Golden Rule of Forecasting: Be Conservative

11 July 2013
GoldenRule 188-clean2.docx

Working Paper Draft

J. Scott Armstrong
The Wharton School, University of Pennsylvania, Philadelphia, PA, USA
and Ehrenberg-Bass Institute
armstrong@wharton.upenn.edu

Kesten C. Green
University of South Australia School of Business
and Ehrenberg-Bass Institute, Adelaide, Australia
kesten.green@unisa.edu.au

Andreas Graefe
Department of Communication Science and Media Research
LMU Munich, Germany
a.graefe@lmu.de

Abstract

The Golden Rule of Forecasting is to be conservative: Forecasts should be consistent with what is known about the past. That is, forecasters should be guided by cumulative knowledge about the situation and by knowledge about forecasting methods validated for the situation. Forecasts from conservative procedures were more accurate than forecasts from less conservative procedures under all conditions in prior studies. Conservatism in forecasting thus has the status of a rule. Gains from conservatism are greater when the situation is uncertain and complex, and especially when bias is likely. Conservative procedures are simple to understand and implement. They can, however, be expensive because forecasters need to acquire comprehensive knowledge about the situation. Complex statistical analyses and access to large databases encourage forecasting without prior knowledge. Forecasters may also be motivated to produce unconservative forecasts to serve political ends, or for publicity. The Golden Rule Checklist can help forecasters to follow accepted evidence-based conservative forecasting practices and thereby improve the accuracy of their forecasts. People who are not forecasting experts can readily understand the guidelines and use them to identify doubtful forecasts.

Key words: accuracy, big data, causal forces, causal models, combining, complexity, contrary series, damped trends, decision-making, decomposition, Delphi, ethics, extrapolation, inconsistent trends, index method, judgmental forecasting, judgmental bootstrapping, regression, risk, simplicity, structured analogies.

Please send suggestions for change, especially relevant papers (by you or others) that we have overlooked in this version.

Introduction

Imagine that you are a manager with a big database who hires a consultant to predict profitable locations for stores. The consultant responds with a model that uses the latest statistical techniques. You do not understand the model; its variables seem ridiculous, and the forecasts absurd. Why? Complex statistical procedures applied to non-experimental data on complex situations are unlikely to identify causal relationships. Human knowledge and reasoning, and experimentation, are needed.

This review of evidence suggests that forecasters should follow the *Golden Rule of Forecasting*: Be conservative. Conservatism in forecasting is the expectation that the future will be much like the past and that future changes will be circumscribed by past changes. Conservatism requires a valid and reliable assessment of the situation, use of cumulative knowledge about causality, and the application of appropriate validated forecasting procedures. In short, know the past and present to predict the future, and meet claims that things are different now with skepticism.

The Golden Rule is relevant to all forecasting problems. The rule is especially important when bias is likely, and when the situation is uncertain and complex. Such situations are common in public policy, as with forecasts of the economy, environmental impacts, and expenditures on welfare, transportation, education, and medicine. Situations in businesses, too, can be complex and uncertain, as with new products and competitive strategy, and firms may face optimistic forecasts to promote investments or a vested interest, or pessimistic forecasts to increase the chances of good news.

This article provides guidance on being conservative in formulating forecasting problems. It then describes evidence on the application of the Golden Rule to judgmental, extrapolative, and causal methods. Finally, it describes how combining forecasts from different methods produces conservative forecasts.

The Golden Rule Checklist in the Exhibit can help forecasters to be conservative when making forecasts. Some of the guidelines in the Checklist are inherently conservative and some encourage the use of conservative procedures. The Exhibit also shows the ranges of error reductions that have been achieved by following each of the guidelines. While the evidence is strong, it is not based on an exhaustive review. Moreover, knowledge on some of the guidelines would benefit from replications.

Exhibit: Golden Rule Checklist
(With evidence on percentage error reduction)

Guideline	*Done or N/A (✓ or ✗)*	*% error reduction*
1. Problem formulation		
1.1. Obtain and use all knowledge		
1.1.1. Obtain all relevant information and sufficient understanding	☐	
1.1.2. Decompose the problem to best use knowledge and judgment	☐	(27–51)
1.1.3. Use evidence-based forecasting methods validated for the situation	☐	(88)
1.2. Avoid bias		
1.2.1. Specify multiple hypotheses or conceal the purpose of the forecast	☐	
1.2.2. Obtain signed ethics statements before and after forecasting	☐	
1.2.3. Structure adjustments for important knowledge outside the model	☐	
1.3. Provide full disclosure to encourage independent audits and replications	☐	
2. Judgmental methods		
2.1. Avoid unaided judgment	☐	
2.2. Frame questions in various ways	☐	
2.3. Combine independent forecasts from heterogeneous experts	☐	(12)
2.4. Obtain reasons for forecasts	☐	
2.5. Ask forecaster why the forecast might be wrong, then to consider revising	☐	
2.6. Use judgmental bootstrapping	☐	(73*)
2.7. Use structured analogies	☐	(39)
3. Extrapolation methods		
3.1. Use all valid and reliable data	☐	
3.2. Decompose by causal forces	☐	(50)
3.3. Damp trend forecasts if the…		
3.3.1. Situation is uncertain or unstable	☐	(5)
3.3.2. Forecast time horizon is longer than the historical series	☐	
3.3.3. Trend goes outside the range of the previous data	☐	
3.3.4. Short- and long-term trend directions are inconsistent	☐	
3.3.5. Series are contrary (trend is inconsistent with causal forces)	☐	(17–43)
3.4. Damp seasonal factors	☐	(2 -20)
4. Causal methods		
4.1. Use prior knowledge to select variables and estimate effects	☐	(20)
4.2. Damp estimated weights	☐	(4)
4.3. Use diverse information, data, and models	☐	(30-39)
4.4. Use all important variables	☐	(48)
5. Combine forecasts from validated methods and diverse data	☐	(<=50)

*More accurate in percent of tests

Human brains are not adapted to solve complex problems with many variables. Think of operating a nuclear power generation plant. For such tasks, checklists are vital. For a review of evidence on the efficacy of checklists, see Arkes, Shaffer, and Dawes (2006). One study reports reductions of bloodstream infections in intensive care units of 103 Michigan hospitals. The reductions occurred after physicians were required to follow five simple rules when inserting catheters: (1) wash hands, (2) clean the patient's skin, (3) use full-barrier precautions when inserting central venous catheters, (4) avoid the femoral site, and (5) remove unnecessary catheters. Adhering to the simple checklist reduced the median infection rate from 2.7 per 1,000 patients to zero after three months. Benefits persisted: sixteen to eighteen months after the checklist was introduced, infection rates had decreased by 66 percent (Pronovost et al. 2006).

Another study reports on the application of a 19-item checklist to surgical procedures for thousands of patients in eight hospitals in eight cities around the world. Following the introduction of the checklist, death rates declined by 47 percent (from 1.5 to 0.8 percent) and complications by 36 percent (from 11 to 7 percent) (Haynes, Weiser, Berry, Lipsitz, Breizat, and Dellinger 2009). Gawande (2010) provides further evidence of the usefulness of checklists in medicine, and in other fields such as aviation and finance.

The Golden Rule Checklist guidelines presented in this article are the product of empirical tests of alternative methods, and logic where sufficient. To ensure the evidence is properly summarized, the authors of this article sent emails to the lead authors of all the articles cited. If no reply was received, they sent a second email. If no reply was received an email was sent to the co-authors. [TBA] The authors also asked whether any relevant evidence had been overlooked. About [TBA] percent of the contacted authors replied. Their responses lead to improvements in the article.

Formulating the Problem

Forecasting problems are everywhere decisions are made, but are often not seen in this light. Forecasters must first formulate the forecasting problem. Proper formulation allows effective use of prior knowledge.

Obtain and use all knowledge (1.1)

Forecasts should be based on cumulative knowledge about the forecasting problem and about forecasting methods. Forecasters typically need to work with domain experts in order to acquire the relevant prior knowledge and data. Managers may not be familiar with what happened in the past, however; and their recollections may be unreliable. Furthermore, managers often believe that they are experiencing new and unprecedented conditions, that the past is no longer relevant, and that no other situation can provide useful information. President Dwight Eisenhower described this belief as follows: "Things are more like they are now than they ever were in the past." This belief that nothing like this has ever happened before poses a challenge to the use of the Golden Rule.

Obtain all relevant information and sufficient understanding (1.1.1): Conservative forecasting requires that the forecaster obtain all relevant information and sufficient understanding of the relevant theories and evidence in order to realistically represent the situation in forecasting models. A superficial knowledge of the situation can lead to large errors if, for example, the forecaster fails to recognize that the state of knowledge is contested and uncertain, or if he is ignorant of key relationships, or of a special event that will have a major impact.

Decompose the problem to best use knowledge and judgment (1.1.2): Decompose the problem in ways that best help the collection of knowledge and relevant data about the situation, and to select proper forecasting methods for each sub-problem. Decomposing the problem enables forecasters to draw upon diverse expertise and more data. Decomposition is more effective to the extent that the errors in the component forecasts are uncorrelated.

Decomposition allows forecasters to better match forecasting methods to the situation by, for example, using causal models to forecast market size and extrapolating market-share data from analogous sites. Or by using extrapolation to forecast business-as-usual demand and judgmental adjustments to forecast the effects of special events such as marketing campaigns or a change in product design.

