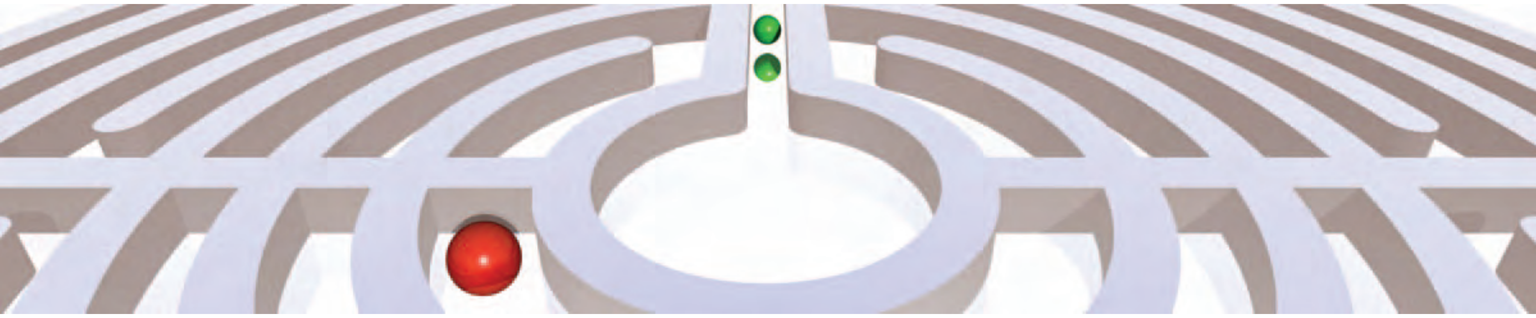
Issue 3 ■ February 2006

FORESIGHT

The International Journal of Applied Forecasting



SPECIAL FEATURES

The Keys to the White House: Forecast for 2008

Improving the Forecasting Process: Two Case Studies

Software: Spotlight on Excel for Data Analysis and Forecasting

A PUBLICATION OF THE INTERNATIONAL INSTITUTE OF FORECASTERS

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

SPECIAL FEATURE

THE KEYS TO THE WHITE HOUSE: FORECAST FOR 2008

IMPROVING THE FORECASTING PROCESS: TWO CASE STUDIES

FORECASTING PRINCIPLES AND METHODS

SOFTWARE: SPOTLIGHT ON EXCEL

BOOK REVIEW

FORESIGHT
The International Journal of Applied Forecasting

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 implementation of forecasting processes
- Forecasting principles and methods
- Integration of forecasting into business planning
- Forecasting books, software and other technology
- Forecasting-application issues in related fields
- Case studies
- Briefings on new research

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: The International Journal of Applied Forecasting
140 Birchwood Drive, Colchester, Vermont 05446, USA

FORESIGHT
The International Journal of Applied Forecasting

(ISSN 1555-9068)

FORESIGHT is published 3 times annually in February, June and October.

A one-year subscription to FORESIGHT includes all THREE issues of the journal plus a FREE digital version of each issue available as a downloadable pdf. Two or more subscriptions are available at a discount rate. The more subscriptions you order, the more you save per issue.

Subscribe to FORESIGHT

(credit card payments accepted)

Online www.forecasters.org

Toll Free Worldwide 866.395.5220

Email Bill Wicker, Managing Editor, billwicker@adelphia.net

Mail check or credit card info
FORESIGHT
140 Birchwood Drive
Colchester, VT 05446 USA

FORESIGHT for IIF Members

Your annual IIF membership fee of $120 ($40 for a student) includes subscriptions to:

FORESIGHT: The International Journal of Applied Forecasting
The International Journal of Forecasting (quarterly)
The Oracle of the IIF (quarterly newsletter)

Members of the IIF also receive a discounted registration fee for the annual International Symposium on Forecasting (ISF).

Bulk Orders and Reprints

Please contact Bill Wicker, billwicker@adelphia.net

Subscription Rate

$95 per year: USA
$99 per year: Canada and Mexico
$105 per year: Other countries

(domestic/in US dollars, includes mailing)

Subscribe/Renew for 2 years and save $15!

$175/USA $183 Canada & Mexico $195/other countries

Multi-Subscription Rates

# of Subscriptions	Price per Subscription
1	$95 each
2-5	$75 each
6-10	$70 each
16-20	$60 each
21+	pricing on request

(domestic/in US dollars, includes mailing) Canada & Mexico, add $4 per year. Other countries, add $10 per year.

CONTRIBUTE TO FORESIGHT
To contribute an article or offer to serve as a reviewer or editor, contact Editor Len Tashman, lentashman@cs.com

FORESIGHT is published by the
International Institute of Forecasters
Business Manager: Pamela N. Stroud
53 Tesla Avenue, Medford, MA 02155
pstroud@comcast.net

FORESIGHT: THE INTERNATIONAL JOURNAL OF APPLIED FORECASTING

YES. Send me a one-year (3 issues) subscription.

☐ $95 / USA ☐ $99 / Canada and Mexico ☐ $105 / other countries

I want to SAVE $15! Subscribe/Renew for TWO YEARS.

☐ $175 / USA ☐ $183 / Canada and Mexico ☐ $195 / other countries

☐ Check (US funds only) enclosed ☐ Credit Card # ______________________

Exp. Date __________ Name on Card ______________ Signature ______________
please print

NAME COMPANY

ADDRESS

CITY/STATE/ZIP (POST CODE)/COUNTRY

EMAIL

PHONE ( ) TOTAL AMOUNT $

MAIL TO FORESIGHT: The International Journal of Applied Forecasting, 140 Birchwood Drive, Colchester, VT 05446 USA
PHONE TOLLFREE WORLDWIDE 866.395.5220
SUBSCRIBE OR RENEW ONLINE www.forecasters.org

subscribe

I would love to suggest a subscription to our library.
Professor Michael E. Smith, Western Carolina University

Here's a new idea!

Add FORESIGHT to your company or university library!

You'll get articles on forecasting applications, useful tutorials, and informative case studies, written for practitioners and students in a concise accessible style.

Ask your librarian to order FORESIGHT today.

SPECIAL LIBRARY RATE / $195 annually

FORESIGHT is published three times annually in February, June and October. A one-year subscription includes all three issues of the journal and a digital version of each issue available as a downloadable pdf.

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

Library Recommendation Form
Please bring to your Librarian along with a copy of Foresight.

TO: Librarian / Library / Acquisition Committee

FROM: ______________________

POSITION: ______________________

DEPARTMENT: ______________________

Please place an order for this publication. I consider it a needed reference source in business forecasting.

Signature ______________________

Date ______________________

SPECIAL LIBRARY RATE $195

To order FORESIGHT
(credit card payments accepted)

Online at: www.forecasters.org

Toll Free Worldwide: 866.395.5220

Email: Bill Wicker, Managing Editor, billwicker@adelphia.net

Mail checks to: FORESIGHT
140 Birchwood Drive
Colchester, VT 05446 USA

SPECIAL FEATURE

THE KEYS TO THE WHITE HOUSE: FORECAST FOR 2008

THE KEYS TO THE WHITE HOUSE: FORECAST FOR 2008

by Allan J. Lichtman

Preview: The Keys to the White House is a historically based prediction system that retrospectively accounts for the popular-vote winners of every American presidential election from 1860 to 1980. The system has forecasted the popular-vote winners of every presidential election from 1984 through 2004. It proves that presidential election results turn primarily on the performance of the party controlling the White House and that politics as usual by the challenging candidate has no impact on results. The system includes no polling data, and it considers performance indicators that transcend economic concerns. Already the Keys are lining up for 2008, demonstrating surprisingly bright prospects for Democrats to recapture the White House.

Allan Lichtman is a professor at American University in Washington, D.C. *The Thirteen Keys to the Presidency* is one of six books he has written. Allan also has published many scholarly and popular articles, has been an expert witness in dozens of civil rights and redistricting cases, and was the national political analyst for CNN Headline News. Allan is now testing his forecasting skills by running for the open U.S. Senate seat in Maryland.

- Since 1860, American presidential election results have followed a common pattern: the American electorate chooses a president according to the performance of the party holding the White House.
- Debates, advertising, television appearances, news coverage, and campaign strategies – the usual grist for the punditry mill – count for virtually nothing on Election Day.
- The incumbent party's performance can be assessed by answers to 13 simple questions I call the Keys. When six or more Keys turn against the incumbent party, I predict that party's candidate will lose the popular vote in the next presidential election.
- The Keys have correctly forecast the popular-vote outcomes of the last six presidential elections. Already the Keys are lining up for 2008, pointing to the likelihood of a change in party control of the White House.

By showing that governing, not campaigning, counts in presidential elections, the Keys suggest an alternative to today's shallow, sound-bite politics that do not benefit the parties, the candidates, the people, or the country.

Table 1. The 13 Keys to the White House

The Keys are statements that favor the reelection of the incumbent party. When five or fewer statements are false, the incumbent party wins. When six or more are false, the challenging party wins.

KEY 1	[Party Mandate] After the midterm elections, the incumbent party holds more seats in the U.S. House of Representatives than it did after the previous midterm elections.
KEY 2	[Contest] There is no serious contest for the incumbent party's nomination.
KEY 3	[Incumbency] The incumbent party's candidate is the sitting president.
KEY 4	[Third Party] There is no significant third-party campaign.
KEY 5	[Short-term Economy] The economy is not in recession during the election campaign.
KEY 6	[Long-term Economy] Real per capita economic growth during the term equals or exceeds mean growth during the previous two terms.
KEY 7	[Policy Change] The incumbent administration effects major changes in national policy.
KEY 8	[Social Unrest] There is no sustained social unrest during the term.
KEY 9	[Scandal] The incumbent administration is untainted by major scandal.
KEY 10	[Foreign/Military Failure] The incumbent administration suffers no major failure in foreign or military affairs.
KEY 11	[Foreign/Military Success] The incumbent administration achieves a major success in foreign or military affairs.
KEY 12	[Incumbent Charisma] The incumbent party's candidate is charismatic or is a national hero.
KEY 13	[Challenger Charisma] The challenging party's candidate is not charismatic or is not a national hero.

A New Vision of Presidential Politics

This paper focuses on how presidential elections really work in the United States. A properly functioning democracy demands not only fair and accurate systems for voting but also a candid, wide-ranging exploration of crucial issues and ideas by the presidential candidates and their parties. Every four years, however, Americans are subjected to shallow and even offensive presidential campaigns. The media, the candidates, the pollsters, and the consultants all believe that elections are exercises in voter manipulation, negative campaigning, bland, scripted lines, and meaningless debates.

In contrast, my study shows that the American electorate chooses a president according to the performance of the party holding the White House. We can measure performance by the consequential events of the previous term: economic boom and bust, foreign policy successes and failures, social unrest, scandal, and policy innovation. If the nation fares well during the term of the incumbent party, that party wins another four years in office; otherwise, the challenging party prevails.

Given the public's cynicism toward politics, I believe that nothing a candidate does conventionally during a campaign will change his or her prospects at the polls. Debates, advertising, television appearances, news coverage, and campaign strategies—the usual grist for the punditry mill—count for virtually nothing on Election Day. The issues that matter are the ones already resolved before the campaign begins. Thus the fate of an incumbent party is largely in its own hands; there is little that the challenging party can do through politics as usual to influence the outcome of a presidential election.

The Keys Model

I base my vision of American politics on the Keys to the White House, a historically based prediction system created by studying every presidential election from 1860 to 2004. I first developed the Keys system in 1981, in collaboration with Vladimir Keilis-Borok, a world-renowned authority on the mathematics of prediction models. The system shows that we can predict the outcomes of presidential elections based on indicators that track the performance and strength of the party holding the White House.

We used pattern recognition methodology on data for American presidential elections since 1860 (the first election with a four-year record of competition between Republicans and Democrats). We developed 13 questions stated as propositions favoring reelection of the incumbent party. When five or fewer of these propositions are false, or are turned against the party holding the White House, that party wins another term in office (Table 1). When six or more Keys are false, the challenging party wins. Unlike many alternative models, the Keys include no polling data, but they are based on the big picture of how well the party in power has fared prior to an upcoming election. For the methodology used in the study, see Keilis-Borok and Lichtman (1981).

The Keys do not presume that voters are driven by economic concerns alone. Voters are open-minded and sophisticated; they decide presidential elections on a wide-ranging assessment of the incumbent party's performance. The Keys reflect multiple components of that performance.

Retrospectively, the Keys have correctly predicted the popular-vote winner of every presidential election from 1860 through 1980. Prospectively, the Keys have predicted well ahead of time the popular-vote winners of every presidential election from 1984 through 2004. For example, they called Vice President George H. W. Bush's victory in the spring of 1988, when he trailed Mike Dukakis by nearly twenty points in the polls and was being written off by the pundits. The vice president defied the polls and the pundits not because he discovered negative ads or refurbished his image, but because voters ratified the performance of the Reagan administration—four years of prosperity, the defusing of the Cold War, and a scandal that faded away. In 1992, George H. W. Bush lost his chance for a second term, as the Keys had predicted, when a sour economy and a lack of domestic accomplishment tarnished his record as president. In April of 2003, the Keys predicted President George W. Bush's 2004 reelection a year and a half before a contest that pollsters found too close to call right up to election eve. Because Bush was a sitting president with no prospective challenger in his own party, and with no serious third-party competitor, his mixed record of accomplishment at home and abroad was sufficient to anticipate his victory in 2004.

As a nationally based system, the Keys cannot diagnose the results in individual states. Thus the Keys are attuned to the popular vote, not the Electoral College results. In three elections since 1860, when the popular vote diverged from the Electoral College tally—1876, 1888, and 2000—the Keys accurately predicted the popular-vote winner, but not the Electoral College results.

The 2004 Forecast

As early as April 2003, the Keys showed that the incumbent Republicans were positioned to regain the White House in 2004. The party in power had four Keys turned against them for 2004, two short of the fatal six negative Keys. The four Keys are discussed here:

- The weak economy during the Bush term, as compared to the boom years of Clinton's two terms, cost the Republicans the Long-Term Economy Key.
- The relatively modest domestic accomplishments of the Bush administration toppled the Policy Change Key.
- The first successful foreign attack on the continental United States since the War of 1812 cost the Republicans the Foreign/Military Failure Key.
- George W. Bush did not measure up to the charisma of Theodore Roosevelt or Ronald Reagan, forfeiting the incumbent Charisma/Hero Key.

All other Keys favored the incumbent party, with the exception of the Short-Term Economy Key, which could have turned against the GOP if the economy had hit a double-dip recession during the election year. The absence of recession meant that the final lineup remained unchanged, with the Republicans still two Keys short of defeat.

This finding had implications for the presidential election. In July 2004, Keilis-Borok and I wrote (http://www.commondreams.org/views04/0728-01.htm) that

Democratic nominee John Kerry had two strategic choices, given the verdict of the Keys. He could follow the usual meaningless routine in the hope that setbacks to the administration would help elect him in November, or he could take a chance on running a daring, innovative, and programmatic campaign. We said that Kerry could achieve a historical breakthrough that would establish not only the basis for a principled choice of our national leader, but also a grassroots mobilization on issues that matter to America's future. We suggested that he lead a debate on critical neglected issues, that he set up a shadow government with suggested choices for key cabinet positions, and that he publish an alternative budget and alternative drafts of international agreements.

Kerry made the wrong choice. He stuck with conventional advisers and strategies, and he suffered the same fate as Michael Dukakis in 1988, becoming a derided losing candidate. With a different choice, with a bold, imaginative, substantive campaign, he would have established himself as a principled opponent to the Bush administration and positioned himself for another presidential run in 2008.

Forecasting 2008

The Keys to the White House begin lining up for and against an incumbent party early in the term, although a final prediction may not be possible until much later. In April 1982, for example, the Keys predicted Ronald Reagan's reelection more than two and a half years before Election Day. And in my 1990 book, *The Thirteen Keys to the Presidency,* completed just a year into George H. W. Bush's term, I noted that "early in the term, Bush looks more like a Carter than a Reagan" (Lichtman, p.419).

According to an early warning from the Keys, the incumbent Republicans are precariously positioned for 2008; the most likely outcome of that election is a Democratic victory in the presidential campaign. As indicated in Table 2, as of the winter of 2005, only three Keys are likely to fall in favor of the incumbent party. Five Keys are uncertain, and five Keys are likely to fall against the incumbent party. Thus the GOP forfeits the White House in 2008 if the likely positive and negative Keys line up as anticipated, and just one of five uncertain Keys falls against it.

Table 2. The 13 Keys to the White House: Standings, December 2005

KEY NUMBER	DESCRIPTION	OUTCOME 2008
KEY 1	Party Mandate	Likely False
KEY 2	Contest	Likely False
KEY 3	Incumbency	Likely False
KEY 4	Third Party	Likely True
KEY 5	Short-term Economy	Uncertain
KEY 6	Long-term Economy	Uncertain
KEY 7	Policy Change	Likely False
KEY 8	Social Unrest	Likely True
KEY 9	Scandal	Uncertain
KEY 10	Foreign/Military Failure	Uncertain
KEY 11	Foreign/Military Success	Uncertain
KEY 12	Incumbent Charisma	Likely False
KEY 13	Challenger Charisma	Likely True

Likely True:	3 KEYS
Likely False:	5 KEYS
Uncertain:	5 KEYS

The following three Keys currently favor the incumbent Republican Party:

- The lack of any prospective third-party challenger with prospects of winning 5 percent of the vote tilts the Third Party Key toward the GOP.
- The absence of 1960s-style social upheaval likely avoids the loss of the Social Unrest Key.
- No prospective Democratic challenger matches the charisma of Franklin D. Roosevelt or John F. Kennedy, probably keeping the Challenger Charisma/Hero Key in line for the incumbents.

The following five Keys are likely to fall against the incumbent party:

- The Democrats need to win just three U.S. House seats in the 2006 midterm elections to topple the Mandate Key.
- The Republicans are likely to battle fiercely in choosing a nominee to replace George W. Bush, forfeiting the Contest Key.
- Bush's inability to run again in 2008 dooms the Incumbency Key.
- With bitter partisan divisions in Congress, Bush is unlikely to achieve the policy revolution needed to secure the Policy Change Key.
- Of all GOP candidates on the horizon, only John McCain, a possible but unlikely nominee, might be able to secure the Incumbent Charisma/Hero Key.

This leaves five Keys that are uncertain.

- Both the Short-Term Economy Key and the Long-Term Economy Key depend on unpredictable future trends in economic growth.
- The Scandal Key might turn against the administration, pending results of investigations into the response to Hurricane Katrina and the release of the identity of CIA agent Valerie Plame.
- Both the Foreign/Military Failure Key and the Foreign/Military Success Key will turn on unforeseeable events abroad and on homeland security within the United States.

The difficult prospects for Republicans in 2008 explain much of today's politics. The 2006 midterm elections are critical because the Mandate Key turns on the outcome. The president and the Republicans in Congress pushed for the "nuclear option" to end judicial filibusters by majority vote because this is likely their last chance to fill the courts with reliable conservatives. Until the collapse of his approval rating after Hurricane Katrina, the president bucked public opinion on the rewriting of Social Security to win the pivotal Policy Change Key for 2008.

Conclusions

If candidates and the media could understand that governing, not campaigning, counts in presidential elections, we could have a new kind of presidential politics. Candidates could abandon attack ads and instead articulate forthrightly and concretely what Americans should be accomplishing during the next four years. Aspirants for the presidency could use campaigns to build grassroots support for their respective agendas. And incumbent presidents could prepare for upcoming elections by focusing on the stewardship of the country, not on the politics of campaigns. We will not reform our politics and get meaningful participation by the American people until we realize that presidential elections turn on how well an administration has governed the country, not on how well candidates have performed in the campaign.

