes. His latest research project is a panel study on prediction markets.

KEY POINTS

• Improvement over time in an organization's forecasting accuracy can have several sources: better methods, better software and hardware, an improved learning curve, and reduced uncertainty in the organization's environment.

• A major longitudinal study suggested that, over the 20-year period ending in 2006, forecast accuracy had not improved but deteriorated, due partly to over-reliance on forecasting software and failure to appreciate the role of organizational processes and training.

• However, this survey is of dubious validity because it compares the performance of different actors at different times. In place of a longitudinal survey, a case study has promise because it concentrates on one company and allows for application of appropriate metrics for forecasting errors.

• In my case study at a large German automotive manufacturer, I used the metric MAD/Mean to trace changes in forecasting accuracy over the 15- year period, 1991-2005. The MAD/Mean overcomes several major deficiencies of the more traditional metric, the MAPE.

• My results for this company reveal little evidence of improvement over this time period, which I found surprising. It is possible that the forecast environment had become more uncertain over this period, offsetting potential internal improvements. For examinations of forecasting accuracy improvement, it is important to separately identify the effects of internal and external factors.

well as through sharing knowledge about appropriate methods (Byrne & Heavey, 2006). Also important is the establishment of specific organizational units for forecasting, as well as the alignment of forecasting and incentive systems.

(4) Uncertainty and volatility. Forecasting methods need to recognize patterns (e.g trend, seasonality, and structural breaks) and how they change over time. Improved pattern recognition may lead to better forecasts. However, environmental changes may make pattern recognition more difficult by altering historical relationships and by inducing greater volatility. These problems afflict the modeling process across the range, from macro-economic forecasting to forecasting for call centers (Minnucci, 2006). And newly influential variables have to be detected and incorporated into models, increasing the challenges faced by market analysts and forecasters.

The four factors are closely intertwined. Advanced statistical methods are of use only if implemented in software. Better forecasts will lead to better decisions only if organizational processes facilitate the use of additional predictive information. In today's business world factors (1) and (2) should not present a constraint to better forecasting. The more substantial issues concern factors (3) and (4) and which of them prevails.

THE MC CARTHY LONGITUDINAL STUDY

In a review of previous studies supplemented by their own survey of forecasting changes over a 20-year period, Teresa McCarthy and colleagues (McCarthy et al., 2006) found that forecast accuracy had deteriorated over time. They surmised that this grim result was attributable to reduced practitioner familiarity with forecasting methods and to failures in training, processes and performance measurement, and rewards (Category 3 above). They also noted a tendency of managers to rely on forecasting software as a primary solution to their forecasting problems. The concentration on software and under-emphasis on training results in users who don't know what the software does and who tend to accept software results unchecked ("black-box" forecasts).

The McCarthy study reports that only a minority of companies tie compensation incentives to forecast results. Different departments within the company are seldom forced or encouraged to align their different forecasts, a problem that is only recently being addressed through Sales and Operations Planning initiatives.

However, just as Stephan Kolassa concludes (in his preceding article in this issue) that benchmarking is difficult to do from external longitudinal surveys, so the assessment of forecast improvements from such surveys faces the same insurmountable challenges.

McCarthy's surveys:

- Were based on questionnaires. It is hard to control who responds to a questionnaire (key informant bias).
- Used accuracy-metric calculations that seem to have been done by the respondents. We do not know how they did them or how reliable the answers are.
- Compared different studies at different points in time. From comparative-static analyses, one cannot be sure to capture dynamics and trends correctly.
- Included different companies. It is not certain that the responding companies were the same over time, leading to selection biases and survivorship bias.

As an alternative to external surveys, a case study of an individual organization has promise. Using original data avoids informant bias and allows for application of appropriate metrics for forecasting errors. The concentration on one company avoids selection biases. And while the results are not scientifically generalizable, the analysis of a typical manufacturer in a mature industry should be indicative of the entire industry and possibly beyond.

THE GERMAN AUTOMOTIVE MANUFACTURER

Working with a large automobile company, I collected monthly data on **actual** and **planned sales volumes** for three car models sold in six countries. The monthly data span the period 1991 to 2005. The cars are sold in several versions, usually as middle class, upper-middle class, and premium models. The typical life cycle of a version is about seven years. The company has enjoyed decades of successful production and sales of millions of cars. In addition to the collection of sales data, I conducted interviews with company managers to qualitatively assess their forecasting and planning methods.

The company does not differentiate between forecasts (in terms of predicting future events) and plans (in terms of targets for employees). However, the sales target data in this company provide an acceptable proxy for forecasts. Compensation incentives for sales force and sales managers are based on accuracy of achieving planned sales, giving sales personnel motivation to prepare plans close to what they believe will be sold. I did a detailed analysis of plan-actual variances and found no pattern or systematic bias, such as "plans are always higher than actuals."

Given the company's long record of experience, successful market positioning, and large reserve of human and IT resources, one would assume the company would have developed a good forecasting track record. And with data spanning a long period, 15 years, we should be able to detect learning effects.

MEASURING FORECAST ACCURACY IMPROVEMENT

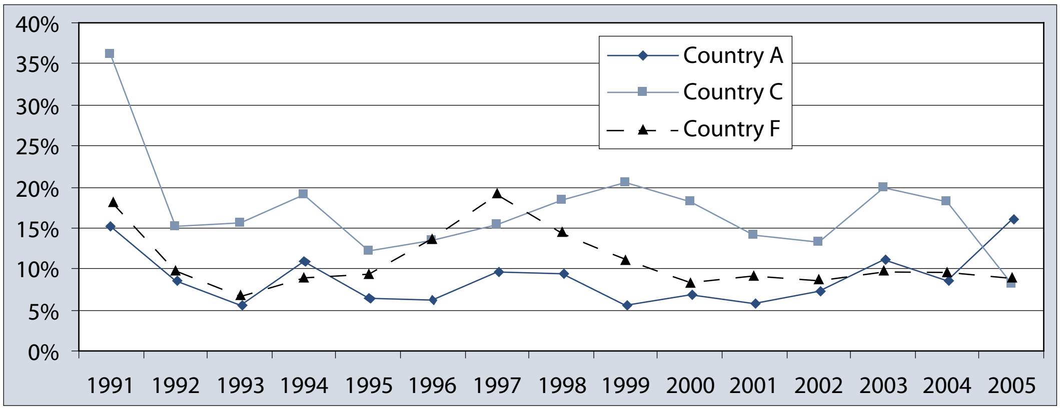
The Metric. While many forecast error metrics are in use, analysis of forecast-accuracy improvement requires a metric that is not denominated in volume (e.g +/- so many cars), since volumes are very different for different model cars. Additionally, there are months of missing data so that the most common percentage error metrics (MAPEs) cannot be calculated. In their *Foresight* article, Kolassa and Schütz (2007) propose a metric that is unit free and appropriate for interrupted data, the MAD/Mean (the mean average deviation divided by mean sales volume). The metric they show can be interpreted as a weighted average percentage error.

For each of three car models, in each of six countries, we calculated an annualized MAD/Mean ratio over the 15-year period 1991-2005. The annualized figure is the yearly average of monthly forecast errors (actual sales

Table 1. Annualized MAD/Mean for Car Model 1

Years	Country A	Country B	Country C	Country D	Country E	Country F
1991	15.1%	16.6%	36.2%	13.9%	20.8%	18.1%
1992	8.6%	5.8%	15.3%	16.8%	21.1%	9.8%
1993	5.5%	16.5%	15.6%	19.3%	20.7%	6.8%
1994	10.9%	11.7%	19.1%	16.1%	19.4%	9.0%
1995	6.4%	9.6%	12.2%	20.3%	19.8%	9.5%
1996	6.2%	16.6%	13.6%	9.0%	14.3%	13.6%
1997	9.6%	12.1%	15.4%	13.8%	16.1%	19.2%
1998	9.4%	23.7%	18.5%	11.5%	27.3%	14.5%
1999	5.5%	13.6%	20.6%	8.8%	19.7%	11.1%
2000	6.8%	13.2%	18.2%	7.4%	9.6%	8.4%
2001	5.7%	8.1%	14.2%	6.9%	12.8%	9.2%
2002	7.3%	10.0%	13.3%	7.2%	8.6%	8.7%
2003	11.0%	15.6%	19.9%	5.3%	16.0%	9.8%
2004	8.6%	21.7%	18.2%	18.0%	10.6%	9.7%
2005	16.0%	17.3%	8.1%	12.4%	18.8%	9.1%

Figure 1. Annualized MAD/Mean Model 1 for Three Countries, Based on Table 1



– target) divided by the yearly average of monthly sales. The results for the first car model are shown in Table 1 and a portion of this table is plotted in Figure 1.

Detecting Trends in Forecast Accuracy Over Time.

A decreasing trend in forecast errors over time should represent a pattern of improvement in forecasting, while an increasing trend in forecast errors would suggest deterioration. Based on the accuracy metric (MAD/Mean) calculated for different countries and different time periods, we tested for indications of these decreasing and increasing trends. Our test results – applied to all three car models – do not indicate an overall trend towards improved forecasts. The specific statistical test employed is described in the on-line appendix. See www.forecasters.org/foresight/documents/Rieg_Issue11.pdf

You can see in Table 1 and Figure 1 that in some years and countries forecast errors are trending downward, but these are not enduring. Similar results were found for car models 2 and 3. We cannot conclude that there has been an improvement in overall forecast accuracy.

Learning Effects from One Product Life Cycle to the Next.

While evidence of improvement in forecast accuracy over time does not emerge, improvement in forecasts due to learning effects could still be possible from one *product life cycle* (PLC) to the other. Each life cycle is roughly seven years long. In a new life cycle, models with new technology and/or design changes are introduced into the markets while the basic car model stays the same.

For each car model and country, we compared forecast errors that occurred in the first month of two successive life cycles. For example, we compared the MAD/Mean for July 1991, the initial month of PLC1, with that of August 1998, the initial month of PLC2. We repeated the comparison for each of the remaining months. The results are shown in Table 2.

We detected downward changes in forecast errors in only 18% of the comparisons, which was essentially the same frequency of upward changes (deterioration in forecast accuracy). So the majority of comparisons revealed no indications of learning from one life cycle to the next.

Internal Vs. External Factors Affecting Forecast Errors.

As we have noted, changes in forecasting accuracy over time can be attributed to 4 types of factors: changes in methods/algorithms, software and

Table 2. Upward or Downward Change between Life Cycles (Up Implies Deterioration)

Mann/Kendall trend test for subsequent product life cycles (plc)

	Model 1			Model 2		Model 3	
Country	plc 1 to plc 2	plc 2 to plc 3	plc 3 to plc 4	plc 1 to plc 2	plc 2 to plc 3	plc 1 to plc 2	plc 2 to plc 3
A	up	up	(no data)	down	up	down	down
B	up	-	(no data)	-	-		down
C	-	-	-	-	-	-	-
D	up	-	(no data)	-	-	up	-
E	-	-	-	down	-	down	-
F	-	down	-	-	-	up	-

LEGEND	meaning	happens	of all tests
up	upward trend, (α = 5%)	7 times =	18%
down	downward trend, (α = 5%)	7 times =	18%
-	no clear trend	24 times =	63%

hardware, people and organizations, and the forecasting environment. The first three are internal changes while the fourth is an external factor. Perhaps our finding that forecast accuracy failed to improve (or show any trend) over time is attributable to offsets between internal and external events, internal improvements being offset by an increasingly uncertain external environment. One lucid example of this was the unforeseen changes in consumer behavior prompted by environmental concerns about large and polluting German vehicles compared to greener Japanese cars with hybrid engines.

In interviews with company officials, however, I was told the company had not made significant changes in the internal factors – in processes, tools, or algorithms. For a company with a long-standing record of corporate success, vast resources, knowledge, and experience, this was surprising. However, since I relied upon interviews held afterwards, I can't rule out whether there were unreported internal changes or, if so, whether any of these were successful.

Analysis of Change in the Forecasting Environment. Changes in the external forecasting environment evolve gradually over time. One way to detect the magnitude of these changes is to compare the variability – degree of fluctuation – of the sales data at different points in time.

Figure 2 shows the actual monthly sales volume as well as a 12-month moving average for car model 1 in two countries. One can see the 7-year product life cycle and a pattern that shows fluctuations but without evidence that these are increasing or decreasing over time. We do see that some seasonal patterns, such as those for Country B, have smaller peaks after 1998. Such changes in patterns are hard to forecast if one has only time series at hand.

One measure of variability is the standard deviation. For each car model and country, we calculated the annual standard deviations of actual sales volumes and then tested this for decreasing or increasing trends. The results were a mixed picture of upward,

Figure 2. The 7-year Life Cycle for Car Model 1

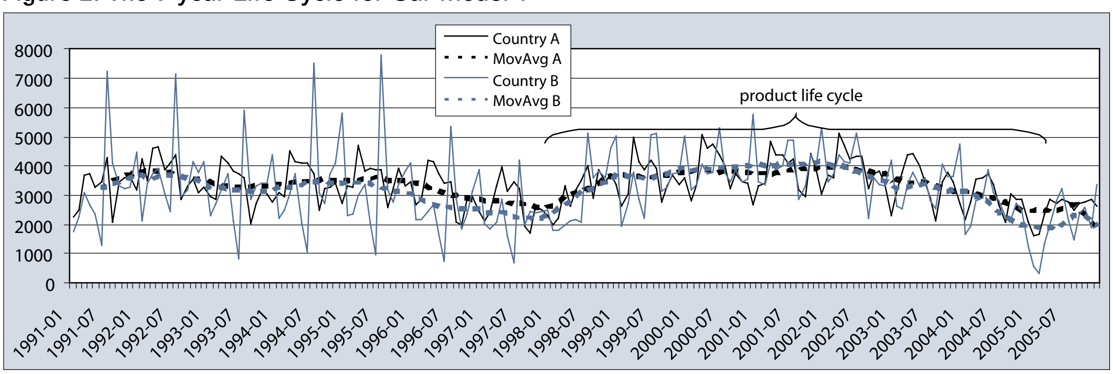
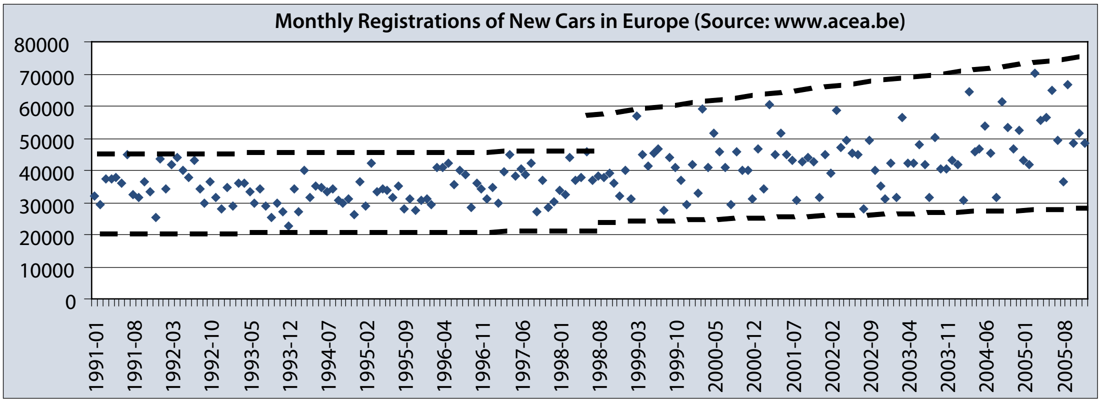


Figure 3. Monthly New Car Registrations in Europe 1991-2005 (Source ACEA)



downward, and no clear trends. Once again, there was no indication of a consistent increase or decrease in the uncertainty of the forecasting environment.

