On the causal effect of fame on citations[1]

Jonathan Brogaard,[2] Joseph Engelberg,[3] Sapnoti Eswar,[4] Edward Van Wesep[5]

November 21, 2020

Abstract

Papers published in economics journals whose first authors are famous have more citations than papers whose second or third authors are famous. As a paper ages, its citation rate varies most with variation in the fame of the first author and less so with the fame of second and third authors. Author order is alphabetical so these patterns are unrelated to underlying quality. The magnitudes we find are large: a three-author paper written by the most prolific author in economics and his two research assistants would receive, on average, more than double the citations if the prolific author were first rather than second or third. The effect is especially pronounced in three, rather than two, author papers, suggesting that burying a famous author in the "*et al*" reduces citations even further.

JEL Codes: J0, J7, Z1

Keywords: Citations, publication, measurement, fame

[1] The authors are indebted to several parties for helpful comments and suggestions, most importantly seminar participants at the University of Colorado Boulder, Texas Christian University, Tony Cookson, Diego Garcia, Todd Gormley, Nandini Gupta, and Ryan Lewis.
[2] Department of Finance, University of Utah. brogaardj@eccles.utah.edu.
[3] Department of Finance, University of San Diego. jengelberg@ucsd.edu.
[4] Department of Finance, University of Cincinnati. eswarsi@ucmail.uc.edu.
[5] Department of Finance, University of Colorado Boulder. edward.vanwesep@colorado.edu.

1. Introduction

An author's fame is often measured by her citation counts. We show that the direction of causality is at least in part reversed: fame *causes* high citation counts. That is, two identical papers, one written by a famous author team and one written by a little-known author team, will receive substantially different citations. This result was famously hypothesized by Merton (1968) and dubbed *The Matthew Effect*. A number of papers written over the last 15 years have hinted at its presence in Economics and Finance academia but, to our knowledge, this is the first study to show that the effect is large and pervasive.

To show this result, an ideal experiment would follow Bertrand and Mullainathan (2004) and the subsequent literature: run an experiment, submitting papers with different authors that are otherwise identical to journals and observing their publication success and citation rates over time. For obvious reasons, this experiment cannot be run.[6]

Instead, we test a joint hypothesis: (i) the first author in a list receives more attention than later authors, and (ii) an author's fame causes her to be cited. The first part of the hypothesis has been extensively studied and confirmed in prior research,[7] and is obvious on its face with three- or four-author papers, references to which often bury later authors with the term *et al*. The joint test is therefore ultimately about the causal effect of fame on citations. Consistent with this hypothesis, we find that papers published in Economics and Finance journals whose first authors are highly cited will receive more citations than papers whose second or third authors are highly cited.

We consider all publications from 1974 to 2017 in a set of 48 top journals in Economics and Finance, and all citations to those publications listed in Web of Science. We define a paper's *citation percentile* as its percentile ranking by citations among all papers published in the same year within that set of journals. For example, consider Kamenica, Mullainathan, and Thaler (2011), entitled "Helping consumers know themselves," and published in the *American Economic Review*. The paper

[6] Similar experiments have been run in the world of fiction. In 1975 and 1979, aspiring writer Chuck Ross sent incorrectly attributed books (or sample pages) of 1969 National Book Award winner *Steps* to eight publishers and was rejected by all. In 2007, David Lassman sent opening chapters and synopses of a number of Jane Austen books to 18 British publishers and was roundly rejected. One publisher was keen to the hoax.

[7] We discuss this research in more depth in Section 2.

had 14 Web of Science citations as of July 2017, placing it in the 57th percentile among all papers written in 2011.[8]

A paper's citation percentile is a measure of its prominence in the field, and an author's count of high citation percentile papers is a measure of her prominence. We define a paper as a *home run* if it is in the top 5% of papers published that year (equivalently, has a citation percentile above 95%). We define an author's *fame* as the count of her high citation percentile papers; the more home runs that an author writes, the more famous she is defined to be. We then regress each paper's citation percentile on the fame of its first author, the fame of its second author, and, for three-authored papers, the fame of its third author.

The null hypothesis is that the coefficients are equal and positive. More famous authors will naturally write more highly-cited papers, but it should not matter whether a more famous author is listed first, second, or third. The alternative "fame" hypothesis is that the coefficient on the first author will be larger than the others.

We find that the first coefficient is indeed larger – much larger. To get a sense of magnitudes, consider again Kamenica, Mullainathan, and Thaler (2011). As of July 2017, the three authors had 3, 11, and 23 home run publications, in that order. Suppose that the author order for this manuscript were reversed. If citations were measuring only the quality of the paper, then it would remain a 57th percentile paper. Instead, according to our estimates, it would have been a 68th percentile paper. It is easy to find examples that generate even larger changes.

The effect is largest for researchers who have written many high-quality papers. That is, if we define a home run to be a top 10% manuscript, the effect shrinks. If we define a home run as top 25%, the effect shrinks again.