References

Keilis-Borok, V. I. & Lichtman, A. J. (2004). What Kerry must do. *Common Dreams*, July 28, 2004, http://www.commondreams.org/views04/0728-01.htm

Keilis-Borok, V. I. & Lichtman, A. J. (1981). Pattern recognition applied to presidential elections in the United States, 1860-1980: The role of integral social, economic, and political traits. *Proceedings of the National Academy of Sciences*, 78, 7230-7234.

Lichtman, A. J. (1990). *The Thirteen Keys to the Presidency*. Lanham, MD: Madison Books.

Contact Info:
Allan J. Lichtman
American University
lichtman@american.edu

INDEX METHODS FOR FORECASTING: AN APPLICATION TO THE AMERICAN PRESIDENTIAL ELECTIONS

by J. Scott Armstrong and Alfred G. Cuzan

Preview: Scott Armstrong and Alfred Cuzan describe Allan Lichtman's Keys Model as an example of an index method of forecasting, which assigns ratings of *favorable*, *unfavorable*, or *indeterminate* to influencing variables. They describe how index methods have been applied in other decision-making contexts, and they discuss when such methods might be useful analytical tools for business forecasters. In the context of presidential election forecasting, they compare the Keys model to several regression models and find that the Keys model stacks up quite well against these more sophisticated alternatives.

J. Scott Armstrong is Professor of Marketing at the Wharton School, University of Pennsylvania. He is a founder of the Journal of Forecasting, the International Journal of Forecasting, and the International Symposium on Forecasting. He is the creator of the Forecasting Principles website, (forecastingprinciples.com) and editor of *Principles of Forecasting: A Handbook for Researchers and Practitioners*. He is one of the first six Honorary Fellows of the International Institute of Forecasters, and was named Society for Marketing Advances/JAI Press Distinguished Marketing Scholar for 2000.

Alfred G. Cuzán joined the faculty at the University of West Florida in 1980. In 1992, he was appointed Chairman of the Department of Government. A Woodrow Wilson Fellow, a Fulbright Scholar, and a Henry Salvatori Fellow, Alfred is the author or co-author of more than forty scholarly items. He has lectured on the impact of fiscal policy on American presidential elections in Argentina, Mexico, and Spain.

- Indexes like Allan Lichtman's Keys model are worthy of the attention of practitioners and researchers. Lichtman's easily understood method can help forecasters when there are (1) many causal variables, (2) good domain knowledge about which variables are important, and (3) limited amounts of data.
- The Keys model has been able to pick the winner of every presidential election since 1860, but we tested how it compared against three traditional regression models in forecasting the percentage of the vote obtained by the incumbent party's candidate. We found that Lichtman's perfect record in forecasting the out-of-sample winner was matched by only one of the three regression models, while its average error was almost as low as those of the best regression models.
- We believe that the Keys model is useful for presidential election forecasting because it uses a different method and different information than do current regression models.

Introduction

Allan Lichtman (2005) reports that the Keys model has picked the winner of every presidential election since 1860, retrospectively through 1980 and prospectively from 1984 to 2004. Given this record, it seems sensible to examine this index *method*. We tested how well the Keys model predicted the winner of the popular vote, and also how closely it forecasted the actual percentage of the two-party vote won by the incumbent ticket. The index method performs well compared with regression models. It also offers the opportunity to incorporate many policy variables. Index methods can be applied to various choice problems faced by organizations.

Index vs. Regression Models

In the early days of forecasting, analysts would sometimes use an index to forecast. They would prepare a list of key variables and determine whether they were favorable (+1), unfavorable (-1), or indeterminate (0) regarding a particular outcome. They would then add the scores and use the total in making forecasts. Thus each variable was assigned the same weight. Applied to forecasting, this use of judgmental indexes has been called an "experience table" or an "index method."

Index methods have been used for various types of forecasting problems, including prediction of the success of prisoners seeking parole. If the candidate exceeded a certain score, he or she was paroled. In an effort to improve parole predictions, Glueck and Glueck (1959) recommended using only the most important variables and assigning differential weights to different variables. This can be done by regression analysis. While regression analysis has been widely adopted, however, little research has been devoted to index methods.

Which approach yields the most accurate forecasts, index methods or regression models? Gough (1962) addressed this issue for parole predictions and found that regression modeling did not improve accuracy. Reviewing the research in this area, Armstrong (1985, p. 230) found three studies in which regression was slightly more accurate (for academic performance, personnel selection, and medicine); however, five studies found that regression was less accurate (three on academic performance, and one each on personnel selection and psychology).

A related approach is to use equal weights in a regression, which brings regression modeling closer to the index method. The equal weights are applied to standardized variables to avoid scaling problems. A large-scale study by Dana and Dawes (2005) found that equal weights forecasts were generally more accurate than regression on a wide variety of cross-sectional data. The gain from equal weights was larger when sample was small and when predictability was poor.

For most problems, regression analysis is limited in that only a few explanatory variables can be put into the model (perhaps three or four variables) because of limited data, measurement errors, and correlations among the explanatory variables (a problem called collinearity). Subjective indexes avoid these estimation problems. Given the many variables and the small amount of data, index methods would seem appropriate for forecasting presidential elections.

Lichtman's Index: The 13 Keys

Allan Lichtman is a historian, and, to the best of our knowledge, the only scholar who has applied an index method to predict the winner of presidential elections. His "Keys to the White House" model consists of 13 explanatory variables. Each variable consists of a statement which, if true, bodes well for the incumbents and, if false, for the opposition, such as KEY 2 (Contest Key): *There is no serious contest for the incumbent party's nomination.*

True (or likely to be true) statements are scored 0, and false statements scored 1. We then count the number of false statements. If fewer than six are false, the incumbents are forecast to win. Conversely, if six or more Keys turn against the incumbents, they are likely to lose.

Some aspects of the Keys model concern us:

1. It uses only 13 variables. One of the benefits of the index method is that there is no limit on the number of variables.
2. Only one of the variables makes a reference to policy (KEY 7, Policy Change: *The incumbent administration effects major changes in national policy*), but it is vague as to the type of policy or the direction of change. One could imagine popular as well as unpopular changes. In ignoring policy variables, however, the Keys model is no worse than most other presidential forecasting models.
3. The assessment as to whether each Key is true or false is done subjectively by one person (Lichtman). For example, what constitutes a "major" change in national policy? Presumably this procedure could be improved by using a panel of experts.
4. The model challenges credibility because to win, the incumbent requires a larger number of favorable factors than does the challenger. To win, the incumbent needs 7 of the 13 Keys in his or her favor. In general, incumbents are thought to have the advantage in political elections.

Using the Lichtman Index to Forecast the Vote Percentage

Most forecasters of presidential elections, economists and political scientists alike, have estimated the percentage of the two-party vote going to the incumbents by using differential weights in regression models. Accordingly, we tested how well Lichtman's method predicts the actual percentage of the two-party vote that will go to incumbents.

We use V to represent the percentage of the two-party vote that will go to the incumbent, and L to represent the Lichtman index, which we define as the total number of Keys favorable to the incumbent. (This is the reverse of Lichtman's coding, as he counts the number of keys that have turned *against* the incumbents.) We fit a regression model relating V to L over the period 1860-2004, alternatively including and omitting the 1912 election, when the Republican Party split in two. (Some researchers, like Fair (2004), whose data series we use, add the William Howard Taft and Theodore Roosevelt vote together for a counterfactual incumbent victory of 54 percent.) We found very little difference in model errors from the inclusion or exclusion of the 1912 election, so we will report results for the inclusive model only.

We obtained the following regression results:

$V = 37.3 * 1.8 L$ where
V = the percentage of the two-party split going to the incumbent
L = the number of Keys favoring the incumbent

Thus the model predicts that an incumbent would start with 37 percent of the vote (even if all Keys are unfavorable) and would add 1.8 percent to this base with each favorable Key. To measure model accuracy, we use two metrics. One is the absolute percentage error—the magnitude of the average errors, whether they are positive or negative. The second metric is the call ratio, which is the percentage of forecasts that correctly pick the election's winner. Retrospectively—that is, when we include all elections in fitting the model, and we look at how closely the model reproduces the historical results—the Keys model came within 3.1 points of the actual percentage going to the incumbent. When we exclude one election at a time and see how the model would have predicted the excluded election (a procedure called a jackknife), the error averaged 3.2 percentage points. This was larger than any of the eight presidential regression models we analyzed for this period; the average error for the other eight models was 2.2 percentage points.

When we retrospectively calculated the percentage of correct predictions by the Keys model (the call ratio), it was 100 percent. (We credit Lichtman's model with a correct call in 1912 because it predicted defeat for the incumbents, even though in Fair's data series they "won" with 54 percent of the vote.) Given the relatively high percentage error compared to other models, the finding that all elections were correctly forecast is surprising. However, one must remember that this is retrospective analysis, a fit of a model to the data. Prior research in other areas has shown a poor relationship between fit and predictive ability (Armstrong 2001, pp. 460-462).

The critical test is how well the models forecast prospectively (that is, for years not included in the estimation sample). In Table 1 we compare the Keys model against three others: Abramowitz (2004), Campbell (2004), and Fair (2004). These are traditional regression models, variations of which have been used in forecasting presidential elections for the better part of two decades. We estimated each of these four models through the 1980 election, which was the final observation included in the original Keys model. Then we used those models to forecast all subsequent elections through that of 2004.

In this prospective test, the Keys model performed well (Table 1). Not only were all election winners picked correctly, but its error was 2.3 percentage points, only slightly higher than the 2.1 percentage point errors for Abramowitz and Campbell, and about half as large as Fair's forecasts. Of the regression models, only Campbell's correctly predicted the winner of all six elections.

Extensions of the Index Method

Index methods do not have to be restricted to equal weights. In a 1772 letter to Priestly (http://homepage3.nifty.com/hiway/dm/franklin.htm) on how to make choices, Ben Franklin described another way, which he called "prudential algebra":

> I endeavor to estimate their respective weights; and where I find two, one on each side, that seem equal, I strike them both out. If I find a reason pro equal to some two reasons con, I strike out the three. If I judge some two reasons con, equal to three reasons pro, I strike out the five; and thus proceeding I find at length where the balance lies....

Given Ben Franklin's excellent record at problem solving, perhaps we should revisit his method, for it provides a useful way to capitalize on the value of expertise.

Table 1. Presidential Elections, 1984-2004 (6 elections)

	Forecast Accuracy	
Model (estimation period)	**Absolute Percentage-Point Error**	**Call Ratio**
Abramowitz (1948-1980)	1.9	67
Campbell (1948-1980)	2.1	100
Fair (1916-1980)	4.5	67
Lichtman (1860-1980)	2.3	100

Index methods can be tailored to the situation at hand. Certainly the needs and interests of the electorate have changed since 1860. The index methods could include all key issues for a given election. For recent elections, the issues could include gay rights, abortion, terrorism, union support, health care, minimum wage, estate taxes, tax rates, and free trade. The position of a candidate could be scored on whether it agrees with the agenda of a certain bloc of voters, such as swing voters.

This list could also include findings related to such personal characteristics as the height of the candidates and whether they look competent. When many elections have been prospectively or retrospectively predicted in this way, the resulting scores could be translated into a percentage vote for the incumbent.

Summary

Although the Keys model has correctly called the winner in 37 consecutive presidential elections, 31 of these were used to fit the model. It is the prospective forecasts of the last six elections that are of prime interest. In these, Lichtman's perfect record was matched by one of the three regression models against which it was compared, and its average error was almost as low as those of the best models. We conclude that the Keys model provides a useful alternative, but there is little reason to prefer it to the exclusion of other models. We expect the Keys model to serve as one of the important components for long-term (at least up to a year) forecasts of presidential elections. It should be especially useful because it uses a different method and different information than do current regression models.

Indexes like the Keys model are worthy of the attention of practitioners and researchers of causal methods. This easily understood method is expected to aid forecasting in situations where there are (1) many causal variables, (2) good domain knowledge about which variables are important and about the direction of effects, and (3) limited amounts of data. These conditions apply where discrete choices must be made, such as for the selection of personnel, retail sites, investment opportunities, product names, or advertising campaigns.

References

Abramowitz, A.I. (2004). When good forecasts go bad: the time-for-change model and the 2004 presidential election. *PS: Political Science and Politics*, 37 (4), 745-746.

Armstrong, J.S. (2001). Evaluating forecasting methods. In J.S. Armstrong, (Ed.), *Principles of Forecasting*. Boston: Kluwer.

Armstrong, J.S. (1985). *Long-range Forecasting: From Crystal Ball to Computer*. New York: John Wiley.

Campbell, J.E. (2004). Forecasting the presidential vote in 2004: placing preference polls in context. *PS: Political Science and Politics*, 37 (4), 763-768.

Dana, J. & Dawes, R. M. (2005). The superiority of simple alternatives to regression for social science predictions. *Journal of Educational and Behavioral Statistics*, 29 (3), 317-331.

Fair, R. (2004). A vote equation and the 2004 election. Available at http://fairmodel.econ.yale.edu/vote2004/vot1104a.pdf

Glueck, S. & Glueck, E. (1959). *Predicting Delinquency and Crime*. Cambridge, MA: Harvard University Press.

Gough, H.G. (1962). Clinical versus statistical prediction in psychology. In L. Postman (Ed.), *Psychology in the Making*. New York: Knopf, 526-584.

Lichtman, A.J. (2005). The keys to the White House: Forecast for 2008. *Foresight: The International Journal of Applied Forecasting*, Issue 3, 5-9.

Contact Info:
J. Scott Armstrong
The Wharton School
University of Pennsylvania
armstrong@wharton.upenn.edu

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

Upcoming

2006 FORESIGHT

"While the future's there for anyone to change
still you know it seems
it would be easier sometimes
to change the past."
from Fountain of Sorrow/Jackson Browne

SPECIAL FEATURE

Forecasting For Call Centers
Feature Editor Jay Minucci

ADVANCING THE ART OF FORECASTING

Forecasting Semi-New Products
by Bill Tonetti

IMPROVING THE FORECASTING PROCESS

The Value of a Structured Forecasting Process
by Simon Clarke, Coca Cola Enterprises

Forecast Process Improvement – Who Gets In the Way and How To Overcome These Barriers
by Mark Moon

ACCURACY IN FORECASTING

Accuracy Metrics For Intermittent Demands

Thomas Edison's Technological and Social Forecasts – A Good Track Record
by Steven Schnaars

BOOK REVIEWS

The Wisdom of Crowds by *New Yorker* columnist, James Surowiecki

Fooled by Randomness by Nassim Nicholas Taleb

SPOTLIGHT ON SOFTWARE

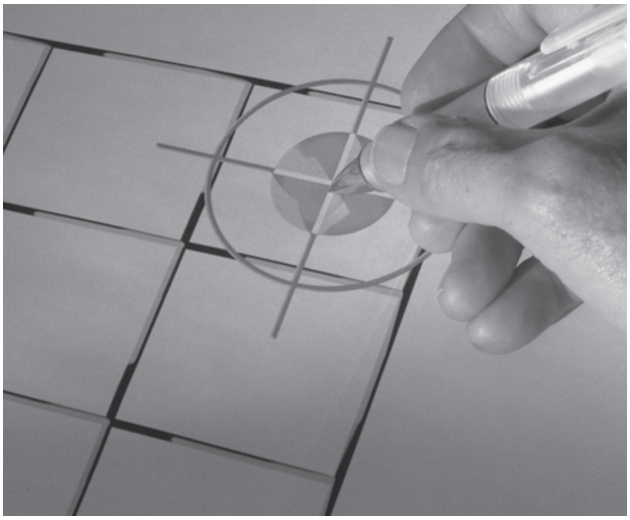
Forecasting In SAP With R3 and APO

FAQ Column by Kesten Green

If you are interested in contributing an article or commentary on any forthcoming article, please contact Len Tashman, FORESIGHT Editor, lentashman@cs.com

2006FORESIGHT2006FORESIGHT2006FORESIGHT2006FORESIGHT2006

IMPROVING THE FORECASTING PROCESS: TWO CASE STUDIES



Preface

The two case studies in this special section of Foresight look behind the issues of forecast accuracy to assess the forecasting process itself.

In "Measuring the Efficiency of an Informal Forecasting Process," Bob Samohyl analyzes a manufacturing company that forecasts orders without a statistical model; this firm forecasts on the basis of executive judgment alone. He shows how the company can quickly determine how much potential there is for improving forecast accuracy even using simple statistical models. He also notes that cost savings are a likely by-product.

The second case study, "Forecasting as a Business Process Diagnostic," by Mario Sepulveda-Guzman, Michael E. Smith, and George W. Mechling, examines a manufacturing company that had developed sophisticated statistical models but had still not obtained satisfactory results for a key product line. The authors raise the interesting question of whether resources would be better spent in a careful audit of the business process being modeled rather than in further modeling efforts. They find that model failure has diagnostic value; it may signal a need to reconstitute the work process. If a company responds to such a signal, it will reap the benefits of improved forecast accuracy.

The section concludes with Ken Kahn's commentary, which puts efforts to improve forecast accuracy into the perspective of the company's ultimate goals, and reinforces the argument that forecast accuracy should not be an end in itself.

Len Tashman,
Foresight Editor

MEASURING THE EFFICIENCY OF AN INFORMAL FORECASTING PROCESS by Robert W. Samohyl

Preview: Forecast efficiency combines considerations of forecast accuracy and the cost of the forecasting process. Bob Samohyl takes us through a case study that shows (1) how a manufacturer unwittingly maintains an inefficient forecasting process in which forecasts are not formally recorded; (2) how forecasts and forecast errors can be calculated, even if these are not recorded by the company; and (3) how the inefficiency can be detected readily by comparing forecast errors against errors from naïve models.

Bob Samohyl (PhD, 1979, Rice University) is from Houston, Texas. He has two daughters, Kristen and Kelly, and five grandchildren. He has lived and worked in Brazil permanently since 1978, and he is presently full professor and vice chairman of the Industrial Engineering Department (EPS) at the Federal University of Santa Catarina (UFSC). He is presently coordinating a project for implementing forecasting techniques in electrical energy companies in Brazil and is a recent vice president of the Brazilian Operations Research Society.

- Informal forecasting by in-house experts ("market feeling") is rarely sufficiently scientific to secure accurate forecasts. The usual results are large forecast errors, excessive inventories, and an unnecessarily costly expenditure of executive time.
- The errors from informal forecasting processes can be calculated on the basis of changes in the stocks of finished goods.
- These errors can be compared to the forecast errors from automatic procedures called *naïve* models. For example, a firm might be able to produce more accurate forecasts on the basis of deliveries made 24 working days earlier, which would be deliveries made on the same day of the week four weeks earlier.
- Better procedures in the firm result in better forecasts and often less effort.

Introduction

The decision to fix production levels of goods and services is the result of a sales forecast. Even though well-planned procedures are increasingly common (Armstrong, 2001), many firms still operate under the illusion that forecasts are a secondary element in production planning and profitability. These firms work under the hypothesis that *market feeling* exercised by in-house experts is sufficiently scientific to secure an accurate view of the future. Firms in this situation usually do not treat the forecasting function as a formal process, and they do not take advantage of the tools of process control to optimize resource utilization.