A somewhat different picture emerges when we look at the data (Figure 3) on new car *registrations*. The dashed lines show upper and lower boundaries. Here there seems to be an increase in volatility beginning in 1998. The data shown are officially recorded, monthly registrations of new vehicles of the company in the case study (source: http://www.acea.be).

So the historical data present a mixed picture, with only the car registration time series revealing a pattern of increasing volatility.

CONCLUSIONS

During the 15-year period 1991-2005, the automotive company was able to improve its forecasts for a few countries, a few models, and a few time periods. However, the overall record does not support a trend toward improved forecast accuracy. Rather the results suggest that forecast-accuracy improvements were transient and vanished as the markets changed. Perhaps, the automobile company should have given more attention to its markets, investing in flexibility to react and adapt quickly.

In this paper, I have offered an analytical framework that can be applied to your own company to depict its forecasting track record over time. If you understand your past forecasting performance, you'll be better prepared to face future challenges in setting and achieving forecast-accuracy goals.

REFERENCES

Armstrong, S. (2006). Findings from evidence-based forecasting: Methods for reducing forecast errors, *International Journal of Forecasting*, 22, 583-598.

Byrne, P.J. & Heavey, C. (2006). The impact of information sharing and forecasting in capacitated industrial supply chain: A case study, *International Journal of Production Economics*, 103, 420-437.

Kolassa, S. & Schütz, W. (2007). Advantages of the MAD/MEAN ratio over the MAPE, *Foresight: The International Journal of Applied Forecasting*, Issue 6, 40-43.

Küsters, U., McCullough, B. & Bell, M. (2006). Forecasting software: Past, present and future, *International Journal of Forecasting*, 22 (3), 599-615.

McCarthy, T., Davis, D., Golicic, S. & Mentzer, J. (2006). The evolution of sales forecasting management: A 20-year longitudinal study of forecasting practices, *Journal of Forecasting*, 25, 303-324.

Makridakis, S. & Hibon, M. (2000). The M3-competition: Results, conclusions and implications, *International Journal of Forecasting*, 16, 451-476.

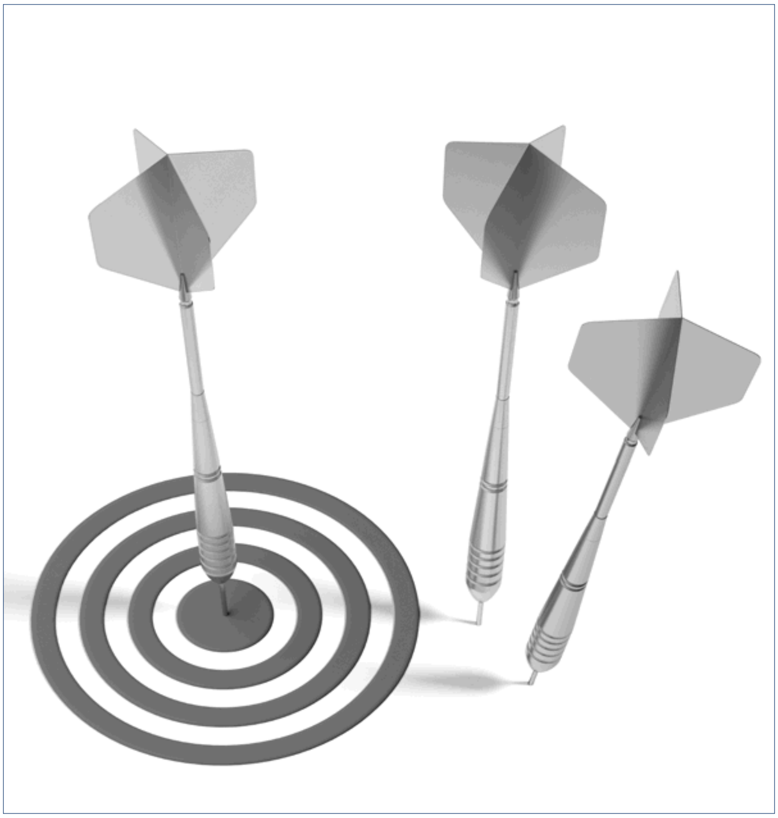
Minnucci, J. (2006). Nano forecasting: Forecasting techniques for short-time intervals, *Foresight: The International Journal of Applied Forecasting*, Issue 4, 6-10.

Sanders, N. & Manrodt, K (2003). Forecasting software in practice: Use, satisfaction, and performance, *Interfaces*, 33 (5), 90-93.

CONTACT
Robert Rieg
Faculty of Business, Aalen University
Robert.Rieg@htw-aalen.de

COMMENTARY ON BENCHMARKING

Teresa McCarthy, Donna Davis, Susan Golicic, and John Mentzer



Robert Rieg and Stephan Kolassa have described what they believe are shortcomings of surveys of forecast accuracy. Each makes reference to our own longitudinal study (McCarthy et al., 2006), and so we welcome this opportunity to reply, as well as to address the broader question of the wisdom of benchmarking forecast accuracy.

Our study explored how sales forecasting management and practice have changed over the past two decades. We replicated and compared results of a survey that was administered to forecasting executives 10 and 20 years prior, while including additional questions to capture new information relevant to the changing business environment. We hoped to provide forecasting managers with a comprehensive view of current and past forecasting practices, to help them understand forecasting trends, and to improve forecasting performance in their own firms.

Our survey explored four overarching dimensions of the forecasting process: forecasting management, techniques, systems, and performance measurement, the last section including data on forecast accuracy. Among the many results presented, our survey revealed that forecast accuracy appears to be deteriorating over time.

SURVEY VALIDITY

Kolassa and Rieg both question the validity of our study's results as benchmarks, partly because we used a survey to collect our data. We agree that surveys have limitations. However, all research methods have their weaknesses. Each also has strengths, and choosing a method to collect and analyze data on any topic always involves a trade-off. McGrath noted years ago that the research process should be regarded "not as a set of problems to be solved, but rather as a set of dilemmas to be lived with," and there is "no one true method that will guarantee success" (1981, p. 179).

The recommended approach is to match the research objective with the most appropriate research method so that the strengths can be maximized and weaknesses minimized. Our research priority was to examine general practices over time in various areas of forecasting management. Therefore, we felt that

- Teresa M. McCarthy, left, is an Associate Professor of Marketing at Bryant University in Smithfield, RI, USA.
- Donna F. Davis is an Assistant Professor of Marketing at Texas Tech University in Lubbock, TX, USA.
- Susan L. Golicic is an Assistant Professor of Supply Chain Management at Colorado State University in Fort Collins, CO, USA.
- John T. Mentzer is the Bruce Chair of Excellence Professor of Marketing and Logistics at the University of Tennessee in Knoxville, TN, USA.

conducting a survey of multiple forecasters from different companies and industries was the best way to obtain this information.

One specific weakness of survey research is the potential for *key informant bias*, since it is difficult to control who actually responds to a questionnaire. We tried to minimize this bias by following recommended survey protocols (Stanton & Rogelberg, 2001). We required a respondent password to complete the survey, ensured that the responses came from companies on our sample list of forecasting executives, and verified that the respondents were in positions affording them knowledge of the forecasting process. It is possible that some of the responses we received were not "truly accurate," but we are confident that informant bias is not a significant challenge to our findings.

Kolassa and Rieg each note that the three surveys did not have identical respondents. We never intended to follow the same companies and industries across the decades. Rather we tried to obtain representative data on the practices companies use in forecasting along with changes in these practices over time. In order to compare practices, we replicated the survey questions about those practices across the surveys and added questions pertaining to new practices that were introduced during the 20-year period.

Kolassa points out that the number of respondents decreased from the earlier studies to the 2006 survey. Unfortunately, response rates to business research in general are declining, due to constraints on practitioners' time and the frequency of requests for their participation in forecasting research. However, lower response rates are acceptable, provided rigorous methods are followed and a satisfactory level of data is obtained (both true of our survey).

Rieg expressed concern about selection bias; that is, any sample of current companies and managers will naturally contain more successes than failures. However, concentration on a single company does not avoid this bias per se, particularly if it is a successful company. Denrell (2005) points out that reducing this bias means working with firms that have failed or are in emerging industries as opposed to a mature industry. Our survey questions sought to provide an accurate picture of current forecasting management, whether the practices were considered sophisticated or dysfunctional. Indeed, our study finds that many aspects of forecasting management have not improved. Instead, we believe they reveal an unsettling downward trend, in spite of increased investments and improved technologies.

BENCHMARKING FORECAST ACCURACY

Both Kolassa and Rieg question the usefulness of benchmarking forecast accuracy as a way to improve forecasting management. We share this concern. Our study does not recommend that reported forecast accuracy results be used as benchmarks.

Direct comparisons of forecast-accuracy levels across firms and industries suffer from several problems. Kolassa raises a key question: "Could we, in principle, collect more or better data and create 'real' benchmarks in forecasting?" We concur that such efforts would be misguided. There is no magic number that qualifies as the correct target for forecast accuracy across organizations, product types, time horizons and/or granularity.

However, we think that collecting and publishing reports of forecast accuracy is nevertheless useful to build knowledge about linkages between forecast accuracy and forecast management. It is also useful to have some indication of increasing or decreasing levels of forecast accuracy for units of analysis beyond a single strategic business unit, such as corporate and industry level analyses.

Forecast-accuracy measurements are key performance indicators for evaluating a firm's forecasting competence. Managers are obliged to set expectations about appropriate forecast-accuracy goals and to measure progress toward those goals. Ultimately, determining the right forecast-accuracy target is an essential link in aligning business processes with the firm's business needs.

While it is advisable for managers to consider their firm's particular business requirements in setting forecast-accuracy targets, it is important to recognize that business competition is a comparative phenomenon. That is, performance is not judged in isolation. To assure survival, a firm must perform better than competitors. Thus managers want to answer the question, "How do we stack up against the competition?" While we agree that reliance on published reports of forecast accuracy is not appropriate for setting an individual firm's forecast-accuracy targets, we believe that such reports may help managers determine if their targets are viable.

Kolassa argues that reports of forecast accuracy across industries are not helpful, due to noncomparability of industry factors. Yet benchmarking research suggests that companies should look outside their own industries to find best practices that can be adapted to help them gain a competitive edge (Zairi & Al-Mashari, 2005). Innovative approaches to managing processes and people often emerge from an external focus structured to identify, transfer, and adapt best practices in industries other than one's own. The aim of benchmarking research is not to set benchmarks for individual firms but to provide a source of data that, when considered in combination with other sources, can inform process improvement efforts. As noted by Kolassa, "failure to provide benchmarks does not mean the results are uninformative to practicing forecasters" (pp. 9-10).

CONCLUSIONS

Ultimately, the conclusions of our research do not differ substantially from the conclusions made by Kolassa and Rieg. For example, Kolassa writes, "Forecast accuracy improvements due to process and organizational changes should be monitored over time" (p. 14). The implication is that forecast accuracy is just one of many elements to consider and monitor when managing the forecasting process. Similarly, Rieg's general premise is that reliance on improved forecasting algorithms, hardware, and software alone without attention to managing the people, processes, and changes in the external environment could restrict improvements in forecast accuracy. Our research supports both of these conclusions.

Business executives and forecasting managers frequently ask, "What should our forecast accuracy be?" The answer: it depends. Decisions on targeted forecast-accuracy levels must consider multiple factors, such as expected customer service levels, the competitive environment, the resources available within the firm, and existing forecast accuracy in the firm (i.e., a continuous process improvement approach). But no single piece of research on its own can understand and explain all of the intricacies of forecasting management. Our survey is only one piece of this puzzle.

REFERENCES

Denrell, J. (2005). Selection bias and the perils of benchmarking, *Harvard Business Review*, 83 (4), 114-119.

McCarthy, T.M., Davis, D.F., Golicic, S.L. & Mentzer, J.T. (2006). The evolution of sales forecasting management: A 20-year longitudinal study of forecasting practices, *Journal of Forecasting*, 25 (5), 303-324.

McGrath, J.E. (1981). Dilemmatics: The study of research choices and dilemmas, *American Behavioral Scientist*, 25 (2), 179-210.

Stanton, J.M. & Rogelberg, S.G. (2001). Using Internet/Intranet web pages to collect organizational research data, *Organizational Research Methods*, 4 (3), 200-217.

Zairi, M.& Al-Mashari, M. (2005). The role of benchmarking in best practice: Management and knowledge sharing, *The Journal of Computer Information Systems*, 45 (4), 14-31.

