

# Refereeing Ever More – Progressing Ever Less

Matthew Spiegel  
Yale School of Management

June 7, 2018

I would like to thank the numerous people that helped put together the data for this editorial. Including . . . I also want to thank the Society for Financial Studies for giving me the opportunity to present my thoughts on the state of academic research and the editorial process during the 2017 North American Cavalcade.

Contact Information: Yale School of Management, P.O. Box 208200, New Haven CT 06521-8200, Email: [matthew.spiegel@yale.edu](mailto:matthew.spiegel@yale.edu).

## Abstract

Over the past several decades academic finance articles have grown longer and longer. Increasing demands by referees continue to make matters worse. Authors currently face long lists nearly pointless “requests” for additional tests. (Reverse causality anyone?) Since larger and larger teams are now needed to handle all of the demands the number of authors on each paper has grown as well. The result has been a system in which papers take years to go from initial idea to publication. This might be acceptable if it led to a faster expansion of the profession’s knowledge base. But that seems difficult to believe. As in the past no single article can definitively show that something is true. What can be done to reverse this trend? Referees can start by limiting their reports to two pages or less. Editors can help by telling authors to ignore most of the “suggestions” in referee reports, especially those that are on pages three and up. Instead of futilely pursuing a quest to eliminate all endogeneity concerns, skip the endogeneity tests. They are mostly data mined at this point anyway. Instead publish shorter papers without the robustness and endogeneity tests. Then if other researchers doubt the veracity of the results let them show they are due to some peculiarity in the data or that the causality runs the other way. Then publish the new paper. This will let authors move on with their research agendas and provide a way to crowd source robustness and endogeneity checks. A system that seems much more likely to produce useful results than the current system of demanding that authors conduct every robustness and endogeneity check some one to three referees can come up with.

Academic finance seems determined to force the collective production of papers that increasingly resemble diary entries; long missives read by no one other than its authors and perhaps two or three referees. There is no evidence that a long running decline in author talent might be to blame. Instead we have seen the rise of “did you check for” bloat. Test after test conducted to rule out seemingly more and more obscure and remote alternative explanations. Imagine a paper claiming that entrepreneurs open restaurants in areas with easy road access. Do you really think the causality goes the other way? Knowing people wish to open restaurants in currently inaccessible places the government authorities build roads to make the areas accessible? Nobody thinks this is likely and nobody wants to see an instrumental variable test ruling it out. Well, nobody except the referee that made the authors check and which then turned into Table A11.<sup>1</sup> These inane demands do nothing but result in the production of tables that are pointless filler. In the end we have long papers most of which nobody reads with results that nobody checks.

Many papers now approach 100 pages with their internet appendices. These are no longer articles. They are monographs. Has the creation of monographs (in the guise of articles) really improved the profession’s rate of knowledge creation? Be honest, when was the last time you read through all of the tables in a published paper? How about the internet appendix? It took a lot of work to produce tables 6 through 23 and those in the appendix labeled A1 to A17. Tables that few people ever look at. This is not helpful if our goal is to advance our understanding of financial markets. It is time we stopped talking about paper bloat and do something about it. Best of all, it is an issue we can all do something about. Coordinated collective action is not needed.

What drives this trend towards excessively long papers? Referees in combination with inattentive editors. Referee reports in excess of four pages are no longer a novelty. I have asked people

---

<sup>1</sup> No, this did not really happen. At least as far as I know. It is meant to be apocryphal. But if you thought it might be a true story, think about what that implies for today’s refereeing process.

why they write such long reports. Often, they tell me they do it because editors expect it. A response that is especially prevalent among younger members of the profession. Sadly, this was predicted by Hirshleifer (2014). He hypothesized that our colleagues, when called upon to referee, try to impress editors by demanding authors make changes. Changes that thoughtful third parties would likely conclude are unnecessary. In essence, referees engage in signal jamming. In Hirshleifer's model, and I fear in real life as well, people believe "good" referees demand changes so they have to as well. These days, papers are typically submitted after years of work. Referees are thus left making "suggestions" to "improve" a paper that in reality does little to advance its thesis. How many times have you been the victim of a referee offering up a preposterous reverse causality explanation that will now take you six months of work to check? If all this helped to produce results that are undeniably true it might be sensible. But that is clearly not the case. It is easy to find articles that reach very different conclusions about the same issue. Subsequent authors then frequently cite papers on both sides of the debate and pronounce it a tie. This is not progress.

