International Journal of Forecasting 28 (2012) 689–694

Contents lists available at SciVerse ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Illusions in regression analysis

1. Introduction

The paper, "Illusion of Predictability: How Regression Statistics Mislead Experts", by Emre Soyer and Robin Hogarth, is dedicated to the memory of Arnold Zellner (1927–2010).[1] I am sure that Arnold would have agreed with me that their paper is a fitting tribute.

Given the widespread use of regression analysis, the implications of the article are important for both the life and social sciences. Employing a simple experiment, Soyer and Hogarth (2012, hereafter "S&H") show that some of the world's leading experts in econometrics can be misled by the standard statistics provided with regression analyses: t, p, F, R^2 and the like.

S&H follow a rich history on the illusions of predictability associated with the use of regression analysis on non-experimental data. A look at the history of regression analysis suggests why illusions of predictability occur and why they have increased over time—to the detriment, as S&H show, of scientific analysis and forecast-ability.[2]

2. Historical view of illusions in regression analysis

Regression analysis entered the social sciences in the 1870s with the pioneering work by Francis Galton, but "least squares" goes back to at least the early 1800s and the German mathematician Karl Gauss, who used the technique for predicting astronomical phenomena.

For most of its history, regression analysis has been a complex, cumbersome, and expensive undertaking. Consider Milton Friedman's experiences more than forty years prior to user-friendly software and the personal computer revolution. Around 1944, as part of the war effort, Friedman was asked to analyze data on the alloys used in turbine engine blades. He used regression analysis to develop a model that predicted the time to failure as a function of stress, temperature, and some metallurgical variables representing the alloy's composition. Obtaining estimates for Friedman's equation by hand and calculating test statistics would have taken a skilled analyst about three months' labor. Fortunately, a large computer, built from many IBM card-sorters and housed in Harvard's air-conditioned gymnasium, could do the calculations. Ignoring the time required for data input, the computer needed 40 hours to calculate the regression estimates and test statistics. Today, a regression of the size and complexity of Friedman's could be executed in about one second.

Friedman was delighted with the results; the model had a high R^2 and the variables were "statistically significant" at conventional levels. As a result, Friedman recommended two new improved alloys, which his model predicted would survive several hundred hours at high temperatures. Tests of the new alloys were carried out by engineers in an MIT laboratory. The result? The first alloy broke in about two hours and the second one in about three. Friedman thus concluded that one should focus on tests of outputs (forecasts) rather than statistically significant inputs. He also stated that "the more complex the regression, the more skeptical I am" (Friedman & Schwartz, 1991).

Given that doing regressions was expensive, it was sensible to rely heavily on *a priori* analyses. Opinions about the proper model (e.g., which variables are important) should not be based on opinions or untested ideas, even if they are offered by famous people. (The names Keynes and Samuelson spring to mind.) Instead, the evidence should be based on meta-analyses. Meta-analyses produce more accurate and less biased summaries than those provided by traditional reviews, as was shown by Cumming (2012).

When possible, meta-analyses should be used when making decisions as to which variables to include; specifying the expected direction of the relationships; and specifying the nature of the functional form, ranges of magnitudes of relationships, and size of expected magnitudes of those relationships. It is also important to determine relationships that can be measured outside the model based on either common knowledge (for example, adjusting for inflation or transforming the data to a per capita basis) or analyses of other data.

[1] For more about Arnold Zellner, see García-Ferrer (2011).

[2] There is also an extensive body of literature in which leading econometricians have warned of the illusions associated with regressions (e.g., see the review by Kennedy, 2002). Karni and Shapiro (1980) made fun of regression analysis procedures in their "data torture" paper.

doi:10.1016/j.ijforecast.2012.02.001

Analysts should use simple pre-specified rules to combine *a priori* estimates with estimates obtained from regression analysis (for example, one might weight each estimate of a relationship equally, then re-run the regression to estimate the coefficients for the other variables). This approach was recommended by Wold and Jureen (1953), who called it "conditional regression". I think of it as a "poor man's Bayesian analysis". I prefer it to formal Bayesian forecasting methods because of its clarity about the nature of each causal relationship and the related evidence (and because I feel a strong need for sleep whenever I try to read a paper on Bayesian forecasting). In this paper, I refer to it as an *a priori* analysis.