CONTACT
Teresa McCarthy
Bryant University
tmccart4@bryant.edu

COMMENTARY ON BENCHMARKING

Jim Hoover, *Foresight* Software Editor

Benchmarking is a concern of critical importance to forecasters and their organizations. The Kolassa and Rieg articles in this issue pose the key questions: Are your forecasts as good as they could be? Is forecast accuracy improving or diminishing over time? And how do you seek relevant information on these questions?

Stephan Kolassa's article gives some needed perspective on the usefulness of published surveys as benchmarks. One problem he identifies is that not all companies measure forecast accuracy using the same metric, but doing so is critical for comparability. Another issue is the low response rate behind the forecasting benchmark surveys. Most broadly, he discusses how differences across product lines, aggregation levels, and time frames for the forecasts all can undermine reliability of the supposed benchmarks.

Adding up the problems in making apples-to-apples comparisons, Stephan concludes that you should not use *external* benchmarks to determine if your own forecasting process is effective. So how *do* you judge its effectiveness? Robert Rieg suggests *internal* benchmarks – measuring changes in a company's own forecast accuracy over time. The internal focus addresses many of the comparability issues. If you are making genuine improvements in forecasting, the results – when measured against prior periods for the same type of item – should show it.

Yet, as Robert argues, external events may cause stable or improving forecast accuracy to deteriorate. In the U.S. Department of Defense, we saw demand for previously stable items rise significantly during the work-up period for the war in Iraq and then decline sharply after the initial ground campaign had ended. Most forecasting methods will have difficulty reacting quickly and appropriately to such external factors.

Robert presents a case study describing a series of forecast cycles in the European automobile industry and the resulting forecast-accuracy outcomes. For this auto manufacturer, after an initial period of improvement, forecast accuracy leveled out and then worsened. Even so, Robert's article illustrates how to make measurements of your own forecast accuracy over time and use them to evaluate and perhaps drive process improvement.

My review of the literature indicates there is very little written on this subject. There are articles on how to measure forecast accuracy, there are principles offered for improving forecast accuracy, and many consultants will tell you they know how to improve forecasting performance. Methodologists tend to show how a new forecasting process fares against other systems when applied to historical data. Why don't they study whether the implementation of a new method improved an organization's forecasting performance over time?

Robert cites a survey by Sanders and Manrodt (2003), which concludes that "the use of appropriate software can lay the groundwork for improved forecasts." I agree. Software can help forecasters avoid some mistakes, such as entering incorrect data, failing to

Jim Hoover is a Captain in the U.S. Navy Supply Corps. He has over 24 years of experience in Supply Chain Management and currently is Chief of Staff at the Naval Supply Systems Headquarters. Previously, as Deputy Commander, Fleet Logistics Operations, he was responsible for world-wide implementation and operation of the Navy's food, parts, retail, fuel, and ammunition supply chains. Jim's software columns and articles appear in Issues 1, 2, 4, and 7 of *Foresight*.

check for outliers, ignoring hierarchical relationships, and the like. So software can potentially improve forecasts. But without measuring your own results over time, how would you know?

I recently worked for an agency with more than $30 billion in annual sales. For years, this organization didn't track forecast accuracy in a systematic way. As part of an expensive implementation of Enterprise Resource Planning (ERP) and a Supply Chain Management System, it decided that tracking forecast accuracy was essential in order to achieve the inventory savings expected from the new system.

Many of the issues Stephan and Robert describe have been experienced by my organization: which specific metrics to use, which forecasting methods to use, how to aggregate individual SKUs' forecasts, how to track the results over time, where to store the forecasts and the actual demands to best allow forecasters access to the results and to propose improvements, and, finally, how to prioritize which of the SKUs they should pursue for accuracy improvements.

My organization had the necessary leadership backbone to attack these problems and still found some seemingly intractable. Despite having a contracted ERP integrator help with the task, it is just now beginning to institute a systematic approach to solving these issues. Progress began when the IT department, operations, and customer relations were brought together to resolve these issues collectively. Committed leadership, oversight, and participation were critical to finally making headway.

CONTACT
Jim Hoover
Naval Supply Systems Command
Hooverjh@aol.com

Special Feature on Forecastability
Some items are more difficult to forecast than others, so that a particular average percent error may represent a successful outcome for some items and a failure of forecasting for others. The forecastability articles will address these issues and more:

- How can we determine the forecastability of an individual time series, so that we have a basis for judging the success of any forecasting method?
- Relative error metrics have been offered that compare the forecast errors of a designated method to those of a naïve method. The no-change (naïve 1) forecast is a virtual standard in business-forecasting software. Should it be? Are there better alternatives?
- Coefficients of variation on detrended, deseasonalized data are being considered by some companies as indicators of forecastability. Is this metric useful?

Forecast Process Improvement
- The Forecasting Mantra
- Sales Forecasting: Improving the Relationship Between Demand People and Supply People

Forecast Accuracy Measurement:
- How to Define the MAPE
- Measuring Forecast-Improvement Initiatives

The World of Forecasting
Predicting Recessions

Forecast Model-Building
Statistical Significance Testing: Is It Useful?

- Software Reviews
- Book Reviews
- Hot New Research Column
- Forecasting Intelligence Column
- Case Studies

OVERCOMING CHALLENGES IN OPERATIONAL FORECASTING PROJECTS

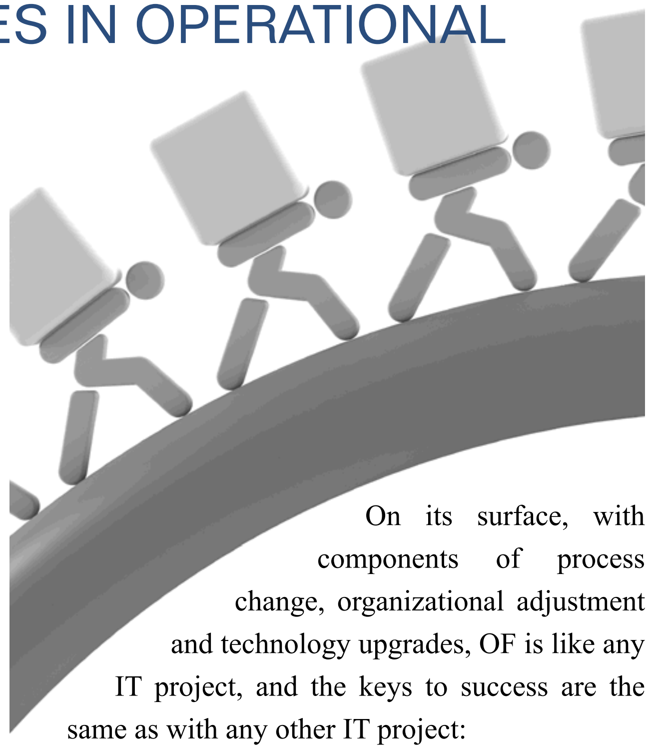
Ian Watson-Jones

PREVIEW

While no one who has attempted to manage a new Operational Forecasting (OF) project will tell you it was a piece of cake, they probably never fully anticipated the breadth of issues that would have to be addressed or the sustained leadership requirements necessary for effecting change that works. Ian Watson-Jones has spearheaded many an OF project. In this feature article, Ian describes the wide range of elements that can undermine project success and offers mighty sensible recommendations for anticipating and overcoming the challenges. Don't miss his checklists of Process, System, and Organization issues.

INTRODUCTION

For industrial and consumer product companies wishing to improve their overall supply chain performance, addressing Operational Forecasting is a common, almost obvious, place to start. It can also be a high-risk undertaking, and many companies fail to consider the potential pitfalls involved. Improving OF is often confused with merely installing forecasting software. If accuracy does not improve, executives see this as reaffirming their sense that the business is too complex and unpredictable to forecast. When inventory reductions or order fill rate increases do not materialize, the impulse is to accuse the software vendor of overpromising and underdelivering. While either conclusion is occasionally the culprit, it is far more likely that the cause of failure was a fundamental mishandling of the project, its conception and execution.



On its surface, with components of process change, organizational adjustment and technology upgrades, OF is like any IT project, and the keys to success are the same as with any other IT project:

- Successful stakeholder management
- A compelling business case
- An efficient, consensus-driven business process
- A plan to improve data quality
- Hardware and software that support new business requirements
- A change-management strategy

Prior projects with a narrower scope or a better-defined set of best practices might safely have taken some of these areas for granted. But it is past success in other efforts that leads OF projects to be underestimated and mishandled. OF has broad organizational scope and complex business processes and is ravenous for meaningful data, all tied up in a project whose value to the company can be difficult to prove.

Ian Watson-Jones is a Senior Managing Consultant in Supply Chain Planning with IBM Global Business Services, the largest provider of Supply Chain Management process and application services in the world. He is also a member of *Foresight*'s Practitioner Advisory Board. When he is not traveling to advise clients, Ian enjoys cycling around the San Francisco Bay area. He solemnly swears that this article is not intended to discourage you from undertaking an OF project.

KEY POINTS

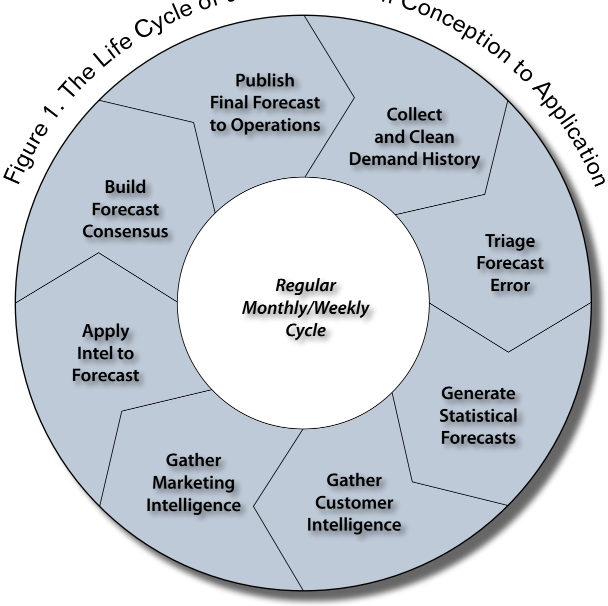
- The keys to success in Operational Forecasting include effective stakeholder management, a proper business case, an efficient, consensus-driven business process, a plan for overcoming data quality issues, and hardware and software support for the new business requirements. But the linchpin is the creation of a process of change management.
- Attempting to satisfy requirements of all stakeholders is impossible. Always select the leanest possible project and support it with straightforward requirements and a limited, highly achievable business case.
- To meet the political challenges, identify key stakeholders and work with them to develop a joint business case. Stakeholders should commit to forecast improvement and agree that they will credit the forecasting project with enabling this improvement.
- Data quality is a risk factor. Anticipate data quality issues and plan accordingly.
- Delay hardware decisions as long as possible to allow for an accurate sizing of the solution to take place. During installation, load a production-sized dataset and see how the system responds. Doing so before going live will help avoid unpleasant surprises.
- Software selection has two basic components: statistical modeling and workflow facilitation. It's best to evaluate the two dimensions separately.
- Change management is the key to ensuring that stakeholders are consulted throughout the OF project. Stakeholders need to clearly understand what is in it for them and how the forecast will improve before they will begin providing meaningful input, contributing to consensus, or accepting the results.

My article deals with each of these dimensions, pointing out the challenges operational forecasting poses and sharing insights from successful projects.

SCOPE OF OPERATIONAL FORECASTING

OF comprises the entire repeating life cycle of a forecast, from conception to application (see Figure 1).

Figure 1. The Life Cycle of a Forecast from Conception to Application



For demand planners, collecting history and developing the statistical forecast are just the beginning of their job. Their work in a forecast cycle is completed only when all stakeholders have provided input and the forecast is finally released to Sales and Operations Planning (S&OP) for use in planning and execution decisions. Additional demand information is gathered from sales, marketing, and customers.

When best practices are in place, the key output will be one enterprise-wide demand plan that is used by all functions to support operational decisions. This plan is determined with participation from Sales, Marketing, Customers, Finance, Operations, and New Product Engineering. At this level of consensus, the demand plan can be confidently applied to supply and inventory planning.

Given the large data volume driven by the entire universe of SKUs, customers, geographies and distribution centers that will need to be managed, I assume that specialized software will be implemented to support this process.

SUCCESSFUL STAKEHOLDER MANAGEMENT

When initiating business process improvement, we need to identify and document all stakeholders' expectations. For OF, this will turn out to be a long list of business functions, geographies and management levels. Each of these will have an opinion on what the priorities of the new process and system should be. Before it is even implemented, a forecasting project can become buried under a list of "must-haves." Some common examples:

Marketing frequently views investment in OF as the chance to utilize every product management technique learned in business school, especially new product forecasting and advanced data-mining methods. In consumer-focused companies, marketing will also be looking for tight integration between OF and the promotion-planning process. Meeting these requests could be a project all its own.

Sales may view the OF project as a means to shift their level of engagement in forecasting. One sales VP – a forecasting true believer – wants to create an infrastructure where regional and customer-focused sales representatives provide a bottom-up forecast for their corner of the business. The flip side is a sales VP who believes his team should be selling, not mired in "annoyances" like forecasting.

Finance would like the operational forecast in dollars or other currencies and yield revenue/margin estimates that offer early warning of financial projections that may be at risk. However, what starts as a simple rough-cut revenue estimate frequently becomes unwieldy, with requests like customer-specific pricing or inventory depreciation models.

Operations will seek the most granular forecasts, within a system that directly interfaces to supply chain planning and inventory management without new allocation requirements (e.g., months to weeks, countries to distribution centers, or channels to customer sites). In build-to-order supply chains, operations may prefer forecasts for intermediate materials rather than finished SKUs. Where lead times are long, operations may request a forecasting freeze period, enforcing near-term stability on the forecast if they lose flexibility to respond to changes.

Customers may be interested in supply-chain integration and collaboration, in which, for example, the customer provides the forecast that the company is required to support, even if no guarantees of accuracy are made. This imposes a series of policy decisions about how forecast information can be shared without inadvertently revealing confidential information.