For time-series data, obtain independent forecasts of the current level and trend, then add the components. Obtain sufficient knowledge to estimate the current situation properly. For example, the initial and the revised estimates of economic indicators often differ considerably. On January 30, 2009, the Bureau of Economic Analysis at the U.S. Department of Commerce estimated a real GDP decrease of 3.8 percent for the fourth quarter of 2008. One month later, the figure was revised to negative 6.2 percent. The final estimate is that GDP fell by 8.9 percent. Revisions of this size are common. One study analyzes deviations between initial and revised estimates of quarterly GDP growth from 1961 to 1996 (Runkle 1998). The figures reveal upward revisions by as much as 7.5 percentage points and downward revisions by as much as 6.2 percentage points. Problems with data are worse when collection is difficult, and where adjustments to data are secret and subject to political interference. These factors are present in the national accounts statistics of many African countries with, for example, an update of the system used in Ghana resulting in a GDP estimate that was 62 percent higher than the previous estimate (Devarajan 2011). In China, a review of gross industrial output figures for the town of Henglan found that the town's economic development and technology information bureau had overstated output by almost four times (McMahon 2013). Such errors affect the forecasts. One study found that about 20 percent of the total error in predicting GNP one-year-ahead arose from errors in estimating current GNP (Zarnowitz 1967).

As a consequence of the uncertainty over data, forecasters should seek alternative estimates of starting values. Consider estimating the current mean by combining the latest observation with estimates from evidence-based methods such as exponentially smoothed levels with a correction for lag, or with the constant from a regression analysis of the series against time.

The importance of starting with an accurate estimate of the current level, sometimes referred to as "nowcasting", was demonstrated with forecasts of U.S. lodging market sales. An econometric model provided 7 one-year-ahead forecasts, 6 two-ahead, and so on for the 7-year period from 1965 through 1971; 28 *ex ante* forecasts in total. The Mean Absolute Percentage Error (MAPE) was 12.7 when the official statistics were used as the starting level, but only 7.5 when the starting level was calculated as an average of the official data and an econometric estimate (Tessier and Armstrong 1977).

Decomposition improves accuracy most when the situation being forecast is uncertain. *Multiplicative decomposition* was used to structure 15 highly uncertain situations in three experiments, and subjects then made judgmental forecasts for each component. The average component forecasts were then multiplied. The procedure reduced median error ratios by 42 percent (MacGregor 2001, Exhibit 2).

When forecasts are made for segments and then added, the procedure is known as *additive decomposition*, or segmentation or bottom-up forecasting. Segments might be a firm's sales of different products or its sales by geographical regions. By considering segments separately, forecasters are able to make use of the local knowledge of experts from different domains (e.g. product experts and region experts).

Segmentation forecasts were more accurate for 74 percent of 192 series used in the M-Competition (Dangerfield and Morris 1992). When seven teams of experts forecast the number of hours needed to complete software projects, the errors of bottom-up forecasts were 51 percent smaller than the errors of direct forecasts of total hours (Jørgensen 2004).

Decomposition is especially useful where there is prior knowledge about causal relationships and large data sets. For example, data from 2,717 gas stations were used to estimate a stepwise regression model starting with 19 variables, and a segmentation model using the same variables. The models were used to forecast sales for 3,000 holdout gas stations. The MAPE for the regression model forecasts was 58 percent , compared to 41 percent for the segmentation method, an error reduction of 29 percent (Armstrong and Andress 1970).

The "Simulmatics Study" forecast voting in the Kennedy-Nixon election of 1960 for each of 480 voter-type segments. The data on voter type were from 50 surveys with more than 100,000 respondents conducted form 1952 to 1958. The study was apparently successful in helping to develop Kennedy's campaign strategy, and the method was used again for the 1964 election (Pool, Abelson and Popkin 1965).

Use evidence-based forecasting methods validated for the situation (1.1.3): Avoid biased methods by using only procedures that have been empirically validated under the conditions that apply in the situation being forecast. Fortunately, there is much evidence on which forecasting methods work best under what conditions. The evidence derives from empirical comparisons of the out-of-sample forecast accuracy of alternative methods. Knowledge about forecasting is summarized in the form of evidence-based conditional principles in the *Principles of Forecasting* handbook, a collaborative effort of 40 forecasting researchers and 123 expert reviewers (Armstrong, 2001c). As far as the authors of this article are aware, the handbook is the only published summary of evidence-based forecasting principles. The *Forecasting Method Selection Tree* at forecastingprimciples.com summarizes how to select which of the 15 forecasting methods are appropriate to the conditions. There is no "best method" overall.

Despite the extensive evidence on forecasting methods and the ready availability of evidence-based principles for forecasting, many forecasters overlook this knowledge. Consider the forecasts that are the basis of the manmade global warming alarm (Randall, et al. 2007). An audit found that the procedures violated 72 of 89 relevant forecasting principles (Green and Armstrong 2007a). In other words, the alarming forecasts are unscientific.

Do not assume that well known methods have been validated. Many statistical forecasting procedures have been proposed simply on the basis of experts' opinions. In general, statisticians seem to be little interested in how well their proposed methods perform in empirical validation tests. A check of the Social Science and Science Citation Indexes (SSCI and SCI) found that four key comparative validation studies on time-series forecasting were cited only three times per year between 1974 and 1991 in all the statistics journals indexed (Fildes and Makridakis 1995). Many thousands of time-series studies were published over that time.

Box and Jenkins (1970) provide an example of a popular but unvalidated statistical method proposed for forecasting. In a 1992 survey of 49 forecasting experts at the 1987 International Symposium on Forecasting, over half reported that the Box-Jenkins method was useful (Collopy and Armstrong1992a). However, despite many journal articles and widespread applications, little validation research has been done. When validation tests were done, Box-Jenkins procedures performed poorly relative to evidence-based procedures. For example, the M2- and M3-Competitions compared the accuracy of Box-Jenkins forecasts to forecasts from two conservative benchmark methods, damped trend and combining forecasts. The combined forecast was the simple average of single exponential smoothing (no trend), Holt's linear exponential smoothing (full trend), and exponential smoothing with damped trend. The M2-Competition included 29 series and 30 time horizons, and the M3 Competition included 3,003 series and 18 time horizons. Averaging across all time series and all forecast horizons, the MAPE of Box-Jenkins forecasts was 38 percent larger than the MAPE of the damped trend forecast in the M2-Competition and 3 percent larger in the M3-Competition. The Box-Jenkins forecast error was 37 percent larger than the combined forecast error in the M2-Competition and 4 percent larger in the M3-Competition (Makridakis, Chatfield, Hibon, Lawrence, Mills, Ord, and Simmons 1993, Exhibit 3; Makridakis and Hibon 2000, Table 6).

Apparently, professional forecasters seldom validate their methods successively against alternative evidence-based methods; otherwise they would advertise their findings. Independent evaluations of one popular commercial program, known as Focus Forecasting, found the forecasts to be substantially less accurate than forecasts from exponential smoothing (Flores and Whybark 1986; Gardner, Anderson-Fletcher, and Wickes 2001) and damped smoothing (Gardner 1997).

Avoid bias (1.2)

Avoiding bias in data and methods is a conservative forecasting practice that provides substantial gains in forecast accuracy. Bias is common in the government sector because failed forecasts typically are not punished. Instead governments often reward biased forecasts by providing additional funding when outcomes are disappointing. Bias also occurs in firms, but the desire for long-run profitability provides an incentive to avoid poor forecasting in many situations.

Specify multiple hypotheses or conceal the purpose of the forecast (1.2.1)**:** Forecasters sometimes depart from prior knowledge due to biases they may not be aware of, such as optimism or using the most easily available data. Financial incentives and the expectations of managers (deference to authority) can also cause forecasters to ignore important prior knowledge. For example, polar bear population forecasting reports requested by the U.S. Fish and Wildlife Service included title pages that read "USGS [U.S. Geological Survey] Science Strategy to Support U.S. Fish and Wildlife Service Polar Bear Listing Decision." A listing decision required forecasts of declining polar bear numbers. The State of Alaska commissioned an audit of the forecasting procedures (Armstrong, Green, and Soon 2008). Consistent with this guideline (1.2.1), the Alaska government officials provided no guidance on the process or desired outcome of the audit or on how they would use it.

Specifying multiple reasonable hypotheses, or possible outcomes, before forecasting allows for the possibility that a non-favored hypothesis may turn out to be superior. For example, an unbiased formulation of the polar bear population problem would be to forecast which population development is most likely: decreasing, little change, or increasing. At the time of the Senate Hearing "Examining Threats and Protections for the Polar Bear" on January 30 2008, the polar bear population had been growing during the previous two decades or more due to hunting restrictions. With the restrictions still in place, one might expect the upward trend to continue, at least in the short-term. However, the government scientists' forecast a rapid *decrease* in the polar bear population. The population has continued to grow since 2008.

Another precaution is to use analysts who are unaware of the purpose of the forecasts and to provide them with written instructions for how to select, clean, transform, and adjust data, and to then obtain the forecasts using the prescribed methods.

Obtain signed ethics statements before and after forecasting (1.2.2): Bias might be deliberate if the purpose of the forecasts is to serve strategic goals. Cost-benefit estimates for large-scale public works projects provide examples. In such cases, benefit forecasts are commonly biased upwards and cost forecasts downwards in order to meet governments' benefit-cost ratio criteria. For example, one study found that first-year demand forecasts for 62 large rail transportation projects were consistently optimistic, with a median overestimate of demand of 96 percent (Flyvbjerg, 2013).

Bias can be reduced by obtaining signed ethics statements from each of those involved in the forecasting project when starting work on a project, and then upon completion. Ideally, this would state that the analyst understands and will follow (has followed) evidence-based forecasting procedures. Laboratory studies have shown that when people reflect on their ethical standards, they act in ways that are more consistent with those ethics than they would have done otherwise (Armstrong, 2010, pp. 89-94; Shu, Mazar, Gino, Ariely, and Bazerman, 2011). Include a requirement to report any relationships—financial or personal—that might lead to a conflict of interest. Conflicts of interest are of less concern if multiple hypotheses are properly tested or if the forecasters are unaware of the purpose of the forecast.

