

REPLICATIONS OF FORECASTING RESEARCH

Heiner Evanschitzky
Professor of Marketing
Department of Marketing, University of Strathclyde
Glasgow, G4 0RQ
United Kingdom
+44-(0)141-548-5802
evanschitzky@strath.ac.uk

J. Scott Armstrong
Professor of Marketing
The Wharton School, University of Pennsylvania
Philadelphia, PA 19104
USA
armstrong@wharton.upenn.edu

March 27, 2009

Paper invited by the *International Journal of Forecasting*.

ABSTRACT

We examined the frequency of replications published in the two leading forecasting journals, the *International Journal of Forecasting (IJF)* and the *Journal of Forecasting (JoF)*. Replications in the *IJF* and *JoF* comprised 9.4% of the empirical papers. This compares with various areas of management science ranging from 2.2% in the *Journal of Marketing Research* to 18.1% in the *American Economic Review*. We also found that 36.2% of replications in forecasting journals provided full support, 44.7% partial support, and 19.1% no support for initial study findings. Given the importance of replications, we recommend steps to encourage replications, such as requiring full-disclosure of methods and data for all published papers, and inviting researchers to replicate specified important papers.

INTRODUCTION

Gardner and Diaz-Saiz (2008) replicated and extended research by Fildes et al. (1998), which was itself an extension of the M-Competition study (Makridakis, et al. 1982). Using changes in the estimation procedures, they concluded the primary conclusion was supported but disagreed with a secondary conclusion. This demonstrates the value of replications in showing where we can gain confidence and in indicating areas in need of further research.

Experts have claimed that replication is vital to scientific progress (e.g., Hunter, 2001). Replications help to ensure that findings can be reproduced. Extensions go beyond that to examine whether the findings can be generalized.

Despite these benefits, relatively few of the papers published in various areas of management science are replications (Hubbard & Vetter, 1996; Evanschitzky et al., 2007). A number of reasons have been used to explain this. First, and perhaps most important, many studies in management science are unimportant, thus it would be senseless to replicate them. Second, authors seldom provide sufficient detail in the paper (or in response to requests) to allow for replication. And, third, reviewers seem to be biased against replications because they think they do not offer something new or because the results are not statistically significant.

Misinterpretations of null hypothesis testing procedures may have undermined the perceived need for replication. Oakes (1986) showed that 42 of 70 (60%) experienced academic psychologists falsely believed that an experimental outcome that is significant at the 0.01 level has a 0.99 probability of being statistically significant if the study were replicated.

Based on these observations, we examine the state of replication research in forecasting and then suggest ways to make further improvements with respect to replication.

THE RECORD OF REPLICATIONS IN THE LEADING FORECASTING JOURNALS

The definitions of the central terms in this study are extensions of those employed by Hubbard and Armstrong's (1994, p. 236). A replication is defined as "a duplication of a previously published empirical

study that is concerned with assessing whether similar findings can be obtained upon repeating the study.” Likewise, a replication with extension is “a duplication of a previously published empirical research project that serves to investigate the ability to generalize earlier research findings.” The latter goal of determining the range of conditions under which the findings do and do not hold up can also be addressed by studies conducting a series of experiments within the same article. However, such works lack independent assessment. Therefore, following Hubbard and Armstrong (1994), we only consider replications that were published as independent papers.

In forecasting journals, we find the following three types of replications: “real replications,” “model comparisons,” and “data re-analyses.” While “real replications” are defined along the Hubbard and Armstrong (1994, p. 236) definition of replications, “model comparisons” are an application of a previously published statistical analysis that is concerned with assessing whether a superior goodness of fit can be obtained comparing the original statistical model with at least one other statistical model, and “data re-analyses” is an application of previously published data that is concerned with assessing whether similar findings can be obtained involving the same data or a sub-sample of the data. For example, McCullough (2000) showed that the selection of a software program for a given method had an important effect on results. The aforementioned three types of replications in forecasting are not mutual exclusive and are frequently used in conjunction.

We did a frequency count of the three types of replications appearing in the *International Journal of Forecasting (IJF)* and the *Journal of Forecasting (JoF)*. The numbers are estimated based on a census of all 766 empirical articles published in both journals for the period 1996-2005. Two independent raters classified all of these works to determine the publication incidence of replications and extensions. The raters were in agreement for 81.2% of their initial ratings. This is similar to the 88% agreement rate reported in Evanschitzky et al. (2007) and 90% in Hubbard and Armstrong (1994).