In the mid-1960s, I was working on my Ph.D. Thesis at MIT. While the costs of regression analysis had plunged, it still involved punch cards and overnight runs. However, the most time-consuming part of my thesis was the *a priori* analysis. Before doing any regression analyses, I gave John Little, my thesis advisor, *a priori* estimates of the coefficients for all of the variables in a demand-forecasting model. As it turned out, these purely *a priori* models provided relatively accurate forecasts on their own. I then used regression analyses of time series, longitudinal, and household data to estimate parameters. These were used to revise the *a priori* estimates. This procedure provided forecasts that were substantially more accurate than those from either extrapolation methods or stepwise regression on the complete set of causal variables that were considered (Armstrong, 1968a,b).

Despite warnings over the past half-century or more (Zellner, 2001, traces this back to Sir Harold Jeffreys in the mid-1900s), *a priori* analysis among academic researchers seems to be giving way to the belief that, given enormous databases, they can use complex methods and analytical measures such as R^2 and t-statistics to *create* models. They even try various transformations or different lags of variables to see which fit the historical data best. Einhorn (1972) 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". Ord (2012) provides a simple demonstration of how standard regression procedures, applied without *a priori* analyses, can lead one astray.

3. Forecast accuracy and confidence

We have ample evidence that regression analysis often provides useful forecasts (Allen & Fildes, 2001; Armstrong, 1985). Regression-based prediction is most effective when dealing with small numbers of variables, and large amounts of reliable and valid data, where changes are expected to be large and predictable, and there are well-established causal relationships—such as the elasticities of income, price, and advertising when forecasting demand. However, there are various illusions that reduce the forecast accuracy and lead to overconfidence in regression analysis. I discuss five of them here.

Complexity illusion: It seems common sense that complex solutions are needed for complex and uncertain problems. However, research findings suggest the opposite. For example, Christ (1960) found that simultaneous equations provided forecasts that were more accurate than those from simpler regression models when tested on artificial data, but not when tested out-of-sample using real data. My summary of the empirical evidence concluded that the increased complexity of regression models typically reduced the forecast accuracy (Armstrong, 1985, pp. 225–232). Zellner (2001) reached the same conclusion in his review of the research. He also found that many users have become disillusioned with complicated models. For example, he reported that "the Federal Reserve Bank of Minneapolis decided to scrap its complicated vector autoregressive (VAR, i.e., Very Awful Regression) models after their poor performance in forecasting turning points, etc".

Evidence favoring simplicity has continued to appear over the past quarter century. Why then is there such a strong interest in complex regression analyses? Perhaps this is due to academics' preference for complex solutions, as described by Hogarth (in press).

Somewhere, I encountered the idea that statistics were supposed to aid communication. Complex regression methods and a flock of diagnostic statistics have taken us in the other direction.

The solution is to rely upon *a priori analysis* when specifying a model. Follow Zellner's (2001) advice and use Occam's Razor.[3] In other words, keep it simple. Start with a very simple model, such as a no-change model, and then only add complexity if there is experimental evidence to support the complication. Also, do not try to estimate relationships for more than three variables in a regression (the findings of Goldstein & Gigerenzer, 2009, are consistent with this rule-of-thumb).

Illusion that regression models are sufficient: Forecasts are often derived only from what is thought to be the best model. This belief has a long history in forecasting.

For solutions, I would like to call your attention to two of the most important findings in forecasting. The first is that the naïve, or no-change, model is often quite accurate. It is to forecasting what the placebo is to medicine. This approach is especially difficult to beat in situations involving complexity and uncertainty. Here, it often helps to shrink each coefficient toward having no effect (but remember to re-run the regression to calibrate the constant term).

The second is the benefit of combining forecasts. That is, finding two or more valid forecasting methods and then calculating averages of their forecasts. For example, produce forecasts using different regression models, and then combine the forecasts. This is especially effective when the methods, models, and data differ substantially. Combining forecasts has reduced errors by between 10% and 58% (depending on the conditions) compared to the average errors of the uncombined individual forecasts (Graefe, Armstrong, Cuzán, & Jones, 2012).

Illusion that regressions provide the best linear unbiased estimators: Statisticians have devoted a lot of time to

[3] Zellner traced this idea back to Sir Harold Jeffreys in the 1930s.

showing that regressions lead to the best estimates of relationships. However, studies have shown that regression estimates produce *ex ante* forecasts that are often *less accurate* than forecasts from "unit weights" models. Schmidt (1971) was one of the first to test this idea, and found that unit weights were superior to regression weights when the regressions were based on many variables and small sample sizes. Dana and Dawes (2004) and Einhorn and Hogarth (1975) show the conditions under which regression is and is not more effective than equal weights.

