Making Progress in Forecasting

Forthcoming in the
International Journal of Forecasting

J. Scott Armstrong
Wharton School, University of Pennsylvania
armstrong@wharton.upenn.edu

Robert Fildes
Centre for Forecasting, Lancaster University
R.Fildes@lancaster.ac.uk

April 17, 2006

Twenty-five years ago, the International Institute of Forecasters was established "to bridge the gap between theory and practice." Its primary vehicle was the *Journal of Forecasting* and is now the *International Journal of Forecasting*. The Institute emphasizes empirical comparisons of reasonable forecasting approaches. Such studies can be used to identify the best forecasting procedures to use under given conditions, a process we call evidence-based forecasting. Unfortunately, evidence-based forecasting meets resistance from academics and practitioners when the findings differ from currently accepted beliefs. As a consequence, although much progress has been made in developing improved forecasting methods, the diffusion of useful forecasting methods has been disappointing. To bridge the gap between theory and practice, we recommend a stronger emphasis on the method of multiple hypotheses and on invited replications of important research. It is then necessary to translate the findings into principles that are easy to understand and apply. The Internet and software provide important opportunities for making the latest findings available to researchers and practitioners. Because researchers and practitioners believe that their areas are unique, we should organize findings so that they are relevant to each area and make them easily available when people search for information about forecasting in their area. Finally, progress depends on our ability to overcome organizational berriers.

Keywords

Barriers to implementation; Evidence-based forecasting; forecasting practice; forecasting software; freeware; replication.

In 1980, Spyros Makridakis invited us along with Bob Carbone to join him in developing the emerging field of forecasting research (Fildes and Nikolopoulos, 2006). In 1981 our group founded the International Institute of Forecasters and launched the International Symposium on Forecasting. We inaugurated the *Journal of Forecasting* (*JoF*) the next year and then, in 1985, the *International Journal of Forecasting (IJF).* These were exciting days. We viewed ourselves as visionaries and set an ambitious objective: "To unify the field and to bridge the gap between theory and practice". We planned to publish research to identify the best procedures for forecasting. In this paper, we look back at our successes and failures in the last 25 years and suggest ways to accelerate future progress.

Foundations of Progress

In 1890, Chamberlin, a geologist and later president of the American Association for the Advancement of Science, examined why some areas of science progress rapidly while others do not. He concluded that those sciences that test multiple reasonable hypotheses were the ones that made the most progress. A properly conducted test of multiple reasonable hypotheses about an important problem should always be of some value: it can tell which hypotheses (or methods) are most appropriate under given conditions.

Despite reprintings of Chamberlin's paper (e.g., Chamberlin 1965), a follow-up by Platt (1964) and much attention to those papers (a March 2006 Google search of titles and last names produced 264 web-site cites to Chamberlin's paper and almost 800 to Platt's), the management sciences seldom follow this advice. For example, in the Armstrong, Brodie & Parsons' (2001) publication audit of over 1,700 empirical papers in six leading marketing journals during 1984-1999, only 13% of used the method of multiple hypotheses, and that percentage had been declining over time. Furthermore, most of the variations among the hypotheses were minor: when coding for "major variations with independent hypotheses," only 3% of the papers qualified. Only 14% of the multiple hypotheses studies examined conditions. In all, only a fraction of one percent of the empirical studies tested substantially different hypotheses and included conditions.

A different coding approach to a publication audit was taken by Reisman and Kirchnick (1994). They examined papers from two reputable management science journals. They found that fewer than 13% of papers compared methods or models on real data.

Using Chamberlin's conclusion along with other findings on scientific publishing (Armstrong, 1982), we developed procedures for the *JoF*. These procedures were later adopted for the *IJF*. The aims of the *IJF* state:

> "The intention is to make forecasting useful and relevant to decision and policy makers who need forecasts. The journal places strong emphasis on empirical studies, evaluation activities, implementation research, and ways of improving the practice of forecasting. For empirical studies, the journal gives preference to papers that compare "multiple hypotheses" (two or more reasonable hypotheses)."