Results of the count show that the most prominent type of “replication” seems to be model comparison (on average 27.3% for *JoF* and *IJF*), followed by data re-analyses (18.9%). The average of

“real” replications was 9.4% for the two forecasting journals. The yearly fluctuations were fairly high for both the *IJF* and the *JoF*. There were no clear trends over the 10-year assessment period.

Comparing the number of replications and extensions in forecasting journals with those being published in other areas of management science, we obtain similar results. Hubbard and Vetter’s (1996) analysis of replications of empirical studies showed that the average replications varied from 2.2% of the empirical papers (*Journal of Marketing Research*) to 18.1% (*American Economic Review*), with a median of 6.6%. Thus, the forecasting journals, in particular the *IJF* at 11.4% (and the *JoF* at 7.3%), match up well with other leading management-science journals.

We also assessed to what extent replications are able to reproduce results from the original work. For the sake of comparability, we only consider the 72 (9.4%) “real” replications published in forecasting journals. Following Evanschitzky et al. (2007), we examined findings that “confirmed,” “partial support,” or provided “no support.” It is worth noting that for 25 of these studies, insufficient information was available to make such a comparison. Hence, when comparing results to Hubbard and Vetter’s (1996) analysis, we only include those real replication for which sufficient information to judge on the degree of support for the original study was available. This reduces the number of studies to 47. Table 1 shows the support provided by “real” replications for the period from 1996 to 2005.

Insert Table 1 about here

Again, we use the Hubbard and Vetter’s (1996) findings for benchmarks. Averaging across journals in accounting, economics, finance, management, and marketing, full support was found in 27% of the studies and partial support in 27%; most important was that 46% of the replication studies did not provide support of initial findings. Compared to these findings, results on the degree of support for initial study findings is higher in the forecasting journals. On average across *IJF* and *JoF*, we find out that replications support initial findings in 36.2% of all cases; partial support is found in 44.7%, while 19.1% of the studies do not support initial findings.

DISCUSSION

A relatively low number of replications across all disciplines of management science in conjunction with the majority of replication attempts failing to provide support of initial findings, we call for an increase in replication research. To aid in doing so, the data and methods used in the original studies should be made available on the Internet concurrent with a paper's publication. This is important because authors often lose track of the data and the methods. The lack of data and methods inhibits replication attempts. Full disclosure of data and methods used in the original studies are a precondition to encouraging replication studies.

Full disclosure requires an effort by authors, but it is time well spent not only for science, but for the authors as well. 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, such as type of article, co-authorship, age of paper, length of paper, and characteristics of authors.) We suspect, for example, that this full disclosure has also contributed to the strong impact of the M-Competition studies; a Google scholar search for "Makridakis and "M-Competition" in January 2009 yielded 271 sites.

Active participation by a journal's editor is needed to ensure that the data and methods are archived prior to publication. The *Journal of Money, Credit and Banking (JMCB)* had a policy that authors must deposit the data and code used for papers they publish. This began in 1982-1984, when all submissions to the *JMCB* were required to submit the data and methods to the journal's editorial office (Dewald, Thursby, & Anderon, 1986). Enforcement dropped off however. McCullough (2007) found that during the 1996-2003 period, only 14 of 186 empirical papers in the *JMCB* provided sufficient materials to allow for a replication. From 2001 through 2004 the rate of archiving dropped to 3% for *JMCB* papers.

A similar editorial policy is applied by the *Journal of Conflict Resolution*. That journal created a dedicated website on which all necessary materials have to be made available for replication purposes. Since the first issues of 2002, the editor requires authors to comply with this policy as condition of

publication (Russett, 2003). Similarly, the *International Journal of Forecasting* has for the past few years had a policy of requesting data and details on the methods used in order to encourage replications. It also has recently instituted a systematic procedure to obtain data and methods from authors prior to publication.

Fortunately for editors, compliance with archiving is now simpler than it has been previously. Authors can be required to provide footnotes showing how to find the data and methods on the Internet. This highlights the availability of the information and also enables the editor to download the materials on the journal's website. This procedure is currently required by such journals as the *American Journal of Political Science*.

Replications should focus on important papers. For example, Ioannidis (2005) found that replication studies were conducted for about 3/4 of highly cited papers in medicine (in a sample from 1990 through 2003). Editors could identify important papers in the field that should be replicated/extended, and then invite designated researchers to publish such replications. If invitations were restricted to important problems, they would be more likely to gain cooperation from the authors of the original study. Furthermore, authors are much more likely to undertake a replication of an important study, especially when it is an invited paper.