Structure adjustments for important knowledge outside the model (1.2.3)
Judgmental adjustment of forecasts should be confined to experts' estimates of the effects of important influences that are not included in the forecasting model (Sanders and Ritzman 2001). The estimates should be made in ignorance of the forecasts from the model, but with knowledge of what variables and other information the model uses (Armstrong and Collopy 1998; Armstrong, Adya, and Collopy 2001). The experts' estimates should be derived in a structured way (Armstrong and Collopy 1998), and the rationale and the process documented and disclosed (Goodwin 2000). Compose the final forecast from the model forecast and the expert's adjustments. If the Golden Rule guidelines are followed, judgmental adjustments should be rare

Nevertheless, when forecasters and managers obtain forecasts from quantitative methods they are often tempted to adjust them. In a survey of forecasters at 96 U.S. corporations, about 45 percent of the respondents claimed that they always made judgmental adjustments to statistical forecasts. Only 9 percent said that they never did. The main reasons the respondents gave for revising quantitative forecasts were to incorporate knowledge of the environment (39 percent), product knowledge (30 percent), and past experience (26 percent) (Sanders and Manrodt 1994).

The reasons forecasters give for making adjustments seem sensible, and structured judgmental adjustments can improve accuracy when experts have good knowledge of the effects of special events and changes in causal forces (Fildes and Goodwin, 2007). Nevertheless, adjustments often introduce biases. A survey of 45 managers in a large conglomerate found that 64 percent of them believed that "forecasts are frequently politically motivated" (Fildes and Hastings 1994). In psychology, extensive research on cross-sectional data led to the conclusion that one should not subjectively adjust forecasts that are obtained from a quantitative model. For example, a summary of relevant research on personnel selection revealed that employers should not meet job candidates because this leads them to adjust the forecasts from statistical models to the detriment of accuracy (Meehl, 1954). Armstrong (1985, pp. 235-238) summarizes seven studies on this issue.

In view of the potential for judgmental adjustments to introduce bias (see Checklist guideline 2.1), adjustments should only be made when the conditions for successful adjustment are met and when bias can be avoided (Goodwin and Fildes 1999; Fildes, Goodwin, Lawrence, and Nikolopoulos 2009). Should important new information arise, reapply the Golden Rule Guidelines using the new information.

Provide full disclosure to encourage independent audits and replications (1.3)

Fully disclose the data and methods used for forecasting, and describe how they were selected. Full disclosure is important not only for identifying and avoiding biases, but also for documenting the data and forecasting procedures such that independent audits and replications are possible.

Failures in full disclosure are often due to oversight, but are sometimes intentional. For example, in preparation for a presentation to a U.S. Senate Science Committee hearing, the first author requested the data that the U.S. Fish and Wildlife Service researchers had used to prepare their forecasts that polar bears were endangered. The researchers refused to provide these data on the grounds that they were using them.

Replications are also important for detecting mistakes. One study found 23 books and articles, most of which were peer-reviewed, that included mistakes in the trend component of exponential smoothing model formulations (Gardner 1984). A follow-up study found mistakes in exponential smoothing programs used in two companies (Gardner 1985).

Judgmental Methods

Judgment is often used for important decisions such as whether to start a war, launch a new product, acquire a company, buy a house, select a CEO, or stimulate the economy. Unfortunately, judgmental methods are prone to numerous biases.

Avoid unaided judgment (2.1)

By unaided judgment we mean unstructured judgmental forecasting that ignores evidence-based procedures. Unaided judgment is not conservative because it is a product of faulty memories, inadequate mental representation of complex phenomena, and unreliable record keeping, to mention only a few of the shortcomings. Moreover, when experts use their unaided judgment, they tend to more easily remember recent, extreme, and vivid events (Kahneman, 2011). As a result, they overweight the importance of such events when making judgmental forecasts, which leads them to overestimate change. Consistent with these conclusions, in a study of over 27,000 political and economic forecasts made over a 20-year period, 284 experts from different fields forecast a change from the status quo 65 percent of the time when changes from the status quo actually occurred only 51 percent of the time (Tetlock 2005, p. 83).

Unaided judges tend to see patterns in the past and predict they will persist, even when they lack reasons for the patterns. Even forecasting experts are tempted to depart from conservatism in this way. For example, when two of the authors asked attendees at the 2012 *International Symposium on Forecasting* to forecast the next 25 years on two 50-year charts showing monthly global average temperature anomalies, about half (26) of 51 respondents drew zigzag lines to resemble the pattern in the historical series—a procedure that is virtually certain to increase forecast error relative to a straight line.

The biases of unaided judgment are magnified if forecasts are made in traditional group meetings. For example, participants may be reluctant to share their opinions so as to avoid conflict or ridicule. Despite its lack of predictive validity, unaided judgment is the method that people turn to when making forecasts for important decisions. Based on experimental evidence, it is difficult to find a method that produces such poor forecasts as unaided group meetings (Armstrong, 2006b).

Frame questions in various ways (2.2)

The way in which experts are asked for forecasts influences their forecasts. Asking for forecasts in ways that are clear and unbiased is difficult. Conservative forecasting requires a strategy for dealing with biases and misunderstandings that may arise from question wording. The long-standing approach for dealing with this problem is to pose the forecasting question in more than one way and to pre-test the different wordings to ensure they are clear to the experts. The final questions should be written and followed exactly, and the responses to the different questions combined.

Combine independent forecasts from heterogeneous experts (2.3)

To increase the amount of information considered and to reduce the effects of biases, combine anonymous forecasts from a heterogeneous group of independent experts. Surprisingly, contributions from experts with only modest domain knowledge can improve the accuracy of a combined forecast (Armstrong, 1980a). Good results can be achieved from combinations of forecasts from eight to twelve experts when their knowledge about the problem is heterogeneous. Adding forecasts from more experts helps, although the rate of improvement in forecast accuracy diminishes substantially (Hogarth, 1978).

One review presented evidence from seven studies that involved combining forecasts from between 4 and 79 experts. Combining forecasts reduced error by 12 percent compared to the typical expert forecast. The error reductions ranged from 7 percent to 19 percent (Armstrong 2001a). Another study analyzed the accuracy of expert forecasts of the outcomes of the three U.S. presidential elections from 2004 to 2012. The error of the combined forecasts from 12 to 15 experts was 12 percent less than that of the forecast of the typical expert (Graefe, Armstrong, Jones Jr., and Cuzán 2013).

Obtain reasons for forecasts (2.4)

Ask forecasters to provide reasons for their forecasts. Doing so will encourage them to draw upon more information as well as to provide full disclosure. The Delphi method elicits independent and anonymous forecasts, along with reasons, from experts. Provide a summary of the forecasts and reasons to the experts and ask them to revise their own forecasts, free from group pressures, in later rounds.

Delphi tends to provide more accurate forecasts than alternative group methods. A review of the literature finds that Delphi outperformed statistical groups (i.e., simple one-round surveys) in twelve studies and was less accurate in two studies, with two ties. Compared to traditional meetings, Delphi was more accurate in five studies and less accurate in one study; two studies showed no difference (Rowe and Wright, 2001). Results from a laboratory experiment on estimation tasks confirm these findings and show that the Delphi method not only outperformed prediction markets, it was also easier to understand (Graefe and Armstrong, 2011).

The Delphi method is well suited when relevant knowledge is distributed among experts. Providing information about reasons increases the accuracy to a greater extent than simply providing the aggregate forecasts without reasons. The Delphi method is easy to implement using freeware available at forecastingprinciples.com. .

Ask a forecaster why his forecast might be wrong, then to consider revising (2.5)

To make judgmental forecasts more conservative, ask people to consider why their forecasts might be wrong. One study asked students to predict the outcome of their job search efforts. In general, the students made forecasts that were more accurate if they were first asked to write reasons why their desired outcome might not occur (Hoch 1985). The forecasts tended to be more accurate because the forecasters were less optimistic.

Use judgmental bootstrapping (2.6)

People are often erratic in applying what they know to a problem. They might get overloaded with information, forgetful, tired, unable to concentrate, or irritable. Judgmental bootstrapping is a method for applying forecasters' implicit rules in a consistent way.

To forecast using judgmental bootstrapping, develop a quantitative model to infer how an expert or a group of experts makes the forecasts. To do so, first present the expert with 20 or more artificial cases in which the values of the variables that experts normally use vary independently of one another. Then ask the expert to make forecasts for each case. Finally, estimate a simple regression of the expert's forecasts against the variables. This is the judgmental bootstrapping model.

A review of eleven studies from various fields—including psychology, education, personnel, marketing, and finance—found that judgmental bootstrapping forecasts were more accurate than forecasts from unaided judgment in eight studies. There was no difference in two studies. In the main, the studies reported accuracy in terms of correlations. One study, however, reported an error reduction of 6.4 percent (Ashton, Ashton, and Davis, 1994). Judgmental bootstrapping can also reveal when forecasters use inappropriate information. For example, looks and height would be irrelevant for a job as a computer programmer. Forecasters can eliminate irrelevant variables from the model and thereby improve the forecasts (Armstrong, 2001b).

Use structured analogies (2.7)

To obtain conservative judgments, a number of experts, working independently, can be asked to identify analogous events from the past, rate their similarity to the current (target) situation, and identify the outcome implied for the target situation by each of the analogies. The forecaster then uses the average of the outcomes implied by each expert's top-rated analogy as the forecast. The method is known as structured analogies.

Structured analogies forecasts are conservative because they are consistent with what happened in similar situations. For example, to forecast trolley ridership in a U.S. city, one might examine the performance of similar trolley systems around the world (see Scheib, 2012). Data on the outcomes of the closest analogies could be used to make forecasts about a proposed trolley system.

Similarly, to forecast whether the California High Speed Rail would cover its costs, a forecaster could ask experts to identify similar high-speed rail systems (HSR), and obtain information on their profitability. The Congressional Research Service did this, and found "Few if any HSR lines anywhere in the world have earned enough revenue to cover both their construction and operating costs, even where population density is far greater than anywhere in the United States (Ryan and Session 2013).

Research on structured analogies is in its infancy, but the findings of substantial improvements in accuracy for situations that are complex and uncertain are encouraging. In one study, eight conflict situations were described to experts. They included union-management disputes, corporate takeover battles, and warfare. Unaided expert predictions of the decisions made in these situations were little more accurate than randomly selecting from a list of feasible decisions. In contrast, the error reduction from using structured analogies (97 forecasts) was 25 percent relative to guessing. Furthermore, the error reduction was as much as 39 percent for the 44 forecasts that were derived from data provided by experts who had identified two or more analogies (Green and Armstrong 2007b).