In this paper, I will describe a simple method for evaluating the efficiency of the forecasting process in a firm where forecasting is relegated to an informal or market-feeling process. By *efficient*, I mean a forecasting process that generates acceptable forecasts quickly and cheaply. This method tests the accuracy of the informal process against several ***naïve*** models. A naïve model is one that forecasts that future demands will be no different from demands in a certain past period.

Case Study

To provide a concrete illustration of my method, I utilize a real case study. Companies that lack a formal forecasting function could learn from the experience presented in this case: a naïve forecast may be at least as accurate as the firm's informal forecast while actually costing the firm a great deal less.

The XYZ Muffler Company uses daily data for its very short-term production line decisions. Table 1 shows a sample of 624 daily observations on its best-selling muffler. The dates include nonholiday Mondays through Saturdays. The data in this table are real, although the company name has been disguised.

The column labeled "delivery to client" is the amount of finished product that actually arrives at the client on that

Table 1. The XYZ Muffler Company Data

Date	Delivery to client	Finished product	Sales
22-Mar-02	161	0	77
23-Mar-02	255	0	76
25-Mar-02	140	0	76
26-Mar-02	126	0	67
27-Mar-02	162	0	0
28-Mar-02	83	0	86
29-Mar-02	139	0	113
30-Mar-02	609	0	373
-	-	-	-
-	-	-	-
-	-	-	-
-	-	-	-
-	-	-	-
05-Mar-04	56	0	54
06-Mar-04	209	320	142
08-Mar-04	184	41	197
09-Mar-04	167	0	200
10-Mar-04	0	0	0
11-Mar-04	51	56	171
12-Mar-04	197	435	199
13-Mar-04	146	237	163

particular day. “Finished product” is that which rolls off the assembly line that day. The “sales” column indicates the amount of orders placed by clients on that day.

The firm’s database lacks an accurate estimate of finished product inventories, as well as data on the length of the delivery period. It is also interesting that the company does not systematically analyze forecast errors; in fact, they do not calculate forecast errors. I will show how an analyst still may be able to calculate forecast errors and implement a more efficient forecasting process, even in this informal forecasting situation.

The firm is capable of producing hundreds of different kinds of mufflers, but here we concentrate efforts on only one, its best seller. The firm is a major supplier of replacement mufflers and sells strictly to auto parts stores. It uses a relatively sophisticated computerized system of Material Requirements Planning (MRP) as part of a larger system for Enterprise Resource Planning (ERP), but it has no formal forecasting system.

Rather, an ad hoc group of upper-level managers from production, sales, and marketing make the forecasts, which are purely judgmental. The group meets at irregular times during the week, normally when information appears that a member of the group judges sufficiently important, or when at least two of the members find themselves drinking espresso in the managers’ lounge. The managers take notes at these meetings, and these are reported to the vice president in charge of production. The vice president makes the final decisions on what, when, and how much to produce.

The Timing of Orders and the Need for Forecasts

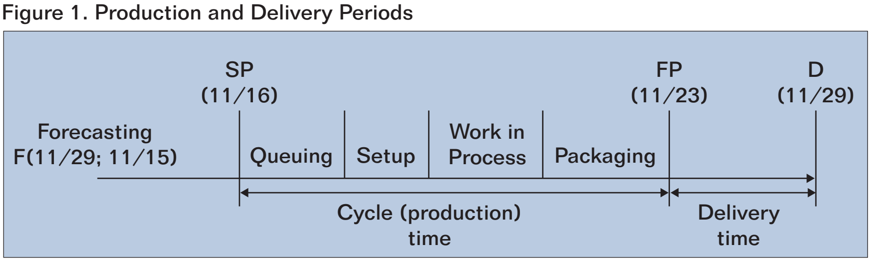
Figure 1 shows the sequence of events from the date of an informal forecast on 11/15 through the resulting delivery on 11/29.

Assume that an order comes in requesting delivery on 11/29, **D(11/29),** and that it takes five working days to make the delivery. The work must be in finished condition on 11/23, **FP(11/23).** Given that it takes six working days (nonholiday Mondays through Saturdays) to produce the muffler, including queuing, setup, work-in-process, and packaging time, production must be scheduled to start at 11/16 **SP(11/16).**

Orders can come at any time, but the order arrival date shows when forecasting is crucial. Consider three cases:

Case 1: The order is received after packaging—that is, between points FP and D. In this case, the firm cannot satisfy client expectations because there is insufficient time to expedite delivery.

Figure 1. Production and Delivery Periods



Case 2: The order is received during production, between SP and FP. While there is insufficient time to produce a new unit for the 11/29 delivery, there is time to deliver product that is already stocked in inventory. This case is well served by the forecasting function of the firm. An accurate forecast would ensure that adequate stocks of finished product are available for delivery. Hopp and Spearman (2000) call this situation the make-to-stock factory environment.

Case 3: The order comes before SP. Now the firm has time to produce new product and deliver to the client on schedule. This case, in which clients are willing to wait a relatively long time for delivery, is called the make-to-order environment. Here the forecasting function would not be as high a priority as it is in Case 2. Nevertheless, when orders arrive very close to this period, additional costs of queuing and setup time loom.

When we consider the queue and setup times, the length of production lead time depends on the current demand. If an order arrives at a moment of slack demand and is very similar to an order already in processing, lead time is relatively short. On the other hand, an order may arrive and enter into a long queue or be sufficiently different from product in process so that the resulting lead time is long. Hence, a forecast of orders is helpful for determining which orders should be placed at the head of the queue. For instance, if the firm forecasts a series of large orders from priority clients for the end of October, the firm should prepare production lines for this eventuality.

In the next section, we consider a situation very common in factories that are technologically current in controlling work processes. When such a firm does not have a hands-on system for verification of actual stock in inventory, it can lose contact with its actual stock of finished product. In this context, we investigate the time relationship between the forecast (F), finished product (FP), and delivery (D) to evaluate the accuracy of the firm's informal forecasts.

Measuring Forecast Accuracy When the Firm Lacks a Formal Forecast Function

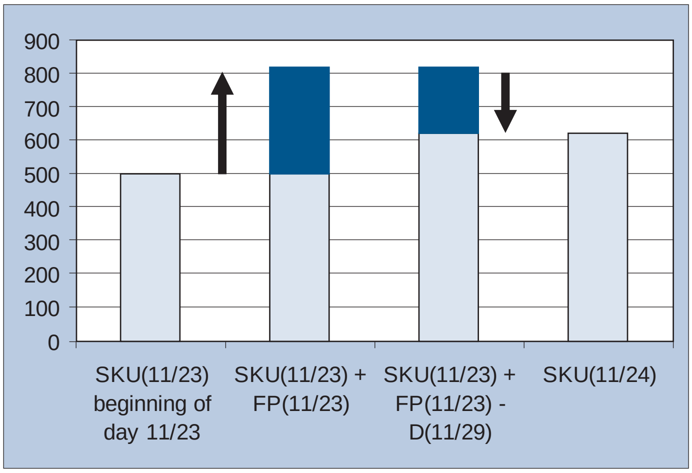
Any forecast is characterized by two dates: the date the forecast is made and the time period being forecast. In Figure 1, the notation F(11/29; 11/15) represents the forecast made on 11/15 for deliveries on 11/29. Given that forecasts are made to produce and deliver product, the latest that a forecast for 11/29 should be made is SP(11/16).

When a forecast is accurate, the inventory of finished product does not change because we sell exactly what the forecast says we need to produce. When forecasts are not accurate, deliveries will not match sales for the requested date, and stocks of finished product will vary. Therefore, if XYZ Muffler had data on the variability of stock, it could use that data as a measure of forecast error.

Even though XYZ Muffler does not have verifiable data on finished goods inventories, we nevertheless can construct a measure of forecast accuracy. The basic premise is that the value of finished product at a certain date, 11/23 in Figure 1, reflects the value of the forecast made some days earlier, say, 11/15, for delivery on 11/29. **Therefore, the amount of finished product on 11/23 can serve as a proxy for the forecast made on 11/15, and the difference between finished product and ultimate delivery of that product becomes a measure of forecast error.**

If the forecast for deliveries on 11/29 is inaccurate, the firm will see its stock of finished goods rise if the forecast overshoots, or fall if the forecast undershoots. In turn, we can measure this variation in stock by the difference between finished product and deliveries. Deliveries to the client usually will not occur on the same date that finished product comes off the assembly line because a certain time interval is required for special packaging, transport, and other handling. So we should not measure forecast error by comparing finished product and same-day deliveries. Instead, we need to reflect logistics reality and allow a certain number of days for delivery lag.

Figure 2. Daily Variation in Finished Product Inventory



The logistics manager revealed that the delivery lag could be zero (same-day delivery) or as many as six days. In Figure 1, we had assumed five working days for delivery. Using the five-working-day assumption for delivery lag, the forecast error can be calculated as

Forecast error (5-day-delivery lag)	= FP(11/23) – D(11/29)
	= Change in FP Inventory (11/24 – 11/23)

In Figure 2, each bar labeled "SKU" measures the number of finished units in stock for a particular day. The first SKU (11/23) is the finished product inventory at the beginning of 11/23. The next bar shows that new finished product, FP(11/23), has entered inventory during the day. The third bar shows the depletion of inventory during the same day as the delivery process begins. The last bar is the stock beginning the next day, 11/24. Hence, the forecast error, the daily variation in stock, is identical to the difference between finished product and deliveries as they leave the factory. If finished product coming off the assembly line was 300 units on 11/23, and 200 units were taken out of stock this same day for delivery five days later, the variation in stock is 100 units, which is precisely the forecast error. From the data in Table 1, we can calculate the daily variations in finished product inventory (i.e., forecast error) for any specified delivery lag.

Then we can calculate an average of the forecast errors—called the MAD, for mean absolute deviation—for each specified delivery lag, from zero to five days. Figure 3 shows the resulting average forecast errors.

The average forecast error is smallest, about 160, when the delivery lag is zero (that is, when delivery of finished product takes place on the same day as its finished production). For more realistic lags, the MAD ranges from 175 to 195 units. Let us use an MAD of 180 as a representative figure.

In this value of 180, we have an estimate of the forecast error as the firm presently operates—without a well-defined forecasting process.

Can this level of forecast error be readily reduced?

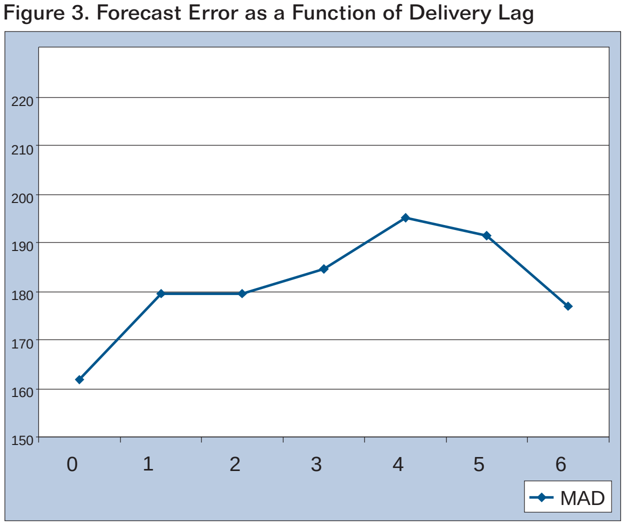
Judging the Size of the Forecast Error

Is a MAD of 180 large or small? It is a question of comparison. We need an objective criterion for judging the size of the firm's forecast errors. According to the muffler company's managers, forecasts that come from the present ad hoc procedures are the best possible, given the uncertainty of the sector and the economy in general. Is their judgment unquestionable?

One way to tell is to compare their forecast error with that of a **naïve** forecast, which is a forecast that there will be no change in deliveries from some prior level.

A forecast must be made before the production cycle begins for a specified delivery date. In terms of Figure 1, the forecast must be made before the start of production at SP(11/16). We used 11/15 as the date the forecast was made,

Figure 3. Forecast Error as a Function of Delivery Lag



F(11/29; 11/15), which is 14 calendar days (12 working days) earlier than the required delivery date.

Let us define a naïve forecast as the value of actual deliveries made a number of days earlier (in this case, 14 days). The idea is that what happened on a certain day in the past might repeat itself in the future, hence the name "naïve." (In financial circles, a naïve model is also called a *random walk*.)

According to Figure 1, a naïve forecast for deliveries on 11/29 would be deliveries made at some past date, for instance on 11/15.

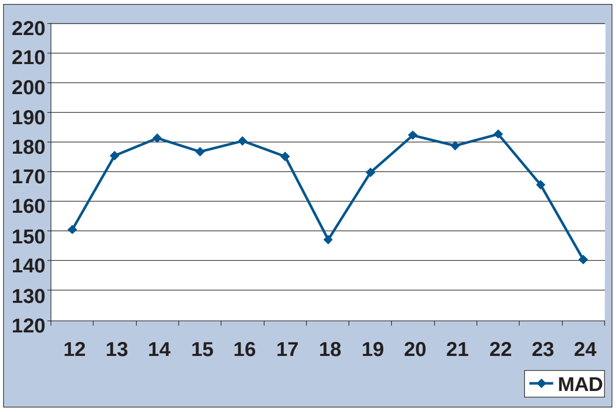
Naïve Forecast, F(11/29; 11/15) = D(11/15)

Alternatively, the forecast could have used actual deliveries on 11/8, 21 calendar days (18 working days) earlier, or even actual deliveries on 11/1, 28 calendar days (24 working days) earlier. Later we will see that using actual deliveries from 28 calendar days ago as our forecast for future deliveries produces surprisingly promising results.

Forecast Error from Several Naïve Models

We calculated forecast errors from a naïve model based on the actual amount of past deliveries made 12 working days earlier, 13 working days earlier, 14 working days earlier,

Figure 4. Forecast Accuracy for Naïve Forecasts



and so on, through deliveries made 24 working days earlier. Figure 4 shows the average forecast error (MAD) for each model.

We can see from the initial point that the naïve forecast based on deliveries made 12 working days earlier has a MAD of 150 units. We call attention to the result of the best naïve forecast, which has an MAD of 140 units, the forecast based on deliveries made 24 working days ago (or 28 calendar days), which is the same day of the week four weeks earlier. The best naïve forecasts have smaller MADs than we calculated in Figure 3 for the firm's informal forecasts. **Forecasts made by simply looking back to the deliveries of 24 working days earlier are more accurate than the forecasts produced by the firm's executive committee.**

These results were rather perplexing for our muffler company. The company was spending considerable resources on executive meetings among high-level managers who should have been expected to understand and predict market behavior. Moreover, the vice president of production should have been able to synthesize the information and determine production levels that matched deliveries.

In addition, since the forecasting function was relegated to an informal status in the firm, the inefficiencies were being hidden. Specific tasks were not being openly defined, and data needed to assess the accuracy of forecasts and the performance of the forecasters were not being gathered, much less analyzed.

Any manufacturer who uses an informal forecasting process should take the time to determine how the firm's forecast accuracy compares with that of a naïve model. If the naïve model is as good or better, management should realize that greater accuracy is possible without necessarily increasing the cost of the forecasting process.

The forecasting literature recommends safeguards against biases in formal forecasting processes (Mentzer and Moon, 2004; Goodwin, 2005). One can only imagine the degree of bias when the forecast function, as it exists in the XYZ Muffler Company, is not just purely judgmental but also informal.

Extensions

In this example, we have seen that more efficient forecasts can result from better procedures and possibly through less effort instead of more. If a naïve model improves a firm's forecast accuracy, the firm should simply switch to naïve models and then consider moving to more advanced methods as a next step. One family of methods that has been widely adopted for production-level forecasting is that of exponential smoothing. Gardner and Anderson (1997) have shown that exponential smoothing outperforms purely naïve procedures. You can find a description of this methodology in any forecasting textbook.

The reliance on automatic forecasting does not imply that executive judgment is banished from the forecasting process. Indeed, the firm could structure its forecasting function to include judgmental considerations. See the Special Feature in Issue 1 of Foresight (June 2005), *When and How to Judgmentally Adjust Statistical Forecasts.*

References

Armstrong, J. S. (2001). *Principles of Forecasting*. Boston: Kluwer Academic Publishers.

Gardner, E. S. & Anderson, E. A. (1997). Focus forecasting reconsidered. *International Journal of Forecasting*, 13, 501-508.

Goodwin, P. (2005). How to integrate management judgment with statistical forecasts. *Foresight: The International Journal of Applied Forecasting*, Issue 1, 8-11.

Hopp, W. & Spearman M. (2000). *Factory Physics*, 2nd ed. New York: McGraw-Hill.

Mentzer, J. T. & Moon, M. A. (2004). *Sales Forecasting Management: A Demand Management Approach*, 2nd ed. Thousand Oaks, California: Sage Publications.

Contact Info:
Robert Wayne Samohyl
Federal University of Santa Catarina
samohyl@deps.ufsc.br

All-in-all I think the journal is a home run. The information is relevant to practitioners and is presented in a way that is not overly academic but with significant credibility.

Thomas Ross, Financial Analyst
Brooks Sports

I have found my first two issues of FORESIGHT very informative...an important forum for practitioners to share their experiences....

Dan Kennedy, Senior Economist
Connecticut Department of Labor

I really like the Special Feature section– it allows someone to get a far deeper understanding than is possible by either a series of extracts on a subject or by an in-depth feature from a single author. Getting several different perspectives in one issue is great!

Simon Clarke, Forecasting Process Mgr.
Coca-Cola Enterprises – North America

I found a book review very useful, especially when I was first starting out as a forecaster. And the 'Pollyvote' article was a nice change from the business forecasting. I thoroughly enjoyed academics taking themselves a little more lightly.

Rob Dhuyvetter, Mgmt. Science Analyst
J. R. Simplot Company

FORECASTING AS A BUSINESS PROCESS DIAGNOSTIC

by Mario Sepulveda-Guzman, Michael E. Smith and George W. Mechling

Preview: This case study describes a manufacturer's failure to develop an adequate cost-forecasting model, and it examines the bases of that failure. The authors see the failure as an indication that the business systems in place need to be redesigned. They make the interesting point that the success or failure of a specific forecasting effort can become a diagnostic of the entire system's performance.

Mario Sepulveda-Guzman is the Commercial Team Leader of a Fortune 100 company that is the world's leading manufacturer of construction and mining equipment. He has more than 11 years of experience in the manufacturing arena and he has held positions in purchasing, design Engineering and quality. He has an MBA from Western Carolina University and a Masters in engineering from the University of Dayton. His research and experience in forecasting methods has helped his firm achieve benefits in excess of $1M.

Michael Smith is Assistant Professor of Management and International Business at Western Carolina University, where he teaches Strategy and Supply Chain Management at the graduate and undergraduate levels. Before completing his PhD, Michael spent more than 15 years in executive management, serving as the chief operating officer for businesses in both the service and manufacturing sectors. His teaching and research, which is aimed at promoting business performance through systemic management of inter- and intra-firm relationships, has been published and presented in numerous venues.

George Mechling is Professor of Management and International Business at Western Carolina University, where he teaches Management Science, Operations Management, and Management of Technological Innovation at the graduate and undergraduate levels. George spent more than 10 years in the manufacturing industry serving as the traffic manager and market and economic conditions analyst for businesses in the steel fabrication sector. He has consulted in the private sector and provided expert-witness testimony. He has many research publications and presentations to his credit.

- The failure to construct satisfactory forecasting models may be a signal that managers need to make changes in underlying work processes.
- We use the case study to illustrate a situation in which the source of forecasting failure is not with the functional forms of the models being estimated, or even with the data, but rather with the business processes.
- Model failure has diagnostic value. It may signal that the next step is to reconstitute the work process or to reexamine the process design to find and correct its flaws. If a company is successful at doing this, it will reap the benefits of improved forecast accuracy.