These days you do not even need to wait for a referee report to be victimized by people demanding you check for exceedingly unlikely alternative explanations. Just go to a seminar or conference. Questions often have nothing to do with trying to understand what was done or what it might imply. Instead they are of the "gotcha" variety. Somebody comes in to present a 90-page paper with table-after-table checking the paper's main table (typically Table 3) against a battery of alternative explanations. (Alternatives that reasonable priors would put at having well under a 0.01% chance of materially affecting anything.) Nevertheless, somebody in the audience will ask, "have you checked for . . ." For what? Unicorns? If a paper has checked any number of remote alternative explanations what makes you think your remote explanation is any more likely to make a difference? But, this is what happens when over refereeing becomes the norm. However, many tests an author has performed anyone can always think of more.

Editors can stop referees from abusing authors if they want to. Instead a typical letter includes the phrase, “please satisfy the referees.” That might be fine if the three reports, each over four pages long, were limited to just four sentences and came up with similar suggestions. Instead authors end up trying to figure out how to navigate among a plethora of demands; many in direct conflict to each other. (I recall a case where one referee on a paper of mine thought section X was the paper’s best and the other wanted me to remove it. The editor just said satisfy the referees. How?) Part of the problem here is that editors are overtaxed at many outlets. Few people can truly edit if they have to deal with an inflow of more than a dozen or so 85 page papers per month along with a blizzard of 4 page plus referee reports. Instead they throw up their hands and just say “satisfy the referees.” What they should do is edit and tell the authors to ignore all but one or two points in each review. Or better yet, tell authors they can totally ignore a report’s “advice” if the author thinks it will make the paper worse.

Paper bloat has also fostered bibliography bloat. Alas, the latter has not increased the number of editorial voices or outlets through which the community regularly communicates. Instead we have seen an increased concentration of citations to articles in the “big 3” finance journals: *The Review of Financial Studies*, *The Journal of Financial Economics* and *The Journal of Finance*. This emphasis on articles published in a small number of outlets further damages progress. Fewer editorial voices mean fewer ideas are debated and tested. Worse, articles with dubious results are often protected from serious criticism leaving the profession to accept results as true, when in fact the profession should know better.

Why are bibliographies becoming ever more focused on the big 3 outlets? This is somewhat more difficult to understand, but two factors may be at play. When I ask people about this they tell me that signaling to the editor plays an important role. Authors are worried that if they cite articles outside of the big 3, then editors will think their paper also belongs in a journal outside the big 3. This is very unlikely to be true. It is difficult enough to get anyone to read a paper let alone catalog its bibliography!

It is true that if all of a paper's citations are from say a physics journal a finance journal may indicate to the author it has gone to the wrong outlet. But, a finance paper sent to a finance journal is not going to be turned down simply because it cites articles from a range of sources. (If you have a letter from an editor indicating otherwise, it is likely the wording is deliberately misleading. The goal being to let you off "easy" and hide the fact that the editor is just not interested in your paper.) Whatever drives the trend towards bibliographies focused on the big 3, it further damages the profession. Different journals offer different editorial voices. By restricting attention to just a few outlets, important ideas are easily quashed. Referees for the big 3 are often drawn from a narrow pool of people that have previously written on a topic. As such, a single person may be asked to review the same paper at multiple outlets. By issuing just two negative reviews they can likely sink a paper's future. That confers too much power in individuals that may have a vested interest is burying dissenting views.