Extrapolation methods

Extrapolation is an inherently conservative approach to forecasting because it is based on data on past behavior. There are, however, a number of threats to conservatism from abuses of extrapolation.

Use all valid and reliable data (3.1)

Conservative forecasting requires that forecasters use all of the valid and reliable data that can be obtained. Practitioners often violate this simple guideline. For example, by selecting a starting point for a time-series forecast or by selecting a subset of cross-sectional data, a forecaster has much influence over the resulting forecast. This allows people to make forecasts that support their prior beliefs, as occurs in climate forecasting (Green, Soon, and Armstrong 2013)

To avoid biases, intentional or unintentional, use analysts who are unaware of the purpose of the study to make any necessary revisions and adjustments to the data. Fully disclose revisions and the reasons for them.

Decompose by causal forces (3.2)

Ask domain experts to assess whether the trend of a series will be growth, decay, supporting, opposing, regressing, or unknown. Growth, for example, means that the causal forces will lead the series to increase, irrespective of the historical trend. When forecasting a time-series that is the product of conflicting causal forces such as growth *and* decay, decompose the series by these forces, and extrapolate each component separately. Doing so allows the forecaster to make use of causal knowledge about each component.

Consider the problem of forecasting highway deaths. The number of deaths tends to increase with the number of miles driven, but to decrease as the safety of vehicles and roads increases. As a consequence of the conflicting forces, the direction of the trend in the fatality rate is uncertain. By decomposing the problem into miles-driven-per-year and deaths-per-mile-driven, the analyst can use knowledge about the individual trends to extrapolate each component. The forecast for the total number of deaths per year is calculated as the product of the two components.

Decomposition by causal forces is applicable if two conditions are met: (1) the problem can be structured in a way that each component series is subject to a single causal force, and (2) it is possible to obtain relatively accurate forecasts of each component. One study examined twelve annual time series for airline and automobile accidents, airline revenues, personal computer sales, and cigarette production. Nine series met the conditions to the extent that the forecasts for the components were more accurate than those for the global series. Successive updating was used to make 575 ex ante forecasts; some for forecast horizons from 1 to 5 years and some for horizons from 1 to 10 years. Forecasting the decomposed series separately reduced the median absolute percentage error (MdAPE) of the combined forecasts by 56 percent relative to forecasts from extrapolating the global series (Armstrong, Collopy, and Yokum, 2005).

Damp trend forecasts (3.3)

When trends are uncertain, extrapolate conservatively. Uncertainty arises if the trend has been variable or unstable, or the forecast horizon is longer than the historical series, or values go outside the range of previous observations, or the recent trend is inconsistent with the long-term trend. Most importantly, uncertainty is high when the trend is inconsistent with causal forces.

***Damp a trend if the situation is uncertain or unstable (3.3.1)*:** If a trend is inconsistent, damp the trend toward zero. Damp more strongly when the forecast horizon is longer because of the increased uncertainty. Damping the trend toward zero yielded an average error reduction of almost five percent in a review of ten studies. The damping was based on the variability in the historical data (Armstrong, 2006a).

Damp a trend if the forecast time horizon is longer than the historical series (3.3.2): Uncertainty is high if the forecast horizon is longer than the length of the historical time series. If making forecasts in such a situation cannot be avoided, increase the damping of the trend toward zero as the forecast horizon increases. The U.S. Fish and Wildlife Service scientists in the aforementioned study of the population of polar bears overlooked the need for damping when they used only five years of historical data to forecast 50 years into the future.

Damp a trend if it goes outside the range of the previous data (3.3.3)**:** Avoid forecasts that go outside of the range of the historical values and trend rates that exceed previous rates. For example, government forecasts for the proposed California High Speed Rail violated this guideline when they forecast that their trains would travel at peak speeds of 220 miles per hour. At the time of writing, the planned speed has not been achieved anywhere else in the world (Vranich, Cox, and Moore, 2013).

Damp a trend if the short- and long-term trend directions are inconsistent (3.3.4)**:** If the direction of the short-term trend is inconsistent with that of the long-term trend, the short-term trend should be damped towards the long-term trend as the forecast horizon lengthens.

Damp a trend if series are contrary (trend is inconsistent with causal forces) (3.3.5)**:** If the causal forces acting on a time series conflict with the observed trend in a time series, a condition called a “contrary series,” damp the trend heavily toward the no-change forecast. Research findings to date suggest a simple guideline: *ignore trends for contrary series.* This is the “contrary-series rule.”

One study compared forecasts from the contrary series rule to forecasts from Holt’s exponential smoothing, a method that ignores causal forces. In a test of 20 contrary series from the M-Competition (Makridakis et al, 1982), the no-change model reduced the MdAPE (Median Absolute Percentage Error) by 18 percent for one-year ahead forecasts, and by 40 percent for six-year ahead forecasts. Additional tests involved contrary series from four other sets of data: Chinese epidemics, unit product sales, number of U.S. Navy personnel, and economic and demographic variables. On average the MdAPE for the no-change forecasts was 17 percent less than Holt’s forecast errors for 943 one-step-ahead forecasts. For 723 long-range forecasts, the error reduction was 43 percent (Armstrong and Collopy 1993).

The most famous application of the contrary series rule is Julian Simon's 1980 bet with Paul Ehrlich on the prices of resources. Ehrlich claimed that resources are limited. As a result, for example, Ehrlich forecast mass starvation by the 1990s as resources ran out. Simon argued that the human ingenuity and effort caused resources to become more plentiful and thus cheaper. Trends over centuries have been consistent with these causal factors, though political forces can and do oppose them. Consistent with the contrary-series rule, Simon argued that the best forecast for any resource was that the real price would not change, and invited Ehrlich to pick resources and a time period for a bet. Ehrlich nominated five metals whose prices had been rising rapidly in recent years, and bet that their prices would be higher in 1990. Simon's no-change price forecasts were more accurate for all five metals over the ten-year period (see Tierney 1990).

Damp seasonal factors (3.4)

For situations that are clearly affected by the season, such as monthly sales of sunscreen or furnace oil, seasonal factors improve forecast accuracy. When the situation is uncertain as well as seasonal, damp the estimated seasonal effects towards 1.0, for multiplicative factors, or toward zero for additive factors. Increase damping when there are only a few years of data or where the causal basis of the seasonal factors is uncertainty.

One study tested the effect of damping the estimated seasonal factors for the 1,428 monthly series of the M3-Competition. Damping was done by the automatic application of statistical rules based on the variability in each time-series. For 59 percent to 65 percent of the series, damped forecasts were more accurate. The gains in accuracy were larger for short-term forecasts, one- to three-months ahead. For series found by statistical tests to have seasonal patterns, damping seasonal factors reduced MAPEs by up to five percent (Miller and Williams 2004).

Uncertainty about seasonal factors may arise from various sources and suggest different approaches to damping. Here are four. (1) Smooth seasonal factors to allow for year-to-year variations by combining with the factors for the month before and the month after the month of interest. (2) Damp factors based on the number of years of data: more damping when there is only a single year, and less damping as each year is added. (3) Damp factors more when causal support for seasonality is weaker. (4) Increase damping of seasonal factors as the forecast horizon lengthens and uncertainty increases. A scheme for implementing these suggestions is provided in Armstrong and Collopy (2000). When the scheme was tested on 62 randomly selected monthly series from the M-Competition (Makridakis et al.1982), damping reduced the MdAPE by 7.2 percent for one-year-ahead forecasts, and by 5.0 percent for 18-months-ahead forecasts. Horizon modified damping reduced error for 56 percent of the series.

Another approach is to damp seasonal factors toward those from similar series. That is, combine the seasonal factors from analogous series (e.g., all types of luxury cars) and then use the resulting factors when forecasting the series of interest (e.g., the Hyundai Genesis). Seasonal factors damped in this way were consistently more accurate, with error reductions of up to 40 percent, in an analysis of 44 series of retail sales data from a large U.K. department store chain (Bunn and Vassilopoulos 1999). Another study combined seasonal crime rates from six precincts in Pittsburgh and found the combined-seasonality forecast errors were about 8 percent smaller than the individual-precinct-seasonality forecast errors (Gorr, Olligschlaeger, and Thompson, 2003).

Causal methods

Regression analysis is currently the most common approach for developing and estimating causal models. Judging by its name, the method sounds conservative. It is conservative in that it regresses to the mean values of the series by damping the coefficients for the predictor variables in response to unexplained variability in the historical data. A regression model will not be conservative, however, if any model variable correlates with important excluded variables over the estimation period. Regression models also do not reflect uncertainty in predicting the causal variables or changes in causal factors, to mention only a few of the problems. For a more detailed discussion of problems with using regression analysis for forecasting, see Armstrong (2012) and Soyer and Hogarth (2012).

Use prior knowledge to select variables and estimate effects (4.1)

Scientific discoveries about causality were made prior to the availability of regression analysis. For example, John Snow discovered the cause of cholera in London in the 1850s as a result of "the clarity of the prior reasoning, the bringing together of many different lines of evidence, and the amount of shoe leather Snow was willing to use to get the data [to test for the causes]" (Freedman (1991, p. 298). Until the latter part of the 20th Century, data collection and statistical analyses remained expensive and forecasters had little choice but to develop their models using well-supported theories and a priori analysis.

In econometrics, a priori analysis uses elasticities. Elasticities are unit-free and easy to interpret. They represent the ratio of the percentage change that occurs in the variable to be forecast in response to a percentage change in the causal variable. For example, a price elasticity of demand of -1.5 would mean that if the price were increased by 10 percent, unit sales would go down by 15 percent. Forecasters should examine prior research in order to determine the likely values of relevant elasticities and their plausible lower and upper bounds. For problems in economic forecasting, for example, one can find estimates of income, price, and advertising elasticities in published meta-analyses. If little prior research exists, use estimates from domain experts and from data on the specific situation. To obtain elasticity estimates from available data, formulate models in multiplicative terms by calculating the logarithm of the data values, and then run the regression analysis.