One useful characteristic of regression estimates is that they become more conservative as the uncertainty increases. Unfortunately though, some aspects of uncertainty are ignored. For example, the coefficients can "get credit" for important excluded variables that happen to be correlated with the predictor variables. Adding variables to the regression cannot solve this problem.

In addition, analysts searching for the best fit, and publication practices favoring statistically significant results, often defeat conservatism. Thus, regressions typically over-estimate changes. Ioannidis (2005, 2008) provides more reasons why regressions over-estimate changes.

One solution is to combine forecasts, and include the naïve model among the alternative models. In particular, for problems involving time series, damp the forecasts more heavily toward naïve model forecasts as the forecast horizon increases, to reflect the effects of the increasing uncertainty in the more distant future.

Illusion of control: Users of regressions assume that by putting variables into the equation they are somehow controlling for these variables. However, this only occurs for experimental data. Adding variables does not mean controlling for them in non-experimental data, because many variables typically co-vary with other predictor variables. The problem becomes worse as more variables are added to the regression. Large sample sizes cannot resolve this problem, so statistics on the numbers of degrees of freedom are misleading.

One solution is to use evidence from experimental studies to estimate the effects, and then adjust the dependent variable for these effects.

"Fit implies accuracy" illusion: Analysts assume that models with a better fit provide more accurate forecasts. This ignores the research showing that the fit bears little relationship to the *ex ante* forecast accuracy, especially for time series. Typically, the fit improves as the complexity increases, while the *ex ante* forecast accuracy decreases—a conclusion that Zellner (2001) traced back to Sir Harold Jeffreys in the 1930s. In addition, analysts use statistics to improve the fit of the model to the data. In one of my Tom Swift studies, Tom used standard procedures, starting with 31 observations and 30 potential variables. He used stepwise regression and only included variables where t was greater than 2.0. Along the way, he dropped three outliers. The final regression had eight variables and an R^2 (adjusted for degrees of freedom) of 0.85. Not bad, considering that the data were from Rand's book of random numbers (Armstrong, 1970).

I traced studies of this illusion back to at least 1956 in an early review of the research on fit and accuracy (Armstrong, 1985). Studies have continued to find that the fit is not a good way of assessing predictive ability (e.g., Pant & Starbuck, 1990).

The obvious solution is to avoid the use of t, p, F, R^2 and the like when using regressions.

4. Using regression for decision-making

Regression analysis provides an objective and systematic way of analyzing data. As a result, decisions based on regression are less likely to be subject to bias, they are consistent, the basis for the decisions can be fully explained, and they are generally useful. The gains are especially well documented when compared to judgmental decisions made based on the same data (Armstrong, 2001; Grove & Meehl, 1996). However, two illusions, those of statistical significance and correlations, can reduce the value of regression analysis.

Statistical significance illusion: S&H incorporate the illusion due to tests of statistical significance. Meehl (1978) concluded that "reliance on merely refuting the null hypothesis…is basically unsound, poor scientific strategy, and one of the worst things that ever happened in the history of psychology".

Schmidt (1996) offered the following challenge: "Can you articulate even one legitimate contribution that significance testing has made (or makes) to the research enterprise (i.e., any way in which it contributes to the development of cumulative scientific knowledge)?" One might also ask whether there is a study in which statistical significance improves decision-making. In contrast, it is easy to find cases where statistical significance has harmed decision-making. Ziliak and McCloskey (2008) document devastating examples taken from across the sciences. To offer another example, Hauer (2004) demonstrates harmful decisions related to automobile traffic safety, such as the "Right-turn-on-red decision". Cumming (2012) describes additional examples of the harm caused by the use of statistical significance.

The commonly recommended solution is to use confidence intervals and avoid the use of statistical significance. Statisticians argue that statistical significance provides the same information as confidence intervals. However, the issue is the way in which people use the information. Significance levels lead to confusion even among leading researchers. Cumming (2012, pp. 13–14) describes an experiment showing that when researchers in psychology, behavioral neuroscience, and medicine were presented with a set of results, only 40% of the 55 participants who used significance levels to guide their interpretation reached correct conclusions. In stark contrast, 95% of the 57 participants who thought in terms of confidence intervals reached correct conclusions.