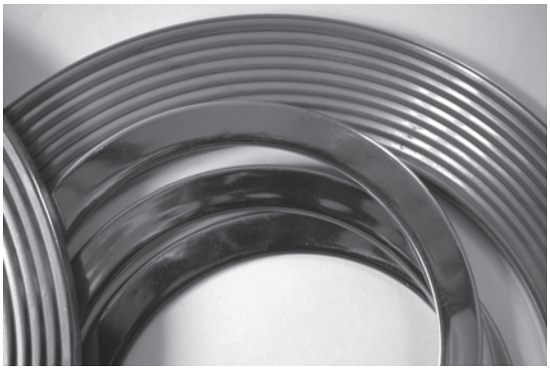
Introduction

Forecasting provides decision makers with rigorously informed speculation about future conditions. But what if the results of such activity are unreliable? Statistical analysts would probably seek to refine the forecasting models and improve the data. But what if the forecasting results are still inadequate after these steps?

Could it be that there are limits to the benefits of forecasting activity? Must decision makers then resign themselves to working with a questionable forecast because that is better than no forecast at all?

We believe that the inability to develop a useful forecasting model may serve as a valuable diagnostic tool that signals the need to make changes in underlying work processes. We base this conclusion on our experience in constructing forecasting models for a large manufacturing organization.

The Organizational Context of the Problem



MMG Gaskets, a subsidiary of Kronecker Products, manufactures gaskets of different sizes, compositions, types, and labor content for engines and hydraulic equipment. Given these differences, the production costs of these gaskets can vary widely and nonlinearly because material content necessarily varies nonlinearly with diameter, which is the standard measure of gasket size for the industry. The absence of cost formulae has often delayed price quotes to potential customers by as much as two weeks, while purchasing and production personnel would research and communicate cost information back to the sales department. Delays of this sort can place an organization at a competitive disadvantage. Therefore, MMG formed a team to undertake the construction of statistical cost models of the firm's gasket production.

Purchasing and production records provided a year's worth of data on more than 500 gasket orders, each of which consisted of a number of features considered relevant to gasket production costs. Plotting these features against the order's total cost per unit showed that many of these paired relationships were direct, continuous, well behaved, and nonlinear. MMG then experimented with various functional forms and variables to generate total cost models.

Attempts at Forecasting

The team made numerous attempts to construct a model, but it failed to achieve useful results. However, these efforts showed that the model was performing significantly better for estimating the cost of the set of gaskets with diameters 200 mm and greater than it was for the set of smaller gaskets.

The researchers learned that the larger gaskets were produced in-house, while the smaller ones were outsourced. When the team tried distinct models for each of the two data sets, it found results that seemed adequate to establish a pricing guide for the larger gaskets. But it had no such luck with the smaller gaskets.

The team attempted to improve upon these poor forecasting results for the smaller gasket set by investigating the organization's business processes. They found that MMG obtained its smaller gaskets from three suppliers with three different cost structures. Further, the methods of production were such that costing for each supplier was variable, as the cost of the dies used to stamp out gaskets varied with the size of the gasket and with degrees of utilization of the dies. These variations in cost were confounding the firm's forecasting efforts.

The team then attempted to statistically standardize the cost allocations by matching cost figures to their respective suppliers and by factoring in the number of units produced. But again the results were disappointing. This time the results showed that the outsourcing decisions were not being made in a systematic manner and that pricing by the suppliers was not systematic.

A Self-Critique by the Forecasting Team

Given the complexity of the problem, forecast analysts might be tempted to use complex functional forms and techniques to tease out some explainable pattern of variation from the data. However, the nonlinear approaches that MMG's forecasting team used are probably sufficient.

MMG's forecasting team consisted of three skilled statisticians, one of whom was also an engineer intimately acquainted with the company's operations and supply chain activities—an insider. Also, the team members had worked closely with each other. So there was little reason to believe that they had not exhausted all reasonable explanations for the observed statistical variation. It is also likely that the time and energy the team invested in this research far exceeded what most organizations would be willing to expend. Some companies might have concluded that a cost-effective solution would not be possible. Would this be the implication for MMG regarding the cost forecasts for its smaller gaskets?

At some point, one must conclude that the problem is not with the functional forms of the models being estimated, or even with the data, but rather with the business processes. Resolving such issues does not fall within the province of forecasting. The formation of systematic work processes requires managerial intervention. The findings of the forecasting team raised managerial issues that should be addressed if forecasting efforts were to have a reasonable chance of succeeding.

Perhaps the MMG team should have recognized the lack of systematic work processes at the outset, but it did not. In this situation, the team's experience suggests that without its forecasting efforts, MMG's problems would have remained hidden.

Implications for Practice

A coherently designed and effectively managed process should behave predictably. There will always be some noise in such a system; however, this noise must remain at an acceptable level. Modeling the process should therefore lead to useful forecasts. Modeling might fail because some assignable cause has disrupted the process or because the process design was incorrect in the first place. An understanding of the process's design is a prerequisite to modeling efforts.

The lesson from the failure of MMG's forecast modeling is that model failure has diagnostic value. Model failure may signal that the next step is to reconstitute the work process or to reexamine the process design to find and correct its flaws. If a company is successful at doing this, it will reap the benefits of improved forecast accuracy.

Business Results Follow Management Intervention

After an investigation of pricing systems for the gaskets, MMG managers found that prices were based on past practices, which varied by supplier. MMG then began to develop purchasing practices that sourced the most expensive gaskets from suppliers that provided pricing advantages. MMG found that new suppliers were providing faster response, showing greater flexibility, and undertaking better R&D on their products. Although implementation of the new purchasing system is still quite recent, cost savings are already being realized and are projected to reach $100,000 per year. As impressive as the cost savings might be, greater savings may result from the recasting of existing supplier relationships. The shift in sourcing practices gained renewed attention from the dominant supplier, as its business with MMG declined. This supplier agreed to negotiate a new three-year contract, which should result in additional savings of approximately $350,000 per year.

A potential forecasting failure uncovered the opportunity for improved business practices. MMG management now seeks to apply the same approach to develop better internal processes for its large gaskets.

Conclusion

There are practical and methodological limits to the efficacy of forecasting. The failure of a forecasting model can become a valuable diagnostic of system malfunction. Likely, the problem is not with the forecasting effort but with the object of the effort, the system itself. A forecasting problem might emerge as a management opportunity to correct an entire production system.

Contact Info:
Mario Sepulveda-Guzman, MBA
Caterpillar Precision Seals
Guzman_Mario_A@cat.com

Michael E. Smith, PhD
Western Carolina University
mesmith@wcu.edu

George W. Mechling, PhD
Western Carolina University
gmechling@wcu.edu

COMMENTARY: PUTTING FORECAST ACCURACY INTO PERSPECTIVE

by Kenneth B. Kahn

Ken Kahn is an associate professor of marketing at the University of Tennessee. His teaching and research interests concern product development, product management, sales forecasting, and interdepartmental integration. He has published in a variety of journals and is the author of *Product Planning Essentials* (Sage Publications). Ken is codirector of UT's Sales Forecasting Management Forum, which emphasizes education and research in sales forecasting and market analysis.

Introduction

The case studies by Samohyl and by Sepulveda-Guzman, Smith, and Mechling (hereafter SS&M) nicely illuminate sales-forecasting realities. In this commentary, I will frame their arguments by posing three questions:

- Should we view *forecast accuracy* as an end in itself or rather a means to an end?
- Do forecast effectiveness and efficiency go hand in hand?
- At what level in the organization should forecasting performance be evaluated?

Is Forecast Accuracy an End or a Means to an End?

Samohyl compares the forecast accuracy of the informal forecasting process to that of a naïve statistical model. In such comparisons, forecast accuracy—measured by Mean Absolute Errors (MAD) or Mean Absolute Percentage Errors (MAPE)—is considered an end in itself. The more accurate model is the better choice for the organization.

In contrast, by pointing to the need for proper linkages between the forecasting process and other business processes, SS&M show that forecast accuracy is a means to an end such as cost saving, or improved customer service. The important theme is that a forecast model failure may provide an opportunity to improve business practices and achieve cost savings. Gillette is a good example of a company that recognizes forecast accuracy as an important intermediate step toward achieving customer service objectives (Covas, 2004). This company's ultimate objective is to provide better service to the customer, and it recognizes that improving accuracy is a means to enable better performance throughout the supply chain

Do Efficiency and Effectiveness Go Hand in Hand?

Efficiency is usually defined as the relationship between performance outcomes and the inputs required to achieve them. Samohyl's concern is that a forecasting process should be efficient, that is the process "generates acceptable forecasts quickly <u>and cheaply</u>." In comparison, the forecasting literature defines **effectiveness** as the organization's ability to achieve its intended goals, given organizational capabilities, competition, consumer preferences, and other environmental conditions (Kerin & Peterson, 1998). So a forecasting process is effective if it meets or exceeds the organization's forecast accuracy goals. But effective forecasting may not be possible to do efficiently because the cost of increasing forecast accuracy may be prohibitive.

Studies (Barghava, Dubelaar, and Ramaswami, 1994; Vorhies and Morgan, 2003) report that there are inherent trade-offs between efficiency and effectiveness, which implies that firms will have difficulty in achieving both goals. Samohyl's company is a case in point. Management spends considerable resources trying to understand and anticipate market behavior, but to little avail. Using a simple approach to forecasting could improve efficiency without harming—and probably improving—forecast accuracy. SS&M agree that more sophisticated modeling methods could have been employed by their company, but what arcane advantage would these methods serve?

At What Level Should We Evaluate Forecast Performance?

At what level in the product hierarchy (e.g., stock-keeping unit at the location-distribution center level, national stock-keeping unit level, product line level, business unit level, company level, etc.) should forecast performance be evaluated? This is a critical question because an organization's efforts to improve performance at one level may inadvertently compromise performance at another level.

Both Samohyl and SS&M focus on the stock-keeping unit (SKU) level, and, in their case studies, I think they were right to do so. I know of companies that look only at top-line numbers, a view that can hide lower-level business problems. For example, a major consumer products firm reported the phenomenally low forecast error of 1.3 percent for month 3, surpassing the company's established forecast error goal. However, the company had spent $286,000 to expedite international shipments. Thus it lost sight of the bottom line.

In general, focusing on the top line imparts a favorable bias to the forecast. A senior management mandate to improve accuracy may prompt an analyst to report forecast accuracy at high levels. Tadepalli (1992) found evidence that if forecasts indicate that goals are unattainable, personnel will tend to "reinterpret" inputs to ensure that goals are met. Mentzer and Moon (2005), in their *Sales Forecasting Management*, observe that "What gets measured, gets rewarded. And what gets rewarded, gets done" (p. 44). How we measure and report performance company-wide has keen implications for the linkage between the forecasting process and business processes.

Conclusion

There is an adage that *it is better to have unanswered questions than to have unquestioned answers*. Management cannot accept an improvement in forecast accuracy without questioning the cost of the improvement and its impacts across the supply chain. The most fundamental issue for forecasters is the extent to which forecasting efforts contribute to overall business results.

References

Bhargava, M., Dubelaar, C. & Ramaswami, S. (1994). Reconciling diverse measures of performance: A conceptual framework and test of a methodology. *Journal of Business Research*, *31*, 235-246.

Covas, M. (2004). Proceedings of the University of Tennessee's Sales Forecasting Management Forum Fall Conference, September 8-10, 2004, Knoxville, Tennessee: *Gillette's Global Value Chain Center of Expertise Operating Framework*.

Kerin, R.A. & Peterson, R.A. (1998). *Strategic Marketing Problems: Cases and Comments*. Upper Saddle River, NJ: Prentice Hall.

Mentzer, J.T. & Moon, M.A. (2005). *Sales Forecasting Management*. Thousand Oaks, CA: Sage Publications.

Tadepalli, R. (1992). Marketing control: Reconceptualization and implementation using the feedforward method. *European Journal of Marketing*, *26*(1), 24-40.

Vorhies, D. W. & Morgan, N. A. (2003). A configuration theory assessment of marketing organization fit with business strategy and its relationship with marketing performance. *Journal of Marketing*, *67*, 100-115.

Contact Info:
Kenneth B. Kahn, PhD
The University of Tennessee
kkahn@utk.edu

ALIGN YOUR BUSINESS TO PROFIT FULLY FROM MARKET DEMAND

JohnGalt Atlas Planning Suite™

featuring Demand Management Engine

For companies that need to address inventory and demand planning, John Galt offers cost effective, easy to implement and comprehensive Supply Chain Planning Solutions.

Atlas Planning Suite components:

- Demand Management Engine
- Inventory Management
- Rough Cut Capacity Planning
- Promotional Performance Management
- Sales and Operation Planning
- Sarbanes Oxley Reporting
- Planning Portal

John Galt's **Atlas Planning Suite** is an **enterprise scalable solution** that allows customers to **balance supply and demand.**

JohnGalt Forecast Xpert Toolkit™

- **Cost effective** forecasting software
- **Define** your forecasting process
- Develop **accurate** forecasts

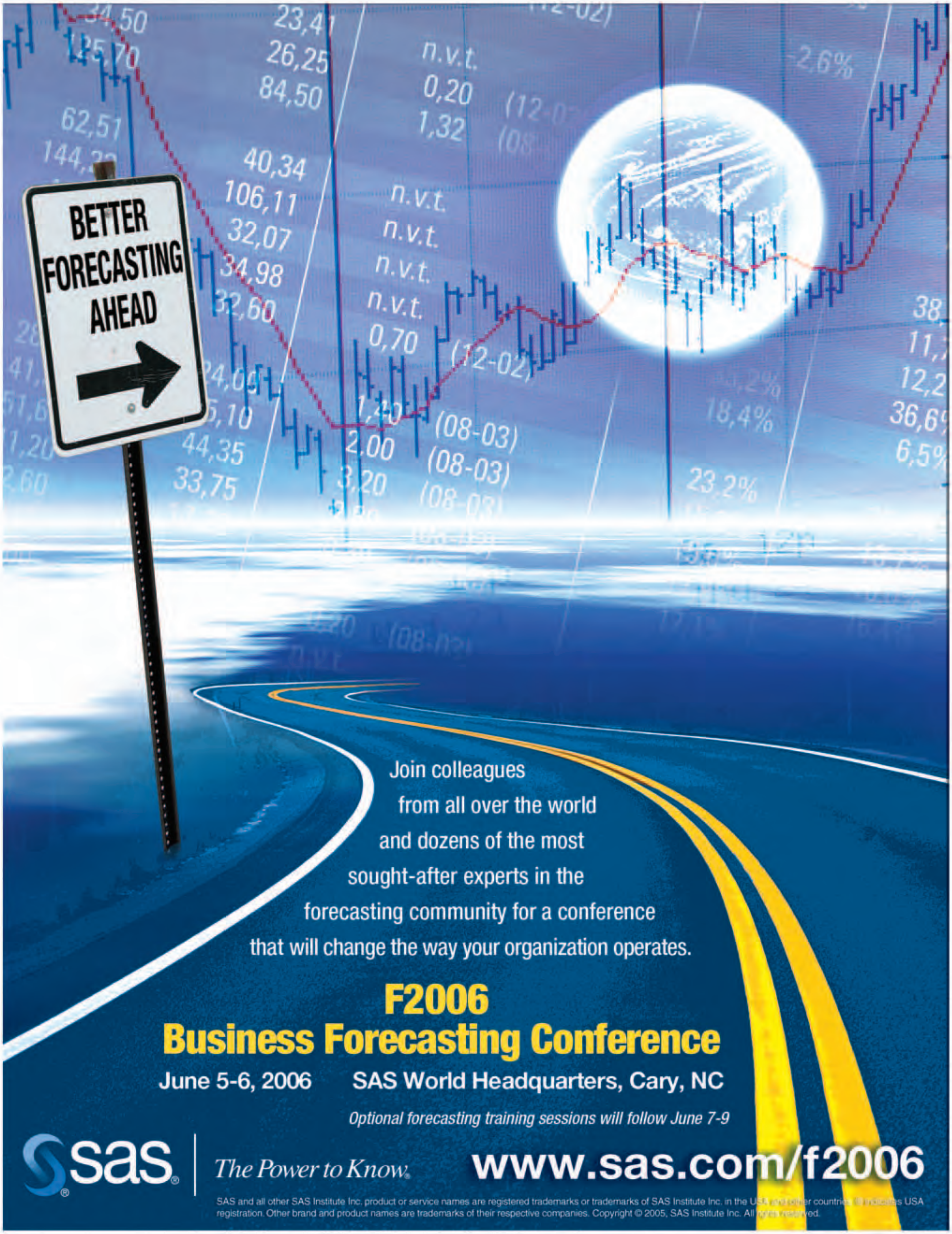
Forecast Xpert Toolkit includes

- ForecastX Wizard 6.0 Add in to Excel
- Business Forecasting Textbook
  J. Holt Wilson - Central Michigan University
  Barry Keating - University of Notre Dame
- Xpert Training in Forecasting Process Definition and Statistical Forecasting

john gAlt solutions, inc.

The Forecast Xperts™ in Enterprise Planning

125 South Clark, Suite 1950 Chicago, Illinois 60603 **Phone:** 312.701.9026 **Fax:** 312.701.9033 **Email:** sales@johngalt.com


BETTER FORECASTING AHEAD
Join colleagues from all over the world and dozens of the most sought-after experts in the forecasting community for a conference that will change the way your organization operates.
F2006
Business Forecasting Conference
June 5-6, 2006
SAS World Headquarters, Cary, NC
Optional forecasting training sessions will follow June 7-9
sas
The Power to Know.
www.sas.com/f2006
SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies. Copyright © 2005, SAS Institute Inc. All rights reserved.

FORECASTING PRINCIPLES AND METHODS

INCREASING THE CREDIBILITY OF YOUR FORECASTS: 7 SUGGESTIONS by Roy L. Pearson

Preview: *Credible*: capable of being believed and worthy of confidence. Credibility is essential for the acceptance of your forecasts. Roy Pearson offers solid advice on enhancing the credibility of forecasts and on reducing forecasting errors. His suggestions are based on years of experience, surveys of academics, and knowledge of best practices.

Roy is Chancellor Professor Emeritus at the College of William and Mary, where for three decades he taught forecasting in the MBA program. From 1984 to 1998, as director of the College's Bureau of Business Research, he regularly published his quarterly forecasts for Virginia and six of its metropolitan areas, and he continues to prepare national, state, and substate forecasts for businesses and government agencies. Roy has served on the Governor's Advisory Board of Economists at the pleasure of five Virginia governors: Robb, Baliles, Wilder, Allen, and Warner. Forecast credibility has been the focus of his research and presentations at professional conferences during the past two years.

To improve forecast credibility:

1. Customize the forecasts for your customers.
2. Analyze seasonality in detail.
3. Consider explanatory modeling, even if it's no more accurate than extrapolation.
4. Base your choice of independent variables on sound logic and "forecastability."
5. Understand the data behind the variables in your model.
6. Weave your numerical forecasts into a story about the future.
7. Include a risk assessment in your forecast.