In the late 1960s, the first author, impressed by the work of econometricians, developed a model to forecast international camera sales by following the original econometrics procedures described above. In particular, he specified a model from a priori analysis *before* analyzing data. He then used data from 1960 to 1967 to estimate another set of coefficients for the model. The final coefficients for the model were then calculated as an average of the prior research-based model estimates with the regression model estimates, a process later referred to as a "poor man's Bayesian regression analysis" (Armstrong and Grohman 1972). The model was tested using a "backcast" of 1954 camera sales for 17 countries, by using only 1960-65 data. The MAPE of the forecasts from the version of the model with statistically estimated coefficients was 30 percent. The MAPE of the forecasts from the final (combined coefficient estimates) model was 23 percent, an error reduction of 23 percent (Armstrong, 1970).

Bayesian analysis provides another way to incorporate prior knowledge into a forecasting model. There is an enormous academic literature on Bayesian analysis, with about 6,000 hits on "Bayesian forecasting" on Google Scholar as of July 2013. Despite the volume of research, the authors of this article are not aware of evidence that ex ante forecasts from Bayesian forecasting are more accurate than those from alternative evidence-based forecasting methods. A study on election forecasting found that a simple average of forecasts from traditional regression analyses produced errors that were 19 percent lower than those from a

Bayesian approach to combining forecasts (Graefe 2013). The lack of evidence in favor of Bayesian analysis is a blessing for those who prefer simple easily understood methods.

Since the 1960s, the stress on prior knowledge in developing causal models has given way to a preference for statistical procedures in the hope that they would replace the need for prior knowledge. As computers make it possible to handle more and more data, and more and more data becomes available, analysts increasingly rely on ever more sophisticated statistical procedures to select predictor variables. Indeed, a survey of leading econometricians in the mid-1970s showed support for the belief that complex statistical procedures yield greater forecast accuracy (Armstrong 1978b). Einhorn (1972) found the trend distressing and concluded, “Just as the alchemists were not successful in turning base metal into gold, the modern researcher cannot rely on the ‘computer’ to turn his data into meaningful and valuable scientific information” (p. 378). Research since then supports Einhorn’s assessment (Armstrong, 2012).

Despite the theoretical and empirical objections, much current practice continues to ignore prior knowledge in favor of selecting models discovered to have statistically significant relationships in a given set of data. Journal editors and reviewers favor statistically significant results, and researchers oblige by searching for the model that best fits available data. These conclusions are consistent with those from a review of why the effects of newly discovered relationships are commonly inflated (Ioannidis, 2008).

Damp estimated weights (4.2)

As with extrapolation forecasts, damping is useful for making causal model forecasts more conservative. One strategy is to damp estimated model coefficients toward zero. The process, which is sometimes referred to as shrinkage, reduces the amount of change that a model will predict, and is thus conservative. Another strategy is to damp the influences of variables (i.e., their weights) so that they are more equal to one another. To do this, express the variable values as differences from their mean divided by the their standard deviation (i.e. normalized variables), and damp the estimated coefficients toward equality. In an extreme version of damping, all variables are assigned equal weights.

An early test of extreme damping in situations with many variables and small sample sizes found that out-of-sample forecasts from equal-weights models are often more accurate than forecasts from models with regression estimated weights (Schmidt, 1971). Dawes and

Corrigan (1974) and others followed with similar results. A more recent study examined forecasts derived from multiple regression models that were developed for twenty prediction problems from statistics textbooks (Czerlinski, Gigerenzer, and Goldstein 1999). The purpose of the textbook problems was to demonstrate the value of multiple regression analysis. When tested on holdout data, however, multiple regression models produced slightly less accurate forecasts (68 percent correct predictions) than the equal weights models (69 percent correct). Tests using five real non-experimental social science datasets and synthetic datasets found that equal-weights models typically produce forecasts that are more accurate than forecasts from models with regression estimates, especially when the estimation sample sizes were small (Dana and Dawes, 2004).

Economists and political scientists have been developing and publishing regression model forecasts of election results since the 1980s. In one study, equal weighting outperformed two of three established regression models—and did equally well as the third—when making out-of-sample forecasts of the 23 U.S. presidential election results from 1916 to 2004 (Cuzán and Bundrick, 2009). An extension of this study analyzed *ex ante* forecasts from nine established models. Equal-weights models yielded more accurate forecasts than six of the nine regression models. On average, the error of the equal-weights model forecasts was four percent lower than the error of the regression model forecasts (Graefe, 2013).

Use diverse information, data, and models (4.3)

When estimating relationships using nonexperimental data, regression models can properly include only a subset of variables—typically about three—no matter what the sample size. But many practical problems involve more than three important variables, and may involve variables for which data are inadequate for regression analysis in that, for example, the causal variable values may vary little in the sample or may be unreliable due to data collection problems. For example, the long-run growth rates of nations might depend on fifty or more variables the values of many of which do not vary over long periods and that may be greatly uncertain in some nations. As a consequence, regression models tend to exclude important knowledge about causality.

One way to deal with the limitations of regression analysis is to develop different models, using different variables and data, and to then to combine the forecasts from each model. In a study on 10-year ahead forecasts of population in 100 counties of North Carolina, the average MAPE for a set of econometric models was 9.5 percent. In contrast, the MAPE for the combined forecast was only 5.8 percent, an error reduction of 39 percent (Namboodiri and Lalu 1971).

Another test involved forecasting U.S. presidential election results. Most of the well-known regression models for this task are based on some measure of the incumbent's performance in handling the economy and one or two other variables. The models differ in the variables and in the data used. Across the six elections from 1992 to 2012, the combined forecasts from all of the published models in each year—the number of which increased from 6 to 22 across the six elections—had a mean absolute error that was 30 percent less than that of the typical model (Graefe, Armstrong, Jones, and Cuzán 2013).

Use all important variables (4.4)

Another approach to developing conservative causal models is to incorporate all important knowledge about causal relationships in one model. To do this, first identify all relevant variables by asking experts or, whenever possible, reviewing prior evidence; preferably experimental evidence. Then code the expected directional influence of the variables on whatever is being forecast (e.g., job performance). For example, assign a coefficient of +1 for variables with a positive impact and zero otherwise. To use the model to forecast, rate the subject of the forecasts (e.g., a job candidate) against each variable. If the reliability of the ratings is low, use several raters and average their ratings. Finally, add the variable ratings to calculate a single index score. For selection problems, the option with the highest index score would be predicted to be the best. For numerical forecasts, estimate a simple linear regression model that estimates the relationship between the index score and the variable to be predicted (e.g., sales of a new movie).

The inclusion of all important variables is possible because the method does not require the estimation of variable coefficients (weights) from data. The approach is called the *index method*, and is based on an insight that came from Benjamin Franklin's "Method for deciding doubtful matters" (Sparks, 1844). Index models might also be called *knowledge models* because of they allow the inclusion of all knowledge about factors affecting the thing being forecast.

The index method has been used to forecast U.S. presidential elections, a situation for which there is knowledge on a large number of predictor variables but few observations. For example, one index model based on 59 biographical variables correctly predicted the winners of 28 of the 30 U.S. presidential elections from 1896 to 2012 (Armstrong and Graefe 2011). Another index model was based on surveys of how voters expect U.S. presidential candidates to handle important issues. The number of issues varied from 23 to 47. The model correctly predicted the election winner in ten of the eleven elections from 1972 to 2012 (Graefe and Armstrong 2013). Forecasts from an index model that included all of the variables used in nine established regression models had an error that was less than half (48 percent) of the error of the typical individual regression model for the ten elections from 1976 to 2012 (Graefe, 2013).

Combine forecasts from validated methods and diverse data

Combining forecasts from evidence-base methods is a conservative strategy. One reason is that different forecasts incorporate different knowledge, so combining increases the use of knowledge. Combining also helps to reduce the effects of mistakes (such as computational errors), the selection of an inappropriate forecasting method, or the use of a poorly specified model. Combining forecasts across methods is particularly valuable if the biases of the component forecasts are uncorrelated.

To reduce the effects of biases, the data and methods should be specified prior to making the forecasts. In particular, forecasters should decide on a weighting scheme for the forecasts prior to forecasting. Many scholars have proposed methods for how to best weight the component forecasts of a combination. However, a review of more than two hundred published papers from the fields of forecasting, psychology, statistics, and management science concluded that using equal weights usually provides the best forecast when combining (Clemen 1989). An updated review has reinforced Clemen’s conclusion (Mancuso and Werner, 2013).

Combine by calculating trimmed averages unless strong evidence suggests that forecasts from some methods are consistently more accurate than those from others in the given situation or subject to similar conditions. The benefits of differential weights under certain conditions were demonstrated for the selection of different extrapolation methods depending on the forecast horizon. For example, rule-based forecasting, which varies the weights on extrapolation methods with the horizon, provided the most accurate forecasts for annual data in the M-Competition (Collopy and Armstrong, 1992b).

The benefits of combining are not intuitively obvious. In a series of experiments with highly qualified MBA students, a majority of participants thought that averaging estimates would deliver only average performance (Larrick and Soll 2006). The gains from combining can, however, be considerable. Moreover, combining can improve accuracy even when the administrator knows which individual method will be most accurate (Graefe, Armstrong, Jones Jr., and Cuzán 2013).

Gains from combining are higher when different evidence-based forecasting methods are used and the forecasts draw upon different data. This proposition was tested in a study involving forecasts of the popular vote shares in the six U.S. presidential elections from 1992 to 2012. Averaging forecasts within and across four established election-forecasting methods (poll projections, prediction markets, expert judgments, and quantitative models) yielded forecasts that were more accurate than those from all of the component methods. The error reduction compared to the typical component method forecasts was as much as 60 percent (Graefe, Armstrong, Jones Jr., and Cuzán, 2013).