Supply chain managers will be looking for a broad range of metrics to track forecasts supplied by each contributor at every available lag and level of aggregation. In addition, they typically want the forecasting system to provide a detailed audit trail, so forecasts that turn into costly mistakes can be backtracked.

Demand planners are interested in the productivity advantages the overall solution provides. They hope that complex requirements from the aforementioned groups can be automated and not impose cumbersome manual processes upon them. They want workflows designed for manage-by-exception capabilities, particularly at companies with high SKU counts.

The IT department will want the simplest overall solution: functionality self-contained in the software, no resource-intensive data preparation, no new requirements for interfaces to other systems. Batch schedules should not overload the off-hours when maintenance functions typically occur.

Recommendations on Stakeholder Management

With this range of stakeholders and expansive vision of forecasting functionality, it is easy to see how projects get out of hand. Trying to satisfy every last request is a recipe for disaster in any scenario. But project funding may be viewed as the one shot at success in an expense-sensitive environment, and requirements not fulfilled on the first try may never be met. Where is the right balance?

The right balance is nearly always for the leanest project possible. Fears of one-time funding are justified when a project does not deliver as promised. However, most executives welcome the chance to continue a process that demonstrates the ability to fund its own improvements with successful results. Forecasting has this potential if the case for investment is realistic and the project is well managed. So the best strategy is a modest project with straightforward requirements and a limited, achievable business case.

In such a project, where the business case has a tighter focus, not all stakeholders can be key users. Key users should be the driving force in designing the new process and possibly the source of the project leadership. But who the key users are will vary significantly by industry and corporate culture. Where businesses are responding to rapidly changing technology or fashions, Marketing is probably the driving force behind the company's success and should get first choice of functionality options. At commodity companies, where relationships and contracts are crucial, Sales is more likely in the driver's seat. Finance will have different levels of influence in different corporate cultures. Operations' role in realizing potential gains from forecast improvement may not make them key users, but their opinion should break any deadlocks on project direction.

COMPELLING BUSINESS CASE

The development of a compelling business case for an OF project faces unique difficulties. Typical process improvement projects are able to demonstrate their value with easily quantifiable, straightforward metrics, such as cycle-time reduction or cost-per-unit improvement. Most will agree that improved forecast accuracy is good for business, but there is no formula that unequivocally links better forecasts to improvements in any standard financial measure. Inventory, lost sales, order fill and transportation costs are all related to forecast quality. Yet there are so many variables contributing to each metric, it is difficult to say what effect improvements in forecast accuracy will have.

The leap of faith from improved forecast accuracy to financial success makes it difficult for executives to justify funding an OF project. Business cases are frequently dismissed for being based on comparable companies that "don't share our unique challenges" or internal surveys that merely "quantify unproven opinions."

Even if a project is successfully funded without belief in a business case, there may be future difficulties. An OF project drifting off schedule is at a higher risk of cancellation without a business case with strong buy-in. Even if completed, it may be difficult to prove the project's eventual success, since nothing can be conclusively tied to the results. If a second phase or follow-on project is needed, the best possible business case would be a successful first phase.

Recommendations on Business Case Development

When developing a business case, consider both its mathematics and its politics. Regardless of how the business case is calculated, the question the budget-keepers are going to ask will not be "How much is it worth?" but "Who is committing to this benefit?" After you identify your primary stakeholders, work with them to develop a joint business case.

A joint business case should have two up-front agreements built in: first, a commitment to change that is triggered by improvements in the forecast; and second, a commitment later on to credit the forecasting project for helping to enable that change. With the first agreement in place, it will be considerably easier to fund the project and justify the temporary assignment of staff from those departments required to assist. The second agreement keeps benefits that could be attributed to forecasting from being chalked up to "better marketing" or "inventory micromanagement."

For example, a business case predicated on forecast-driven inventory reduction will be far more convincing if it is presented by the distribution manager who pledges to use that additional accuracy to reduce DC safety stock levels. A business case built on recouping lost sales will be far more compelling if it is presented by the VP of Sales clutching a list of sales orders that were cancelled due to limited availability. For any business case, work with financial analysts to make sure that benefits are expressed in the precise measures of working capital and operational expense that executives use to make decisions.

After project completion, the links between forecasting and key business metrics will fade. There will be competition to claim credit for an inventory reduction or an order-fill improvement. Being able to demonstrate promises made and kept, and having allies who will share some credit for metric improvements, will keep the power of forecasting visible and may be the main selling point in the business case for funding the next phase of improvements.

AN EFFICIENT, CONSENSUS-DRIVEN BUSINESS PROCESS

The final step in effectively responding to stakeholder requirements comes when the team designs and validates the detailed business process. Here the team decides which functionality is in scope and how each function will be realized.

The time required for this phase is frequently underestimated at the outset of the project for various reasons:

- In sophisticated forecasting systems, some decisions that are made during the initial implementation cannot be changed without repeating much of the implementation process.
- Discussing and documenting a cross-functional process can expose organizational disagreements about accountability for forecast, inventory, and order-fill performance.
- Trying to integrate with processes such as sales reporting, promotion planning, inventory planning, or production scheduling may expose a lack of standards in calendars, units of measure, or the meaning of key data.

These conflicts are not well understood or anticipated and therefore not factored into project estimates. Indeed, the high-level process appears simple to design. This expectation is reflected in a project plan that assumes the design phase will proceed smoothly and quickly (Figure 2).

Figure 2. The Initial Plan

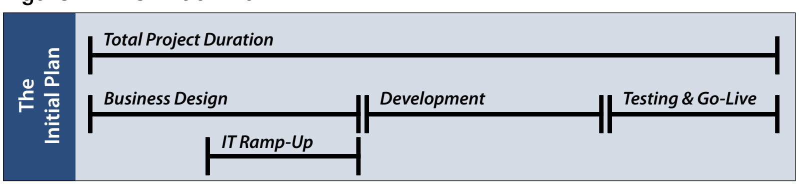


Figure 3. The Frequent Reality

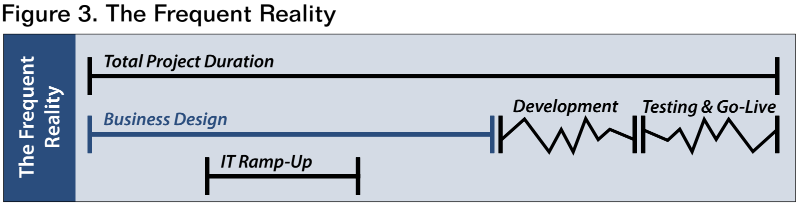
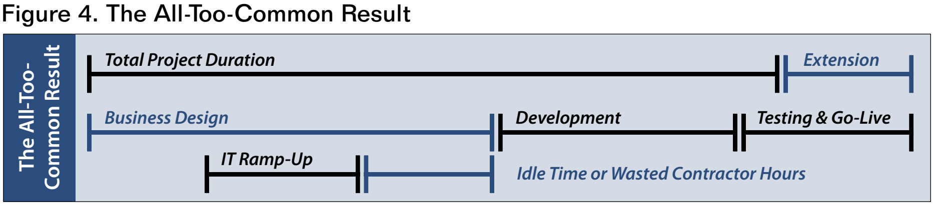


Figure 4. The All-Too-Common Result



However, developing the details of the business process and system design frequently exposes barely resolvable conflicts. Primary stakeholders must buy in that their requirements are being fulfilled, and secondary stakeholders must be convinced that the process can be successful when their requirements are not. The selection of the forecasting software and the influence of the software on the business process must be debated. As the business design phase drags on longer than anticipated, the project plan for the remainder of the project becomes compressed (Figure 3).

When it turns out that development and testing estimates were accurate, regardless of the duration of the process and system design, the whole timeline gets extended, and the project develops a bad reputation with executives before it has a chance to show its worth. (Figure 4).

Recommendations on Creation of a Consensus-Driven Business Process

Allow much more time for business design than you might for less complex projects. If there is flexibility, schedule two projects: one focused on process and system design, one focused on technical implementation that will not begin until design is complete. It may take longer but will be less risky and expensive.

When you do take on the design phase, there will be issues beyond the core business process and system design. Organizational strategy decisions may overshadow the decision-making process, with some stakeholders unable to objectively evaluate the business process and system until the organizational issues are resolved.

As you take on the design phase, consider this issue checklist. While it cannot predict every conflict, an OF design is probably incomplete until it has answered the relevant questions here. Answering these questions from the start will help keep the design from dragging on later.

PROCESS CHECKLIST

✓Definition of Demand. Will history be represented by *shipments* or *sales orders*? Which date field best represents demand: promise date, request date, scheduled date, or arrival date? Are backorders included? Is there downstream retail demand that may be more predictable?

✓Monthly vs. Weekly Frequency. Can we update the forecast monthly? Do operational complexities require weekly updating?

✓Unit of Measure Standardization. If items are sold in multiple-pack sizes, does the forecast represent bulk packages or the individual units within? For example, do one 10-pack and one 50-pack represent 1 demand unit for two different SKUs, or 60 demand units of the same SKU? Will Operations be able to plan efficiently based on the choice made?

✓Calendars and Week/Month Interaction. Will forecast buckets be monthly or weekly? If forecasting in weekly buckets, how will that reconcile with a fiscal calendar that calls for calendar months? If forecasting in monthly buckets, how will that affect Operations?

✓Key Intersection between Demand and Supply Planning. What hierarchy intersection will be used for manufacturing? Does Operations need a total SKU forecast or a breakdown by customer shipping location? Are demand planners willing to be evaluated for accuracy at the level of detail required by Operations?

✓ Definition of Forecast Accuracy Metrics. Which metric will be used to evaluate the effectiveness of the new OF process? Will item forecasts be weighted? What about measuring forecast stability, as operations so often requests?

✓ History Adjustment Processes. What events are so unusual as to be considered outliers? How can promotional demand be separated from base demand? Can we track one-time orders consistently enough to apply automatic adjustments? Can the impact of lost sales be used to adjust history? How do adjustments stay in sync if history is gradually corrected for accuracy?

SYSTEM CHECKLIST

✓Hierarchy Dimensions and Levels. How many hierarchy dimensions (e.g., product, customer, location, DC) will be implemented in the system? How many levels in each dimension (e.g., SKU, family, category, target market, division in a product dimension)? How many different aggregation paths through each dimension?

✓Hierarchy Data Sources and Maintenance Processes. How will hierarchy dimensions based on changing market strategies be maintained? How will the forecasting system be kept current with the latest information on customer priority, market segmentation, and corporate organization?

✓Volume of Manual Processes. How many different tasks require manual work by the demand planner for the process to be successful? Is the process allowing time for the demand planners to analyze and improve forecasts, or is their schedule crammed with data manipulation and management?

✓ History Data Availability and Quality. Given the strict new definition of historical demand that everyone has agreed to use, how much history actually exists that meets that criterion?

✓ Frequency and Type of Forecast Archives. How many different versions of the forecast from how many different stakeholders will be saved for later evaluation? Will monthly or weekly snapshots be saved of the pure statistical forecast? the pre-S&OP forecast? the post-S&OP forecast? the sales forecast? the marketing forecast? customer forecasts?

✓ What-if Capabilities. Should the system allow for multiple forecast scenarios to be maintained and compared? Do these multiple scenarios cover the entire hierarchy? Can these scenarios be archived for future forecast-error measurement?

ORGANIZATON CHECKLIST

✓ Accountability for Forecast Error and Inventory. Which functional area is responsible for forecast error? Which function is responsible for inventory performance? If they are one and the same, how can metric manipulation be discouraged? If they are different functions, how can finger-pointing be avoided?

✓ Detailed Forecast Responsibility. To whom do demand planners report? Sales? Marketing? Finance? Operations? An independent Supply Chain organization?

✓ Override Authority. Will executives or functions such as marketing have the ability to override the forecast without the agreement of the demand planner? Who will be accountable for the accuracy of those changes?

SUPPORTING DATA QUALITY

Data quality is one of the best-known risk factors for almost any IT project, and yet the impact is nearly always underestimated. In OF, the most common understanding of data quality requirements is to make sure that historical sales or shipment data are correct and a system is in place to correct errors. But this overlooks the data dependency at nearly every other step of the process, not just at its periodic starting point of introducing new history.

Consider this annotation of the OF cycle (Figure 5, next page), which indicates likely data quality hiccups at every step of the process.

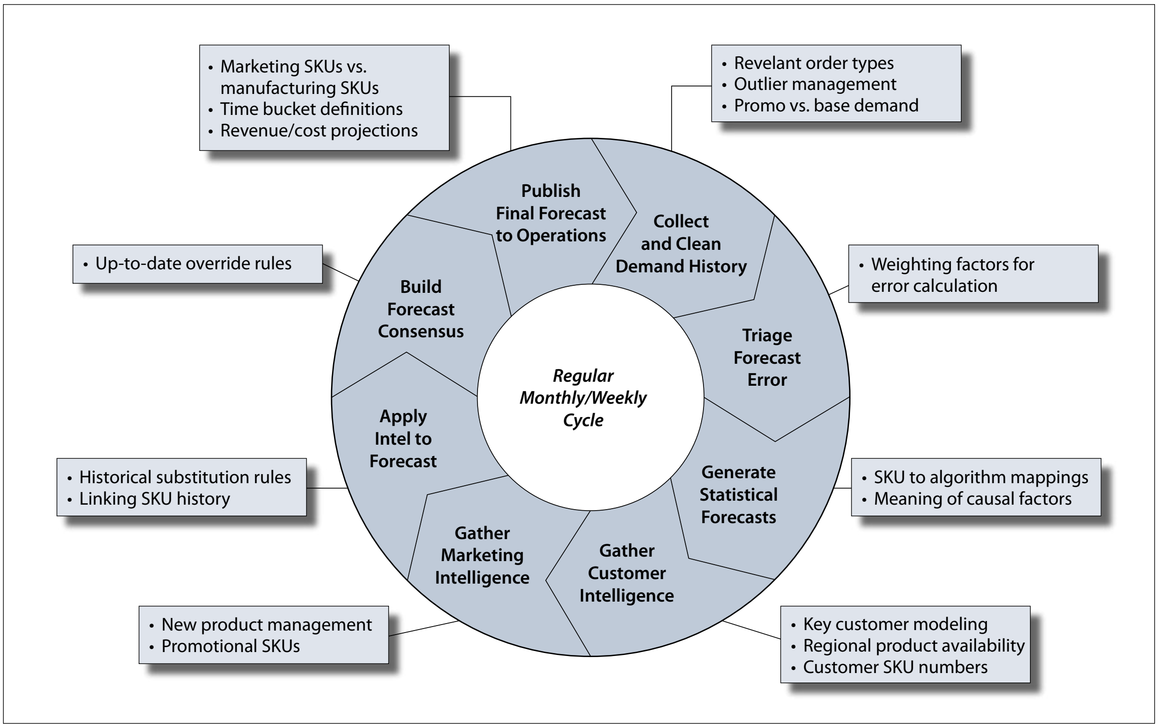
Product and customer masters used in order and inventory management systems are not usually clean enough for forecasting. They may contain a litany of one-off solutions to temporary customer service issues that will never be repeated and should certainly not be forecast: temporary customer locations, temporary SKUs, third-party handling, nonsellable products, accounting adjustments, etc. In addition, much of the forecasting process will require centralized maintenance of data that is not typically maintained or never existed before. SKU-to-algorithm mappings and historical substitution relationships need to evolve from spreadsheets or tribal knowledge to a data format the forecasting system understands. Relationships between forecasted SKUs and SKUs that are understood by customers or manufacturing must always be up to date.