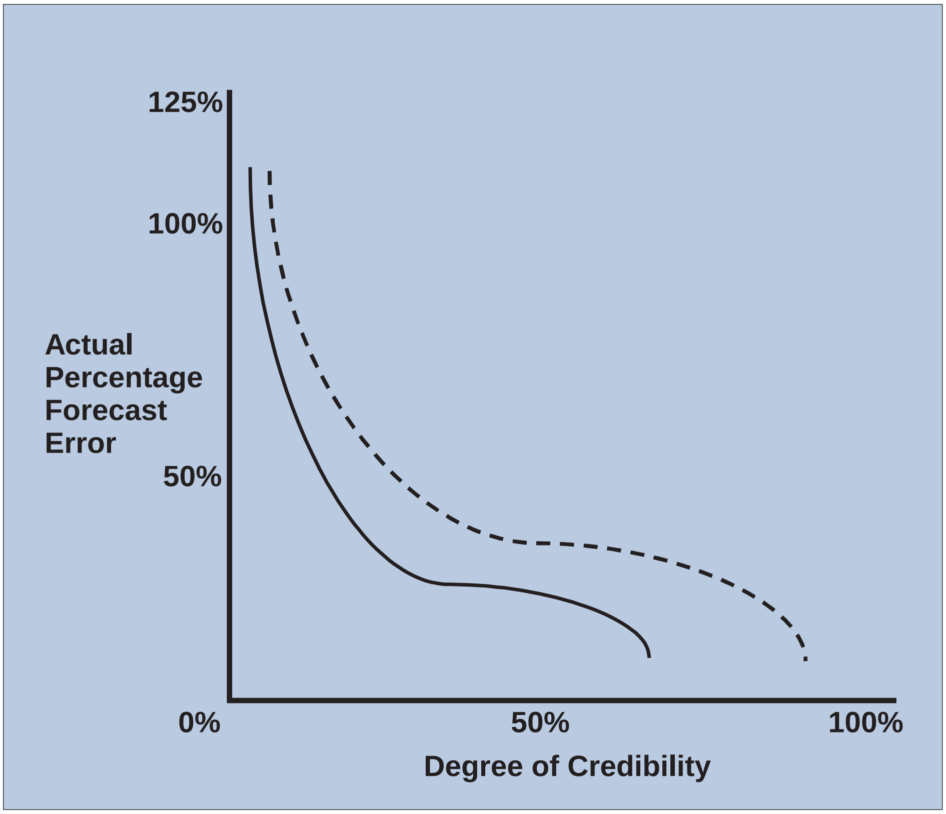
Forecast Error and Credibility

Figure 1 illustrates two important dimensions of effective forecasting: reduced forecast error and increased credibility. The ranges are different, as there is no upper limit to how wrong you might be, while 100 percent is the maximum score for credibility. The two dimensions are inversely related. With very high errors, especially if they are worse than the errors from a *naïve* (no-change) forecast, credibility will be low regardless of how effectively you have prepared and explained the forecasts. With consistently low errors, credibility may be high even if your presentation is unremarkable.

The scholarly writings on forecasting focus predominantly on reducing forecast error. But forecasting is about more than achieving a low error rate. In addition, the forecaster must plan the forecasting effort and present oral or written reports to maximize credibility. Doing so will shift the curve upward and to the right in Figure 1, giving your forecasts more bang for the buck.

Here are my main recommendations, chosen because I and others find that these elements are too often undervalued or neglected. While many of them may reduce forecast error, I offer them principally to enhance forecast credibility.

Figure 1: Forecast Error and Credibility



❶ *Customize the Forecasts for Your Customers*

Every business forecaster should be able to say truthfully that "My forecasts drive company decisions." These decisions are made by the users of the forecasts, who are the forecaster's customers. The forecast will play little or no role in these decisions unless (1) it provides information users need; (2) users understand it; and (3) they conclude that it is credible.

So you must know your customers and address what they really need to know. How will they use the forecasts? What are their key concerns? How global or narrow is their perspective? How knowledgeable are they about the product being forecasted, about its history and key driving forces? You should periodically survey your customers. That will pay off in credibility. A survey may also complement the insight gleaned from your historical data.

❷ *Analyze Seasonality in Detail*

A high percentage of your forecasting activity probably goes into making monthly forecasts for 12 to 18 months ahead. While seasonality, the monthly variation driven by normal weather, holiday, and trading-day patterns, is a key analytical component of the forecast, it is too often underanalyzed. It is true that in the long run, the trends and cyclical behavior of the data dominate the seasonality; however, for 12-to-18 month forecasts, seasonality cannot be treated as a second-class citizen.

What really matters is that the focus of analysis is matched to the forecast horizon. Try an experiment. If you have at least five years of monthly data for a product or product group, ask your software to seasonally adjust the data. You can do this with the Census X-12-ARIMA program and with many other software packages. You can also access a free module from the Census Bureau at http://www.census.gov/srd/www/x12a/. Then, using a spreadsheet, calculate the variance of both the unadjusted series and the seasonally adjusted series. Finally, calculate the percentage reduction from the former to the latter. That percentage indicates how much of the variance is driven by seasonality. Expect the variance to be quite low for series that track the overall economy.

Do the same calculation for each calendar year. For most products, especially consumer goods, the average yearly variance attributable to seasonality will be well over 50 percent.

To illustrate, I took unadjusted and adjusted monthly data for 1992–2004 from the U.S. Census Web site for 37 national retail sales series. Using all 13 years, the variance due to seasonality, including holiday and trading-day effects, averaged 38 percent of the total variance for these 37 series. So the variance due to trends and cyclical forces—the trend-cycle, as Census calls it—dominated the seasonality. For each of the 13 individual years, however, seasonality averaged 95 percent of the yearly variation, with a low of 80 percent and high of 100 percent. For a one-year forecasting horizon, variation due to seasonality clearly dominates.

Seasonality analysis with the Census program has a side benefit: it shows me whether seasonality has been changing over time. For the 37 retail sales series, using only seasonal indexes, the absolute percentage changes in the monthly indexes from 1992 to same month in 2004 were:

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Mean	3.6%	4.8%	2.8%	2.0%	1.3%	1.9%	2.2%	2.5%	2.2%	2.1%	3.3%	5.4%
Median	2.4%	2.6%	2.0%	1.5%	0.7%	1.2%	1.4%	0.9%	1.5%	1.6%	2.4%	4.7%
Max	10.9%	23.9%	7.4%	9.0%	7.3%	10.4%	11.8%	14.6%	11.6%	8.2%	16.9%	15.0%

For many retail store categories, changes have been dramatic. For example, the December 2004 seasonal index, compared to the 1992 index, is down 15 percent for warehouse clubs and superstores, and the November index for electronic shopping and mail-order house retailers is down 16.9 percent. Both of these indexes still are over 130, but well below what they used to be. In the other direction, the February 2004 seasonal index for jewelry stores is up 23.9 percent having risen steadily from 80.5 to 99.7. As social, economic, and demographic changes occur, so do changes in the seasonality of demand. Recognizing changes in seasonal patterns can improve the modeling of the data and lead to reduced forecast error.

Another advantage of using a Census-type seasonal analysis is being able to do trading-day adjustments, which is important for a wide range of goods and services. Without adjusting, you can have a significant amount of seemingly random variation left in your model because the number of each day of the week in each month of the year follows a very long-term pattern. For example, the specific days of the week in your 2005 calendar will not be matched again until 2011, 2022, and 2033. Identifying your trading-day pattern can be useful for your forecast users. It can also reduce your forecast errors.

Note how important seasonality is for your forecast horizon. Whether you use exponential smoothing or regression as your forecasting method, separately analyzing seasonality and trading-day effects may be an effective way to increase your credibility.

❸ *Consider Explanatory Modeling, Even If It's No More Accurate Than Extrapolation*

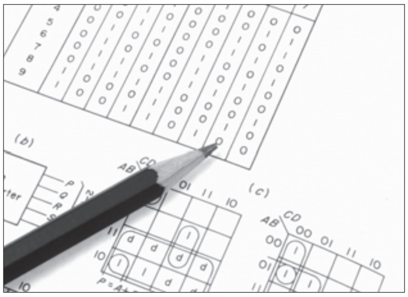
How do you decide whether to forecast by an extrapolation method, such as exponential smoothing, or by building an explanatory model, of which multiple regression is the standard? Some researchers have compared the forecasting accuracy of the two approaches. In one important study, Allen and Fildes (2001) concluded that "Overall, econometric forecasts are more accurate than extrapolative, although the difference is not great, considering that some of the extrapolative forecasts are naive no-change forecasts" (p. 344).

If accuracy is the sole consideration, the small expected advantage of using multiple regression may not be worth the time and cost. However, for establishing credibility with your forecast users, you may need to answer specific questions. What forces are driving the forecast? What will happen if a certain event occurs? For such questions, you often need a causal model, one that relates the forecasted variable to causal forces, especially ones of concern to your customers, such as energy prices, interest rates, consumer confidence, and company or industry advertising. So multiple regression modeling may still be worth the time and cost, even if it fails to significantly reduce forecast errors.

❹ *(a) Choose Your Independent Variables Based on Sound Logic*

Even a nearly perfect statistical fit does not prove a meaningful causal relationship. The good fit may be occurring purely by chance; in that case, when the model is asked to forecast, it will yield large forecast errors. To demonstrate spurious fits, I selected 24 absolutely irrelevant independent variables for a model to predict monthly sales of U.S. health and personal care stores. I chose variables such as Italian stock prices, inflation in the United Kingdom, and U.S. burlap and rubber prices. Then I selected regression models with seven independent variables. Taking 24 variables seven at a time, there are 346,104 possible combinations of seven independent variables. By pure chance, hundreds of these models will show a nearly perfect statistical fit, with an adjusted R-square of 99% or more. A forecaster who found one of these models should not be willing to risk his or her job by using it to forecast. Rather, the choice of independent variables needs to be based on reasonable theory or common sense. If you cannot think of a plausible reason for a causal linkage to your dependent variable, there is a good chance that there is none.

Moreover, sound logic usually requires lagging some variables because not all relationships are concurrent. For example, building permits precede



housing starts, which in turn precede construction spending. Leading indicators exist for national output, income, and employment, and also for many industry and company variables. The sound logic behind the lagged relationship can increase the credibility of the forecasts, and precedence is a big plus in asserting causality. Using properly lagged variables can also reduce forecast error and provide advance warning of turning points.

4 *(b) Base Your Choice of Independent Variables on "Forecastability"*

How accurately can you forecast the independent variable candidates? If an independent variable cannot be forecast any better than a forecast of no change in the future, it won't do you any good to include it in the model. For example, stock market indexes consistently have followed an unpredictable random walk (i.e., they lack "forecastability").

More generally, the decision to retain a variable in your model depends not only on the probable forecast error for the variable but also its explanatory value in your model. The net effect may be to improve your forecast's accuracy, even when the expected forecast errors for the independent variable are relatively high. Bassin (2005) discusses methods for testing the impact of including a variable, and he provides helpful references.

I offer two additional suggestions. First, try to include a larger number of national economic variables than you have used in the past. Some of these work well for forecasting product line sales, even in a metropolitan area. No area in the nation is an economic island; rather, all areas are influenced by national and international forces. Moreover, the U.S. government goes to great lengths to revise the historical macroeconomic data and to keep them consistent over time. For example, the monthly retail sales series at the Census Web site have been revised back to 1992 in order to reflect NAICS code categories (North American Industry Classification System www.census.gov/epcd/www/naics.html). Furthermore, the national series are available with little delay: monthly retail sales and employment data are posted in the following month.

Most importantly, there are many free sources of forecasts of many national data series. If you prepare monthly forecasts, you might be concerned that the free forecasts are typically for quarters, not months. That leads to my second suggestion: convert their quarterly forecasts into monthly ones that you can plug in to your monthly models. Doing so can dramatically expand your options for including external forecasts in your model. Frequency conversion functions are built into some forecasting software, such as EViews and the Matlab toolboxes.

To illustrate the accuracy of quarter-to-month conversions, I converted ten years of the quarterly personal income data from the GDP accounts to monthly values (I used the quadratic-match-average option in EViews). I then compared the converted monthly values to the actual monthly values reported by the Bureau of Economic Analysis, and I found that the monthly mean absolute percent difference was less than 1/10 of one percent. In practice, I use the actual monthly historical data; then I take free quarterly forecasts of personal income and convert these into monthly forecasts. Try it; I think you will like the results.

In addition to selecting logical economic variables, I recommend using dummy variables to capture the effects of unusual events. These variables may be particularly appropriate in the coming months to identify the impacts of recent hurricanes—impacts that are not necessarily reflected in energy prices or employment data. And should there be a recurrence of the event, the coefficients on the dummy variables provide an estimate of their impact. We know that hurricanes will recur, so analyzing past impacts and developing a hurricane scenario can be a beneficial exercise. Home Depot has done this for several years.

5 *Understand the Data behind the Variables in Your Model*

Once you have selected certain variables for your modeling, determine how the data on these variables have been derived. Who compiles the data, and how do they define the variable? How do they collect the data—from full reporting, from sampling, or from extrapolation from a base period? Are the data consistent over time? This is an important issue with the conversions from SIC (Standard Industrial Classification, www.osha.gov/oshstats/sicser.html) to NAICS codes, and also with the changes in definitions of metropolitan areas.

Economic models make extensive use of relative prices (the price of your product divided by a broad price index). But what is the best denominator? Choices include the GDP deflator, a Consumer Price Index for a specific type of consumer expenditure, some component of the Producer Price Index, or an index you construct based on the prices charged by your major competitors. Being able to justify your choice will add credibility to your forecasts.

❻ *Weave Your Numerical Forecasts into a Story about the Future*

Plan your presentation as a story about the future. Memorable stories have three parts: a character, a problem, and a resolution of the problem. For business forecasters, your character is the product being forecast; your problem is the product's future; your forecast helps resolve the problem.

Develop the character by describing the product's history and the forces that you believe shaped that history. Even informed readers will have differing views about the product's past, and especially its present, and they will need to understand your point of view. But don't dwell predominantly on the past. Give your perspective, but remember that the forecast is the message. My rule of thumb is to spend no more than one-third of the presentation on the past and present; you should spend the remaining two-thirds on the forecast. Practitioners in my surveys generally favored a 50-50 split.

The problem in the story is the product's future. Do not attempt to downplay the problem, believing that reassurance will enhance your credibility. It won't. The opposite outcome is more likely. Can you recall an interesting story in which the problem faced by the main character was trivial? The future always is uncertain, and you need to be honest and clear about the risks and alternative outcomes. You might discuss the causal forces related to your forecast, along with their relative importance. For example, if your sales forecast is highly sensitive to consumer income or confidence, but less sensitive to competitors' prices, including these insights can increase understanding and credibility.

You now have a lead-in to how your forecast can help resolve the problem. Why is your forecast the most likely path over the forecast horizon? If that path is contrary to its recent direction, or to the general trend in the economy, what are the reasons for its divergence?

In answering such questions, you are telling a story about future events and their impact on your product and company. Do not undermine your story by showing equations or pages full of numbers, even to a roomful of econometricians. Put the technical details in an appendix to your report, where they can be viewed as a supplement to your story. I recommend the KISS 'N' KIN approach to the story: Keep It Sufficiently Simple and Keep It Nonmathematical.

For models on how to tell your story, read some of the free monthly publications published by financial institutions. Consult Wachovia's *Monthly Economic Outlook* at http://www.wachovia.com/corp_inst/page/0,,13_54,00.html and Diane Swonk's *Themes on the Economy*, at Mesirow Financial's Web site, http://www.mesirowfinancial.com/media/dswonk/default.jsp. The character here is the U.S. economy, and such publications tell a story about its current condition and how it will change in the next year or so. The stories they tell are short and clear, and they contain only a few numbers supported by graphs. The full array of numbers is placed at the end of each publication.

Numbers are not your story, but they are the foundation for it. A helpful guide to writing and talking about numbers is *The Chicago Guide to Writing about Numbers* (Miller, 2004), which offers 12 basic principles.

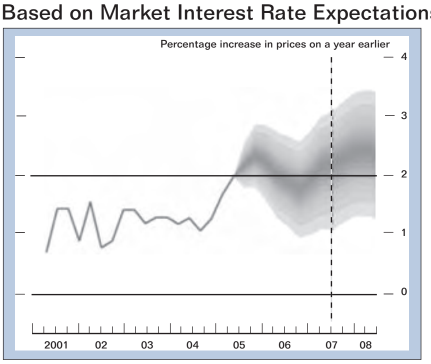
❼ *Include a Risk Assessment for Your Forecast*

Your forecast will be wrong—the question is by how much and in which direction. You need to show probable forecast errors. You can base these errors on past forecast results, on out-of-sample metrics for your forecast model, and on what-if simulations. For regression models, be sure to incorporate the expected errors for the independent variable forecasts. Otherwise, you are likely to understate the probable errors by 100% or more.

Prediction Intervals (PIs) show a probability range for future outcomes, but I find nothing sacred about the 95-percent PIs that are emphasized in statistics courses. I frequently use 50-percent PIs, and my business audiences find them reasonable. Other practitioners I have surveyed reported using 50 percent, 60 percent, or 65 percent.

I like the fan chart initiated by the Bank of England for its inflation reports, available at http://www.bankofengland.co.uk/publications/inflationreport/2005.htm. Chart 1 shows a fan chart for inflation forecasts. The chart displays PIs in 10-percent increments, encompassing a 90-percent prediction interval. The U.S. Congressional Budget Office is also using this approach for current budget surplus or deficit forecasts.

Chart 1. Current CPI Inflation Projection Based on Market Interest Rate Expectations



How much of your presentation should be devoted to this risk analysis? Forecasters I have questioned recommend devoting about 10 percent of the presentation to discussing risks.

Conclusion

I hope you try some of these suggestions. I believe they will enhance the credibility and accuracy of your forecasts. They might also make forecasting more fun.

References

Allen, G. & Fildes, R. (2001). Econometric forecasting. In J. S. Armstrong (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners* (pp. 303-362). Boston: Kluwer Academic Publishers.

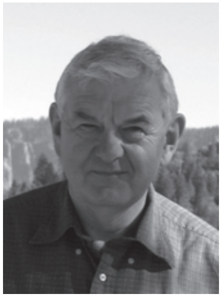
Bassin, W. M. (2005). To include or not to include an explanatory variable: That is the question. *Foresight: The International Journal of Applied Forecasting*, Issue 2, 33-36.

Miller, Jane E. (2004). *The Chicago Guide to Writing about Numbers*. Chicago: University of Chicago Press.

Contact Info:
Roy L. Pearson
College of William and Mary
roy.pearson@business.wm.edu

CREDIT SCORING: THE STATE OF THE ART by Lyn C. Thomas

Preview: Credit scoring is the most successful and widely used application of forecasting in the whole financial sector. The name refers to the techniques that help lenders decide whether or not to approve loan applications. In this article, Lyn Thomas discusses the origins of credit scoring, describes the major techniques in use, and examines the recent advances in the field designed to deal with new regulations and a more competitive consumer credit market.



Lyn Thomas, Professor of Management Science at the University of Southampton, UK, has written or edited three books and numerous research papers in the area of credit scoring. Lyn founded the Credit Research Centre at the University of Edinburgh and has helped organize the biennial international conferences run there for the past 18 years. He is a past president of the UK Operational Research Society and a Fellow of the Royal Society of Edinburgh.

- Credit scoring has provided the underpinning for enormous growth in consumer credit over the past 50 years.
- Traditional credit-scoring models use classification techniques to predict which borrowers will default.
- Changes in lending objectives and in regulatory and market conditions are providing new challenges for credit-risk assessment, including the need to accurately assess default risk.
- Lenders formerly asked which borrowers will default. Now they ask *when* a borrower will default. This change has prompted the use of new types of models.
- Recent modeling advances address the need for default risk scoring, profit scoring, and acceptance scoring.