Discussion

This review of experimental evidence finds that conservative forecasting practices provide consistent and substantial improvements in accuracy over forecasts from non-conservative practices. The gains in accuracy are summarized in the Exhibit. With the opportunities for improved performance that have been identified by researchers, and the efforts that have been made to communicate these opportunities via conferences, journals, and websites, one might expect substantial improvements in the practice of forecasting. Practice does not appear to have improved, however: it is difficult to find studies that refute Ascher's (1978) conclusion that, in practice, forecast accuracy does not improve over time. Ascher based his conclusion on his review of forecasting for population, economics, energy, transportation, and technology.

In part, findings on the importance of conservatism in forecasting have been slow to make their way into forecasting practice because few forecasting practitioners are aware of the evidence on which methods work best in what situations. Another reason is that low-cost computation, complex statistical methods, and large databases have seduced forecasters away from conservative forecasting procedures.

One way to improve forecasting practice would be to implement conservative forecasting procedures, as default options, in forecasting software products. Unfortunately, software providers have little incentive to provide users with tools that they can easily understand. That said, some software providers have told the authors of this article that they would be happy to work with clients who wanted to have the more effective methods made available in their software. For example, it would be a simple and inexpensive matter for them to include the contrary series rule.

It is possible that complex statistical methods may have benefits in some situations. Nevertheless, with the risk that they will mislead forecasters ~~hide mistakes,~~ and confuse clients, they should be avoided unless experimental evidence supports their use in a situation. To date, no such evidence has been forthcoming and so forecasting practitioners have no need to learn these techniques.

In contrast, large databases *can* play an important role in forecasting by using the method of decomposition. Decomposition enables forecasters to make proper use of large bodies of data with information on many important variables, and that may include interactions or multicollinearity, or causal priorities among predictor variables, or non-linear effects.

The following sections discuss other reasons for the lack of improvement in forecasting practice. In particular, the motivations of forecasters and of forecast users.

A preference for complexity: Revisiting the Rain Dance

Rainmaker Theory 1: “The rainmaker gets so involved with
the dance that he sometimes forgets that he has to make rain.”
Rainmaker Theory 2: “Yes, I know it didn’t
rain–but didn’t you like the dance?

The Rain Dance theories were introduced into the forecasting literature in the 1970s to explain why forecasters prefer complexity to simplicity: “The rain dance has something for everyone. The dancer gets paid. The clients get to watch a good dance. The decision maker gets to shift the problem off to someone else in a socially acceptable way, as in “Who can blame him? He hired the best dancer in the business” (Armstrong, 1985, p. 16–17, 152, 261, 429–30).

The rain dance features intelligent people from impressive organizations. They speak in code and their machines do wondrous things when they are fed enormous databases. Conservative procedures seem slow and mundane by comparison. Conservative forecasting methods are unlikely to impress clients because they are typically easy to understand. Imagine a forecaster suggesting using the no-change model to a client. The client might be reluctant to pay for such advice. In contrast, complex methods are more persuasive. People wrongly believe that complex methods are necessary to solve complex problems. When confronted with evidence that simple methods provide forecasts that are more accurate than do complex ones, people resist the findings (Hogarth 2012).

Reports by experts who use complex math are held in higher regard, even when the mathematics is completely unrelated to the problem (Ericksson 2012). In addition, when experts explain something in a complex way, people have more confidence than if it is explained simply (Armstrong 1980b).

Complex methods can help to shelter a forecaster from criticism. They can make it difficult for forecasting clients to hold forecasters to account if a forecast is wrong. In contrast, conservative methods are often easy to understand, and are thus easy to criticize in case of a poor or unpopular forecast.

The first author encountered the rain dance when he was a tenure-less young researcher working with a small team of consultants who were asked to test the forecast validity of a complex model provided to a firm at high cost by an outside vendor. The model predicted the effects of advertising on market share. Confidential interviews with analysts in the firm revealed that none of the analysts understood how the model worked. The team then showed that an almost perfect correspondence with predictions from the vendor's model was achieved by regressing the firm's market share against advertising expenditures. When the team delivered their report, the firm terminated their consulting relationship with the team. The termination was expected, as the firm had been clear that the team's report should be favorable toward the vendor's model. Two of the team members wrote a paper (with the participants' names disguised) to demonstrate how to evaluate forecasting models. The paper was submitted to the *Journal of Marketing*, and was accepted. The vendor threatened legal action if the paper were not withdrawn. The head of the two authors' academic department advised them that it would be wise to withdraw the paper. The paper was published (Armstrong and Shapiro 1974). There were consequences for the authors' careers.

Desire for novelty: Novelty is a route to fame. Forecasts that violate the Golden Rule create attention as they go against the crowd. They help a forecaster to create a reputation if they come true. If false, the forecasts are often quickly forgotten, or sometimes remembered as wise warnings or inspiring visions of what might have been.

Forecasts are sometimes presented as new and interesting discoveries. Consider, for example, forecasts that eating a certain food or avoiding a certain ingredient will increase your life span. Given the improbability of identifying causal relationships from non-experimental data, complex statistical analysis of large datasets is a self-perpetuating academic activity, with further research either challenging or confirming a previous study. And so the journals are filled. Forecasts from conservative procedures, by contrast, are less likely to be newsworthy.

Politics of forecasting: Organizations often view forecasts as motivational tools, rather than as attempts to know what will happen given the adoption of a particular strategy. They may believe that an unusual forecast will help them to take advantage of an apparent new trend. This fallacious reasoning confuses forecasting with planning. Forecasters should confine themselves to forecasting what is likely to happen given the strategies that decision-makers wish to consider.

Conservative forecasting procedures are likely to produce forecasts that displease clients. For example, on August 4, 2011, the front page of *The New York Times* announced a new educational program, partly funded by New York City, for disadvantaged young black and Latino males. Predictions that the program would be successful were apparently based on expert opinions. Prior experimental evidence suggested otherwise. For example, between 1939 and 1944 the Cambridge-Somerville experiment gave counseling and training to a randomly selected half of a group of more than 500 troubled young men. Thirty years later, the counseled and trained men reported that they had benefited immensely from the program. However, the evidence suggested otherwise. Those who had been in the program had less-prestigious jobs, lower job-satisfaction, and higher crime rates. They also had higher rates of alcoholism and sickness, and died younger (McCord 1978). No attempt was made to explain why the results would differ in 2011. A letter sent to *The New York Times* to inform the public about the implausibility of the forecast that the program would be successful failed to meet the paper's criterion of "all the news that's fit to print."

Economic forecasts commonly show systematic biases depending on the political agenda of the institution that published the forecast. Politicians want support for public works projects. Managers in firms want to invest other people's money. These biases are more prevalent for long-term forecasts in the public sector, since the institutions and politicians are rarely evaluated based on the accuracy of their forecasts, especially their long-term forecasts.

Role of the no-change model

The ultimate conservative model is the no-change model. It should be the benchmark model against which to validate any forecasting model. The no-change model differs depending on the situation. It may be based on the long-run level, the extrapolation of a long-run trend, a base rate, or the usual behavior in similar situations.

The no-change model is a surprisingly good benchmark for most forecasting problems. It is incorporated in the primary metric for comparing the accuracy of alternative forecasting methods, the RAE (Relative Absolute Error). Evidence on the value of the RAE is provided in Armstrong and Collopy (1992).

One simple approach to the Golden Rule is to combine a forecast with a no-change forecast. More weight should be placed on the no-change model if uncertainty is high. Uncertainty typically increases with the complexity of the problem and the length of the forecast horizon.

Consider the common problem of sales forecasting. One study compared the accuracy of forecasts from the no-change model with forecasts from six full-trend extrapolation methods; 180 forecasts of fifteen economic time-series, consisting of prices of resources, production, and indicators such as unemployment claims. On average, the no-change model yielded the most accurate forecasts. The MAPE of forecasts from the no-change model was half that of the most complex extrapolation method tested, "generalized adaptive filtering" (Schnaars and Bavuso 1986).

The behavior of short-term stock market prices is complex and uncertain. As a result, one would expect that forecasts from the no-change model would be difficult to beat. Not surprisingly, then, researchers attempts since the 1930s to beat the current market price have proven unsuccessful for those who lack inside information. Malkiel (2012) has documented this phenomenon over the years in a book that is now in its tenth edition.

Long-term climate change is another ideal candidate for the no-trend model as it is a complex problem with high uncertainty. Thus, the no-change model, extrapolating without change the most recent year's global mean temperature, is an obvious benchmark method. In contrast, the IPCC relies on the outputs of complex and very expensive computer models to create "scenarios." From these, they derived a forecast that global average temperatures will increase at a rate 0.03ºC per annum due to human carbon dioxide emissions (IPCC, 1990, p. xi;

IPCC, 1992, p. 17). A validation study tested the accuracy of the IPCC forecasts against forecasts from the no-change model using data from 1851 to 1974, which is roughly the period of the Industrial Revolution. The validation study compared the forecasts from the two approaches from one-year ahead to 100-years ahead, successively updated each year from 1851. The IPCC forecast errors were over seven times larger than those for the no-change model for the 7,550 forecast comparisons. The no-change model was especially more effective for the longer forecasts horizons: Its errors were 92 percent smaller for 305 long-range forecasts—from 91 through to 100 years ahead (Green, Armstrong, and Soon 2009).

As implied above, forecasters may resist employing the no-change model because it is hard to justify big consulting fees, and the forecasts from the model lack novelty. There is little appeal for a forecast where the implication is often "don't just do something, stand there!"

Golden Rule Checklist: A possible solution

The Golden Rule Checklist provides an evidence-based standard against which forecasting procedures can be examined. Applying it requires little training and, with only 27 guidelines, someone who is familiar with a forecasting report can make a quick assessment of which of the Golden Rule guidelines are relevant and whether the forecasters followed them.

For example, the first two authors used the Checklist to assess the forecasting procedures described in the Intergovernmental Panel on Climate Change's Fourth Assessment Report (Randall, et al. 2007). They each spent ten minutes on the task, concurred that 25 of the 27 guidelines were relevant and that none were followed.