Correlation illusion: We all claim to understand that correlation is not causation. Correlations might occur because A causes B, or because B causes A, or because they are each related to C, or they could be spurious. However, when presented with sophisticated and complex regressions, people often forget that; and researchers in medicine, economics, psychology, finance, marketing, sociology, and so on, fill journals and newspapers with interesting but erroneous – and even costly – findings.

In one study, we had an opportunity to compare the findings from experiments with those from analyses of non-experimental data for 24 causal statements. The directional effects differed for 8 of the 24 comparisons (Armstrong & Patnaik, 2009). My conclusion is that analyses of non-experimental data are often misleading.

This illusion has led people to make poor decisions about such things as what to eat or drink (e.g., coffee, once bad, is now good for health), what medical procedures to use (e.g., the frequently recommended PSA test for prostate cancer has now been shown to be harmful), and what economic policies the government should adopt in recessions (e.g., trusting the government to be more efficient than the market).

According to Zellner (2001), Sir Harold Jeffreys warned of this illusion, and, in 1961, referred to it as the "most fundamental fallacy of all".

The solution is to base causality on meta-analyses of experimental studies.

5. Discussion

An obvious conclusion from the study by S&H is to de-emphasize descriptive statistics for regression packages. Software developers should provide statistics on the abilities of alternative methods to produce accurate forecasts on holdout samples as the default option. They could still allow users to push a button to access the traditional regression statistics, but a warning label should be provided near the button.

S&H, echoing Friedman, emphasize that scientific theories should be tested for their predictive abilities, relative to other methods. Ord (2012) deplores the fact that few regression packages assist in such analyses. It would be helpful if software providers would focus on *ex ante* testing by making it easy to simulate the forecasting situation. For cross-sectional forecasts, use jackknifing—that is, use all but one data point to estimate the model, then predict for the excluded observation, and repeat until each observation in the data has been predicted. For time series, withhold data, then use successive updating and report the accuracy for each forecast horizon. These testing procedures are less likely to lead to overconfidence, because they include the uncertainty from errors due to over-fitting, in addition to the errors in forecasting the predictor variables.

Software packages should provide statistics that allow meaningful comparisons among methods to be made (Armstrong & Collopy, 1992). The MdRAE (Median Relative Absolute Error) was designed for such comparisons, and many software packages now provide this statistic, so it should be among the default statistics, along with a link to the literature on this topic. Do not provide the RMSE (Root Mean Square Error), as it is unreliable and uninformative.

Allen and Fildes (2001) note that since 1985 there has been a substantial increase in the amount of attention paid to *ex ante* forecast comparisons in published papers. This is consistent with the aims stated at the founding of the *Journal of Forecasting* in 1982, and, subsequently, the *International Journal of Forecasting*.

Regression analysis is clearly one of the most important tools available to researchers. However, it is not the only game in town. Researchers made scientific discoveries about causality prior to the availability of regression analysis, as was shown by Freedman (1991) in his paper aptly titled "Statistical models and shoe leather". He demonstrates how major gains were made in epidemiology in the 1800s. For example, John Snow's discovery of the cause of cholera in London in the 1850s came about from "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". These three characteristics of good science, as described by Freedman, are missing from most regression analyses that I see in journals today.

We would be wise to recall a method that Ben Franklin used for addressing the issue of how to make decisions when many variables are involved (Sparks, 1844).[4] He suggested that one should list listing the variables related to the choice between two options, identify which option is better for each variable, weight the variables, and then add. Pick the option that has the highest score. Andreas Graefe and I have built upon Franklin's advice in developing what we call the *index method* for forecasting. The method relies on *a priori* analysis (preferably experimental findings) alone to determine which variables are important and the direction of the effect for each variable. Franklin suggested differential weights, but the literature discussed above suggests that unit weights are a good place to start. Regression analyses can then be used to estimate the effects of an index score. The index model allows analysts to take into account "the knowledge present in a field", as recommended by Zellner (2001). The few tests to date suggest that the index method provides useful forecasts when there are many important variables and substantial prior knowledge (e.g., see Armstrong & Graefe, 2011).

It can be expensive to do *a priori* analyses for complex situations. For example, I am currently involved in a project for forecasting the effectiveness of advertisements by using the index method. I have spent many hours over a 16-year period summarizing the knowledge. This led to a total of 195 principles (causal condition/action statements), each based on a meta-analysis when there was more than one source of evidence. The vast majority of them were sufficiently complex that neither prior experience nor regression analyses were able to discover them. They were formulated thanks to a century of experimental and non-experimental studies. None of these evidence-based principles were found in the advertising textbooks and handbooks that I analyzed. Many are counter-intuitive, and are often violated by advertisers (Armstrong, 2011). Early findings suggest that an index model can provide useful forecasts in this situation. In contrast, regression analyses have met with repeated failures in this area because there may be over 50 principles used – or misused – in a given ad.