Introduction and Brief History

If asked which aspect of their lives is being forecast the most frequently, few people would respond that it is the chance they will default on a loan. Yet that is the case. Since its introduction 50 years ago in 1956, credit scoring, also called application scoring, has become ubiquitous in consumer finance. Its techniques help lenders decide whether to lend to a new applicant. The related approach, called behavioral scoring, is concerned with the operating and marketing policies applied to an existing customer, such as a request for an increase in the customer's credit limit.

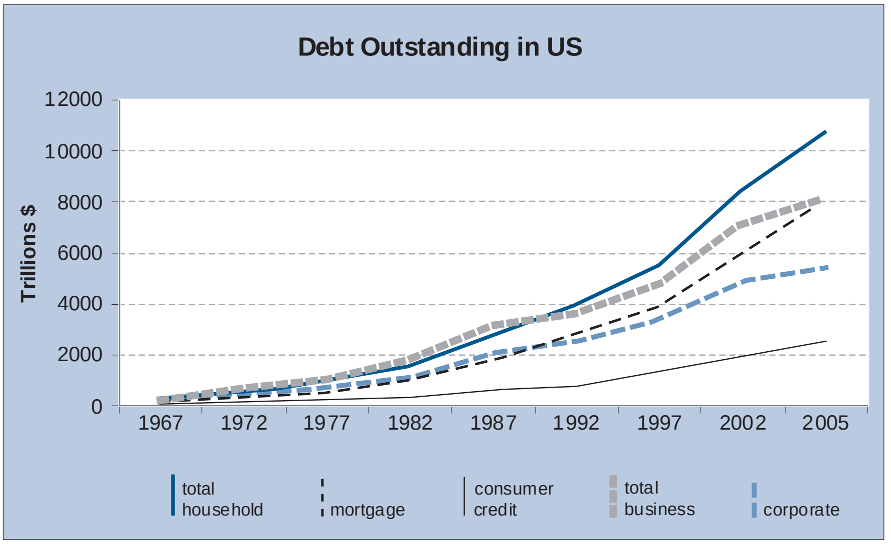
A forecast of whether you will default in the next year is being made every month by almost every organization that is lending money to you—banks, mortgage companies, credit card organizations, utility and insurance companies, and retail stores. The typical consumer is being scored more than one hundred times a year. The reason for this is the explosive growth in consumer credit. Figure 1 shows that (a) growth in the United States in lending to households overtook lending to businesses in the mid-1980s and (b) there is more money lent on housing than on corporate credit.

Until recently, the only aim of credit scoring was to support consumer lending decisions, and this has shaped its philosophy and methodology. Credit scoring was introduced originally to make consumer lending decisions more consistent and less affected by subjective bias. Moreover, an automatic method for assessing default risk became a necessity with the increase in the volumes of applicants that accompanied growth in consumer credit, particularly with the advent of credit cards in the late 1960s. The underlying philosophy was pragmatic, so that any

method that predicted well could be used. The underlying assumption was that this prediction would change little over time, so relationships that had been valid in the recent past would also be valid in the immediate future.

Two types of information are used in lending decisions. One is the information on the individual. For credit scoring, this is the information from the credit application form and also from credit reference bureaus. The second type is the mass of data on previous applicants, who could number in the millions. Credit-scoring techniques use this mass of information to identify the crucial features in the applicant's information. In behavioral scoring models, the individual's information shows how she or he has performed in the last year, and this rating is compared with the mass of data on the performance of other customers over a fixed time period, as well as their default status some time after that period.

Figure 1: US Household, Consumer and Business Debt Growth



Source: Federal Reserve Board

The Traditional Techniques of Credit Scoring

The objective of credit scoring is to assess a very specific default risk: the chance that an applicant will miss three consecutive payments in the next twelve months. The methodology of credit scoring is based on classification techniques. That is, the lender uses a sample of previous applicants and relates their subsequent default statuses to the information provided on their application forms and obtained from credit bureaus.

The statistical method that has served as the industry norm for classification is logistic regression, though linear regression, classification trees, and linear programming are also used (Thomas, Edelman and Crook, 2002). The differences among them are less important than the common ground. In all these credit-scoring methods, the following procedures apply:

- A sample of previous applicants is taken, which can vary from a few thousand to hundreds of thousands.
- For each applicant, credit performance in the first year or so is recorded, and it is judged acceptable (good credit risk) or unacceptable (bad credit risk).
- Also recorded are input characteristics—each applicant's credit bureau rating and information on his or her application form, such as age, income, professional status, years at present address, owner or renter, etc.
- The statistical procedure then utilizes the input characteristics to estimate a probability that the applicant will default on the loan.
- The lender chooses the cutoff score (i.e., the default probability below which applicants are accepted). The lender does so either subjectively or on the basis of evidence from a holdout sample of applicants that gives the default rate among applicants whose estimated default probability is below the cutoff.

The performance difference among the various statistical models does not seem significant. Baesens et al. (2003) made a detailed comparison of 17 methods on many combinations of credit data sets and classification measures, thereby determining the optimal method for each data set. They discovered, however, that the other methods did not often perform significantly worse than the optimal method.

Current Pressures

As we enter the second 50 years of credit scoring, we see pressure for changes in the forecasting approach to consumer lending. The pressure has three main sources.

First, the lenders themselves are changing the objective of the credit-lending decision from one of minimizing default risk to more general business objectives that create shareholder value in the firm. They are seeking to make decisions that maximize the profitability of a customer or an applicant. Doing so is more complicated than it might seem: there are many decisions that affect profit, such as which variant of a loan product to offer and which operating and marketing policies to apply to the borrower. It is not simply the accept/reject decision that default-based scoring supports.

Second, in Western countries, the market for consumer credit is becoming mature and, in some cases, almost saturated. Borrowers can often choose among many lenders willing to offer them loans. Third, the emergence of the Internet has made it a lot easier for borrowers to compare the terms of different loans. As a result, there is an emphasis by lenders on customer retention and new-customer acquisition.

Consumers are expecting more customization of their products in general, and loan products are no exception. Think of the different features one can have on a credit card: the overdraft limit, the interest rate charges, an air miles or bonus point scheme, cashbacks related to turnover, initial discount rates, free insurance or protection on certain purchases, foreign exchange or travel bookings, and card design.

The third source of pressure on traditional credit scoring comes from the New Basel Accord on banking supervision, which is due for implementation in 2007 or 2008. The Accord changes the regulations on how much capital a lending organization in the European Union must set aside to cover the unexpected risks inherent in its lending. Whereas the existing rules require banks to set aside a certain fixed percentage of each loan, the Accord allows an Internal Rating Based (IRB) approach where banks can use their own models to estimate the probability of default, as well as the monetary loss if there is a default for each segment of their loans. These estimates are inputs into a formula that defines the amount of capital to be set aside.

To implement the IRB approach in consumer lending, credit scoring is exactly what is needed. Credit scoring is an internal ratings-based approach that provides estimates of the probability of default. However, the Accord demands standards of accuracy in default-risk estimation, standards which put a different emphasis on how the performance of credit-scoring systems should be measured.

Assessing the Accuracy of a Credit-Scoring Model

Traditionally a credit-scoring model has been measured in terms of its predictive and its discriminative powers. Predictive power refers to the quality of the resulting decisions, which are functions of both the scoring system and the choice of a cutoff score. Statistical measures of predictive power are calculated from tables comparing actual default status against predicted status. For example, among borrowers who did default, what percentage had been predicted to do so? Discrimination looks at how well the score separates the good from the bad credit risks in terms of borrower rankings, and it uses statistical measures appropriate for rank ordering.

The Basel Accord, however, makes it necessary to calibrate the system's accuracy in estimating the probability of default; in doing so, it is stimulating the search for new ways to evaluate credit-scoring systems.

The three sources of pressure on lenders—the move from default scoring to profit scoring, the need to customize products, and the New Basel Accord regulations—are in different ways encouraging the use of new methods in credit scoring. One such method that is new to credit scoring but is well established in other forecasting areas is survival analysis.

1. Survival Analysis for Scoring Default Risk

In the survival analysis approach, the key question changes from "Which borrower will default?" to "When will a borrower default?" Survival analysis models the time until an event occurs. It was initially applied to mortality data, then in industrial engineering to the lifetime of equipment.

Today it is used in many medical applications. The key parameter is the hazard rate, the chance of the default event occurring in the next instant in time, given that it has not yet occurred. Proportional hazard models develop an estimate of the hazard rate on the basis of (a) the borrower's input characteristics and (b) a "baseline" hazard function, which describes the time until default for a borrower who has standard characteristics. It is obtained by filtering out individual differences in the default experience of previous applicants.

The advantages of modeling default risk in this way are considerable when compared to the standard classification approaches. There is no need to specify some arbitrary time horizon, such as defaulting within one year. Survival analysis can model the performance of the borrower over the whole duration of the loan, permitting much longer time intervals than the 12 months used to assess the borrower's status in the standard classification approaches. Also, in traditional credit scoring, much of the data become irrelevant as consumers cease borrowing for reasons other than default. With survival analysis, all the available data can be used.

Survival models are useful as well when the goal is not simply credit scoring but also profit scoring.

2. Profit Scoring

Profit-scoring models calculate the net present value (NPV) of expected future profits on a loan after allowing for the possibility of borrower default. So one needs to estimate when default will occur to calculate the profitability of the loan.

Profit scoring also needs to deal with competing risks, as different types of events can impact profit. Repayments can stop not only if the borrower defaults but also if he or she prepays or pays off the loan early because of a move to another lender. This is the main cause of unprofitable mortgage loans**.** For profit scoring, we can build a hazard model for prepayment or attrition (when the borrower moves to another lender). This new model can employ the very same data used by the hazard model for default risk.

Over long periods of time, economic conditions may vary, and this can affect the default probability. One can extend the survival-based models to incorporate time-varying economic variables and economic forecasts. The model will then estimate a risk score (hazard rate) that is a combination of the borrower's characteristics and the economic forecasts.

3. Acceptance Scoring

To customize their products for borrowers, lenders could use past data that shows which types of applicants accepted which types of products. They could then model how likely a new applicant is to accept a particular product. If there are only a few products and the customer is definitely going to choose one of them, discrete choice models are appropriate.

In the market for credit cards, the lender is offering only one product, a credit card, but there are a large number of possible variants of the product that can be offered, and the customer can decide not to accept any of them. Moreover, lenders tend to offer only one variant of the product; if the customer refuses that variant, the customer then tends to turn to another lender.

Consider the very simple case of a credit card where the only variable feature is the credit limit. (The analysis extends easily to as many features as are required.) The lender has to decide on the credit limit to offer and would like to determine what the minimum limit is that would entice the applicant to accept the card. This is the task of acceptance scoring.

Acceptance scoring is difficult to do from data on past applicants. If someone accepted a credit card with a credit limit of $2000, all we know is that his or her minimum acceptance level is below $2000. Similarly, if the prospective customer refused a credit card with a credit limit of $2000, all we know is that this minimum acceptance level is above $2000. The hazard-survival models cannot deal with situations where the threshold credit limit is unknown, but there is another survival analysis technique that can handle such situations. It is called the accelerated life model.

In this case there is again a "baseline" distribution that describes how likely the average person is to accept the offer for each credit limit. The characteristics of an individual are then combined into an acceptance multiplier. Someone with an acceptance multiplier of 2 will accept a credit card with a $2000 credit limit with the same probability as the average person accepts a credit card with a $1000 limit. Likewise, the individual will accept a credit card with a $10,000 (2 x $5000) credit limit with the same probability that the average person accepts one with a $5000 credit limit. In this way, one can estimate how likely the applicant is to accept the different types of credit cards offered. Hence the name *acceptance scoring*. The profitability of an applicant to the lender depends not only on the profit generated by each variant of the credit card but also on how likely the applicant is to accept that variant.

Low Default Portfolios in the New Basel Accord

One area of concern for those developing credit-scoring systems for use in the New Basel Accord is how to cope with low-default portfolios (LDP). In the relatively benign economic environment in some countries over the last ten years, there have been very few defaults in some types of consumer lending, especially lending for home purchases. Using a time horizon of one year, there would not be enough defaulted loans to build a robust model. So one has to use as long a data history as possible.

Allowing defaults at any time during the history introduces two new problems. First, loan and applicant characteristics change over time, and those characteristics that are more common in the newer loans might be considered superior simply because the newer loans have had less time to go bad. Second, estimating the default rate over 12 months, which is what the Basel Accord requires, is difficult if one has been using a sample in which most of the defaults occurred well after 12 months. Both these problems can be overcome if one uses a survival analysis approach rather than a fixed-time-horizon classification model.

Conclusion

Credit scoring is a major tool in forecasting financial risk. Once a lender starts using statistical or mathematical models to estimate risks of default, he never returns to judgment-based decisions. In this review I have outlined the pressures from customers, lenders, and regulators that have led to new approaches for credit scoring. I have described how one such approach, survival analysis, can be used for profit scoring and acceptance scoring, and also for dealing with problems raised by the New Basel Accord.

If these approaches are successful, there will be major impacts on the credit industry and on consumers as well. For the industry, those with the best models of consumer behavior will make the best profits and will have the appropriate level of regulatory capital set aside. Thus there will be strategic advantages in having models which best analyze the wealth of available data. Firms that are confident in their models will start cherry picking (going for the most profitable customers). The subsequent changes in pricing structures, led by risk-based pricing, will bring more diversity of loan products. There will be even more opportunities for the astute consumer to borrow at favorable rates, but there will also be an underclass of consumers who will be priced out of the market, though they might not realize this until they are heavily indebted.

The New Basel Accord has raised the profile of credit scoring among the leading banks as bankers came to realize how effective credit scoring can be in controlling consumer credit risk. However, the Accord also highlighted the current deficiencies of credit scoring models, in that they concentrate more on ranking customers than on accurately predicting default probabilities. Additionally, because credit scorecards do not make use of economic variables, they age very quickly. After 50 years of successful forecasting, credit scoring continues to evolve to meet these new challenges.

References

Baesens, B., van Gestel, T., Vianene, S., Stepanova, M., Suykens, J. & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54, 627-635.

Thomas, L.C., Edelman, D.B. & Crook, J.N. (2002). *Credit Scoring and Its Applications*. Philadelphia: Society for Industrial and Applied Mathematics (Siam).

Contact Info:
Lyn C. Thomas
University of Southampton, UK
L.Thomas@soton.ac.uk

SOFTWARE: SPOTLIGHT ON EXCEL

PREFACE

INCORRECT NONLINEAR TREND CURVES IN EXCEL by Rick Hesse

THE UNRELIABILITY OF EXCEL'S STATISTICAL PROCEDURES by Bruce D. McCullough

ON THE USE AND ABUSE OF MICROSOFT EXCEL by Paul J. Fields

PREFACE: Cautions In Using Excel For Data Analysis and Forecasting

According to major surveys of organizational forecasting practices, there is continued widespread use of spreadsheets for forecasting, despite major advances during the last 20 years in the availability, performance, and ease of use of business-forecasting software. In this software section, we examine the reliability and capability of Microsoft Excel as a statistical tool.

The section comprises three articles by analysts who have looked closely into the functionality of Excel:

1. Rick Hesse leads off the section with "Incorrect Nonlinear Trend Curves in Excel." Excel offers a number of choices, including linear and nonlinear functions, for fitting a trend line to data and projecting that trend forward. Rick shows that Excel calculates nonlinear trends in a "quick and dirty" manner, with results that can be off the mark, and he illustrates the correct approach for such calculations. Rick also enumerates ways in which Excel's Data Analysis Tools fail to provide an up-to-date menu of statistical routines.

2. In "The Unreliability of Excel's Statistical Procedures," Bruce McCullough documents serious flaws in Excel's statistical algorithms, and he advises that Excel should not substitute for a commercial statistical program. Bruce has done extensive testing of statistical algorithms in both professional statistical packages and in Excel. He reports that many Excel algorithms are faulty and that Microsoft's attempts to correct these errors have been far from successful.

3. In "On The Use and Abuse of Microsoft Excel," Paul Fields argues that while Excel has serious inadequacies as a statistical modeling tool, you do not need to throw it out. An analyst can still make safe and effective use of Excel—the keys are to learn what Excel can and cannot do, to master the skills to use it, and to install add-ins to enhance its capabilities.

Starting with these three papers, *Foresight* will pay close attention to Excel as a statistical and forecasting tool. Write to us about your own experience and assessment. We hope to publish an ongoing dialogue on forecasting with spreadsheets.

Len Tashman,
Foresight Editor

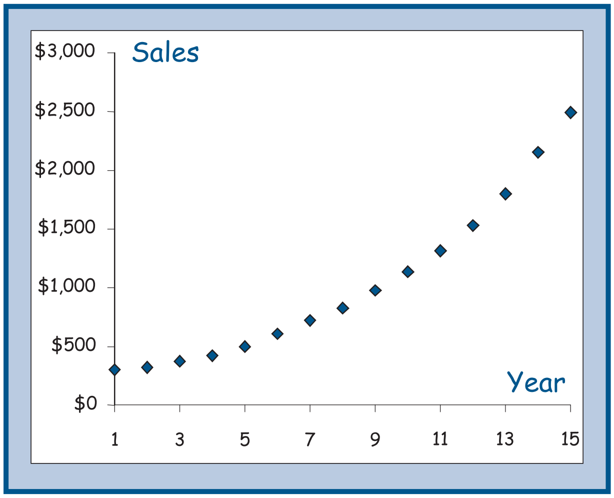
INCORRECT NONLINEAR TREND CURVES IN EXCEL by Rick Hesse

Preview: Many software programs, including Excel, make it easy to fit exponential trends (that is compound interest growth) to time series data. However, with Excel and some other products, there is a big problem: the exponential functions are done incorrectly because they use logarithmic transformations. Rick illustrates the right way to fit exponential trends, and he shows how misleading the Excel procedure can be.

Rick Hesse is Professor and Chair of Decision Sciences at Pepperdine University's Graziadio School of Business and Management. Since 1982 he has been writing a quarterly column in *Decision Line*, "In the Classroom," that has featured teaching tips mainly about spreadsheet use for solving quantitative models. Rick has written articles in *Interfaces, Operations Research, Decision Sciences*, and several textbooks. He has received numerous teaching awards, including the Outstanding Civilian Service Medal from the Department of the Army and Teacher of the Year at the San Diego State School of Business. He has served as a consultant for companies such as ITT, Pratt & Whitney, Semtech, US Airways, GEICO, UPS, and Bluebird.

- To calculate the coefficients of nonlinear growth curves, such as the exponential growth curve and the power curve, Excel does a "quick and dirty" fit based on log transformations. The results are often erroneous.
- A proper approach is a nonlinear fit, and the difference can be substantial.
- A proper nonlinear calculation is made by many forecasting programs, but it can also be implemented using the Solver function in Excel.
- More generally, Excel's Data Analysis Tools do not offer an up-to-date menu of statistical routines.

Figure 1. XY (Scatter) Chart for Data



Introduction

The exponential growth curve is a commonly used nonlinear function. When we say exponential growth, we mean a constant **rate** of growth. The exponential growth curve is equivalent to the return on principal at a compounded interest rate of i per year for n years [$R = P(1+i)^n$]. Consider the illustrative data in Table 1. When plotted in Figure 1, the exponential growth is apparent as a sales level that is increasing faster and faster.

In the "early" days of statistics, before calculators and computers, nonlinear calculations were so gruesome and abhorrent that shortcuts were often sought to ease the pain of calculation. Some statisticians thought that you could take the logarithms of the data, find a linear fit of the logs, and then convert back by putting the answers to the power "e." But doing this would be a very serious mathematical mistake because what minimizes the sum of the logarithms does not minimize the logarithm of the sum. Because logarithms are nothing more than exponents, transforming numbers to the logarithmic values is the same as dealing with the exponents of the data rather than the original data itself. It sounds and looks appealing, but the math just doesn't work out, and the errors are not truly minimized.