Recommendations for Supporting Data Quality

Minimizing master-data quality problems means minimizing the number of new master-data requirements. Consider the example of a mapping table that links, possibly through some additional geographic or customer dimension, generic SKUs forecast by demand planners to specific SKUs planned by manufacturing. If such a table does not already exist in the organization, creating it for the forecasting project is a data-quality problem waiting to happen. Even if the table can be successfully built to support the initial go-live, any efficiencies you've gained will be lost to the never-ending maintenance needs of the table itself. During the design phase, consider the requirement for new master data in the pros and cons of any design decision.

However, for a majority of the data required for OF, the only reliable way to prevent quality issues would be to go back in time and fix all the data from the start. Failing

Figure 5. Operational Forecasting Cycle



that, it is best to assume that data quality issues will affect your project and plan accordingly. This means working with the data specialists on the implementation team to see full data extracts early and often. Allow time for complete datasets to be simulated in testing environments prior to go-live. Dedicate the time of team members who actually understand the data, typically the business users who are already in high demand, to review data extracts and make corrections. A good data specialist on the team can work with business users to identify exception criteria and automate some quality checks, but nothing can eliminate the need for business users to invest the time.

WHEN TO MAKE THE HARDWARE/SOFTWARE DECISIONS

Selecting and purchasing the computer hardware and software required to support the forecasting system are frequently afterthoughts. More accurately, they are decisions often made early in the project but not revisited as the design is refined and actual data loads are calculated. However, design decisions that are made during the project can result in underpowered hardware that creates a negative impression of the overall solution.

Forecasting software is designed around a series of hierarchies. The complexity of these hierarchies has more impact on the performance of the system than any other decision. Consider this hypothetical example:

A company chooses to model its 1,000 SKUs in its hierarchies; five major customers and three distribution centers would have a worst-case of 18,000 separate forecasts required. Adding a dimension for the four sales regions quadruples the worst-case to 72,000 forecasts. Adding the three potential manufacturing sites triples the worst-case again to 216,000. Choosing to forecast at the Customer DC level instead of the aggregate Customer level might increase the worst-case to >800,000 combinations.

Recommendations for the Timing of the New Hardware and Software

Nothing that is done before the solution design is complete can possibly account for this level of potential variation in hardware/software needs, so try delaying acquisition decisions as long as possible. Allow for an accurate sizing of the solution to take place.

During the installation, avoid unpleasant go-live surprises by loading a production-sized dataset and seeing how the system responds. Too many companies attempt this for the first time the weekend before they plan to develop their first forecast and leave themselves no choice but to endure disappointing performance.

SELECTING SOFTWARE: INTEGRATED VS. BEST OF BREED

The market for forecasting software is diverse, and there are a wide variety of vendors offering solutions. Each has strengths and weaknesses, but while there are a lot of different vendors, the capabilities that their solutions provide can be broken into two basic categories: statistical modeling and workflow facilitation.

Statistical modeling refers to the core analytical capability of the software – the available algorithms, the approach to multidimensional hierarchies, and the fit-to-product life cycles in your industry. This is usually the back-end part of the software that the majority of users do not see.

What users do see every day is the *workflow facilitation* capabilities of their software: Are judgment-based overrides allowed? How are overrides tracked and kept separate from the statistical forecast? How can multiple forecasts from multiple users be compared and tracked for accuracy? What reporting and exception identification capabilities speed the work of the Demand Planners? Can the available hierarchies provide a meaningful view of the business?

While the software should satisfy both statistical modeling and workflow facilitation requirements, very few companies really need both capabilities within the same tool.

Recommendations for Software Selection

The main decision is whether to choose a best-of-breed approach, where two different vendors provide the statistical modeling and workflow facilitation capabilities, or a single integrated solution, where one software tool provides both. A best-of-breed approach may be cheaper to implement or may allow a company to take advantage of industry-specific capabilities in one of the tools. Most OF processes separate statistical modeling activities from day-to-day forecast management, which minimizes the need for an integrated solution. Even so, a best-of-breed approach is not realistic in all cases. In very mature OF processes executed by sophisticated demand planners, an integrated solution will probably be required.

A best-of-breed solution will be feasible in companies where:

- Statistical forecasts are updated in an overnight or weekend batch.
- Demand planners' main responsibility is as a clearinghouse for collaboration.
- The majority of demand planners do not have statistical modeling skills.
- There is a need to disconnect statistical forecasting hierarchies and business hierarchies.

On the other hand, a single integrated solution will be required when:

- Sophisticated demand planners frequently switch between modeling, collaborating and overriding.
- Planner or collaborator input regularly affects results of statistical modeling.
- Forecasting software is being acquired as part of a large-scale supply chain software upgrade program.
- Simplification of IT portfolio is paramount.

Any forecasting project is unlikely to be the last one; if stakeholders have positive views of a past effort, continued investment and improvement will be that much easier to achieve.

Separately evaluate your requirements for statistical modeling and workflow facilitation and see how this influences your choices among the vendors you are considering. Also, given the constant consolidation and divestment in the software industry, make sure that anything marketed as a single integrated solution is not actually a solution cobbled together from two recent acquisitions.

CHANGE MANAGEMENT

Trepidation about process and organizational change is understandable. Occasionally it is misplaced, but OF stakeholders are right to worry about the impact of the project. Serious disruptions to the supply chain and customer service can occur if forecast accuracy suddenly dives or if order-of-magnitude errors slip through the cracks.

In the absence of *change management*, stakeholders might initially be asked to participate in interviews or workshops to lay out their requirements and expectations for the new forecasting process. But there may not be any follow through, and, courting disaster, the next stakeholders may be consulted only when the project is nearly ready to go live. In between, a business case is finalized, business processes are designed, the system is configured, scenarios are tested, and training materials are developed and delivered.

While out of the loop, in the new process these stakeholders may, to their chagrin, have been assigned new tasks and responsibilities that they never agreed to. But with the project nearly ready to go live, further changes are difficult and expensive. This conflict can happen in corporate cultures that generally embrace change if the project team takes the flexibility of their team for granted and does not consistently communicate and solicit feedback.

Recommendations for Change Management

Stakeholder engagement and business case management are not just activities to be done at the start of the project and then forgotten. The business case is a key element of an OF project and should have a long life span – from project conception to post-project evaluation of success. The same logic applies to stakeholder engagement – it too should precede and outlast the project.

Proper change management will insure that, after an initial gathering to solicit ideas, stakeholders should have the opportunity to review and approve the final design. While development is in process, demonstrations of key functionality should be held so that stakeholders can help insure that they match the initial vision and the current business environment. Conceptual training should be delivered to most stakeholders so they know what to expect and what will be expected of them after go-live. Finally, ongoing business case measurement after go-live should be shared with key stakeholders so they remain supporters even as the project itself fades into the past. Any forecasting project is unlikely to be the last one; if stakeholders have positive views of a past effort, continued investment and improvement will be that much easier to achieve.

In some corporate cultures, the project team will have to overcome cynicism about the chances for success or active disengagement by functional personnel trying to avoid involvement or protecting their own agendas. Sales and marketing teams that have long viewed forecasting as someone else's problem will not embrace their new responsibilities in contributing to forecast consensus. An operations organization that has been quietly executing its own demand plan, independent of whatever sales and marketing have published, will be skeptical of the new forecast.

In these cases, a more concentrated campaign of communications and build-up of executive support will be needed. Unsupportive stakeholders need to understand how they will benefit and how the forecast will improve before they will begin providing meaningful input, contributing to consensus, or accepting the results. As with forecasting itself, there are top-down and bottom-up approaches to this, and combining both is the best strategy. Going top-down, work with open-minded leaders on developing the business case and communication strategies for the more skeptical. The bottom-up approach means furiously working to get influential and vocal planners on board with the new process and system and encouraging them to make sincere efforts to manage their current and future workloads, using simulations with real data that demonstrate the potential for the new process.

CONCLUSIONS

This article has examined challenges facing OF project management and offered recommendations for anticipating and overcoming these challenges. The obstacles are daunting:

- Numerous stakeholders with competing agendas
- A leap of faith to link anticipated results to business benefits
- Difficult design issues with competing best practices
- A broad range of overlapping software options
- The need for high data quality where procedures are typically lax
- The change management required in introducing a new enterprise-wide demand signal

Despite these hurdles, it is important to stay focused on how worthwhile this investment can be. Investing in OF has the potential for return year after year. It provides the key element needed to run an effective S&OP process and is the springboard for other supply chain initiatives, including distribution planning, master planning, promotion planning, and inventory optimization.

When considering investing in OF, or managing a project in progress, use the lessons outlined here to assess your readiness to proceed, to make preemptive plans for the predictable problems, and to allot sufficient time and resources for the unpredictable, but likely, difficulties that lie ahead.

Other *Foresight* Articles on Improving the Operational Forecasting Process

- Special features on cost of forecast error, Issues 7 and 8.
- Borneman, J. The forecaster as leader of the forecasting process, Issue 7, 41-44.
- Clarke, S. Managing the introduction of a structured forecast process, Issue 4, 21-25.
- Mello, J. The impact of corporate culture on sales forecasting, Issue 2, 12-15.
- Mello, J. & Esper, T. S&OP, forecasting, and the knowledge-creating company, Issue 7, 23-27.
- Moon, M. Breaking down barriers to forecast process improvement, Issue 4, 26-30.
- Oliva, R. & Watson, N. Managing functional biases in organizational forecasts, Issue 5, 27-31.
- Sepulveda-Guzman, M., Smith, M. & Mechling, G. Forecasting as a business process diagnostic, Issue 3, 22-24.

CONTACT
Ian Watson-Jones
IBM Global Business Services
Ian.watson-jones@us.ibm.com

COMMENTARIES ON OVERCOMING CHALLENGES IN OPERATIONAL FORECASTING PROJECTS

Patrick Wader

Ian Watson-Jones's article provides valuable advice to anyone involved in operational forecasting. Based on years of experience in OF, I agree with most of Ian's thoughts and wish to expand on his main points.

1. Process Definition. Ian warns against underestimating the time needed to solve process issues prior to implementation. I completely agree. A fresh OF project must first eliminate multiple process questions and provide a thorough definition of planning for the entire supply chain. Otherwise forecasting tools could become twisted to support ordering, promoting marketing calendars, or production planning. An overall process definition at the outset could lead to a more suitable tool. Therefore, define beforehand the expected inputs and the projected treatment of outputs further down the planning chain. Then pick your tool.

Often systems or data flow processes neglect the organizational view of those who actually perform the process steps. Specifying who does what will help to fully define those steps. Divisional structures are prevalent in many companies, so remember that an understanding of "demand planner" needs to be shared across divisions, if one process and software solution is to be rolled out. Unfortunately, this common understanding is rare. Mandatory organizational change management may take twice to three times longer than the project itself.

As Director of Program Management worldwide for Robert Bosch GmbH in Germany, Patrick is in charge of rolling out IT solutions globally for one Bosch division. He has led various forecasting projects in international companies as well as in research. Patrick holds a PhD from Aachen University of Technology, Germany. Better even than forecasting, he likes skiing, sailing, and recitations from *The Big Lebowski*.

Mark A. Moon

Ian Watson-Jones has written a very useful and readable article on managing forecasting-improvement projects. He has identified a number of key requisites for success in project management, including effective stakeholder management, a proper business case, an efficient, consensus-driven business process, a plan for overcoming data quality issues, and hardware/software support for the business requirements. Perhaps his most important contribution, however, is the recognition that *change management* is the linchpin of successful forecasting process improvement. Over the past twelve years, I have worked with more than thirty companies on forecasting process-improvement projects. Without exception, I found that those that succeed are projects where change management is a clear priority.

There are three additional points, however, that I would offer to supplement Ian's list: the critical role of demand/supply integration (DSI), the importance of sales and marketing buy-in, and the need for a process-improvement *champion* to ensure overall project success.

Demand/Supply Integration. While Ian does refer to S&OP processes, I would like to emphasize the crucial nature of this overall business-planning process, which I refer to as demand/supply integration (DSI), and the role that demand forecasting plays in its success. The

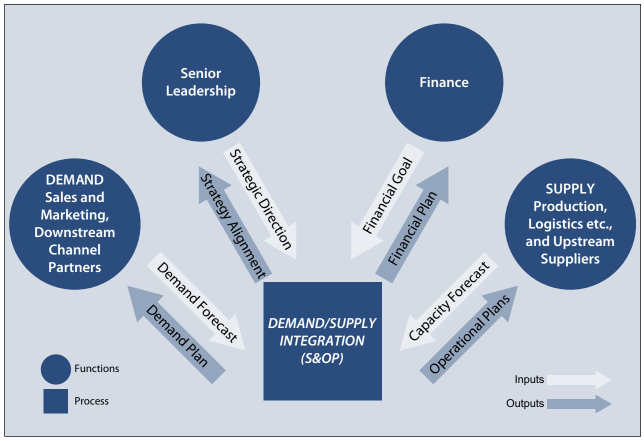
Mark Moon is Director of the Sales Forecasting Management Forum and Associate Professor of Marketing at the University of Tennessee, Knoxville. Mark and UTK colleague Tom Mentzer are co-authors of *Sales Forecasting Management: A Demand Management Approach* (Sage, 2005). See Mark's previous *Foresight* article, "Breaking Down Barriers to Forecast Process Improvement," in Issue 4 (June 2006).