We can stop this trend by acknowledging that longer papers with multiple robustness and endogeneity tests are no more likely to yield definitive results than shorter ones. We all know that academics are smart and persistent. If a particular result will get a paper published they will find a way to cut the data and run their tests to show it passes whatever test a referee demands. The issue of  $p$  hacking has gained traction of late. So should the issue of instrumental variable hacking (Jiang (2017)). No referee is going to check an author's code or figure out if the results are due to a particular way the data was "cleaned." Even if authors never engage in nefarious activities the vast number of academics working on problems guarantees most results are effectively "group" hacked. That is just life.

Instead of demanding a long list of robustness and endogeneity tests, stop with the basic result. Put a reasonable page cap on all submitted papers. (I would shoot for 25, but realistically how about we just try for 60?) Tell referees that any demands for robustness or endogeneity tests will be ignored, unless they are incredibly easy to implement. We should publish shorter papers. If someone thinks a paper's results are either not particularly robust or yield results with the causality inverted let them test

their hypothesis. Should it turn out the competing researcher is right, publish that paper too. The profession should then cease citing the original paper and move on, referencing the newer one. That is what is known as progress. It also lets the profession crowd source the problem of checking the veracity of an article's results. That is far more likely to winnow out the ones driven by spurious correlations, reverse causality or coding errors than the current system of cat and mouse, in which the mice have forever in which to figure out how to evade the cat. In the end all we get are tired mice and unfed cats.

### What Has Happened?

Where has referee induced bloat brought us? A lot of defensive writing. It starts with bibliographies that have ballooned in length. In what follows all but one graph will feature the trends at the big 3 journals. Why those and those alone? Well for one it takes a lot of time to pull the data journal-by-journal! More importantly, or perhaps more unfortunately, the path laid down by these journals is the one the profession is nearly guaranteed to follow elsewhere.