I noticed that many calculators had written instructions for doing this incorrect transformation, but many manufactur-

Table 1. Sales

Year	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Sales (k)	$301	$320	$372	$423	$500	$608	$721	$826	$978	$1,135	$1,315	$1,530	$1,800	$2,152	$2,491

ers have now eliminated these instructions. Having written articles about this very problem 20 years ago (Hesse, 1983 and 1987), I was surprised to find the problem persisting in Excel. What makes this error all the more serious is the ease with which users can simply click on the graph of the data to get an incorrect curve fit. Many good forecasting programs, such as SAS and SPSS, do not make this mistake, so obviously this problem of using log transformations has been recognized by some of our colleagues.

Here is an example that shows the correct fit and how it compares to the incorrectly fitted curve produced by Excel.

The Incorrect Procedure in Excel

The pattern of exponential growth, illustrated by Figure 1, takes the mathematical form $Y = ae^{bT}+c$, where b is the growth rate, a is the intercept, and c is the asymptote or lower limit. For now, we will assume that c = 0 (which Excel does). T is time (year).

A right click on any of the data points on the graph brings up the Trendline menu shown in Figure 2. The tab for options can also be clicked, and you may request the exponential trend formula on the graph.

The resulting trendline is reported in Figure 3 as $y = 237.5583 * e^{0.1561*T}$. The curve looks like a good fit, but looks can be deceiving.

Excel makes this calculation by taking the natural logarithms of the Y data, fitting a straight line regression, and then raising the parameters to the power "e," as shown in Figure 4. Notice that the exponent of 0.1561 (which implies a growth rate of 16.89% per year, calculated by raising "e" to the power .1561 and then subtracting 1) is the same in Figures 3 and 4. The multiplier of 237.5583 in Figure 3 is the intercept of 5.4704 in Figure 4 raised to the power "e." Thus Figure 4 illustrates that Figure 3 produces the incorrect values of the parameters for the exponential fit.

The Correct Procedure

How should the calculation be made? While correct solutions are available in many forecasting programs, I will use the Excel Solver function to show how it can be done. I compared the Solver solution with that from the SAS Gauss-Newton routine and found them to be the same.

With Solver, we can perform a *gradient search* for the optimal coefficients of a nonlinear model. Gradient search is analogous to finding the lowest point in a valley, but doing so in the fog. Imagine walking along the valley, always going downhill until you reach the bottom of the valley. If there is only one bottom point, we are assured that we will eventually find it (a global minimum).

Figure 2. Excel Trendline Menu

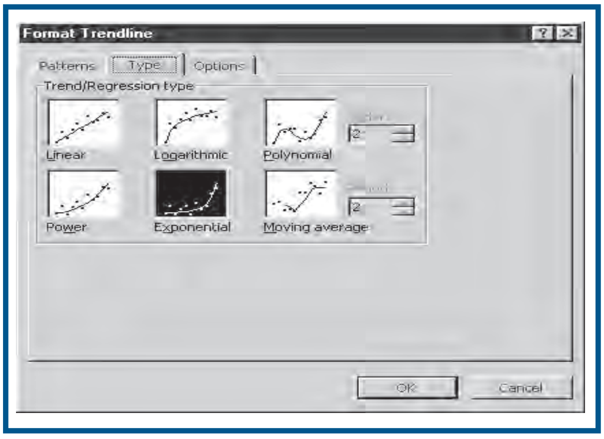


Figure 3. Excel Trendline and Formula

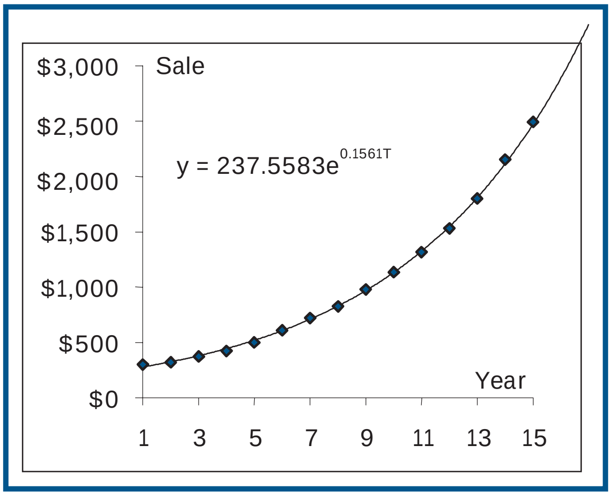
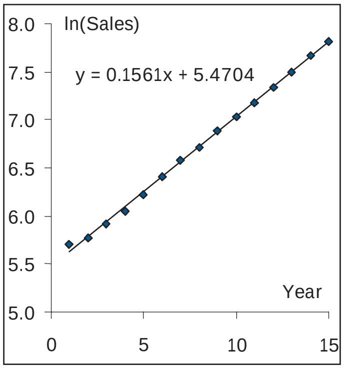


Figure 4. Derivation of Excel's Incorrect Exponential Fit

Year	ln(Sales)
1	5.707
2	5.768
3	5.919
4	6.047
5	6.215
6	6.410
7	6.581
8	6.717
9	6.886
10	7.034
11	7.182
12	7.333
13	7.496
14	7.674
15	7.820



The optimal coefficients in a model are those that minimize some model-error metric such as the root mean square error (RMSE). Finding the minimum value of the RMSE is similar to finding the valley bottom, except that this valley of errors can have some false bottoms—points near but not at the floor. To safeguard against a false bottom (a local minimum), we may need to either run the Solver a few times or use different starting points. Also, the results tend to vary in the 6th or so significant figure, depending on which version of Excel you use or which processing chip is on your computer. However, the results are repeatable within tolerance.

Solver is run by selecting Tools and then Solver from the Excel menu. You begin by inputting starting values of the intercept and slope coefficients a and b. I set the intercept equal to the mean of the sales figure and I set the slope equal to zero. I then asked Solver to find the values for a and b that would minimize the RMSE, whose formula is in cell C6. These optimal values are reported in cells B5 and B6.

Let us now compare the correct results from Solver with the wrong results from the Exponential Trendline fit, shown in Figure 5. The RMSE from the Excel Trendline result (16.360 in cell E6) is 13.4% higher (worse) than the Solver RMSE (14.425 in cell C6). The forecasts from the Solver start out a bit lower than the Excel forecasts because the intercept is lower, but they end up higher because the Solver estimate of the growth rate is higher. Notice that the fore-

Figure 5. Comparing the Exponential Curve Fits

	A	B	C	D	E	F
1	Nonlinear Exponential Regression Model $Y = ae^{bT}$					
2	Sales					
3		Solver		Excel		
4	Parameters	Regression	Solver	Regression	Excel	
5	a =	233.5189	RMSE	237.5583	RMSE	n
6	b =	0.1578	14.4249	0.1561	16.360	15
7		17.09%	Growth Rate	16.89%	13.4%	<== % Worse
8	Quarter	Sales	Solver	Solver	Excel	Excel
9	T	Y	Forecast	Error	Forecast	Error
10	1	$301	$273.44	($27.56)	$277.69	($23.31)
11	2	$320	$320.18	$0.18	$324.61	$4.61
12	3	$372	$374.91	$2.91	$379.45	$7.45
13	4	$423	$438.99	$15.99	$443.55	$20.55
14	5	$500	$514.03	$14.03	$518.49	$18.49
15	6	$608	$60	[illegible]	[illegible]	$1.91)
16	7	$721	$70	[illegible]	[illegible]	12.52)
17	8	$826	$82	[illegible]	[illegible]	2.17
18	9	$978	$96	[illegible]	[illegible]	$9.91)
19	10	$1,135	$1,1	[illegible]	[illegible]	3.36)
20	11	$1,315	$1,3	[illegible]	[illegible]	7.83
21	12	$1,530	$1,5	[illegible]	[illegible]	16.31
22	13	$1,800	$1,8	[illegible]	[illegible]	7.56
23	14	$2,152	$2,127.15	($24.85)	$2,112.93	($39.07)
24	15	$2,491	$2,490.77	($0.23)	$2,469.90	($21.10)
25	16		$2,916.53		$2,887.18	
26	17		$3,415.08		$3,374.95	

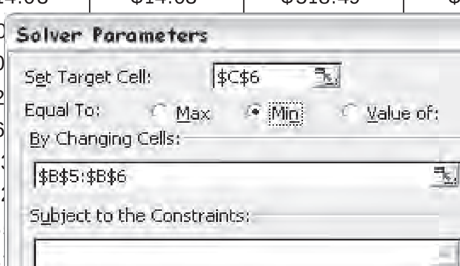


Figure 6. Exponential Fit with Constant "c"

	A	B	C	D	E	F
1	Nonlinear Exponential Regression Model $Y = ae^{bT}$					
2	Sales					
3		Solver		Excel		
4	Parameters	Regression	Solver	Regression	Excel	
5	a =	221.7886	RMSE	237.5583	RMSE	n
6	b =	0.1608	13.9369	0.1561	16.360	15
7	c =	19.9793			17.4%	<== % Worse
8	Quarter	Sales	Solver	Solver	Excel	Excel
9	T	Y	Forecast	Error	Forecast	Error
10	1	$301	$280.47	($20.53)	$277.69	($23.31)
11	2	$320	$325.93	$5.93	$324.61	$4.61
24	15	$2,491	$2,495.98	$4.98	$2,469.90	($21.10)
25	16		$2,928.05		$2,887.18	
26	17		$3,435.52		$3,374.95	

casts for time periods 16 and 17 are moving up faster than the Excel forecasts.

An even better fit results when the formula allows the Solver to find the best asymptote (lowest value), instead of assuming that c = 0 for the nonlinear equation $Y = ae^{bT}+c$. The Solver will now find the optimal values of a, b, and c.. Figure 6 shows that the fit is a = 221.7886, b = 0.1608, and c = 19.9793, with an RMSE = 13.937, which means that Excel's RMSE is now 17.4% worse than the optimal RMSE. Again, the Solver forecasts rise faster for periods 16 and 17 than the Excel forecasts.

Other Nonlinear Functions

The exponential trend is not the only nonlinear function for which Excel uses the erroneous log transformation. The same problem occurs with the power curve, which is also shown as an option in Figure 2. The power curve has the algebraic form of $Y = aX^b + c$. When X stands for time, you can recognize this as Moore's Law. Gordon Moore, co-founder of Intel, predicted in 1965 that components on a computer board would double every year or so. He thought this would hold true for another 10 years, but 40 years later it still holds true.

For both the exponential and power curves, some statisticians will tell you that taking the logs of the data works well when the data is "well behaved." It's true that when the data almost exactly fits an exponential or power curve, taking the logs works pretty well, but forecasters hardly ever have data that well behaved. Of course an incorrect method works if there are few or no errors!

Just recently, one of my in-career MBA students switched jobs within the electronics industry, and he had to implement a Weibull reliability curve fit (another type of nonlinear curve). His new company lacked the sophisticated software that found the true fit, so he was forced to rely on logarithms, and something went very wrong. Fortunately, he corrected the errors by using Solver in the manner I have shown, and he was able to come up with the correct results in time for his presentation. A few weeks later, his company obtained the expensive curve-fitting software, and he verified that the Solver results were on the mark.

Other Concerns With Excel

The statistical part of Excel—the so-called Data Analysis Tools (DAT)—has always been a weak link. It was a weak set of routines when created in 1995, and it has not been updated since, even though Microsoft has come out with at least four new Excel versions in ten years.

Here are a few of my complaints:

1. Some DAT routines are not interactive. True, if we change a data point in Figure 4, the trend line and equations are automatically recomputed. But for a histogram, ANOVA, or t-tests, the results are hard coded: changing data does not change the results.

2. Some functions require rigid formatting of the data. For example, to obtain a histogram, you must put the data in one column, and you cannot use the rectangular area identified by the user. For a regression, the X variables must be in contiguous columns.

3. Certain calculations can be made only from the original data, not from intermediate calculations. For example, you cannot do a t-test by inputting the sample mean and standard deviation.

4. Certain functions are incorrect. For example, the RANK command does not properly account for ties.

5. Exploratory data analysis graphs are not offered. It would not be that difficult to develop a Box and Whiskers plot using quartiles and then include it in the Chart Gallery.

Other users have registered similar complaints over the years about statistical functions and accuracy in Excel, but, according to McCullough and Wilson (2005), Microsoft is impervious to these complaints.

Conclusion

Microsoft Excel is giving incorrect nonlinear fits for both exponential and power curves. The error stems from the use of the logarithmic transformation—taking logarithms to transform the data, then fitting a linear model, and finally retransforming the results to the original data. Microsoft should at least warn the user by identifying the exponential and power curve trend lines as "quick and dirty." What is more difficult is convincing people not to use these options for nonlinear fits.

References

Hesse, R. (1987). Son of log transformation or return of the living dead. *Decision Line*, *18*(1).

Hesse, R. (1983). Too quick and too dirty: Least squares for exponential curves. *Decision Line*, *14*(3).

McCullough, B.D. & Wilson, B. (2005). On the accuracy of statistical procedures in Microsoft Excel 2003. *Computational Statistics and Data Analysis*, *49*(4), 1244-1252.

Contact Info:
Rick Hesse
Graziadio School of Business and Management
Pepperdine University
R.Hesse@Pepperdine.edu

THE UNRELIABLITY OF EXCEL'S STATISTICAL PROCEDURES

by Bruce D. McCullough

Bruce McCullough is on the faculty of Decision Sciences at Drexel University in Philadelphia and is Software Editor for the International Journal of Forecasting. He has written extensively on the accuracy of statistical and econometric software, and his paper "Is it safe to assume that software is accurate?" won the Best Paper 2000-2001 award in the International Journal of Forecasting.

Introduction

In the small world where computer science overlaps with statistics, it was well known that Microsoft Excel was riddled with statistical errors. It was so well known that no one bothered to write about it. In the larger world, however, it remained Microsoft's dark secret. Professional statisticians wrote textbooks with titles like "Statistics with Excel," and a generation of students learned to do statistics with Excel. "Surely," the student reasoned, "it is safe to use Excel for statistics. If it weren't, my professor would have chosen a different software package." So these students went on to use Excel in the business world. It is quite conceivable that more statistical calculations are performed in Excel than in any statistical software package.

Testing the Accuracy of Statistical Software

Several years ago, I developed a methodology for testing the accuracy of statistical software (McCullough, 1998 and 1999), and I applied this method to some major statistical packages, including SAS, SPSS, and S-Plus. I found a few errors in each of them (McCullough, 1999). A coauthor and I applied the same methodology to Excel 97 (McCullough and Wilson, 1999), and we found numerous errors. So egregious were these errors that we advised people who conduct statistical analyses of data not to use Excel.

The scope of these errors is not minor. My methodology analyzes three areas: random number generation, estimation (which has four components: univariate, ANOVA, linear regression, and nonlinear regression), and statistical distributions (for example, tabulating the normal distribution or calculating *p*-values). Excel failed in all three areas.

In the estimation area, we found Excel wanting in all four components. When we applied Excel Solver to 27 problems in the nonlinear least squares regression suite, Solver gave incorrect answers 21 times. In fact, it missed completely 21 times. For example, it returned a coefficient of 454.12 when the correct answer is 238.94. Rick Hesse and others have found errors in specific functions that I did not examine, such as the LINEST, TREND, LOGEST, and GROWTH worksheet functions.

Microsoft's Track Record

It's not as if Microsoft would have to develop new algorithms to solve these problems. For most of the inaccuracies, good algorithms have already been developed and are well known in the statistical community. Microsoft simply used bad algorithms to begin with, and it never bothered to replace them with good algorithms. Revision after revision, in Excel 4.0, Excel 5.0, Excel 95 through Excel 97 and beyond, Microsoft has allowed the errors to persist—unbeknownst to its legions of users.

So unbelievable was Microsoft's cavalier attitude toward accuracy that I came to believe (McCullough, 2002) the company might be catering to a demand for inaccurate statistical software. There is simply no other way to explain Microsoft's lack of response. Contrast Microsoft's behavior with that of a responsible software company such as SAS. When SAS becomes aware of an error, it publishes the error on its Web site, often with a workaround, so that users can avoid the problem. SAS fixes the problem quickly, often by the next minor release, and almost always by the next major release. And SAS fixes problems correctly.

In its Excel XP release, Microsoft attempted to fix some statistical problems, but it did not do a good job

(McCullough and Wilson, 2002). This failure presaged Microsoft's attempt at a major overhaul with Excel 2003. While it fixed many functions, it failed to fix many others.

Perhaps most embarrassing was Microsoft's attempt to install a new random number generator (RNG). In its natural state, the RNG should produce numbers between zero and one. Microsoft chose a very well-known RNG (called the Wichmann-Hill RNG), but could not make it work right: Excel would occasionally spit out negative numbers. What makes this so embarrassing is that the source code for this algorithm is very easy to obtain. Hence it is fair to say that Microsoft did not correctly implement an algorithm for which source code is widely available. Nor did it do adequate testing before releasing the product. In our analysis of Excel 2003, we wrote that "Excel 2003 is an improvement over previous versions, but not enough has been done that its use for statistical purposes can be recommended" (McCullough and Wilson, 2005, p. 1244). Assuming that Microsoft will make another attempt to fix Excel, given Microsoft's track record, it will not be enough for the company to say that it has "fixed" errors. Microsoft will have to prove that it has fixed them correctly.

Warnings, Faults, and Workarounds

Professional statisticians continue to write books with titles like "Statistics with Excel," but they now warn students not to bet their jobs on Excel's accuracy. They advise students to use a real statistical package when they need to do statistics.

If Dante had to conjure a new circle for the 21st century, it would contain persons condemned to do statistics with Excel. What are these poor, unfortunate souls to do? To their succor has come a retired engineer who, in a tour de force, has catalogued Excel's statistical errors and offered many workarounds. These can be found at David A. Heiser's Web site entitled "Microsoft Excel 2000 and 2003: Faults, Problems, Workarounds, and Fixes," which is located at

http://www.daheiser.info/excel/frontpage.html

In this issue of *Foresight*, Rick Hesse provides another example of Microsoft's decision to use a bad algorithm and its refusal to fix this problem over the years. Fortunately for those who have to use Excel, Professor Hesse also provides a workaround. Note that while Professor Hesse does use Excel Solver, he has verified the results using SAS.

References

McCullough, B. D. (2002). *Proceedings of the 2001 Joint Statistical Meeting* [CD-ROM]: *Does Microsoft fix errors in Excel?* Alexandria, VA: American Statistical Association.

McCullough, B. D. (1999). Assessing the reliability of statistical software: Part II. *The American Statistician, 53*(2), 149-159.

McCullough, B. D. (1998). Assessing the reliability of statistical software: Part I. *The American Statistician, 52*(4), 358-366.

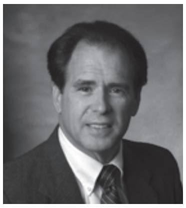
McCullough, B. D. & Wilson, B. (2005). On the accuracy of statistical procedures in Microsoft Excel 2003. *Computational Statistics and Data Analysis, 49*(4), 1244-1252.

McCullough, B. D. & Wilson, B. (2002). On the accuracy of statistical procedures in Microsoft Excel 2000 and Excel XP. *Computational Statistics and Data Analysis, 40*(4), 713-721.