2. Stakeholder Management. Who owns the forecast? The answer will quickly tell you who feels involved with, and who feels left out of, the process. If sales owns the forecast, do they also own the inventory that derives from this plan? If demand planners do, what is the responsibility of marketing/sales? If operations is directly involved, how do you prevent forecasting to capacity constraints? Ian and I agree on sharply defining requirements by stakeholder group, having clear expectations, and starting with a lean project.

3. Business Case. You can realize the benefits of improved forecasting only if planning/execution processes are stable and clear input/output relations are established. In a make-to-stock environment, input to inventory planning feeds production and procurement planning. If these processes are "flexible" (i.e., forecasts are untrusted and can be washed down to other planners' "beliefs"), no fill rate and inventory-level effects will be provable, and you'll be able to measure only the effect on forecast errors. MAPEs alone rarely impress executives.

I stress Ian's point that we must begin with the leanest project possible. Most supply chains are not linear. One central warehouse delivers to many local warehouses that have overlapping distribution zones. Two plants may produce identical SKUs forming complex supply networks, which makes it more difficult to show effects of improved forecasting and planning. We deal with this difficulty by cutting out pilot regions and supply chains – the smaller the better. Focus first on the products of one division, in one market. It is easier to show the effect of better demand forecasting for one product line in a small market at the end of a distribution chain than to do this for all products in a major market with complex, overlapping distribution structures. Also, it's not wise to try to improve planning processes while simultaneously changing physical structures. Historic data will not fit easily, and problems in supply-chain execution and consequent mistrust of players can alter planning inputs.

Figure 1. Demand/Supply Integration (DSI)



University of Tennessee vision of DSI is depicted in Figure 1.

While it is beyond the scope of this commentary to explain this diagram in detail, suffice it to say that demand forecasting serves as a critical element to an overall, corporate-wide effort to integrate demand with supply. Process-improvement efforts are most effective when they are incorporated into a broad, enterprise-level initiative. It is true that this integration may make achieving buy-in and arriving at consensus more complicated. Still, successful companies understand that effective forecasting practices are of little value unless the forecasts are balanced against capacity constraints, financial goals, and corporate objectives.

Buy-In from Sales and Marketing. My second element is the critical nature of buy-in from the demand side of the organization – namely, sales and marketing. In my experience, the most common source of resistance to process-improvement efforts in forecasting are found in the sales and marketing groups. The supply-side functions in most enterprises are usually the ones that feel the most immediate pain of inaccurate forecasts. Manufacturing groups suffer from eleventh-hour changes to accommodate unexpected demand. Transportation groups suffer from unbudgeted expedited shipping. Procurement groups suffer from last-minute purchases of components or

4. Data Quality. Ian notes that data quality is key to all steps of the forecasting process. Two thoughts: First, purge historical data of errors; then mark past events to get a clean baseline that can be used for future event forecasting. Don't simply throw out event effects from historic data. Second, identify early how you define forecasting accuracy and what system and database you will use to track it. Record your baseline accuracy before rolling out a new solution. Building the database solution (e.g., a business warehouse) requires extra effort and can quickly become its own IT project.

5. Hardware/Software Issues. The author advises delaying hardware and software decisions as long as possible. I agree in principle, but prefer to separate software from hardware considerations. OF projects are often driven to implement software that a vendor proposes, usually with little market research on comparing existing solutions. The bigger the company, the more important it is to check various criteria before going with one specific program and vendor. Instead, derive a well-defined set of tightly managed business requirements from the target planning process. These requirements also reveal the workflow facilitation and process support needed in the planning stages. Consider:

- Is the scope of the project truly limited to forecasting, or does it touch upon various planning steps?
- How many users will there be? Where? Requiring what languages?
- Is the vendor big and stable enough to support a worldwide rollout?
- Does the software perform batch processing or real-time calculations? If batch processing, what are the expected computation times/downtimes? What comes on top for interface runs? Can one solution run worldwide, 24/7?

After you decide what program and vendor fit your target processes best, your technical requirements usually suggest appropriate hardware and software choices.

raw materials, often at higher spot-market prices. And supply-chain executives suffer from holding excessive inventory to cover for volatile, but unforecasted, fluctuations in demand.

These supply-side functions usually require very little convincing that forecasting process improvements are necessary. However, we find that the demand-side functions need to be more engaged in the process. Time and again I have heard sales executives say, "I want my people selling, not forecasting!" However, those sales and marketing people are the closest to customers and markets and are best positioned to recognize and report on future demand issues. Thus the change-management challenge is often the most acute when engaging the sales and marketing functions in forecasting process improvement.

Ian observes that both top-down and bottom-up change management is needed in process-improvement efforts. In the case of sales and marketing buy-in, I believe that top-down support (and enforcement, if necessary) is absolutely essential: senior-executive support must be present.

It is useful to follow Ian's advice to "manage these stakeholders" by demonstrating what improved forecasting can do for them. Although sales and marketing people may need to become more fully engaged, they will be persuaded by evidence that fill rates will be improved, customer-service levels will be enhanced, and product will be more available to ship on time if forecasting can be more effective.

Forecasting Champion. My final consideration is the need for an effective *forecasting champion*. Our research team discovered that forecasting improvement efforts are most effective when they are shepherded by effective change agents (Mentzer et al., 1997). Often, when I am working with a company on a forecasting improvement project, I will ask numerous people the following question: "If your CEO wakes up in

6. Change Management. I concur that stakeholders must be involved in all stages of the project. From the start of business design to the rollout of a forecasting solution, change management has to ensure continuous support. Do not plan your first rollout with the same duration as your fifth; allow for a beginning buffer. Meanwhile, pick pilot projects where you anticipate little resistance to change. Don't try to replace an existing, long-term forecasting process at one division if you haven't proven an enhanced approach elsewhere, forming strong allies through desirable bottom-line effects.

SUMMARY

The ingredients to a successful OF project are a clearly defined process and change management that targets a small pilot area. Unfortunately, many OF projects are software driven without clear process requirements. The bigger the company, the greater the need to avoid this approach. A successful road map might be:

1. Focus the planning process, defining all major acting roles.
2. Derive business and technical specs.
3. Pick software/vendor, then hardware.
4. Pick small-scale pilot area.
5. Implement quickly, prove success.
6. Only then: *roll out.*

CONTACT
Patrick Wader
wader@gmx.net

the middle of the night with a forecasting nightmare, whom does he or she call?" If I ask that question to ten people, and I get ten different names in response, then I know that the company has a problem in terms of its organizational focus on forecasting. As discussed in the Mentzer article, a forecasting champion serves the role of process designer, cross-functional orchestrator, resource collector, and overall cheerleader, keeping forecasting and demand/supply integration on the radar screen of key organizational constituents. Companies that want to improve these processes must first commit to establishing such a go-to position, then choose an effective agent to fill that role and, finally, provide the resources and support that this champion will need to be effective.

To summarize, I would add 3 additional recommendations to Ian's valuable list:

1. Put the project in the context of an overall DSI strategy.
2. Work hard to win the support of demand-side constituents, particularly sales and marketing.
3. Choose and support a forecasting champion as the point person in the process-improvement effort.

Attention to these issues may increase your organization's chances of success dramatically.

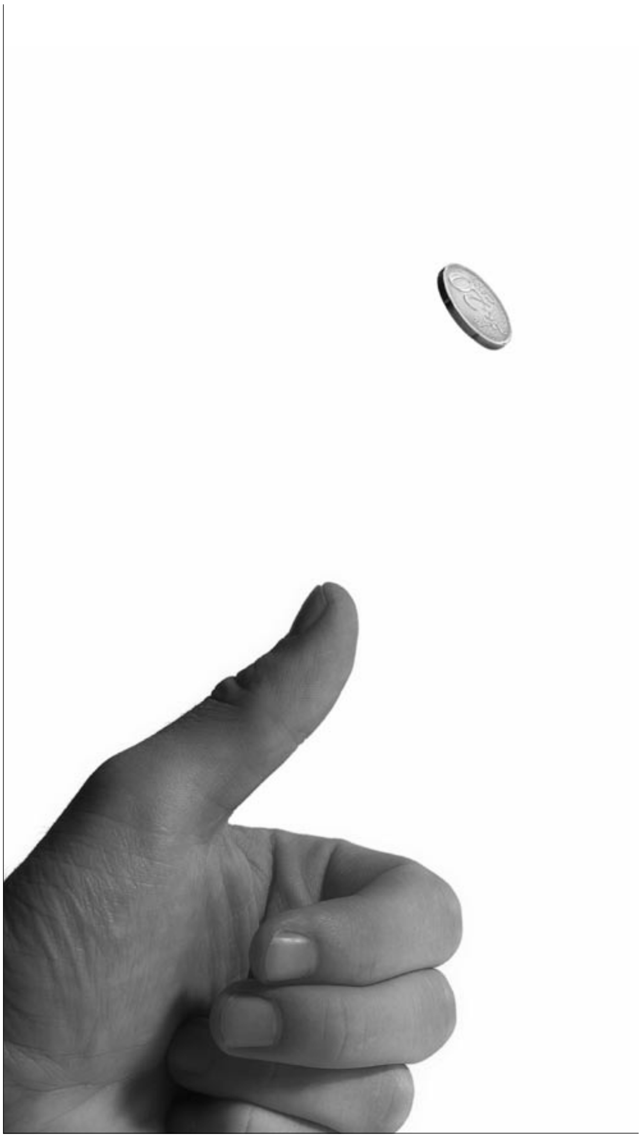
REFERENCE

Mentzer, J.T., Moon, M.A., Kent, J.L. & Smith, C.D. (1997). The need for a forecasting champion, *Journal of Business Forecasting*, 16 (Fall), 3-8.

CONTACT
Mark A. Moon
Department of Marketing and Logistics
University of Tennessee
mmoon@utk.edu

REGRESSION MODEL FORECASTS OF THE U.S. PRESIDENTIAL ELECTION

Randall Jones, Jr. and Alfred Cuzán



In the Summer 2008 issue of *Foresight*, we described regression models that have been used to forecast American presidential elections over the past three decades. Most of the analysts who created these models are still forecasting presidential elections and have made their forecasts for 2008, which are now available and are reported here.

Many of the models have been modified over time, but the structure of four of them has remained relatively stable since 1996. Those four models were featured in the previous article and appear first in the accompanying table of forecasts. All of them incorporate economic growth and, with the exception of Ray Fair's model, consider some measure of public opinion toward either the current administration or the presidential candidates themselves. These indicators are common among the remaining models in the table as well, some of which also take into account the cyclical nature of elections, wars during election years, primary election results, government spending, and time in office.

As evident in the table, all of the regression models but one forecast a loss by the incumbent party's candidate, Republican Senator John McCain. Across all of the models, his median share of the major-party

Randall J. Jones, Jr. is Professor of Political Science at the University of Central Oklahoma, the author of *Who Will Be in the White House? Predicting Presidential Elections (2002)*, and, with Alfred Cuzán and Scott Armstrong, creator of the Pollyvote Web site, www.politicalforecasting.com.

Alfred G. Cuzán is Chairman of the Department of Government at the University of West Florida. He has written widely on election forecasting, including an article he coauthored with Randall on forecasting midterm elections to the U.S. House of Representatives that appeared in the Fall 2006 issue of *Foresight*.

vote is estimated to be 48%. The median forecast for Democratic Senator Barack Obama is 52%. This is about the same result as that obtained from the four stable models and all other models, respectively, taken separately. The regression forecasts are close to the predictions by other methods, as reported at Pollyvote.com. Thus, it is nearly a consensus prediction that the Democrats will likely win the presidency in 2008. However, the expected margin is relatively small, so a McCain victory cannot be ruled out. To paraphrase the inimitable Yogi Berra (party affiliation unknown), this election ain't over till it's over.

Regression Forecasts of the Major-Party Vote in the 2008 American Presidential Election

AUTHOR	DATE OF FORECAST	PERCENT OF TWO-PARTY VOTE	
		McCain	Obama
Ray Fair*	July 31	48.5	51.5
Alan Abramowitz	August 28	45.7	54.3
Christopher Wlezien & Robert Erikson	August 28	47.8	52.2
James Campbell	September 8	52.7	47.3
Allan Lichtman	August 7, 2007	46.0	54.0
Helmut Norpoth	January 15	49.9	50.1
Douglas Hibbs	June 7	48.2	51.8
Karl Klarner	July 28	47.0	53.0
Alfred Cuzán & Charles Bundrick	August 2	48.0	52.0
Thomas Holbrook	August 28	44.3	55.7
Michael Lewis-Beck & Charles Tien	August 28	49.9	50.1
Brad Lockerbie	August 28	41.8	58.2
Andreas Graefe & Scott Armstrong	September 3	48.8	51.2
Median (mean), first four models	August 28	48.2 (48.7)	51.8 (51.3)
Median (mean), all other models	August 2	48.0 (47.1)	52.0 (52.9)
Median (mean), all models	August 2	48.0 (47.6)	52.0 (52.4)

The first four forecasts are from models that have undergone little revision through several election cycles. Unless otherwise noted, forecasts were issued in 2008.

*Fair updates his forecast quarterly. His earliest forecast, announced on November 1, 2006, was for the Republican to receive 46.5% of the major-party vote.