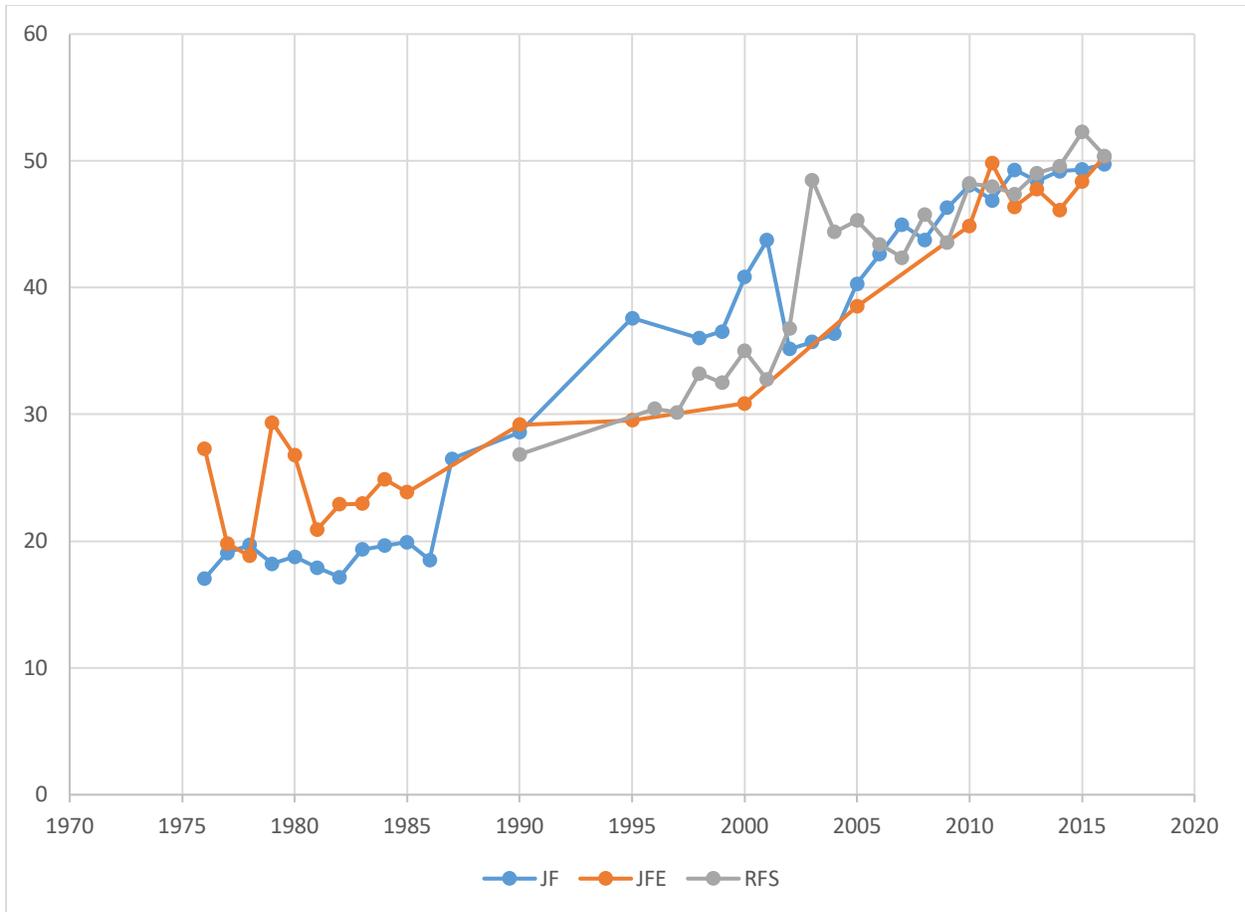


Figure 1: Average Bibliographic Entries per Article

Figure 1 displays the average length of article bibliographies by journal-by-year. The upward drift is unmistakable. It has also been quite dramatic. Around 1980 papers had on average just 20 citations and today it is about 50. A rise of about 150%! What is driving this trend? When authors have to defend against every conceivable “what if” to receive a coveted acceptance letter they are going to have to cite a lot of papers as a defensive move. If your referee might think it is important to check for A, you need to do the check and you will want to make sure to cite the referee’s work in the area. Start a list of checks A, B, C, etc. and pretty soon you have more than doubled the number of citations. But have authors really read all of the papers? Unlikely. In many cases it is clear that authors not even read the abstracts. Anyone reading a citation list that includes their work has found cases where those

referencing it did nothing but pull articles they never read from other bibliographies. After all, the important point is to cite a potential referee. The paper’s merit? Who cares. At this point an article’s citation count likely reflects a combination of the author’s power in the profession (i.e. likelihood of refereeing your paper) and if the paper has become a “standard cite” meaning it is in everybody’s bibliography because it is in everyone else’s bibliography.

While you might expect longer bibliographies to produce a wider range of editorial voices, you would be wrong. Yes, bibliography counts are up. But so is citation concentration.