Organizations can use the Golden Rule Checklist to audit forecasting procedures, whether the forecasts are produced in-house or by consultants. A failure to follow the evidence-based guidelines could be the basis for penalties within the firm or for obtaining damages from the forecast provider. Such failures could also provide a basis for challenging government policies and regulations.

An expectation of perfect forecast accuracy is unreasonable. Perhaps as a consequence, there have been few lawsuits claiming damages arising from poor forecasts, and in these few the plaintiffs almost always fail. A recent case in Italy of seismologists' non-prediction of an earthquake is an exception, but the case may yet be overturned. Instead of punishing inaccurate forecasts, stronger cases for damages could be made on the identification of incompetent forecasting practice; that is, forecasting without following evidence-based guidelines. The

Golden Rule Checklist could provide the basis for such cases. One would hope that those who make forecasts would be motivated by this possibility to learn how to use evidence-based forecasting procedures.

As it stands, the Golden rule checklist can produce substantially more accurate forecasts than currently used methods for most forecasting problems, but especially for complex situations when there is much prior knowledge. As the evidence shows, there is little risk and much to gain by using conservative methods. To ensure objectivity, it would be wise to have independent auditors use the checklists and provide solutions.

Conclusions

Unfortunately, the trend over the past half-century toward the use of complex statistical methods and large databases has led forecasters to ignore cumulative knowledge, which can require much effort by forecasters to obtain, and to avoid simple evidence-based forecasting procedures. This trend, along with political incentives, has apparently nullified the gains that have been made possible in the ability to forecast.

The review of the evidence shows that the conservatism increases forecast accuracy consistently and by substantial amounts. The benefits of conservatism are available when formulating the forecasting problem and making forecasts with judgmental, extrapolation, and causal methods. Combining forecasts from different evidence-based methods provides forecasts that are more conservative than forecasts from single methods.

The evidence-based *Golden Rule Checklist* described in this article offers a simple low-cost path to conservative forecasting. It is also available at goldenruleofforecasting.com. The Checklist is especially useful when "big data" are available, which might otherwise encourage misunderstanding of the situation and, ironically, forecaster overconfidence. Those who commission forecasts are empowered by the Checklist to audit the proposed procedures in order to ensure that they are conservative.

Forecasters should be much more conservative than they currently are. To address this problem, use the Golden Rule Checklist. Use of each guideline should help. The full benefits of conservative forecasting, however, come at the cost of more shoe leather.

What is remarkable is that the Golden Rule has improved accuracy in all the studies that we have found. It does so no matter what is being forecast, how the guidelines were applied, how many of guidelines were used, when the studies were done, the length of the forecast horizon, the amount and quality of the data, or what criteria were used for accuracy. In addition, all of the guidelines are understandable to reasonably intelligent people, so it is easy to identify a rain-man. On the negative side, forecasters must devote effort to learn about the past (the shoe leather being replaced by Internet searches) and about evidence-based forecasting methods. In addition, clients who believe in a shaman will suffer a loss of innocence.

Acknowledgements: Useful suggestions were received from . . . Hal Arkes, Roy Batchelor, Robert Fildes, Nigel Harvey, Paul Goodwin, …

12,060 words

GoldenRule 188-clean2.docx

References

Arkes, H. R., Shaffer, V. A., & Dawes, R. M. (2006). Comparing holistic and disaggregated ratings in the evaluation of scientific presentations. *Journal of Behavioral Decision Making, 19*(5), 429-439.

Armstrong, J. S. (1970). An application of econometric models to international marketing. *Journal of Marketing Research, 7*(2), 190-198.

Armstrong, J. S. (1980a). The Seer-Sucker theory: The value of experts in forecasting. *Technology Review, 1980*(83), 16-24.

Armstrong, J. S. (1980b). Unintelligible management research and academic prestige. *Interfaces, 10*(2), 80-86.

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

Armstrong, J. S. (2001a). Combining forecasts. In J. S. Armstrong (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners* (pp. 417-439). New York: Springer.

Armstrong, J. S. (2001b). Judgmental bootstrapping: Inferring experts' rules for forecasting. In J. S. Armstrong (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners* (pp. 171-192). New York: Springer.

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

Armstrong, J. S. (2006a). Findings from evidence-based forecasting: Methods for reducing forecast error. *International Journal of Forecasting, 22*(3), 583-598.

Armstrong, J. S. (2006b). How to make better forecasts and decisions: Avoid face-to-face meetings. *Foresight: The International Journal of Applied Forecasting, 5*(2006), 3-8.

Armstrong, J. S. (2010). *Persuasive Advertising*: Palgrave MacMillan.

Armstrong, J. S. (2012). Illusions in regression analysis. *International Journal of Forecasting, 28*(3), 689-694.

Armstrong, J. S., Adya, M., & Collopy, F. (2001). Rule-based forecasting: Using judgment in time-series extrapolation. In J. S. Armstrong (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners* (pp. 259-282). New York: Springer.

Armstrong, J. S., & Andress, J. G. (1970). Exploratory analysis of marketing data: Trees vs. regression. *Journal of Marketing Research, 7*(4), 487-492.

Armstrong, J. S., & Collopy, F. (1992). Error Measures for Generalizing About Forecasting Methods: Empirical Comparisons, *International Journal of Forecasting*, 8, 69-80.

Armstrong, J. S., & Collopy, F. (1993). Causal forces: Structuring knowledge for time-series extrapolation. *Journal of Forecasting, 12*(2), 103-115.

Armstrong, J. S. & Collopy, F. (1998). Integration of statistical methods and judgment for time series forecasting: Principles from empirical research. In G. Wright & P. Goodwin (Eds.), *Forecasting with Judgment* (pp.263-393). Chichester: Wiley.

Armstrong, J. S., & Collopy, F. (2000). *Speculations about seasonal factors*. Internet publication Available from http://www.forecastingprinciples.com/paperpdf/seasonalfactors.pdf.

Armstrong, J. S., Collopy, F., & Yokum, J. T. (2005). Decomposition by causal forces: a procedure for forecasting complex time series. *International Journal of Forecasting, 21*(1), 25-36.

Armstrong, J. S., & Graefe, A. (2011). Predicting elections from biographical information about candidates: A test of the index method. *Journal of Business Research, 64*(7), 699-706.

Armstrong, J. S., Green, K. C., & Soon, W. (2008). Polar bear population forecasts: A public-policy forecasting audit. *Interfaces, 38*(5), 382-405.

Armstrong, J. S., & Grohman, M. C. (1972). A comparative study of methods for long-range market forecasting. *Management Science*, *19*(2), 211-221.

Armstrong, J. S., & Shapiro, A. C. (1974). Analyzing quantitative models. *Journal of Marketing, 38*(2), 61-66.

Ascher, W. (1978). *Forecasting: An Appraisal for Policy-makers and Planners*. Baltimore: The Johns Hopkins University Press.

Ashton, A. H., Ashton, R. H., & Davis, M. N. (1994). White-collar robotics: Leveraging managerial decision making. *California Management Review, 37*(1), 83-109.

Box, G. E., & Jenkins, G. (1970). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.

Bunn, D. W., & Vassilopoulos, A. I. (1999). Comparison of seasonal estimation methods in multi-item short-term forecasting. *International Journal of Forecasting, 15*(4), 431-443.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting, 5*(4), 559-583.

Collopy, F., & Armstrong, J. S. (1992a). Expert opinions about extrapolation and the mystery of the overlooked discontinuities. *International Journal of Forecasting, 8*(4), 575-582.

Collopy, F., & Armstrong, J. S. (1992b). Rule-based forecasting: Development and validation of an expert systems approach to combining time series extrapolations. *Management Science, 38*(10), 1394-1414.

Cuzán, A. G., & Bundrick, C. M. (2009). Predicting presidential elections with equally weighted regressors in Fair's equation and the Fiscal Model. *Political Analysis, 17*(3), 333-340.

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd & A. R. Group (Eds.), *Simple Heuristics That Make Us Smart*. Oxford: Oxford University Press

Dana, J., & Dawes, R. M. (2004). The superiority of simple alternatives to regression for social science predictions. *Journal of Educational and Behavioral Statistics, 29*(3), 317-331.

Dangerfield, B. J., & Morris, J. S. (1992). Top-down or bottom-up: Aggregate versus disaggregate extrapolations. *International Journal of Forecasting, 8*(2), 233-241.

Devarajan, S. (2011). Africa's statistical tragedy. A blog by Shanta Devarajan, World Bank Chief Economist for Africa, http://blogs.worldbank.org/africacan/africa-s-statistical-tragedy

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin, 81*(2), 95-106.

Einhorn, H. J. (1972). Alchemy in the behavioral sciences. *Public Opinion Quarterly, 36*(3), 367-378.

Eriksson, K. (2012). The nonsense math effect, *Judgment and Decision Making*, 7(6), 746-749.

Fildes, R., & Goodwin, P. (2007). Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces, 37*(6), 570-576.

Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: An empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting, 25*(1), 3-23.

Fildes, R., & Hastings, R. (1994). The organization and improvement of farket Forecasting. *The Journal of the Operational Research Society, 45*(1), 1-16.

Fildes, R., & Makridakis, S. (1995). The impact of empirical accuracy studies on time series analysis and forecasting. *International Statistical Review / Revue Internationale de Statistique, 63*(3), 289-308.

Flores, B. E, & Whybark, C. D. (1986). A comparison of focus forecasting with averaging and exponential smoothing. *Production and Inventory Management*, *27*(3), 96-103.

Flyvbjerg, B. (2013). Quality control and due diligence in project management: Getting decisions right by taking the outside view. *International Journal of Project Management, 31*(5), 760-774.

Freedman, D. A. (1991). Statistical models and shoe leather. *Sociological Methodology, 21*(1), 201-313.

Gardner, E. S. Jr. & Anderson E. A. (1997), Focus forecasting reconsidered, *International Journal of Forecasting*, *13*(4), 501-508.