The hypotheses should state the conditions under which the tests are to be carried out, for example by specifying the data characteristics (e.g. high uncertainty, large changes expected, known causality, short series, outliers, start-up series, seasonality, discontinuities). Doing so is critical for purposes of generalizing to new situations. The extensive research efforts expended on finding the one best method have ended in failure.

The testing of reasonable hypotheses implies that current methods should be included as well as established benchmarks. It is important to test the ability of a model to forecast out-of-sample, so as to simulate the forecasting situation. However, tests of conditional forecasts – which assume that future values are known - are useful for assessing the ability to forecast the effects of policy changes.

The testing procedure should contain explicit pre-determined criteria as to what constitutes 'success' or better performance. Ideally, all relevant criteria should be included (e.g., evaluations of the value of forecast accuracy as well as understandability and ease of use). Measures of the magnitude of errors should be included.

The multiple hypotheses approach can be challenging, as there is a need to use each approach in an effective way; that is, one should not set up straw-men by using inappropriate methods. Also, once the research is completed, the findings can be difficult to publish, especially when they refute current beliefs. Experimental studies have shown that reviewers for scientific journals routinely reject findings that conflict with current beliefs about important topics. In Mahoney's (1977) experiment, half of the reviewers received a "submitted paper" in which the results agreed with the beliefs commonly held by these reviewers, and half received a paper with results that contradicted the current beliefs. The methodology was identical for both versions. The version that refuted current beliefs was much more likely to be rejected by reviewers; the rejections were primarily due to "poor methodology."

Armstrong and Hubbard's (1991) survey of the editors of 20 psychology journals found that few controversial papers are published. For all 20 journals, the editors reported only one useful and surprising paper that was unanimously accepted by referees in the previous two years. That represents one paper for 40 journal-years.

Resistance by researchers was apparent in the commentary on the precursor to the M-Competition (Makridakis & Hibon, 1979). In retrospect, many of their comments were overly harsh with respect to some of the M-Competition findings, such as the lack of value for some commonly accepted extrapolation methods. As Makridakis explains (Fildes & Nikolopoulos, 2006), the results were not welcomed by editors or referees, and there was an immediate rejection from the *Journal of the American Statistical Association*.

We reasoned that by actively seeking papers with surprising and useful findings, our forecasting journals could make useful contributions to the science and practice of forecasting.

Progress in the First 25 Years of the International Institute of Forecasters

In Quebec, in 1981, we held the first International Symposium on Forecasting (ISF) in an effort to obtain papers for the initial volume of the *Journal of Forecasting (JoF)*. Following our philosophy of encouraging submissions that would challenge existing beliefs, we decided not to review and screen papers for the Symposium. With minor exceptions, it is a practice that continues. We believe that researchers can judge when they have something worthwhile to present and the attendees are able to judge what is relevant to their own interests. The response to the first ISF was so successful (175 papers and 250 participants) that we planned a second one, this time in Istanbul. After the Istanbul ISF, it was clear that the symposia provided a useful vehicle to reach our objectives. The third symposium in Philadelphia attracted over 1,000 attendees and was addressed by two Nobel Prize laureates, Larry Klein and Wassily Leontief. Twenty-five years later, the annual ISFs continue to prosper and to attract the world's leading researchers.

The publication of the *JoF* was met with considerable enthusiasm. By 1983, its citation impact factor of 1.7 was second out of the 25 ranked management journals, and it was first out of the 26 ranked planning journals (*SSCI Journal Citation Reports*, 1983). Given its high level of subscriptions at the time, reportedly over 1,000 subscriptions, it was profitable for our publisher, John Wiley - - so much so that it eventually caused conflict over how to share the funds. In 1985, the editors and the publisher parted ways. (The break-up was partly due to our poor forecasting about the expected outcome of our negotiation strategy.) As a consequence, we founded the *IJF*. By 1988, the citation impact factor of the *IJF* (.84) was nearly equal to that for *Management Science*.

As editors, we did not review papers that failed to meet the objectives of the *IJF*. We gave a clear passage to those that did. This meant that on occasion we ignored reviewers' recommendations to reject papers that used the method of multiple reasonable hypotheses. This applied, for example, to Lawrence, Edmundson and O'Connor (1985); it is a useful paper that also became one to the most influential papers as listed in Fildes (2006).