As S&H suggest, further experimentation is needed. We need experiments for assessing the ability of alternative techniques to improve the accuracy when tested on

[4] Benjamin Franklin's letter to Joseph Priestley can be found at: http://www.procon.org/viewbackgroundresource.asp?resourceID=1474.

large samples of forecasts of holdout samples. Journal editors should commission such studies. The work of Allen and Fildes (2001) provides an obvious starting point for research topics, as they developed principles for the effective use of regressions based on existing knowledge. They also describe the evidence for the principles. For example, on the matter of evidence for one of the diagnostic statistics, they state (p. 311), "The mountains of articles on autocorrelation testing contrast with the near absence of studies on the impact of autocorrelation correction on [*ex ante*] forecast performance".

Unfortunately, software developers typically fail to incorporate evidence-based findings about which methods work best, and the statisticians who work on new forecasting procedures seldom cite studies that test the relative effectiveness of various methods for forecasting (Fildes & Makridakis, 1995). To make the latest findings easily available to forecasters, the International Institute of Forecasters supports the ForPrin.com website. The goal is to present forecasting principles in such a way that they can be understood by anyone who uses forecasting techniques such as regressions.

S&H recommend visual presentation, and Ziliak (2012) supports them. This seems like a promising avenue. Consider, for example, the effectiveness of the communication provided by Anscombe's Quartet (see Wikipedia).

6. Conclusions

In summary, do not use regression to search for causal relationships, and do not try to predict by using variables that were not specified in the *a priori* analysis. Thus, avoid data mining, stepwise regression, and related methods.

Regression analysis can play an important role in the analysis of non-experimental data. Various illusions can reduce the accuracy of regression analysis, lead to a false sense of confidence, and harm decision-making. Over the past century or so, various effective solutions for dealing with the illusions have been developed. The basic problem is that the solutions are often ignored in practice. S&H shows that this is true even among the world's leading researchers. Researchers might benefit from a systematic check of their use of regressions to ensure that they have taken steps to avoid the illusions. Reviewers could help to make researchers aware of solutions to the illusions. Software providers should inform their users.

To me, S&H's key recommendation is to conduct experiments comparing different approaches to developing and using models. It is remarkable that so little experimentation has been done over the past century to determine which regression methods work best under which conditions.

Arnold Zellner would not have been surprised by these conclusions.

Acknowledgments

Peer review was vital for this paper. Four people reviewed two versions of this paper: P. Geoffrey Allen, Robert Fildes, Kesten C. Green, and Stephen T. Ziliak. Excellent suggestions were also received from Kay A. Armstrong, Jason Dana, Antonio García-Ferrer, Andreas Graefe, Daniel G. Goldstein, Geoff Cumming, Paul Goodwin, Robin Hogarth, Philippe Jacquart, Keith Ord, and Christophe Van den Bulte. However, this does not imply that the reviewers agree with all of the conclusions in this paper.

References

Allen, P. G., & Fildes, R. (2001). Econometric forecasting. In J. S. Armstrong (Ed.), *Principles of forecasting*. Boston: Kluwer Academic Publishers.

Armstrong, J. S. (1968a). Long-range forecasting for a consumer durable in an international market. *Ph.D. Thesis*. MIT.

Armstrong, J. S. (1968b). Long-range forecasting for international markets: the use of causal models. In R. L. King (Ed.), *Marketing and the new science of planning*. Chicago: American Marketing Association. Available at: http://goo.gl/oKYUH.

Armstrong, J. S. (1970). How to avoid exploratory research. *Journal of Advertising Research*, *10*(4), 27–30.

Armstrong, J. S. (1985). *Long-range forecasting*. New York: John Wiley.

Armstrong, J. S. (2001). Judgmental bootstrapping: inferring experts' rules for forecasting. In J. S. Armstrong (Ed.), *Principles of forecasting*. Boston: Kluwer Academic Publishers.

Armstrong, J. S. (2011). Evidence-based advertising: an application to persuasion. *International Journal of Advertising*, *30*(5), 743–767.