**Exhibit 8**

**Figure 1: One-Year Impact Factor<sup>1</sup> Comparison, Year 2005 to 2015**

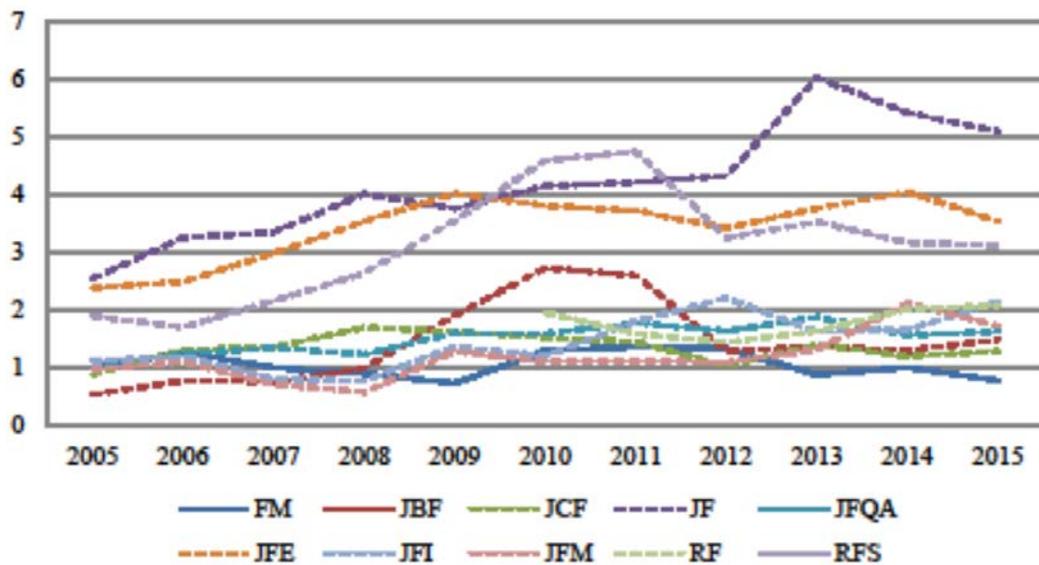


Figure 2: Courtesy of Utpal Bhattacharya (2017).

Bibliography inflation seems to have primarily benefited the big 3. Intuitively, one would have thought electronic searches should have increased the diversity of referenced outlets. Prior to the 1990s the only way to discover an article’s existence was to peruse paper journals in the library. Now, a simple electronic search brings up any number of articles on a particular topic across a wide range of journals. How is it possible that these papers are not making it into bibliographies? As noted earlier, many authors think citing papers outside the big 3 will blacklist their paper from publication in them. Another

possibility is lazy searching. With a few additional commands a list of papers produced by a search can be limited to a small set of outlets.

You might think bibliographies might have a limited length simply because authors would fear looking silly if they cited papers they have not read or do not really understand. Alas, Berninger, Garr, Kiesel and Schiereck's (2017) find that obtuse prose actually increases a paper's citation count. They hypothesize that this comes from authors hoping to look smart by attaching their paper to articles few understand. Unfortunately, given how often fellow academics tell me that they do things like write long referee reports or limit citations primarily to the big three in order to impress editors, their hypothesis seems all too credible.

Whatever drives concentrated citations, it is harming the profession's ability to shift the frontiers of knowledge. If you want diversity of thought, you need diversity of editorial outlets. With just three primary outlets it is easy for a small pool of referees with a vested interest in a particular result to blacklist articles questioning the prevailing orthodoxy. The academic finance community clearly feels a certain amount of frustration with this. If you doubt it, consider how successful Ivo Welch's journal *The Critical Finance Review* has been. It routinely publishes articles questioning existing results. It also publishes *debates*! You know, articles with opposing viewpoints about an issue. Debates. I am sure you have heard of them. Not within most finance journals perhaps, but if you look around you will see them elsewhere. The advantage of publishing debates is that readers can judge for themselves each side's merits. While that, in theory, is possible even if the articles are published years apart it is a lot more difficult. Plus, the recency of the arguments matter. While a dissenting paper may rightfully point out a problem within an existing literature, that problem may have been or may be easily be addressed. A real time debate can bring that out in ways sequential articles written over the years are unlikely to.

Overly long lists of referee demands, combined with dueling opinions and attempts to preemptively fend them off has naturally produced overly long articles. The time series shows just how dramatic this has been.