McCullough, B. D. & Wilson, B. (1999). On the accuracy of statistical procedures in Microsoft Excel 97. *Computational Statistics and Data Analysis, 31*(1), 27-37.

Contact Info:
Bruce D. McCullough
Department of Decision Sciences
and Department of Economics
Drexel University
bdmccullough@drexel.edu

ON THE USE AND ABUSE OF MICROSOFT EXCEL by Paul J. Fields



Paul Fields, PhD, is on the faculty at Brigham Young University in the Department of Statistics. Prior to returning to academia, he worked in industry, where he developed multistage demand-forecasting techniques for production planning and inventory management of perishable goods. His research interests center on forecast-error variance reduction and forecasting method selection.

Introduction

A recent survey of 240 U.S. corporations (Sanders and Manrodt, 2003) found that 48 percent used a spreadsheet program instead of specialized forecasting software to make their forecasts. Surprisingly, only 11 percent used forecasting software, even though forecasting packages typically produce better results. For example, companies using forecasting software reported nearly 7 percent lower average errors than spreadsheet users.

The study also found that 85 percent of the respondents considered "easy to use" and "easily understandable results" the most important features of a forecasting system. Interestingly, these are the features many people find to be the most attractive attributes of a spreadsheet. Therefore, we might infer from the survey results that when companies choose to forecast with a spreadsheet instead of forecasting software, they are trading somewhat decreased forecasting performance for a spreadsheet's ease of use.

In light of these findings, we need to consider carefully the proper use and possible abuse of a spreadsheet program in forecasting. As with any tool, a spreadsheet program such as Microsoft Excel can be used either properly or improperly. To use it properly, you first need to recognize what Excel is and what it is not.

What Excel Is

Excel is simply an advanced scientific calculator with the data visible and the formulas accessible for examination. In this regard, Excel is not a "black box," as a statistical software package can be. The internal operations of a software program are often opaque to the user. With Excel, it is a tremendous advantage that you can see what is going on in the calculations. This feature is especially helpful the first time you are working with a data set, or when you are reviewing someone else's work.

As a spreadsheet program, Excel has an advantage over a calculator in that the worksheets can be saved and reused. You can solve a problem once and then reuse the solution when faced with that problem again. This can save you a great deal of time and effort. As we know, time is money! Another way to view Excel is to recognize that it is also a "communication facilitator." Thanks to Microsoft, Excel is on 90 percent (or some high percentage) of computers in the world. Because Excel is so ubiquitous, it is easy to exchange data using Excel. It is easier and somewhat less prone to error to use its standard format when exchanging data. This applies to exchanging information with other software packages as well. With Excel, you can easily share the results of your analysis in a format that others can recognize and use. Excel enables us to speak the same computing language.

What Excel Is Not

Excel is not a statistical software package, and it is in no way a substitute for one. Bruce McCullough showed us in the previous article that some of the statistical functions in Excel are unreliable. Therefore, when reliable statistical computations are needed, i.e., when there is money on the line, you should be cautious in using Excel's built-in functions. It is better to write your own equations into a worksheet to perform the calculations you need. If you have programming skills, you can write macros to automate the computation routines to do the calculations correctly.

Good Practice

To avoid abusing a tool, good practice is to *use the right tool for the right job*. Here are three helpful guidelines.

1. Match the tool to the job.
2. Know how to use the tool for each job.
3. Increase the capabilities of the tool to do more jobs and do them better.

Let's apply these guidelines to forecasting with Excel.

The Tool and the Job

Excel is a way to get a quick and dirty answer. When you need a better answer, you should use a bona fide statistical package. Consider an analogy: it is efficient to use a rifle when a rifle will do the job, and to use a cannon when a cannon is needed. For many problems, a "computational .22" like Excel is adequate. However, you should bring out the big guns when you need real computing power. In that case, you probably will need to enlist the assistance of someone who knows how to shoot cannons—a statistician. But when a cannon is not necessary, you can stick to a small-caliber gun like Excel.

Excel can also be very helpful when you are learning to do quantitative analysis. It can help you learn new skills. Excel can be a good interim step before you learn to use a full-blown statistically based forecasting package.

Skills with the Tool

Besides the unreliability of the computations in Excel, Microsoft's other "error" with Excel is in making it too easy for people to do quantitative analyses they do not fully understand. The trend-line fitting routines, for instance, are simplistic, which lays a trap for naïve users to misuse the routines. Users may think they are doing an analysis properly when in fact they are unknowingly applying the wrong model to a data set. Rick Hesse provides an excellent example of such a problem in this section of *Foresight*.

Modeling skill is a prerequisite to using any modeling tool. Consider another analogy: you would not jump into your car and expect to drive properly without taking driver's training. The Help function in Excel is no more a way to learn to forecast than the owner's manual in your car is a way to learn to drive. Similar to the owner's manual, Excel's Help function shows you where the controls are; you already need to know how to drive the car. If you do not, you are likely to crash.

To optimize your use of Excel, it is best to learn to build worksheets with equations you input yourself. A worksheet is well suited for time-series forecasting analysis because the consecutive nature of time series data is preserved and can be exploited in the "columns and rows" structure of a worksheet. By learning to build your own forecasting worksheets, you can reliably use Excel to facilitate your forecasting process.

If you also learn to use Microsoft's Visual Basic for Applications (VBA), you can use Excel's VBA capability to write macros to automate complex forecasting computations. This will further extend your ability and open up new vistas of forecasting possibilities.

Increase the Tool's Capabilities

In addition to your own customized worksheets and macros, Excel add-ins can greatly expand Excel's capabilities. For example, as with other built-in functions, the exponential smoothing routine in Excel is very basic and is useful only for simple problems. Consequently, in practice it is inadequate. Excel's capabilities are greatly enhanced with commercial add-ins that can perform many of the commonly used forecasting methods. This way, Excel's ease of use is maintained, the reliability of its computations assured, and its range of capability increased.

What a Forecaster Can Do

Although Excel has some inadequacies, it can still be used for forecasting if you perform the following tasks.

- Build worksheets with correct forecasting equations.
- Write macros using reliable algorithms for forecasting.
- Buy proven forecasting add-ins from a reputable software developer.

You do not need to throw out Excel when forecasting. You *can* have your cake and eat it, too! As you would with any tool, learn what Excel can and cannot do, master the skills to use it, and finally install add-ins to enhance its capabilities. In this way, you can have reliable forecasts along with the attractive features of a spreadsheet.

Reference

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

Contact Info:
Paul J. Fields
Brigham Young University
pjfields@stat.byu.edu

BOOK REVIEW

Dow 36,000: The New Strategy for Profiting from the Coming Rise in the Stock Market
reviewed by Roy Batchelor

Roy Batchelor is HSBC Professor of Banking and Finance at Cass Business School, City University of London, and visiting professor of finance at a business school in Paris. He has worked as a government scientist, modeling and predicting stream flow in rivers; at the National Institute of Economic and Social Research, forecasting the UK economy; and at a London stockbroker, where he tried to predict the stock market. He has published widely in applied economics and finance and is co-editor with Pami Dua of *Financial Forecasting* (Edward Elgar, 2003).

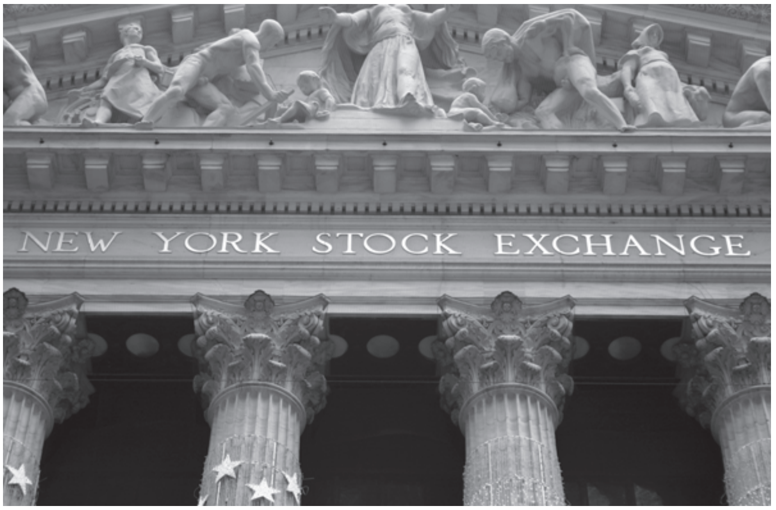
■ James Glassman and Kevin Hassett (1999).

Dow 36,000: The New Strategy for Profiting from the Coming Rise in the Stock Market. New York: Crown Business. ISBN: 0812931459.

It's not often that I review a book that was published five years ago and is currently remaindered at 20 cents on Amazon.com. The forecast in this book is hard to miss—it's right there in very big print on the front cover—and Glassman and Hassett had the courtesy to give a date as well as a number. Writing in late 1999, when the Dow Jones Industrial Average was just breaking through the 10,000 barrier, they wrote that "A sensible target for Dow 36,000 is early 2005, but it could be reached much earlier" (p. 140). The forecast error is pretty easy to spot. Early 2005 has come and gone, and with the Dow still hovering around 10,000 we know that the authors were overoptimistic to the tune of 260 percent. This is a big error in anyone's book, one that deserves some kind of autopsy.

Where did this forecast come from? Recall that by 1999 the U.S. economy had been recession free for almost a decade. There was serious talk of a "New Economy" in which globalization, technical advances, and telecommunications would permanently raise economic growth. The expansion of the Internet enabled small investors to acquire financial information easily and to deal cheaply. There was a great demand for popular treatises on how the financial markets work, particularly for books that gave investors reasons to BUY. Enter *Dow 36,000*, a well-written oeuvre aimed at a general audience. The book features helpful closing chapters explaining exactly how to get started in buying shares. And enter alongside it a whole raft of gee-whiz manuals touting flaky valuation models aimed at helping the new generation of investors put positive valuations on a new generation of businesses, many of which had no prospect of ever rolling into profit. It was seriously suggested, for example, that in the absence of any cash flows, Internet companies be valued in terms of their hit rates or advertising content.

So funds poured into the technology and telecommunications companies that were supposed to drive the new economy. The NASDAQ index that contains many of these stocks grew particularly strongly, and even the "Old Economy" Dow grew fivefold between 1989 and 1999, for an average annual growth rate of over 15%. Even so, extrapolating the trend in the Dow through 2005 would bring it to about 25,000, nowhere near the Glassman and Hassett figure. Almost all commentators on the 1990s' stock market boom, and on the subsequent tech-stock-driven boom and bust, attribute most of the action to a "rational bubble." Even if a price is rising for no good reason, it is perfectly reasonable—although risky—to buy in expectation of further gains. The buying makes prices go up until smart investors start selling.



What is remarkable about the Glassman and Hassett argument is that their forecast is not based on the continuation of a bubble, or on any flaky valuation model. Their forecast instead comes from traditional methods of equity analysis in terms of the "fundamental value" of a share, or what they call the "Perfectly Sensible Price" or PSP. The fundamental value of a company is the value in today's money of the expected flow of dividends and other cash payments to shareholders. Suppose that a company is expected to pay a dividend of $4 per share in a year's time, growing in line with earnings at five percent per year thereafter, and that investors require a return of nine percent per year to compensate for the risk of the share. In this case, the fundamental value of the share is $4/(9%-5%) = $100, or 25 *times* the dividend. Note that share prices are very sensitive to changes in the denominator of this calculation. A fall of 3 points in the required return on the share (in our example, from 9% to 6%) would lead to a fundamental value of $4/(6%-5%) = $400, implying a fourfold rise in the share price to a level that is 100 *times* the dividend.

Glassman and Hassett argue that the rapid rise in the Dow through the 1990s was caused by a fall in the equity-risk premium component of the required return, as investors discovered that shares were not as risky as previously thought. By 1999 the authors reckoned that only part of this adjustment had occurred, and that the Dow was indeed heading to a level of about 100 *times* dividends. This explains their target of 36,000 and their advice to buy and hold equities for the long term.

The book sold well to the investing public, and Glassman, who is a stock market columnist for the *Washington Post*, made regular appearances on U.S. television to promote and defend his forecast. However, the book reviewed very badly, at least by people who really understood the theory. Exchanges between Glassman and academics like Paul Krugman (a Princeton professor and New York Times columnist) became vituperative in the extreme.

A major problem is that in the course of the book Glassman and Hassett slide from a prediction that the price/dividend ratio should be 100 to the prediction that the price/earnings ratio should be 100. Half or less of earnings are paid out in dividends, so perhaps the book should have been called *Dow 18,000*.

Another problem is that in the authors' rapid growth scenario, the risk premium on equity would not remain low. We saw above that when earnings growth is close to the discount rate, share prices are very sensitive to small changes in interest rates and expectations about future earnings. Investors would certainly require compensation for this increased risk.

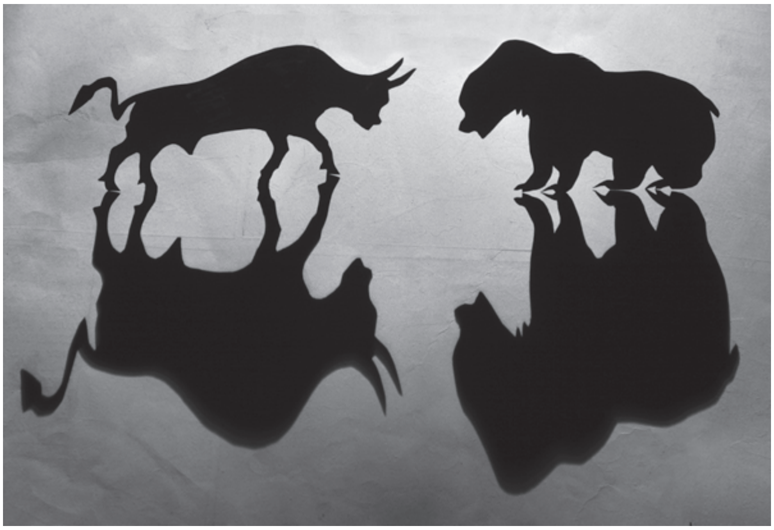
In any case, the discounted dividend model starts to lose plausibility at low discount rates, which imply that a large fraction of the share's current value is due to dividends expected over 20, 30, and 40 years. Professional analysts can barely beat naïve models in one- and two-year-ahead forecasts. Earnings 20 years ahead are anybody's guess, though it is a safe bet that many of the companies currently prominent in the index will not then exist in their current forms.

Nor do the authors address the wider economic implications of *Dow 36,000*. Alan Greenspan set his face against raising interest rates pre-emptively to puncture the stock market bubble in 2000, preferring instead to talk expectations down. Even so, the Fed Funds rate was gradually edged up from 5% to almost 7% between late 1999 and mid 2000, and if the stock market had continued to rise, and household wealth and spending had continued to increase, the Fed would have been forced to raise interest rates yet more aggressively.

Finally, we can now see that the New Economy has not materialized, so the earnings estimates in Glassman and Hassett have proved wildly optimistic.

Academics were rankled by the idea that the workhorse model of modern finance theory had been hijacked in the interests of creating controversy and selling a book. [Note to the editor: Can you *hijack* a horse? Or *rustle* a model?] Anyhow, the dividend discount model is common intellectual property, and in 1999 there were plenty of other analysts using it to call the U.S. stock market. Most of these observers concluded that the U.S. market was overvalued by 20% to 40%. For a clear and sensible use of the model by a respected analyst who has been publishing forecasts for many years, look at the stock valuation charts on Ed Yardeni's Web site: www.yardeni.com. The model suggests that the bubble had burst and the Dow had reverted to fundamental value by 2002. It also suggests that the market is currently 20% undervalued, but don't bet on it.

Unabashed, Glassman published another book in 2002, *The Secret Code of the Superior Investor*, which argued that the Dow had bottomed out. He has not reinstated his 36,000 target. Mind you, he is not the only popular writer abusing economic theory to produce headline-grabbing forecasts. Should you, despite my earnest advice, feel bound to buy *Dow 36,000* at its current bargain-basement price, I recommend that for balance you read almost any of the doom-laden, best-selling works of Ravi Batra—for example, *Great Depression of 1990* (written in 1987 and never recanted), *Surviving the Great Depression of 1990* (written in 1989 and also never recanted), or the more recent *Crash of the Millennium: Surviving the Coming Inflationary Depression* (written in 1999).



Contact Info:
Roy Batchelor
Cass Business School
City University of London
r.a.batchelor@city.ac.uk

The International Institute of Forecasters Certificate of Forecasting Practice

The International Institute of Forecasters (IIF) is launching a certification program. The IIF invites potential course providers to submit appropriate courses for evaluation by the IIF; a successful validation will allow the course provider to award an IIF Certificate of Forecasting Practice on successful completion of the course.

- The course should consist of approximately 200 study hours (total student time commitment).
- The provider of the certificate should pay the IIF a fee ($200) for each person registering for the certificate.
- The provider should describe the entry standards into the proposed certificate course.
- Assessment may be through a variety of mechanisms and need not necessarily depend on formal examination.
- Teaching of the material may include a variety of delivery modes and should include project work and case studies.
- The course must cover the following topics:

  TOPIC
  Introductory data collection & analysis
  Basic Statistics
  The Organization and Management of Forecasting
  Forecasting the economy and its impact on the firm
  Extrapolation methods
  Introductory econometrics
  Judgmental approaches
  Choosing between forecasting methods
  Combining forecasts

- An illustrative list of topics that could be included in a certificate follows:

  TOPIC
  Introductory marketing and economics
  Scenario forecasting
  Advanced time series
  Advanced econometrics
  Financial forecasting
  New product forecasting
  Market research (qualitative)
  Qualitative forecasting methods
  Information systems and data bases
  Macroeconomic Forecasting

think forecasting potential.

In order to gain IIF Certification, course providers should send full course details to:
P. Geoffrey Allen
Secretary-Treasurer of the IIF
Department of Resource Economics
University of Massachusetts
80 Campus Center Way
Amherst, MA 01003-9246 USA
Phone: +1 (413) 545-5715
Fax: +1 (413) 545-5853
allen@resecon.umass.edu

KEYNOTE AND FEATURED SPEAKERS
Greg Allenby
Ohio State University, USA
Chris Chatfield
University of Bath, UK
Robert Engle
New York University, USA
Ronald Lee
University of California, Berkeley, USA
George Tiao
University of Chicago, USA
Peter C. Young
Lancaster University, UK
Arnold Zellner
University of Chicago, USA
GENERAL CHAIR
Antonio García-Ferrer
Universidad Autónoma de Madrid
e-mail: antonio.garcia@uam.es
PROGRAM CO-CHAIRS
Pilar Poncela
Universidad Autónoma de Madrid
e-mail: pilar.poncela@uam.es
Esther Ruiz
Universidad Carlos III de Madrid
ortega@est-econ.uc3m.es
COMMITTEE
Antonio Aznar
Universidad de Zaragoza
José Luis Gallego
Universidad de Cantabria
Agustín Maravall
Banco de España
Elías Moreno
Universidad de Granada
Alfonso Novales
Universidad Complutense
Daniel Peña
Universidad Carlos III
Gabriel Perez-Quirós
Banco de España
Universidad de Alicante
26th International Symposium on Forecasting
JUNE, 11-14 | 2006
WWW.ISF2006.ORG
PALACIO DE LA MAGDALENA
Forecasting challenges in a changing world environment
SANTANDER | SPAIN
IIF
THE INTERNATIONAL INSTITUTE OF FORECASTERS

FORESIGHT: The International Journal of Applied Forecasting
140 Birchwood Drive
Colchester, Vermont 05446
USA