

Documentation of the File Drawer Problem at Finance Conferences: A Follow-Up Study

Manoela N. Morais
Department of Finance
Lubin School of Business
Pace University
New York, NY 10038
mm78341n@pace.edu

Matthew R. Morey
Department of Finance
Lubin School of Business
Pace University
New York, NY 10038
mmorey@pace.edu

Revised May 18, 2020

Abstract

The “file drawer problem” is a publication bias where journal editors are much more likely to accept empirical papers with statistically significant results than those with statistically non-significant results. As a result, papers that have non-significant results are not published and relegated to the file drawer, never to be seen by others. In a previous paper, Morey and Yadav (2018), examined the file drawer problem in finance journals and found evidence that strongly suggests that such a publication bias exists in finance journals. In this follow-up study, we examine the prevalence of the file drawer problem at finance conferences. As such we are first paper in finance that we know of to attempt such an analysis. To do this we examine every single empirical paper presented at the annual Financial Management Association (FMA) conference from 2014-2018. In an examination of 3425 empirical papers we found less than one-half of one percent of these papers had statistically non-significant results. These results suggest that there is also a significant file drawer problem at finance conferences.

Keywords: File Drawer problem, finance conference, statistical significance, publication bias

1. Introduction

The “file drawer problem” is a publication bias where editors of journals are much more likely to accept empirical papers with statistically significant results than those with non-significant results. As a result, papers that have non-significant results are not published and relegated to the file drawer, never to be seen by others. This takes place because the editors are involved in a race for “citation-based impact numbers” and papers with significant results are much more likely to be cited than papers with non-significant results.¹ Due to this bias, a researcher who finds non-significant results often will place them in their file drawer, never to be seen by others. As a result, the published literature is left with a biased set of papers that almost only show statistical significant results.

The presence of the file drawer problem also means that researchers have to find statistically significant results in order to publish them. This -- combined with the need to publish in order to receive tenure, promotion, and grants -- may lead to “p-hacking,” where researchers manipulate and cherry-pick results in order to gain significance. The importance of p-hacking is that many of the results of published papers may not be correct.

In 2018 we published a paper in the *Journal of Investing* (Morey and Yadav (2018)) that examined this bias in finance journals.² To do this we examined 29 finance journals for five years (2010-2014). These journals included A-ranked journals as well as B- and C-ranked journals. In an examination of over 5740 empirical papers we found only 121, or 2.1 percent of the papers, had statistically non-significant results. Indeed, for some journals there was not a single article that had non-significant results. Furthermore, our findings indicate that the file drawer problem was just as

¹ See Fanelli (2013) who finds that papers with statistically significant results are more likely to be cited.

² Note that other papers have investigated publication bias in finance journals. These include Gómez-Bezares and Gómez-Bezares (2006), Harvey, Liu and Zhu (2016), Harvey (2017) and Kim and Ji (2015).

bad, if not worse, in lower-ranked journals as it is in top-ranked journals. The percentage of papers with non-significant results is actually somewhat smaller in B- and C-ranked journals than the A-ranked journals. These results strongly suggested that there is a publication bias in finance journals.

In this paper we present a follow-up study to our previous paper in which we examine the prevalence of the file drawer problem in another area: academic finance conferences. Until this paper, this issue has never been examined in academic finance to our knowledge. We feel this is an important contribution for at least two reasons. First, if we find evidence of the same type of file-drawer problem at finance conferences it just further documents the severity of the problem in the field. Such a result would indicate that even to get into a conference (usually a much lower bar than to be accepted into a journal) authors would likely have to find statistically significant results. Indeed, in some ways such a finding would be even more pernicious than findings of a file-drawer problem at journals. Conferences are where young researchers learn, socialize, and establish their own reputations. Indeed, it is where they first meet many of their colleagues who study the same topics. A finding of a file-drawer problem at conferences would mean that researchers need statistically significant results just to be able to develop themselves in the field.

A second reason for this study is that we can investigate if finance conferences are mitigating the file drawer problem or not. It may be that because finance conferences let in more papers than finance journals³ there may be more willingness to accept empirical papers that do not have statistically significant results. Hence, it may be that conferences provide more of an outlet for statistically non-significant findings than journals. Indeed, in abstracts submitted to medical conferences, Callahan et al. (1998) and Timmer (2002) have consistently found little to no bias

³ Consider, for example, that the 2018 Financial Management Association annual conference had 787 papers while the Journal of Finance published 72 articles that same year.

against studies with statistically non-significant results. With this paper we can investigate if the same holds in finance.

To do this study we examine the largest annual academic finance conference in the U.S (by number of papers), the annual Financial Management Association (FMA) conference, for five consecutive years (2014-2018). For each year we inspect every single empirical paper that was presented at the conference. We then determine whether those papers had statistically significant results or not.

The rest of this paper is organized as follows. In Section 2 we explain our data collection process and results. In Section 3 we conclude.

2. Data and Results

To determine if a paper at the FMA had statistically non-significant results we use the following approach. First using the FMA webpage (www.fma.org, which provides a listing of all the papers presented in each annual FMA conference back to 2014) we examined every paper for each FMA conference program for the years 2014 to 2018. We then included in our sample every paper that was a stand-alone research paper. This meant that we did not include roundtables, panel sessions, pedagogy series, and presentations that were made for a debate setting. In Table 1, the far left hand column shows the total number of stand-alone research papers at each of the 2014-2018 annual FMA conferences. The total number of these papers was 3817 over the five-year period. Hence, the average number of stand-alone research papers is about 763 a year. Again, considering that the Journal of Finance publishes about 72 papers a year, this 763 is a large number. This large number of papers is again one of the reasons for our study: the large number of papers may mean that papers with statistically non-significant are more likely to be included than in journals.