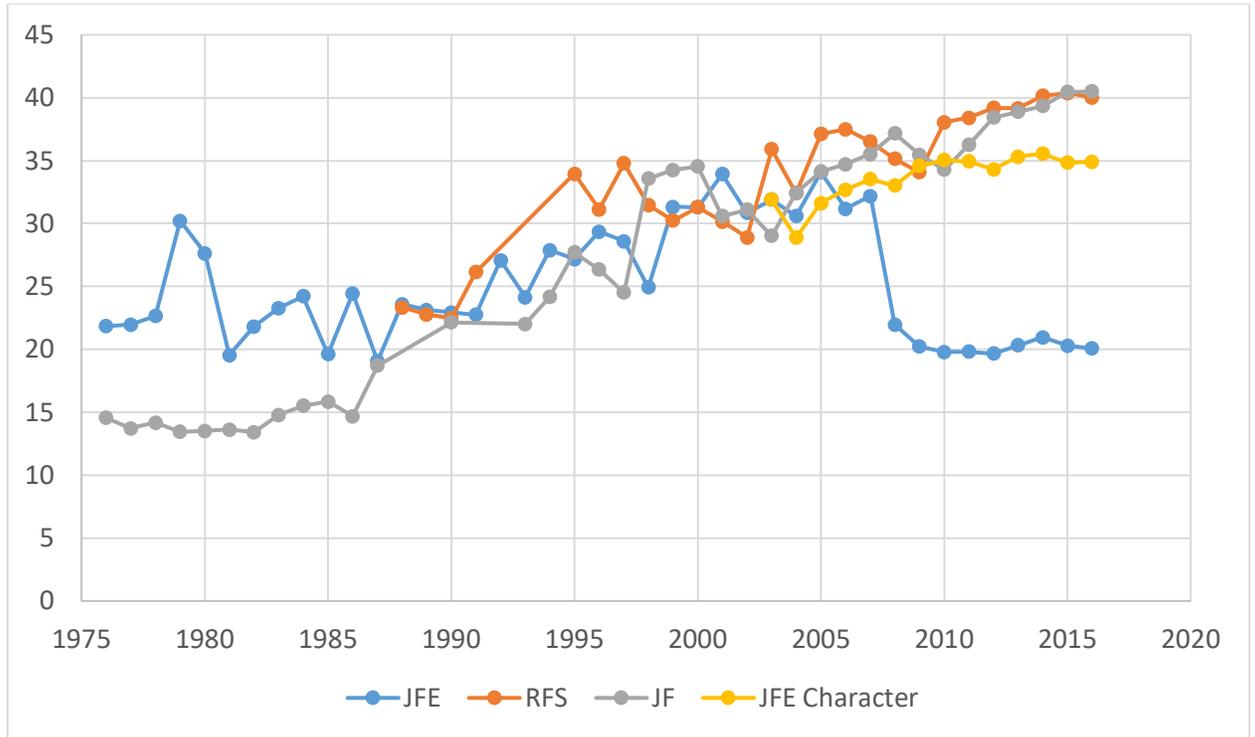


Figure 3: Average Number of pages per article. JFE also displayed by normalized character count due to 2003 format change.

Figure 3 displays the average number of pages in articles published by the big 3. In addition, it includes a normalized character count for the *JFE* to accommodate a 2003 print format change. (The normalization sets the average character count per paper equal to the journal’s average number of pages per paper basis for 2003.) Back in the early 1980’s papers in the *JF* were under 15 pages long and those in the *JFE* in the low 20’s. Now articles in the *JF* are more than twice as long, coming in at just over 40 pages! The increase in the *JFE* has been nearly as dramatic with its articles passing 35 pages in length, after allowing for the 2003 print format change.

Sadly, Figure 3 actually underestimates the actual trend in article length. Prior to the mid-2000s papers did not come with internet appendices. These are now routine. While these pages may only appear as part of the on line edition, they took time for the authors to produce. If it was time well spent it would be one thing. But, paper bloat means it is time spent doing ever more frivolous tests. People, of course, know this and are therefore unlikely to ever look at these additional tables. (Assuming they read past the paper's introduction at all.) For authors and journals, the only thing all this additional work does is produce a referee's letter saying "I am satisfied with the paper and recommend it for publication." A phrase that a reasonable review process would see typed out when the paper was 25 manuscript pages rather than 100.