The M-competition (Makridakis, et al, 1982) is a good example of a multiple hypotheses study on an important topic. It included the leading methods for forecasting by extrapolation and the forecasts were tested on a holdout *ex ante* sample that was known only to the administrator of the competition. Given its importance, we saw no need to put the paper through a traditional review; the issue was how to improve it and to get it into print quickly. This paper was followed by peer review on the M-Competition (Armstrong & Lusk, 1983).

We relied heavily on invited papers for the *JoF* and *IJF*. To do this, we searched journals and conference proceedings for authors who had tested alternative methods and invited them to publish papers. This strategy of inviting papers worked well. As we found later (Armstrong & Pagell, 2003), invited papers have had about 20 times as strong an "impact" as those accepted via traditional channels. This impact factor was based on the number of ISI citations and also on whether the paper had been identified as useful in the development of forecasting principles. Invited papers are also less expensive to process – for authors, editors, and reviewers. For example, there is no wasted reviewer effort, as virtually all of the invited papers are published.

When opportunities arose, we moved swiftly. For example, Allan Murphy called one of us to describe work being done on combining forecasts by a faculty member at the University of Oregon. It summarized research showing that equal-weights combining performed well in comparison with more sophisticated methods. By the end of the conversation, we agreed that it should be a featured talk with commentators at the next International Symposium on Forecasting (June 1989), and that it would be published in the *IJF*. The resulting paper, Clemen (1989), was published along with commentary and became one of the most important and highly cited papers in the *IJF*.

The objectives of the journals and the Institute have remained consistent, stressing the importance of forecasting research in bridging the gap between theoretical contributions and organizational practice. The reviewing procedures have withstood the test of time, with one major exception: By asking reviewers for their recommendations on whether a paper should be published, we screen out important papers as well as bad papers. For example, Gans and Shepard (1994), when interviewing eminent economists, learned that these economists' best papers were typically rejected by reviewers when first submitted. It is also common for Nobel Prize winners to talk about how their findings were initially rejected.

For reasons that are not entirely clear, the forecasting journals have lost some ground over the years. For example, the *IJF* and the *JoF* are no longer amongst the most frequently cited journals in the category of 'Business, Management and Planning.' The *IJF* impact factor for 2004 was .51; New York

University professor Bill Starbuck reports that the *IJF* ranked 386 of 508 journals he regards as relevant to business (pages.stern.nyu.edu/~wstarbuc).

In addition, if you contrast the earliest issues with recent issues, you might agree that fewer of the recent papers seem to contain important findings about how to improve forecasting. For example, in recent years (2001-2004), 27 of 57 papers using a multiple hypothesis approach have focussed on financial modelling, an area where valuable findings have been rare.

We suspect that the reviewing process has shifted too much toward "fairness," using reviewers' recommendations as the primary basis for acceptance. The *IJF* has rejected and delayed what we felt to be some of our best papers, For example, an empirical comparison of alternative methods for extrapolation (Armstrong, Collopy & Yokum 2005) was under review for 13 years by the IJF before it was finally published. The additional studies that were added to this paper did not alter the basic conclusions.

When multiple-hypotheses studies produce surprising findings about important topics, resistance continues even after papers are published. This was shown by Fildes and Makridakis' (1995) analysis of papers that cited the M-Competition studies. Few forecasting studies published in statistical journals have cited the M-Competition findings.

There is usually a gap between influential academic innovations and evidence on their value. Research often proceeds for years without multiple hypotheses studies to assess their value. Fildes (2006) shows that while new forecasting methods, such as neural nets, have been developed in the past 25 years, those developed outside the forecasting community typically were not evaluated through multiple hypotheses studies until many years later.

Diffusion of useful methods has been slow. Textbooks have ignored many of the key findings from empirical work (Cox & Loomis, 2001). For example, few books mention damped trend as an extrapolative method despite it being shown to improve accuracy in multiple hypotheses studies since 1985.