A mid-way policy would be to invite researchers to submit proposals for replication studies. Replications that are submitted through traditional channels might be treated favorably for some manner of publication. This might be a short printed version along with the author's reply, with details on the Internet, to be followed by open peer review. Important replications would receive more print space. Additional emphasis can be provided by appointing a replications editor or a separate section; this has been done since 2003 by the *Journal of Applied Econometrics*.

In those cases where data have not been archived (which is, unfortunately the typical case) those who desire to do further work on a topic should frame their request for cooperation from authors by positioning the study as an extension (which is the normal case) rather than as a replication as the latter implies a search for mistakes.

Current technology makes it relatively easy to link to the relevant published replications. For example, Armstrong, Coviello, and Safranek (1993) conducted a replication/extension of research on the escalation bias phenomena. A Google Scholar search on “escalation bias” showed 178 listings in January 2009, and the first two were related to the Armstrong, Coviello, and Safranek (1993) replication. A more reliable tracking will be obtained by entering the title of the paper and the terms “replication” or “extension” or “meta-analysis.” Such a procedure could be used by authors, reviewers, and editors.

CONCLUSIONS

In comparison with empirical studies published in other areas of the management sciences, replications in the *International Journal of Forecasting* (11.4% of empirical papers published) and the *Journal of Forecasting* (7.3%) are above the median of 6.6%. It is difficult to say what the optimum number of replications studies that should be published is. However, given that many replications do not support the original findings (about 1/5 of the empirical papers on forecasting), we suggest that it would be helpful to publish more replications of important studies. Journals can aid this by requiring and archiving full disclosure and by inviting researchers to replicate specified papers.

REFERENCES

- Armstrong, J. S., Coviello, N., & Safranek, B. (1993). Escalation bias: Does it extend to marketing? *Journal of the Academy of Marketing Science*, 21, 247-253.
- Dewald, W. G., Thursby, J., & Anderson, R. G. (1986). Replication in empirical economics: The Journal of Money Credit and Banking project. *American Economic Review*, 76, 587-603.
- Evanschitzky, H., Baumgarth, C., Hubbard, R., & Armstrong, J. S. (2007). Replication research's disturbing trend. *Journal of Business Research*, 60, 411-415.
- Fildes, R., Hibon, M., Makridakis, S., & Meade, N. (1998). Generalizing about univariate forecasting methods: Further empirical evidence. *International Journal of Forecasting*, 14, 339-358.
- Gardner, E. S. & Diaz-Saiz, J. (2008). Exponential smoothing in the telecommunications data. *International Journal of Forecasting*, 24, 170-174.
- Gleditsch, N. P., Metelits, C. & Strand, H. (2003). Posting your data: Will you be scooped or will you be famous? *International Studies Perspectives*, 4, 89-97.
- Hubbard, R. & Armstrong, J. S. (1992). Are null results becoming an endangered species in marketing? *Marketing Letters*, 3, 127-136
- 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, 153-164.
- Hunter, J. E. (2001). The desperate need for replications. *Journal of Consumer Research*, 28, 149-158.
- Ioannidis J. P. A. (2005). Contradicted and initially stronger effects in highly cited clinical research. *Journal of the American Medical Association*. 294, 218-228.
- 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, 111-153.
- McCullough, B. D. (2000). Is it safe to assume that software is accurate? *International Journal of Forecasting*, 16, 383-397.

McCullough, B. D. (2007). Got replicability? The *Journal of Money, Credit and Banking* archive. *Econ Journal Watch*, 4, 326-337.

Oakes, M. (1986). *Statistical Inference: A Commentary for the Social and Behavioral Sciences*. New York, NY: Wiley.

Russett, B. (2003). The *Journal of Conflict Resolution*'s policy on replication. *International Studies Perspectives*, 4, 88-89.

Acknowledgments

We thank all those authors who replied to our request to ensure that they were accurately cited. In addition, Nils Petter Gleditsch, Raymond Hubbard, and Bruce D. McCullough provided many useful suggestions on early drafts of the paper.

TABLE 1**Support provided by replications: Percentages of paper (number of papers)**

	<i>IJF</i> (32)	<i>JoF</i> (15)	<i>Average for both</i> <i>journals</i> (47)	<i>Hubbard and</i> <i>Vetter (1996)</i>
confirmed	34.4 (11)	40.0 (6)	36.2 (17)	27.0 (72)
partial support	46.9 (15)	40.0 (6)	44.7 (21)	27.0 (73)
no support	18.7 (6)	20.0 (3)	19.1 (9)	46.0 (121)