While paper bloat is a problem for all academics it is a particular problem for junior faculty. Even though page counts have doubled over time, tenure clocks have not. Even worse, the trend towards concentrated bibliographies has made producing an impactful set of articles all the more difficult. Producing a good, but perhaps controversial paper that ends up outside the big 3 is no longer considered a "success." In large part because it will likely receive a limited number of citations. What is an academic to do? Fortunately, for those that like having a lot of co-authors at least, there is a solution! For whatever reason, tenure committees in finance do not divide up a paper's credit among its many authors. Instead each seems to receive 100% credit for it. Incentives matter and the reaction has been to increase the number of authors per paper.

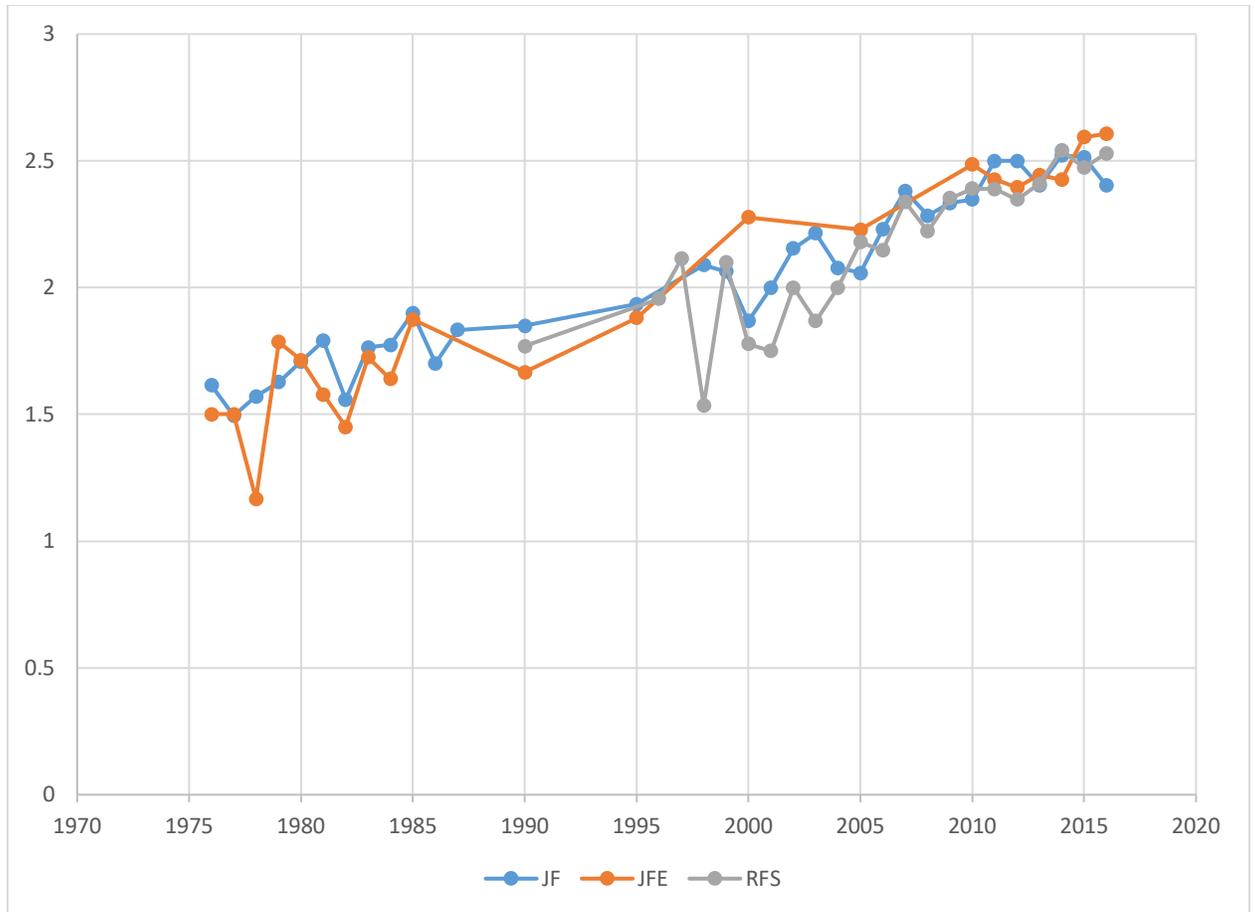


Figure 4: Average number of authors per paper.