In our analysis, we found that for a 169 of these 3817 papers (4.4 percent) there was no link available on the FMA website to the underlying paper listed on the FMA program. Hence, we did not include these papers in our sample. With the remaining 3648 papers we then examined each paper to determine if the paper was an empirical or a solely theoretical paper. Out of the 3648 remaining papers there were 223 solely theoretical papers, leaving us with 3425 empirical papers.

Then with each of the 3425 empirical papers we determined if the result of the paper was significantly significant or not. This process involved reading the abstract and results section of the each of these papers to determine if the paper had statistically significant results or not. To do this we defined any empirical result as being statistically significant if the results were at the ten percent level of statistical significance or stronger (a t-statistic of 1.68 or higher). We included papers that used linear regression, time-series regression, non-linear regression and Bayesian regressions as well as simple descriptive statistics as long as the authors tested for significance.

The results of our analysis are in Table 1. Going from left to right, the Table shows the number of papers in each annual conference, the number of papers without any link in the FMA website, the remaining number of empirical papers and theoretical papers, and the number of empirical papers that had non-significant results (along with the percentage of the total empirical papers found that that had statistically non-significant results).

Overall, the results show that of the 3425 empirical papers presented at the conference from 2004-2018 only 14 had statistically non-significant results. Hence, over a five year period, only a remarkable 0.41 percent of the empirical papers examined at the biggest U.S. annual finance conference had statistically non-significant results. This result is actually less than the 2.1 percent of papers with statistically non-significant results found in our previous paper. Hence, from our

results, it would see the file drawer problem at least as serious at finance conferences as it is in finance journals.

3. Conclusions

In this follow-up study to Morey and Yadav (2018) we conduct the first attempt that we know of at investigating the file drawer problem at academic finance conferences. We show compelling evidence that there is a file drawer problem at finance conferences. In an examination of all the papers listed in the programs of the 2014-2018 annual FMA conference (the largest academic finance conference in the U.S. by number of papers) we find less than one-half of one percent of the papers had statistically non-significant results. Indeed, for 3425 empirical papers we found only 14 had statistically non-significant results. Our results indicate that the file drawer problem is just as bad, if not worse, at finance conferences than it is in journals. The percentage of papers with statistically non-significant results is actually lower at finance conferences than that found in academic finance journals.

As stated in the introduction, the concern with the file drawer problem is twofold. First, attendees of conferences will only see a biased sample of papers. Papers that show statistical non-significance are relegated to the file drawer and hence are never seen by the conference participant. Second, and maybe more important, the file drawer problem creates a strong incentive for researchers to practice p-hacking, where researchers manipulate results in order to gain statistically significant results. Indeed, our findings indicate that just to be included at conferences—the very places where finance professionals learn about new advances in the field and make social connections -- authors should basically only submit papers with statistically significant results. While we cannot say that someone submitting a paper with statistically non-significant results will

be definitely rejected by the FMA conference, the upshot of our result is that this seems likely to happen. To rectify this problem, conferences like the FMA should consider special sessions that encourage results with statistically non-significant results.

Table 1: Percentage of Papers at the 2014-2018 Annual Financial Management Association (FMA) Conference with Statistically Non-Significant Results.

FMA Annual Conferences	Number of Papers	Number of Papers with no available link	Number of Empirical Papers	Number of Non-Empirical Papers	Number of Empirical Papers with Statistically Non-Significant Results	% of Empirical Papers with Statistically Non-Significant Results
2014	735	12	674	49	4	0.59%
2015	712	32	635	45	4	0.63%
2016	831	42	748	41	3	0.40%
2017	752	45	670	37	1	0.15%
2018	787	38	698	51	2	0.29%
Total	3817	169	3425	223	14	0.41%

References

- M. L. Callaham, R. L. Wears, E. J. Weber, C. Barton, G. Young, 1998, “Positive-outcome bias and other limitations in the outcome of research abstracts submitted to a scientific meeting.” *Journal of the American Medical Association*, 280, 254-257.
- Fanelli, Daniele, 2013, “Positive results receive more citations, but only in some disciplines”, *Scientometrics* 94, 701–709.
- Gómez-Bezares, Fernando and Fernando R. Gómez-Bezares, 2006, “On the Use of Hypothesis Testing and Other Problems with Financial Research”, *Journal of Investing*, 15, 4, 64-67.
- Harvey, Campbell, Yan Liu, and Heqing Zhu, 2016, “...and the Cross Section of Expected Returns”. *Review of Financial Studies*, 29, 1, 5-68.
- Harvey, Campbell, 2017, “Presidential Address: The Scientific Outlook in Financial Economics”, *Journal of Finance*, forthcoming.
- Kim, Jae, and Phillip Ji, 2015, “Significance Testing in Empirical Finance”, *Journal of Empirical Finance*, 34, 1-14.
- Morey, Matthew R. and Sonu Yadav, 2018, “Documentation of the File Drawer Problem In Academic Finance Journals” (with Sonu Yadav), *The Journal of Investing*, 27:1, 143-147.
- A. Timmer, R. J. Hilsden, J. Cole, D. Hailey, L. R. Sutherland, 2002, “Publication bias in gastroenterological research—a retrospective cohort study based on abstracts submitted to a scientific meeting.” *BMC Medical Research Methodology* 2, 7.