Software companies are perhaps the predominant influence on organizational forecasting practices. They influence the forecasting methods that are considered for use by practitioners and academics, and they implement the methods. However, software companies have been slow to adopt methods that have been shown to improve accuracy. For example, damped trend is included in only a small number of forecasting software programs and causal forces are ignored. Also, findings from the judgmental forecasting literature have had little influence on the design of software for forecasting (Fildes et al, 2006)

Prospects for Progress

There is a gap between academic innovations and demonstrated effectiveness. There is also a gap between demonstrated effectiveness and what is used. How can these gaps be bridged?

Closing the gap between innovations and effectiveness

More emphasis should be given to the use of multiple hypotheses for the development of evidence-based forecasting. Currently, such papers represent a very small percentage of the papers published on forecasting. Armstrong and Pagell (2003) estimated this to be about 3%. Armstrong (2006), using a different approach, estimated this to be about 2%.

Researchers should devote more effort to conducting empirical studies of multiple reasonable hypotheses for important problems. These studies should include the conditions under which the evidence-based principles should be applied. Even such exemplary studies as the M-Competitions have failed to include explicit conditions in their hypotheses.

Journals should encourage the publication of papers that refute current beliefs and methods, rather than to reject and delay them. An editorial in the first issue of the *Journal of Forecasting* addressed this issue (Makridakis, et al. 1982). It announced the "Note to Referees, " a procedure whereby authors could submit the design of their study for review. That is, they could withhold the findings. This announcement still appears on the back page of the cover of each issue of the *International Journal of Forecasting*. Unfortunately, the procedure is almost never used.

To encourage research that uses the method of multiple hypotheses for important problems, editors can invite papers. This approach removes uncertainty for the author, saves reviewing time for the journals (less reviewing is required because typically you have to review five papers to get one paper whereas is nearly one-to-one for invited papers), and leads to a higher impact in terms of usefulness. The International Institutes of Forecaster's grant scheme (supported by SAS, Inc) has attempted to stimulate such research by offering financial support. Invited papers would go through a review process and the authors would be asked to respond to the reviewers, explaining their responses to the editor. However, the final decisions on changes would rest with the author. Open peer review might also be published for these papers

Another approach would be to avoid asking a reviewer to make a recommendation as to whether a paper should be published. Rather, reviewers would be asked how to improve papers. This approach has been considered by the editors of the *International Journal of Forecasting*.

Journals could examine submitted papers with the assumption that those using multiple reasonable hypotheses for important problems should be published. The editor can allocate space among papers depending on usefulness of the findings and the quality of the evidence.

Studies are seldom replicated in forecasting. The need for replications was illustrated by Hubbard and Vetter (1996). They found that few replications were published in the management sciences (6.2% of empirical studies) and that publication time-lags between the initial paper and the replication are long (4.3 years on average). Most importantly, they found conflicts between the original findings and those from the replication for almost half of the published replications.

To address this problem, journal editors might, upon publication, earmark papers that have important findings in need of replications. In addition, replications submitted through traditional channels should be guaranteed some manner of publication, assuming that they are not clearly inept. This might be a short printed version of the replication findings along with a reply by the original author, with details on the Internet, followed by continuing open peer review. Given that the peer reviews would be published, a rational author could withdraw a paper that has been shown to be inept. Important replications where the conditions for success have been explored would receive more print space.

To aid in replications, full disclosure about the data and methods should be made available on the Internet concurrent with a paper's publication. The level of detail should be sufficient to allow for replication. Item 21 of the *Instructions to Authors* printed on the back of every issue of the *IJF* requires authors to file their data and methods on the Internet on the International Institute of Forecasters website. When we checked this site, we found only the M-Competition data. Fortunately for editors, compliance with archiving is now simpler than previously. Authors can be required to provide footnotes showing how to find the data and methods on the Internet. This makes it easy for readers and those considering replications to find what they need to carry out a replication. Replication footnotes are currently required for publication in such journals as the *American Journal of Political Science*. To ensure that the sources are not lost, the sources should also be downloaded and made available on the journal's web site.

Researchers gain by providing full disclosure. Gleditsch, Metelits and Strand (2003), in their analysis of 416 papers published in the *Journal of Peace Research,* found that papers that *offered data in any form* were cited twice as often as comparable papers without such an offer. (Their study controlled for many variables.) The availability of the M-Competition data, for example, led to many extensions and to a large number of citations for the M-Competition studies.