Back in the 1970s there were about 1.75 authors on a typical paper. That meant there were many solo authored articles and relatively few with three or more. Now the typical paper has 2.5 authors. Indeed, the single authored paper is becoming a rarity. The February 2018 issue of the JFE has 2 out of 10 while neither the March 2018 issue of the RFS nor the December 2017 issue of the JF has any. How far is the profession from having four, five or even six authors as the norm? What choice do people have? Nobody can put together a publication record longer than an article or two prior to tenure if it has to end up at 100 plus pages. The only option is to divvy up the work and rely on the profession's practice of allocating 100% credit to every author.

One might hope that if the finance journals are demanding ever longer papers, perhaps an escape hatch can be found among mainstream economics journals. Alas, no. Card and Della Vigna (2013) find that in the top five economic journals papers are three times longer now than in the 1970s and the number of authors rose from 1.3 in 1970 to 2.3 in 2012. Worse, it seems these trends are being rewarded by others in the profession. They also report that increasing a paper's length and number of authors adds to its citation count. It would appear there is no escaping these trends by trying outlets outside of mainstream finance.

### Do Something About It!

Remember you can help stop the trend towards ever more referee demands. How? For a start stop writing long reports yourself. See how easy that is? Really you can do it! Make a pact with yourself, no report you write will exceed a page and a half. All "suggestions" will be fairly easy to implement. Editors will not blacklist you for this. Good ones may even send you a bottle of champagne to acknowledge your accomplishment. Will that stop *you* from being subject to the current system. Alas, not directly at least. But, if more people take up the cudgel it will decrease the odds you are sent to the referee torture chamber. Plus, I think many people are "doing unto others what was done unto them." By doing less unto others maybe they will pay it forward and do less unto others themselves. Rather than ask if the paper has covered very single test you can cook up, just ask if published would anybody read it? If the answer is yes, good enough. S-T-O-P!

Turning referee reports back into referee reports and letting authors publish shorter articles will help the profession move forward. Right now ideas are put forward and then receive very little scrutiny. Testing the veracity of a paper's claims is currently a losing proposition. Initially, you will need to write an incredibly long article. Then you will send it off to an outlet. The editor in turn will probably select referees with a vested interest in the old results. Why bother? In terms of scientific progress as a

profession we should instead encourage people to disrupt the prevailing orthodoxy. That requires letting debates occur which shorter papers will help foster.

You may be wondering what to do with all the free time you will have from writing shorter referee reports. Here is an idea, cite papers outside the big 3. By doing this you will improve the impact factor attributed to rival journals. If those rise, we will increase the number of journals that “count.” This in turn will increase competition for insightful (as opposed to long pedantic) articles. Competition is good. (You might recall teaching that in a course of yours. It is actually true.) Competition will force journals to do more to cater to authors. That in turn will lead to a natural reduction in the demands editor will let referees get away with. Just think not only will the profession benefit but you will too.

## Bibliography

Berninger, Marc, Edward Gaar, Florian Kiesel and Dirk Schiereck, 2017, “Confused but Convinced: Article Complexity and the Number of Citations,” working paper Department of Business Administration, Economics and Law, Technische Universität Darmstadt.

Bhattacharya, Utpal, 2017, “Editors Mid-year Report 2017,” email sent to editors and advisory editors of *Financial Management*.

Card, David and Stefano Della Vigna, 2013, “Nine Facts about Top Journals in Economics,” *Journal of Economic Literature* 51(1), 144-161.

Hirshleifer, David, 2014, “Editorial: Cosmetic Surgery in the Academic Review Process,” *Review of Financial Studies* 28(3), 637-649.

Jiang, Wei, 2017, “Have Instrumental Variables Brought us Closer to the Truth,” *The Review of Corporate Finance Studies*, 6(2) 127-140.