CONTACT
Randall J. Jones, Jr.
University of Central Oklahoma
ranjones@ucok.edu

Alfred G. Cuzán
The University of West Florida
acuzan@uwf.edu

FORECAST PRO UNLIMITED

AN OFF-THE-SHELF SOLUTION FOR LARGE-VOLUME FORECASTING

Ulrich Küsters and Janko Thyson

Editor's Note: *Foresight* invited three vendors to participate in an evaluation of off-the-shelf forecasting software: Business Forecast Systems (BFS), developer of Forecast Pro products; Smart Software (SmartForecasts products); and John Galt (Forecast X products). Only BFS accepted the invitation. While we appreciate the cooperation of BFS, *Foresight* has serious concerns about a pattern of reluctance by vendors to open their products to public scrutiny.

Pricing and Contact Information*
Forecast Pro Basic: $795
Forecast Pro XE: $1,295
Forecast Pro Unlimited: $4,995

*USD, single user licenses, for English editions.

Quantity and Academic discounts are available.

Business Forecast Systems, Inc.
68 Leonard Street, Belmont, MA 02478 USA
Phone: 617.484.5050 Fax: 617.484.9219
Email: info@forecastpro.com/Online: **www.forecastpro.com**

INTRODUCTION

In this software column, we review Version 5.0 of *Forecast Pro Unlimited* (FPU), a Windows-based product developed by Business Forecast Systems, Inc. (BFS). The program is a successor to the Forecast Pro Batch system, which has been on the market for more than a decade. Compared to *Forecast Pro Extended Edition* – BFS's "hands-on" product – *Forecast Pro Unlimited* provides a smaller method spectrum while offering enhanced data management facilities. FPU is not a planning system, however.

The *Forecast Pro Unlimited* package is mainly for automatic and semiautomatic generation of forecasts for large product hierarchies. The program is designed for users with a business background but who needn't have an in-depth knowledge of statistical forecasting. *Forecast Pro Unlimited* is aimed at large, product-oriented companies in such economic sectors as manufacturing, pharmaceuticals, and telecommunications. In these organizations, forecasting and planning are typically based on monthly data with a horizon up to 18 months, as well as on weekly data with horizons up to 6 weeks.

FORECAST METHODS SPECTRUM

From a statistical point of view, *Forecast Pro Unlimited* offers established and powerful forecasting methods, including exponential smoothing and Box-Jenkins models. It also provides a number of simple forecasting methods, such as the random walk, moving averages, and trend curves. These models are useful for forecasting very short series and often serve as naïve benchmarks for more sophisticated models.

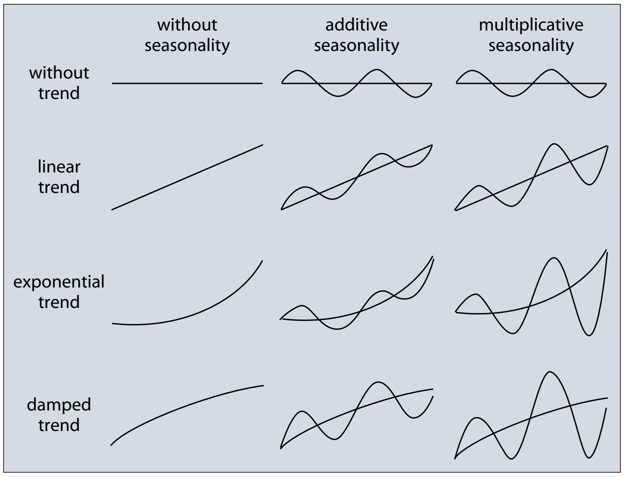
Exponential Smoothing

Forecast Pro Unlimited provides a complete implementation of the Gardner family of exponential smoothing methods (Gardner, 1985). This family combines linear, damped, and exponential-trend methods with additive and multiplicative seasonal indexes (see Figure 1).

Ulrich Küsters is Head of Statistics and Operations Research at the Ingolstadt School of Management. He is coauthor of *The Forecasting Report* (1999) on commercial forecasting systems and of "Forecasting Software: Past, Present and Future" in the special issue (2006) of the *International Journal of Forecasting: Twenty-Five Years of Forecasting.*

Janko Thyson is a research assistant and PhD student in Statistics and Operations Research at the Ingolstadt School of Management.

Figure 1. The Exponential Smoothing Family



Forecast Pro Unlimited offers event models as a useful extension of the exponential smoothing family. This feature allows the inclusion of one or more variables to represent special events, such as holidays and promotions that are not tied to the calendar. For example, the Easter holiday can fall in different weeks of the year and in fact shifts between March and April. Event indexes can also provide estimates of trading-day effects, which are particularly important at the retail level.

The use of event models assumes that the impact of future events of the same type will be consistent in size and shape with their most recently estimated effects. Otherwise, these effects could not be propagated into the future. The event-index approach is a unique feature of *Forecast Pro Unlimited* (as well as of *Forecast Pro Extended Edition*) among exponential smoothing programs.

Forecast Pro Unlimited also offers procedures for forecasting intermittent demands – mainly Croston's method (1972) – but also discrete models based on Poisson and the negative binomial distributions.

A very important and practical feature is the use of product hierarchies to improve estimation of seasonal indexes. *Forecast Pro Unlimited* permits the estimation of seasonal indexes in a higher node (e.g., a product-group aggregate), which can be imposed on lower nodes (individual products or SKUs) of the hierarchy. This approach is useful in situations where data at lower nodes do not include enough cycles to directly estimate seasonal indexes, but the aggregate data have a sufficient time span to do so.

Forecast Pro Unlimited can calculate prediction intervals and safety stock requirements from exponential smoothing models. These calculations derive from approaches recommended in Yar and Chatfield (1990), as well as from ARIMA-equivalent models. The recent advances based on state space models with single source of errors (SSOE) – see Hyndman, Koehler, Ord and Snyder, 2005 – have not been implemented, because the SSOE approach has not yet been enhanced to product hierarchies.

Box-Jenkins

Forecast Pro Unlimited provides a full implementation of the univariate Box-Jenkins methodology (ARIMA), including models for nonseasonal and seasonal data and all commonly used power transformations like the logarithm. Model identification and estimation is essentially automatic. The program does not offer model identification tools, such as ACF and PACF graphs and explicit unit root tests like Dickey and Fuller's ADF test needed for manual selection of a model. Not all the details of how the automatic modeling works are published. However, we do know that the program utilizes backcasting as the initialization technique for estimating parameters. Furthermore, it relies on an information criterion for model selection.

Expert Selection

The program's most valuable feature is the *Expert Selection Mode*, which automatically chooses the "most appropriate" forecasting technique from among the options available in the program. Although the selection details are not transparent, the *Expert* applies several statistical tests and compares different models on the time series being forecasted.

It is common for vendors of commercial forecasting systems to treat such details as proprietary, and, given this, forecasting software cannot be completely evaluated from a theoretical and conceptual perspective. From a practical view, though, *Forecast Pro Unlimited* offers a number of advantages over competing forecasting systems, and these should give the user some confidence in its forecasting performance.

First, the methods it offers comprise a substantial subset of proven and established procedures commonly used in product hierarchy forecasting of manufacturing and sales data. Second, a variant of *Forecast Pro Unlimited* was tested in the famous M3 competition (Makridakis & Hibon, 2000) with favorable results. And third, users can perform rolling out-of-sample evaluations on their own data.

This procedure allows the user to *hold out* a number of the most recent data points from a time series. The program then applies a forecasting method to the remaining data (the fit set) and generates forecasts for the hold-out period; this enables the user to assess the forecast accuracy of the method, not merely the goodness of fit. The process repeats itself after one of the held-out data points is moved to the fit set (hence, "rolling" from the test to the fit set). The result is a tabulation of the accuracy of the forecasting method at different forecasting horizons. The procedure can be applied both to methods selected by the user and to the automatic forecasts of the *Expert*. The program restricts the number of hold-out data points to a value of 99, limiting its usefulness on daily data. For a primer on out-of-sample evaluations, see Len Tashman's article (2000).

The program allows the user to manually select a forecasting method, but it should be noted that some of the available methods can be applied only if the data meet certain requirements. An example is the requirement of at least s+5 observations for seasonal exponential smoothing with linear trend and multiplicative seasonal indexes (s, the seasonal period, equals 12 for monthly data), which can be forced by the modifier "\EXSM=LM". If this requirement is not met, the program issues a warning and resets the model to "\EXSM=LN", dropping seasonal terms. From a statistical point of view, this makes sense.

Most planners consider Holt-Winters as a model with linear trend and seasonal indexes. This can be forced by the modifier "\WINTERS". Note, however, that the *Expert* switches to a linear nonseasonal Holt model without any warning whenever the number of available observations is less than s.

PRODUCT HIERARCHIES, DATA MANAGEMENT, AND WORKFLOW

The primary challenge of production and sales planning is the selection of an appropriate forecasting method for each product or item. This can be a formidable task when there are huge numbers of items to forecast, and the forecasts must be translated into planning figures. While there are several *forecasting* software solutions on the market, very few of them capably perform the database functions needed to reasonably represent a firm's hierarchical data. *Planning systems*, by contrast, usually offer excellent data management facilities for product hierarchies but perform forecasting less than adequately.

Product Hierarchies

As we noted earlier, *Forecast Pro Unlimited* is not a planning system. While the program contains some planning features (like manual forecast overrides), it concentrates on forecasting. The user must manually set up the product hierarchy. In the simplest case, the hierarchy might be flat, but complex hierarchies can require a large number of nodes where each comprises a combination of product group, item, organizational unit, region, and sales channel. The hierarchy must be linked to one or more datasets, which are usually flat ASCII files or flat MS Excel spreadsheets. According to BFS, more than 97% of their customers use one of these. However, it is also possible to use a structured relational database with an ODBC interface, like MS Access.

So how can a production or sales planner effectively use *Forecast Pro Unlimited* as a forecasting engine?

These four steps must be performed:

1. Define a data source, including all time series to be forecasted. This dataset has to be updated outside of *Forecast Pro Unlimited*.

2. Within *Forecast Pro Unlimited*, define all options, including data modes, data sources, forecast horizons, outlier detection and correction procedures, confidence probabilities, output structures and formats, thresholds, etc. or accept the program defaults.

3. Define the product hierarchy, including *modifiers* that instruct the program on how to reconcile the forecasts among the different node levels. A typical reconciliation approach is called top down and placing a top-down modifier at a node (e.g., product group level) results in a prorating of forecasts generated for this level to all nodes below (e.g., products, SKUs).

As soon as the hierarchy is defined, the *Expert* takes over and generates forecasts for all items.

4. After saving all internally defined options and data definitions in a project file, the user can set up a batch file which can be called up every time the data source is updated. The forecasts and related information can be written to a database when using ODBC, or they can be retrieved as MS Excel spreadsheets and/or flat files.

A user needs to think carefully about the definition of the product hierarchy. It is common in organizations that different units (e.g., sales, production, logistics) require different, but internally consistent, forecasts. As a result, the forecasts generated by *Forecast Pro Unlimited* have to be aggregated based on different attributes. While the production department is mainly interested in forecasts of the number of units of simple SKUs, the sales department needs value-based forecasts aggregated for certain combinations of regions, sales channels, and product groups. The transport department often requires forecasts of weights and/or volumes, and so on.

In such cases, a user could define a very detailed product hierarchy, where terminal nodes consist of all logical combinations of relevant attributes. But doing so usually results in intermittent demand at the lowest nodes, which are forecasted in automatic mode by Croston's method. The Croston method, though, is not able to detect seasonal effects and trends which are usually present in the aggregates. In such cases, it is advisable to utilize the top-down modifier to force incorporation of the trend and seasonality in a product aggregate.

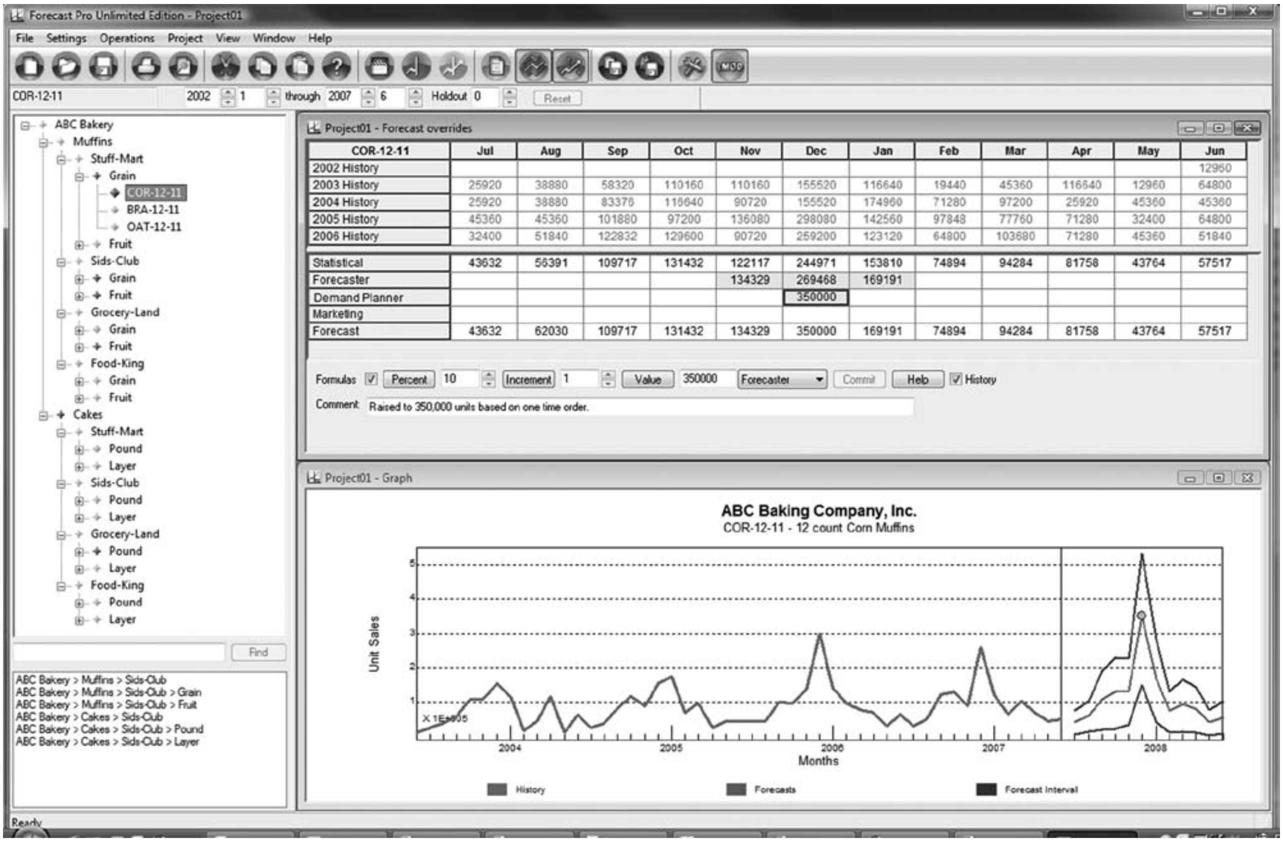
Take note: attribute-based aggregation cannot be done within *Forecast Pro Unlimited*. For such purposes, it is advisable either to use a relational data base management system (RDBMS) or an appropriately configured on line analytical processing (OLAP) tool.