Closing the gap between demonstrated effectiveness and what is used

Armstrong and Pagell (2003) showed that academic journals have been the dominant source of useful findings in forecasting. They estimated that 89% of findings have been announced first in academic journals, with an additional 3% in working papers intended for publication; books provided 6%, and 2% were from conference proceedings. Only a small fraction of one percent came from trade publications.

Journals will continue to play a major role in the future. Certainly they are vital for archiving. They also provide a signal of quality. Their importance is likely to increase with use of the Internet, as it is now easier to find and obtain journal articles. In addition, journals could provide continuous and open peer review along with statistics on downloads, time per download, ratings, and citations. We fear that it will be the rare paper that encourages such engagement however.

In order to reach a large audience, findings must be understandable, readily available, and easy to implement. Most journals fail on all three accounts.

Of papers published in forecasting, few have any immediate value to practising forecasters. As a result, it is expensive and time-consuming to search the literature for useful findings. Authors can reduce the search time if they would summarize their findings and methods in their abstracts. Armstrong and Pagell (2003) found that only 17% of a recent two-year sample of papers in the *IJF* mentioned the findings and methods in the abstract; in the *JoF*, only 9% described them.

Researchers with useful findings can make full-text copies available online (e.g., on their own sites, or on sites such as RePEc and SSRN). If a publisher is not agreeable to posting full text, authors can post the working paper that was submitted to the journal.

Perhaps the biggest need is to summarize the useful research in the form of principles (guidelines, advice, actions, recommendations) so that practitioners can apply them properly. To address this problem, in 2005 the IIF founded a journal, *Foresight: The International Journal of Applied Forecasting*. The goal is to translate research findings into useful and intelligible advice. Textbooks writers should also take up the challenge to summarize useful knowledge in an understandable manner; as shown by Cox and Loomis (2001), even the best textbooks ignore most principles, and few textbooks provide evidence to support their recommendations.

The principles must be easily accessible on the Internet. The IIF supports forecastingprinciples.com, with its avowed goal of using evidence-based forecasting to "summarize all useful knowledge about forecasting so that it can be used by researchers, practitioners, and educators." Researchers are invited to submit evidence that has been excluded or that is contrary to the principles, and action will be taken

so that the site contains all useful knowledge. Important findings that are understandable and actionable can be posted on the site.

As of 2006, visits to forecastingprinciples.com were running at the rate of about 100,000 per year (with additional traffic on the subsidiary sites). Considering the number of people around the world who are making important forecasts, the site is reaching only a small portion of the market. We must find out how to reach more people.

Much of what has been learned about evidence-based forecasting can be integrated into software, freeware, and decision-support systems. If software includes the latest principles, the principles will be applied unless the user over-rides them. Although dedicated business forecasting packages do a better job of incorporating research findings than do general-purpose statistical packages, even the best packages overlook many principles, as documented by Tashman and Hoover (2001). Furthermore, within the field of supply-chain management, where forecasting support systems are extensively used, the forecasting procedures are suspect (Fildes et al, 2006). A list of needed improvements in forecasting software is posted at forecastingprinciples.com as "assessment of the use of forecasting principles" under Software.

Researchers can aid the diffusion of useful findings by implementing their findings in software or freeware packages made available on the Internet. For example, Don Miller and Dan Williams provide freeware for damping seasonality (from their paper, Miller & Williams, 2004) at forecastingprinciples.com. The Delphi program provides another example: funded by the IIF, this freeware program is visited about 7,000 times per year.

Within the software-development community, resistance to new findings can be strong. With so much investment tied up in a particular software product, a provider may regard a new method as a commercial threat. We know of at least two cases in which forecasting software providers threatened legal suits because their programs fared poorly in published multiple hypotheses studies.

Tashman and Hoover (2001) found that many of the larger commercial software providers refuse to have their products reviewed. Their reluctance may be well founded: McCullough (2000) analyzed the same data by the same methods but with software packages from different suppliers; he found disturbing differences (again showing the importance of replications). For example, Excel, which is probably the most commonly used software for forecasting, contains numerous errors. Microsoft has been slow to acknowledge and correct these errors (McCullough, 2006).