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

Armstrong, J. S., & Patnaik, S. (2009). Using quasi-experimental data to develop principles for persuasive advertising. *Journal of Advertising Research*, *49*(2), 170–175.

Christ, C. F. (1960). Simultaneous equation estimation: any verdict yet? *Econometrica*, *28*, 835–845.

Cumming, G. (2012). *Understanding the new statistics: effect sizes, confidence intervals and meta-analysis*. New York: Routledge.

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.

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

Einhorn, H. J., & Hogarth, R. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance*, *13*, 171–192.

Fildes, R., & Makridakis, S. (1995). The impact of empirical accuracy papers on time series analysis and forecasting. *International Statistical Review*, *63*, 289–308.

Freedman, D. A. (1991). Statistical models and shoe leather. *Sociological Methodology*, *21*, 291–313.

Friedman, M. A., & Schwartz, A. J. (1991). Alternative approaches to analyzing economic data. *American Economic Review*, *81*, 39–49.

García-Ferrer, A. (2011). In memory of Arnold Zellner, a great person and scientist. *International Journal of Forecasting*, *27*, 961–967.

Goldstein, D. G., & Gigerenzer, G. (2009). Fast and frugal forecasting. *International Journal of Forecasting*, *25*, 760–772.

Graefe, A., Armstrong, J. S., Cuzán, A. G., & Jones, R. J. Jr. (2012). *Combining Forecasts: An Application to Election Forecasts*. Working Paper.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: the clinical-statistical controversy. *Psychology, Public Policy, and Law*, *2*, 293–323.

Hauer, E. (2004). The harm done by tests of significance. *Accident Analysis and Prevention*, *36*, 495–500.

Hogarth, R. M. (in press). When simple is hard to accept. In P. M. Todd & G. Gigerenzer (Eds.), *Ecological rationality: intelligence in the world*. Oxford: Oxford University Press.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*, 696–701.

Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology*, *19*, 640–648.

Karni, E., & Shapiro, B. K. (1980). Tales of horror from ivory towers. *Journal of Political Economy*, *88*(1), 210–212.

Kennedy, P. (2002). Sinning in the basement: what are the rules? the ten commandments of applied econometrics. *Journal of Economic Surveys*, *16*, 569–589.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*, 806–834.
Ord, K. (2012). The illusion of predictability: a call to action. *International Journal of Forecasting*, *28*(3), 717–718.
Pant, P. N., & Starbuck, W. H. (1990). Innocents in the forest: forecasting and research methods. *Journal of Management*, *16*, 433–446.
Schmidt, F. L. (1971). The relative efficiency of regression and simple unit predictor weights in applied differential psychology. *Educational and Psychological Measurement*, *31*, 699–714.
Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: implications for training of researchers. *Psychological Methods*, *1*, 115–129.
Soyer, E., & Hogarth, R. (2012). The Illusion of predictability: how regression statistics mislead experts. *International Journal of Forecasting*, *28*(3), 695–711.
Sparks, J. (1844). *The works of Benjamin Franklin*. Boston: Charles Tappan Publisher.
Wold, H., & Jureen, L. (1953). *Demand analysis*. New York: John Wiley.
Zellner, A. (2001). Keep it sophisticatedly simple. In H. Keuzenkamp, & M. McAleer (Eds.), *Simplicity, inference, and modelling: keeping it sophisticatedly simple*. Cambridge: Cambridge University Press.
Ziliak, S. T. (2012). Visualizing uncertainty: on Soyer's and Hogarth's "The illusion of predictability: how regression statistics mislead experts". *International Journal of Forecasting*, *28*(3), 712–714.
Ziliak, S. T., & McCloskey, D. N. (2008). *The cult of statistical significance: how the standard error costs us jobs, justice, and lives*. Ann Arbor: The University of Michigan Press.

J. Scott Armstrong (Ph.D., MIT, 1968), Professor of Marketing at the Wharton School, University of Pennsylvania, is a founder of the *Journal of Forecasting*, the *International Journal of Forecasting*, and the International Symposium on Forecasting. He is the most cited professor in the history of the Wharton School's Marketing Department. He is the creator of forecastingprinciples.com and editor of *Principles of Forecasting* (Kluwer 2001). In 1996, he was selected as one of the first "Honorary Fellows" by the International Institute of Forecasters.

J. Scott Armstrong
The Wharton School, University of Pennsylvania, Philadelphia PA 19104, United States
E-mail address: armstrong@wharton.upenn.edu.