Overrides

These steps have described how *Forecast Pro Unlimited* can be used as a forecasting engine. The user can also generate forecasts, confidence intervals, reports, and graphs directly out of FPU without invoking the batch file. In this case, the forecaster inspects all results, including outlier modifications, directly within the program. Furthermore, it is also possible to override an automatically generated forecast judgmentally by applying percentage or absolute adjustments, as well as direct value overrides. One of the great advantages of *Forecast Pro Unlimited* is the fact that data modifiers and overrides defined in a previous session are not lost if the project file has been saved.

Figure 2 shows a screenshot from *Forecast Pro Unlimited*. The left pane, with the *tree view control*, depicts a partially unfolded product hierarchy of the "Bakery" dataset, which comes with the program as a comprehensive example. This bakery has two main products, muffins and cakes, each of which is distributed through a variety of channels (e.g., "Grocery Land"). The bottom pane of the graphic shows the history of total bakery sales – the top node – along with the fitted values (which show how the method tracks the historical data), the forecasts, and their 90 percent

Figure 2. Screenshot of Forecast Pro Unlimited



confidence limits. The upper pane shows the historical values, statistical forecasts, and overrides.

When more than one user override is defined, the system takes the last override as the final forecast. Relational overrides (increment or percent) can either be defined on the statistical forecast (labeled as the "statistical" row in Figure 2) or on any other override row. Weighted averages of the respective overrides cannot be "committed" – that is, stored as final forecasts.

When overrides are committed, the section "Out-of-Sample Rolling Evaluation" of the *Forecast Report* always relies on the statistical forecasts. In contrast, the section "Out-of-Sample Static Evaluation" relies on the committed overrides to the statistical forecasts. Only in the case where no overrides are committed do the two sections coincide.

Forecast Pro Unlimited re-centers the confidence intervals on the final forecasts. However, it does not increase the confidence interval width to reflect the additional risk usually imposed by overriding.

The program allows the configuration of additional panes. It is possible, for example, to show the results of the estimation steps as well as the results of the automatic outlier detection and correction procedures (usually deactivated by default).

Forecast-Accuracy Comparisons

A useful feature, as we noted above, is the evaluation of forecast accuracy through out-of-sample evaluations, which are enabled by holding out certain portions of the time series. These evaluations can be used to compare different forecasting methods or method options, including manual overrides committed by different parties of the organization. Unfortunately, *Forecast Pro Unlimited* does not offer a systematic way to manage and compare competing forecasts.

To do so, the user has to rely on external tools, like MS Excel. In our opinion, the user will rarely bother.

Forecast Pro Unlimited provides only a metric that benchmarks a forecast against a naïve forecast – which is the *random walk* forecast of no change from one period to the next.

Coping with Outliers

Forecast Pro Unlimited is one of the few systems that provides automatic outlier detection and outlier value replacement. However, its technique is rudimentary, as it is based on the iterative replacement of outliers by fitted values. A model-based technique comparable to Ruey Tsay's outlier analysis for Box-Jenkins models (1986) would be preferable.

Factory Defaults

We would caution users that care should be taken when using the factory defaults set by BFS. For example, the default setting for the switch "Allow negative forecasts" in the options menu is set to "OFF." While it is often reasonable to disallow negative forecasts in sales and production forecasting, it might lead to surprising results for data that contain negative values. For example, daily data on newspaper sales often include negative values due to remissions. In such cases, this default could not only misforecast but also generate phantom outliers.

Units vs. Dollars

Unfortunately, *Forecast Pro Unlimited* cannot store and forecast data in units and monetary values simultaneously. The user has to decide either to store all data as units or as monetary values.

Interfacing with MS Excel

MS Excel users will like the powerful export functions available in this package. *Forecast Pro Unlimited* reports such as the *Formatted Forecast Report* and the *Numeric Output Report* can be immediately post-processed in MS Excel. As the program implicitly uses an MS Excel spreadsheet table structure in all reports, results can be easily copied and pasted into MS Excel from screen to screen.

Interfacing with Non-Microsoft Tools

It is still a difficult task to integrate a forecasting system like FPU within the planning processes when non-Microsoft tools like SAP and i2 are used. Certified interfaces are often required. As *Forecast Pro Unlimited* is mainly a stand-alone system, data, forecasts, and planning figures have to be transferred by one of the certified interfaces. This usually requires substantial programming.

Resets

A planner who is not fully familiar with FPU might run the risk of doing unnecessary work. For example, when an option like the hold-out period or the threshold for outlier analysis is changed by the planner, FPU warns him that it will eliminate all modifiers and overrides. Even when accepting this warning message, these project settings are still available and can be restored. However, the user is left alone to arrive at the conclusion that a new data read-in and a reforecast are necessary to restore modifiers and overrides.

RELATION TO FORECAST PRO EXTENDED EDITION (XE)

BFS, the vendor of *Forecast Pro Unlimited*, offers another system, *Forecast Pro XE*, which is targeted to analysts with more sophisticated business forecasting skills. The XE package offers several tools and procedures that are not included in *Forecast Pro Unlimited*; most notable of these are dynamic regression models that allow the estimation of causal relationships between explanatory variables (regressors) and the dependent variable. These relations can be concurrent and/or lagged, allowing the user to specify early-warning indicators as explanatory variables. XE also offers an application of the seasonal decomposition method X11 developed by the U.S. Bureau of the Census. As some methods are not implemented in *Forecast Pro Unlimited*, users who experiment with their data in XE cannot transfer all models to the other program.

At first glance, it might seem unnecessary for *Forecast Pro Unlimited* users to analyze their data with XE, but there is one major exception. While *Forecast Pro Unlimited* allows the choice of specific forecasting methods to individual nodes in a product hierarchy, it does not possess any tools to analyze the adequacy of these prespecified models. As mentioned before, *Forecast Pro Unlimited* lacks key graphical facilities for model identification, like residual ACF and PACF plots. Given this deficiency, any user who is going to apply specified models in *Forecast Pro Unlimited* is well advised to analyze the data carefully with XE, resorting to all available model identification tools to check the adequacy of the chosen model.

Users who migrate from XE to *Forecast Pro Unlimited* will encounter the problem that scripts generated for XE cannot be transferred directly. The main purpose of scripts in *Forecast Pro Unlimited* is the definition of groups and the specification of data files, whereas XE scripts also allow the direct declaration of forecast modifiers. This, in turn, might yield substantial migration efforts, as all XE modifiers have to be transferred manually to a *Forecast Pro Unlimited* project.

Another issue concerns the number of time series which can be forecast. While *Forecast Pro Unlimited* has virtually no restrictions beyond memory and disk-space limitations, XE can be used to forecast only up to 100 demand streams at a time. *Forecast Pro Basic*, BFS's entry-level product, allows only 10.

CONCLUSIONS

With the development of *Forecast Pro Unlimited*, BFS has delivered a forecasting software package tailored to the specific needs of practitioners who have to forecast a large number of time series repeatedly but who do not have the time or knowledge to deal with the statistical details of forecasting methods. The program's *Expert Selection Mode* ranks among the best currently available and is a reliable source of forecasts for the majority of real business data as it is based on accepted statistical methodologies.

For the most part, the user interface allows for an intuitive workflow. Its main strengths are the excellent representation and flexible handling of product hierarchies and its powerful export features. Users of XE who are thinking about migrating to *Forecast Pro Unlimited* need to assess the scope of necessary adoptions because *Forecast Pro Unlimited* differs in such important aspects as script structure, model specification, and model spectrum. If possible, it would make sense to rely on both XE and *Forecast Pro Unlimited*, as the combination of their key features empowers users to handle batch forecasting jobs as well as building and optimizing manually elaborated statistical models for their respective forecasting problems.

Acknowledgements: We are obliged to Jim Hoover and Stephan Scholze for suggestions on a preliminary version of this paper. Errors and omissions are our sole responsibility.

REFERENCES

Croston, J.D. (1972). Forecasting and stock control for intermittent demands, *Operational Research Quarterly*, 24, 639-641.

Gardner Jr., E.S. (1985). Exponential smoothing: The state of the art, *Journal of Forecasting*, 4, 1-28.

Hyndman, R.J., Koehler, A.B., Ord, J.K. & Snyder, R.D. (2005). Prediction intervals for exponential smoothing using two new classes of state space models, *Journal of Forecasting*, 24, 17-37.

Makridakis, S & Hibon, M. (2000). The M3-Competition: Results, conclusions and implications, *International Journal of Forecasting*, 16, 451-476.

Tashman, L.J. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review, *International Journal of Forecasting*, 16, 437-450.

Tsay, R.S. (1986). Time series model specification in the presence of outliers, *Journal of the American Statistical Association*, 81, 132-141.

Yar, M. & Chatfield, C. (1990). Prediction intervals for the Holt-Winters forecasting procedure, *International Journal of Forecasting*, 6, 127-137.

CONTACT
Ulrich Küsters and Janko Thyson
Catholic University of Eichstätt-Ingolstadt
ulrich.kuesters@ku-eichstaett.de
janko.thyson@ku-eichstaett.de

Mohsen Hamoudia, France Telecom Group

What is your current job and position ?

I am Head of Strategic Marketing of the Corporate Large Projects Division within Orange Business Services, France Telecom Group. My division encompasses all communication products including mobility solutions, internet advanced solutions, and voice and data. Working with area managers, I prepare 3-year forecasts for revenue, margin, resources, costs, traffic, and bandwidth, by worldwide region, product and business unit.

How did your forecasting career start?

My career in forecasting was not due to chance or opportunity. It was a choice I made during my studies in the Masters of Statistics and Economics programs at the University of Paris. My first position was as a forecast manager at Air Inter, the French domestic airline, which was ultimately merged into the Air France Group. There I was responsible for forecasting passenger traffic on the French domestic network. The introduction of the TGV (high-speed rail) produced a huge change in the airline business model but initially left us without tools to predict the impacts. With the help of two forecasting colleagues, however, I developed models that helped us see our competitive future.

> MY CAREER IN FORECASTING WAS NOT DUE TO CHANCE OR OPPORTUNITY. IT WAS A CHOICE I MADE DURING MY STUDIES.

What do you find rewarding about the forecaster's job?

It is that the forecaster plays such an important organizational role, providing visibility into the organization's future, anticipating major changes that can affect the organization's development, and improving the decision-making process with solid analytical support.

What have been your career highlights?

In addition to my position with Air Inter, my highlights include four years (1997-2001) as Director of the Forecasts and Economic Studies Division within France Telecom Long Distance. Managing a 12-person team, I experienced firsthand the opening of the European telecom market to full competition and an explosion of new products and services.

Another wonderful experience for me has been my teaching, especially my tenure (since 2000) as an Associate Professor at ESDES, the Business School of the University of Lyon. There I have taught courses in forecasting and quantitative methods and have supervised several master theses.

Now that it's over, how do you feel about all the work you did in organizing the 2008 ISF in Nice? Were you gratified with the results?

The organization of ISF2008 was a good and wonderful challenge. We tried to attract top speakers and original papers and to encourage more student participation. We also wanted to show that the IIF is a very professional organization. The feedback from attendees tells me that these objectives were generally reached, despite the non-cooperation of the espresso machine and air conditioning in some meeting rooms!

What is your next big project?

I've started to write a book on ICT Forecasting which will address forecasting methods in telecommunications and also examine new developments such as the information society and social networks. Additionally, I am involved in organizing the regional ITS (International Telecommunication Society) in Perth, Australia, in 2009.

Any other mountains to climb?

Actually, one of my main hobbies is alpine mountaineering. In the last 25 years, I've climbed several summits in the Pyrenees (Monte Perdido, Monte Aneto) and the Alps (Cervino, Mont Rose, Mont Blanc, the Grandes Jaurasses). Next up hopefully are the Andes. Hey, IIF members, why not hold an ISF in Argentina or in Chile?

Contact: Mohsen.hamoudia@orange-ftgroup.com

Forecasting Summit

Forecasting Summit

THE Conference Where Forecasters Converge to Share Knowledge and Ideas

enjamin Moore & Co. Federal Reserve Bank Capital One The Wharton School Coca-Cola SC Johnson GlaxoSmithKline F

Register now to:

- Learn best practices from leading practitioners and renowned experts
- Acquire new skills that will help advance your career
- Gain insights for dealing with real-world forecasting issues
- Exchange knowledge and ideas about forecasting

Upcoming Conference:

February 23-25, 2009
Hilton in the Walt Disney World® Resort
Orlando, Florida USA

Contact us for a free brochure with full schedule

forecasting-summit.com

Phone: 617-484-5050
E-mail: info@forecasting-summit.com

IIF *Forecasting Summit is presented in cooperation with the International Institute of Forecasters*

FORESIGHT
The International Journal of Applied Forecasting
A PUBLICATION OF THE INTERNATIONAL INSTITUTE OF FORECASTERS

BUSINESS OFFICE: 53 TESLA AVENUE
MEDFORD, MASSACHUSETTS 02155 USA
www.forecasters.org/foresight