Our efforts to date have largely overlooked how people search for information about forecasting methods. Typically, researchers search within their own field. As a result we need to reach forecasters

in crime, economics, finance, medicine, military, demography, earnings, environment, population, production and inventory control, natural resources, politics, retail sales, sports, technology, transportation, weather, or perhaps even within narrower areas such as automotive, earthquakes, epidemics, pharmaceuticals, hurricanes, movies, terrorism, and utilities. To address this, the IIF supports efforts to develop special interest groups (SIGs) at forecastingprinciples.com. Once developed, SIGs must be announced to those interested. The political forecasting SIG is good example; In an April 2006 Google search on "forecasting" and "elections," it was first out of 3.4 million sites. The site attracted about 90,000 visits its first year because it provided forecasts for the U.S. presidential election. Other SIGs, such as those for conflicts (#1 of 1.7 million sites in a search for "forecasting" and "conflicts") and crime (#1 of one million sites) have also been successful.

Implementation of improved methods and processes is not just a technical issue of more focussed and more accessible research with better software. Many organisational barriers discourage improvement. In effect, the new methods suggest that those using the current methods are deficient. This was illustrated by the Oakland Athletics adoption of improved procedures for predicting the success of baseball players. The General Manager based his selection on causal models in contrast with the traditional system based on unaided judgment by baseball scouts. When other teams learned what he had done, he was widely criticized (Lewis 2004, pp.287-301), although other sports teams are now starting to use the forecasting methods mentioned in the book. Where success (as here) is widely recognized, implementing new methods becomes easier.

The IIF should try to support those attempting to introduce new forecasting procedures. Shared case studies, such as the Oakland Athletics, reflecting on the principles underlying success should help. This might be done by having a stream of users groups and problem solving sessions at the annual symposia, encouraging people to seek advice and support of the IIF Discussion Group, and involving researchers in practitioner meetings. In addition, informal networks and cooperative research might help with implementation. Finally, to aid credibility, practitioners could be invited to be certified in forecasting.

Conclusions

After twenty-five years, the forecasting journals have influenced both the theory and practice of forecasting. In particular, adherence to the method of multiple hypotheses by the *IJF* and *JoF* has produced useful findings.

Traditional journal publications procedures do not focus on publishing useful research. Success is often a bi-product arising accidentally from the usual editorial activities. We suggest that journals invite papers that use multiple hypotheses for important issues. Furthermore, when papers are submitted

through traditional channels, reviewers should only be asked how to improve them; the editors should decide which papers are most useful. Editors should invite replications of important papers. To aid in replication, we suggest that published papers provide footnotes to websites with full information about the data and methods.

Perhaps the most important way to implement principles is through software and freeware. While much success has been achieved in the past quarter century, most forecasting principles have been ignored by the available software and freeware. We recommend that the Institute encourages the development of well-designed software by providing financial support.

We believe that people are unlikely to venture outside their field when they encounter forecasting problems. Thus, the *IJF* should continue to invite papers from areas that use forecasting, such as demographics, medicine, weather, and sports. Special issues and workshops can target key areas. Given the current interest, for example, there would appear to be a need for an issue on climate forecasting.

Special Interest Group web pages are needed for relevant disciplines so that people can access the information that they need and do so rapidly and at little cost. The International Institute of Forecasters could provide funding to aid in the development of SIG sites. Researchers will find that by providing such a service to their fields, they will become highly visible.

The IIF can aid practitioners who are facing organisational problems when adopting improved methods and processes. This can be aided by discuss common problems at conferences, the IIF Discussion group, and user groups. *Foresight* and the *Oracle* can help to generate discussion on key issues for practitioners. Certification can add credibility for those trying to introduce new methods.

A Personal Endnote

When we set off to create an institute and related journals, we never doubted the direction we needed to take. Success seemed readily achievable. Like many journeys of discovery, there were failures. The goal of bridging the gap between academicians and practitioners has been elusive, but current technology allows for progress.