Gardner, E. S. Jr,. Anderson-Fletcher, E. A., & Wickes, A. M. (2001). Further results on focus forecasting vs. exponential smoothing, *International Journal of Forecasting*, *17*(2), 287-293.

Gawande, A. (2010). *The Checklist Manifesto: How to Get Things Right*. New York: Metropolitan Books.

Goodwin, P. (2000). Improving the voluntary integration of statistical forecasts and judgment. *International Journal of Forecasting*, *16*(1), 85-99.

Goodwin, P., & Fildes, R.. (1999) Judgmental forecasts of time series affected by special events: Does providing a statistical forecast improve accuracy? *Journal of Behavioral Decision Making*, *12*(1), 37-53.

Gorr, W., Olligschlaeger, A., & Thompson, Y. (2003). Short-term forecasting of crime. *International Journal of Forecasting, 19*(4), 579-594.

Graefe, A. (2013). Equal-weights models: A low-cost and accurate approach to forecasting. *Working paper*.

Graefe, A., & Armstrong, J. S. (2011). Comparing face-to-face meetings, nominal groups, Delphi and prediction markets on an estimation task. *International Journal of Forecasting*, 27(1), 183-195.

Graefe, A., & Armstrong, J. S. (2013). Forecasting elections from voters' perceptions of candidates' ability to handle issues. *Journal of Behavioral Decision Making*, *26*(3), 295-303.

Graefe, A., Armstrong, J. S., Jones Jr., R. J., & Cuzán, A. G. (2013). Combining forecasts: An application to elections. *Forthcoming in the International Journal of Forecasting*, ssrn.com/abstract=1902850.

Green, K. C., & Armstrong, J. S. (2007a). Global warming: Forecasts by scientists versus scientific forecasts. *Energy & Environment, 18*(7-8), 997-1021.

Green, K. C., & Armstrong, J. S. (2007b). Structured analogies for forecasting. *International Journal of Forecasting, 23*(3), 365-376.

Green, K. C., Armstrong, J. S., & Soon, W. (2009). Validity of climate change forecasting for public policy decision making. *International Journal of Forecasting, 25*(4), 826-832.

Green, K. C., Soon, W., & Armstrong, J. S. (2013). Evidence-based forecasting for climate change. *Working paper*.

Haynes, A. B., Weiser, T. G., Berry, W. R., Lipsitz, S. R., Breizat, A. H. S., Dellinger, E. P., in Lapitan, M. C. M. (2009). A surgical safety checklist to reduce morbidity and mortality in a global population. *New England Journal of Medicine, 360*(5), 491-499.

Hoch, S. J. (1985). Counterfactual reasoning and accuracy in predicting personal events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(4), 719-731.

Hogarth, R. M. (1978). A note on aggregating opinions. *Organizational Behavior and Human Performance, 21*(1), 40-46.

Hogarth, R. M. (2012). When simple is hard to accept. In P. M. Todd, G. Gigerenzer & A. R. Group (Eds.), *Ecological Rationality: Intelligence in the World* (pp. 61-79). Oxford: Oxford University Press

Ioannidis, J. P. A. (2008). Why Most Discovered True Associations Are Inflated. *Epidemiology, 19*(5), 640-648.

IPCC (1990). In J. T. Houghton, G. J. Jenkins, & J. J. Ephraums (Eds.), *Climate change: The IPCC scientific assessment*. Cambridge, United Kingdom: Cambridge University Press.

IPCC (1992). In J. T. Houghton, B. A. Callander, & S. K. Varney (Eds.), *Climate change 1992: The supplementary report to the IPCC scientific assessment*. Cambridge, United Kingdom: Cambridge University Press.

Jørgensen, M. (2004). Top-down and bottom-up expert estimation of software development effort. *Information and Software Technology, 46*(1), 3-16.

Juréen, L., & Wold, H. (1953). *Demand Analysis: a Study in Econometrics*: New York; Toronto: Wiley.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrier, Straus and Giroux.

Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science, 52*(1), 111-127.

MacGregor, D. (2001). Decomposition for judgmental forecasting and estimation. In J. S. Armstrong (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners* (pp. 107-123). New York: Springer

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J. Parzen, E., & Winkler, R. (1982). The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition. *Journal of Forecasting, 1*(2), 111-153.

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K., & Simmons, L. F. (1993). The M2-competition: A real-time judgmentally based forecasting study. *International Journal of Forecasting, 9*(1), 5-22.

Makridakis, S., & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting, 16*(4), 451-476.

Makridakis, S., Hibon, M., & Moser, C. (1979). Accuracy of forecasting: An empirical investigation. *Journal of the Royal Statistical Society. Series A (General), 142*(2), 97-145.

McMahon, D. (2013). How China fudges its numbers. *The Wall Street Journal*, 19 June. http://blogs.wsj.com/chinarealtime/2013/06/19/a-rare-look-into-how-china-fudges-its-numbers/

Malkiel. Burton G. (2012). *A Random Walk Down Wall Street: The Time-Tested Strategy for Successful Investing (Tenth Edition)* New York: W.W. Norton.

Mancuso, A. C. B., & Werner, L. (2013). Review of combining forecasts approaches. *Independent Journal of Management & Production*, 4(1), 248-277.

Meehl, P. E. (1954). *Clinical versus statistical prediction*. Minneapolis: University of Minnesota Press.

Meese, R., & Geweke, J. (1984). A comparison of autoregressive univariate forecasting procedures for macroeconomic time series. *Journal of Business & Economic Statistics, 2*(3), 191-200.

Miller, D. M., & Williams, D. (2004). Damping seasonal factors: Shrinkage estimators for the X-12-ARIMA program. *International Journal of Forecasting, 20*(4), 529-549.

Namboodiri, N.K., & Lalu, N.M. (1971). The average of several simple regression estimates as an alternative to the multiple regression estimate in postcensal and intercensal population estimation: A case study. *Rural Sociology,* 36, 187-194.

Newbold, P., & Granger, C. W. J. (1974). Experience with Forecasting Univariate Time Series and the Combination of Forecasts. *Journal of the Royal Statistical Society. Series A (General), 137*(2), 131-165.

Pool, Ithiel de sola, Abelson R. P. & Popkin, S. L. (1965). *Candidates, Issues and Strategies*. Cambridge: MIT Press.

Pronovost, P., Needham, D., Berenholtz, S., Sinopoli, D., Chu, H., Cosgrove, S., Sexton, B., Hyzy, R., Welsh, R., Roth, G., Bander, J., Kepros, J., & Goeschel, C. (2006). An intervention to decrease catheter-related bloodstream infections in the ICU." *New England Journal of Medicine*, 355, 2725-2732.

Randall, D.A., Wood, R.A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., Stouffer, R. J., Sumi, A., & Taylor, K.E. (2007). Climate Models and Their Evaluation. In S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, & H. L. Miller (Eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. XXX). Cambridge, UK and New York, NY, USA: Cambridge University Press.

Rowe, G., & Wright, G. (2001). Expert opinions in forecasting: The role of the Delphi technique. In J. S. Armstrong (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners* (pp. 125-144). New York: Springer.

Runkle, D. E. (1998). Revisionist history: how data revisions distort economic policy research. *Federal Reserve Bank of Minneapolis Quarterly Review, 22*(4), 3-12.

Ryan, P., & Session, J. (2013). Sessions, Ryan Call For Halt On Taxpayer Funding For Risky High-Speed Rail Project. *U.S. Senate Budget Committee, http://www.budget.senate.gov/republican/public/index.cfm/2013/3/sessions-ryan-call-for-halt-on-taxpayer-funding-for-risky-high-speed-rail-project*.

Sanders, N. R., & Manrodt, K. B. (1994). Forecasting practices in US corporations: Survey results. *Interfaces, 24*(2), 92-100.

Sanders N. R., & Ritzman L. P. (2001). Judgmental adjustment of statistical forecasts. In J. S. Armstrong (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners* (pp. 405-416). New York: Springer.

Scheib, S. L. (2012). The streetcar swindle, *Reason*, http://reason.com/archives/2012/09/27/the-streetcar-swindle.

Schmidt, F. L. (1971). The relative efficiency of regression and simple unit predictor weights in applied differential psychology. *Educational and Psychological Measurement, 31*(3), 699-714.

Schnaars, S. P., & Bavuso, R. J. (1986). Extrapolation models on very short-term forecasts. *Journal of Business Research, 14*(1), 27-36.

Shu, L. L., Mazar, N., Gino, F., Ariely, D., & Bazerman, M. H. (2011). *When to sign on the dotted line? Signing first makes ethics salient and decreases dishonest self-reports*. Working paper. Available at http://www.researchgate.net/publication/228138457_When_to_Sign_on_the_Dotted_Line_Signing_First_Makes_Ethics_Salient_and_Decreases_Dishonest_Self-Reports/file/79e4150d9f2cec58a9.pdf.

Soyer, E., & Hogarth, R. M. (2012). Illusion of predictability: How regressions statistics mislead experts. *International Journal of Forecasting, 28*(3), 695-711.

Sparks, J. (1844). *The Works of Benjamin Franklin* (Vol. 8). Boston: Charles Tappan Publisher.

Tessier, T. H., & Armstrong, J. S. (1977). *Improving current sales estimates with econometric models*. Working paper, Available from http://www.forecastingprinciples.com/paperpdf/improvingsalesestimates.pdf.

Tetlock, P. C. (2005). *Expert political judgment*. Princeton: Princeton University Press.

Tierney, J. (1990). Betting on the planet. *New York Times*, http://www.nytimes.com/1990/12/02/magazine/betting-on-the-planet.html?pagewanted=all&src=pm.

Vranich, J., Cox, W., & Moore, A. T. (2013). California high speed rail: An updated due diligence report. *Reason Foundation*. Available at: http://reason.org/files/california_high_speed_rail_report.pdf.

Zarnowitz, V. (1967), *An appraisal of short-term economic forecasts*, Occasional Paper 104, New York: National Bureau of Economic Research.