The journey certainly has formed the basis of a stimulating career for both of us, though the importance of forecasting has not always been appreciated by our colleagues. We hope and expect that those following us -- the journal editors, the conference organisers, and the researchers -- will continue to appreciate the elation and agony when making empirical comparisons that relate to important and practical methods but this is an aspiration – not a prediction.

Acknowledgements: Paul Goodwin, Kesten C. Green, Rob Hyndman, Bruce McÇullough, Don Miller. Keith Ord and Lars-Erik Öller provided useful comments. April Anderson and Bryan LaFrance provided editorial assistance.

References

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

Armstrong, J. S. (1982). Research on scientific journals: Implications for editors and authors," *Journal of Forecasting*, 1, 83-104.

Armstrong, J. S., Brodie, R. J. & Parsons, A. G. (2001), Hypotheses in marketing science: Literature review and publication audit, *Marketing Letters*, 12 (2), 171-187.

Armstrong, J. S. & Hubbard, R. (1991). Does the need for agreement among reviewers inhibit the publication of controversial findings? *Behavioral and Brain Sciences*, 14, 136-137.

Armstrong, J. S. & Lusk, E. (1983). Commentary on the Makridakis time-series competition (M-Competition), with comments by Gardner, Geurts, Lopes, Markland, McLaughlin, Newbold, Pack, and replies by Andersen, Carbone, Fildes, Parzen, Newton, Winkler & Makridakis. *Journal of Forecasting*, 2, 259-311.

Armstrong, J. S. & Pagell, R. (2003). Reaping benefits from management research: Lessons from the forecasting principles project. *Interfaces*, 33 (6), 89-111.

Chamberlin, T. C. (1965). The method of multiple working hypotheses. *Science,* 148*,* 754-759. (Reprint of an 1890 paper).

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

Cox, J.E. Jr., & Loomis, D.G. (2001). Diffusion of forecasting principles through books. In J. S. Armstrong (ed.), *Principles of Forecasting*. Boston: Kluwer Academic Publishers.

Fildes R., Goodwin, P. & Lawrence, M. (2006) The design features of forecasting support systems and their effectiveness, *Decision Support Systems*, (in press).

Fildes, R. (2006). The forecasting journals and their contribution to forecasting research. *International Journal of Forecasting*, 22 (this issue)

Fildes, R. & Nikolopoulos, K. (2006). An interview with Spyros Makridakis. *International Journal of Forecasting,* 22 (this issue).

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

Gans, J. S. & Shepherd G. B. (1994). How are the mighty fallen: Rejected classic articles by leading economists. *Journal of Economic Perspectives*, *8* (1), 165-179.

Gleditsch, N. P., Metelits, C. & Strand H. (2003). Posting your data: Will you be scooped or will you be famous? *International Studies Perspectives*, 4 (1), 89-97.

Hubbard, R. & Vetter, D. E. (1996). An empirical comparison of published replication research in accounting, economics, finance, management and marketing. *Journal of Business Research*, 35 (2), 153-164.

Lawrence, M., Edmundson, R. H. & O'Connor, M. (1985). An examination of judgmental extrapolation of time series. *International Journal of Forecasting*, 1, 25-35.

Lewis, M. (2004), *Moneyball*. New York: W.W, Norton.

Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, 1, 161- 175.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, R. & Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, 1, 111-153.

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

McCullough, B. D. (2000). Is it safe to assume that software is accurate? *International Journal of Forecasting*, 16, 349-357.

McCullough, B. D. (2006). The unreliability of Excel's statistical procedures. *Foresight: The*

International Journal of Applied Forecasting, 3 (February), 44-45.

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

Platt, J. R. (1964). Strong inference. *Science*, 146, 347-353.

Reisman, A. & Kirschnick, F. (1994). The devolution of OR/MS: Implications from a statistical content analysis of papers in flagship journals. *Operations Research,* 42 (4), 557-558.

Tashman, L. J. & Hoover, J. (2001). Diffusion of forecasting principles through software. In J. S. Armstrong (ed.), *Principles of Forecasting*. Boston: Kluwer Academic Publishers.

IJF/ProgressIntro46