If we restrict attention to citations appearing in our set of 48 journals (i.e., dropping citations from less informed authors), we can track citations to individual papers over time. For each year after a paper is published, we observe the fame of all authors in that year and regress the citation percentile of the paper on each author's fame as well as paper fixed-effects. This allows us to hold constant the set of authors and the long-run average citation percentile of each paper and evaluate

[8] If this seems low, it is because Web of Science is a lagging citation indicator and the data were pulled in 2017.

how its citation rate changes as each author's fame increases or decreases. We find that papers whose first authors become more famous see a significantly higher increase in citations relative to papers whose second or third authors become more famous.

All of these results are robust to measuring citations using $\log(1 + citations)$ rather than *citation percentile*, to measuring an author's fame using her contemporaneous citations or cumulative citations, and to measuring a paper's citation level using contemporaneous citations or cumulative citations. In sum, author fame causes citations, and the magnitude is large.

There are several reasons to be interested in the causal effects of fame on citations, the most important being that they are used as a (supposedly) objective measure of impact in the promotion and tenure decisions of academics. Our results place a lower bound on the extent to which fame magnifies a researcher's citation count, and that lower bound is high. Academics can acquire fame in many ways other than producing exceptional research, for example by traveling more to conferences or seminars, taking on editorial responsibilities at a journal, or appearing regularly in mainstream media. Fame can also arise from bias. If some groups are disproportionately put into positions of prominence, those groups will also be disproportionately cited. Using citations as a measure of a researcher's impact is likely to be injecting substantial bias into promotion and tenure decisions.

2. Literature

Our study connects to a number of bibliometric literatures, both within Economics and Finance and without. Most closely related to our work, Simcoe and Waguespack (2011) evaluate publication rates for submissions to the Internet Engineering Task Force (IETF). Publication rates of submissions co-authored by high-status individuals were 77% lower when the high-status author's name was buried in an *et al* in the email announcing the submission.

One could imagine the results in our study being smaller or larger than this. On the one hand, submissions to the IETF are open to the public, so it may be more natural for reviewers to quickly screen new submissions by author name, as opposed to a setting when authors decide whom to cite. Citations are presumably related to important work on which a paper builds, as opposed to added after a quick scan of the literature. On the other hand, the decision to implement a new protocol (akin to publishing a paper) is more important than the decision to cite, so one might expect a smaller effect than in our study.

There is substantial work relating to the effect of alphabetical order on academic success.[9] Einav and Yariv (2006) look at the patterns of academic prominence for individuals with different last names. They find that academics with late names are less prevalent at top economics departments among tenured versus untenured faculty. This is not the case for lower-ranked departments. Their finding does not hold in psychology, a field that does not assign author order alphabetically. They also find that late names are less likely to be fellows of the econometric society. Along these lines, Efthyvoulou (2008) finds that faculty with earlier last names are more likely to be at top departments, to have their work downloaded, and to be cited. Van Praag and van Praag (2008) find that early name authors publish more papers in top economics journals.

There is strong evidence that people read lists from top to bottom so items listed first are disproportionately visible. Arsenault and Larivière (2015) document that papers whose first authors have early last names receive more citations. Huang (2015) shows that scientific papers with earlier first authors are more cited, but papers with earlier second, third, etc., authors are not. The latter result suggests that an association between author ability and last name is unlikely to explain the primary result, though the difference in roles between a first author and other authors in a scientific publication are typically quite large. The effect is more pronounced for papers with more co-authors, suggesting a culling of lists that get too long though, again, scientific papers with more co-authors can differ substantially from those with fewer.[10]

Perhaps most cleanly, Feenberg, Ganguli, Gaulet, and Gruber (2017) show that among papers published at the top of the NBER weekly digest, which at the time listed papers according to the

---

[9] Weber (2018) surveys the literature on alphabetical listing of authors on papers and its effects. The author summarizes the key facts which are: (1) alphabetical listing of authors gives an unfair advantage to authors with last name initials early on in the alphabet, and (2) researchers react strategically to this form of discrimination. The survey documents that first authors are likely to be given more credit for joint work, early surname authors are more likely to work at top departments, and are more likely to receive awards, early surname authors are more likely to have an advantage in publishing papers, and an advantage in downloads and abstract views. Researchers react strategically to this kind of discrimination. Authors with late last names work less in large teams than early surnames. Authors with late surnames are more likely to write papers on their best ideas alone, are more likely to disrespect the alphabetical norm, and are more likely to manipulate their names to move up in the alphabet.

[10] Aad *et al* (2015), for example, has 5,154 authors, and most of their roles were not similar to those of the principal investigators'.

first author's last name, were more downloaded, viewed, and cited than those listed at the bottom. The NBER has adopted random ordering in response. Haque and Ginsparg (2009) find the same result in the ArXiv paper repository.

Ray and Robson (2018) provide a mechanism for improving on the situation in which one's name affects one's success that allows credible signaling of author contributions to a paper, and is able to invade an environment in which ordering is currently alphabetical.

Hamermesh (2018) is the most recent of a line of work evaluating aspects of the citation process in economics. Among a wide variety of interesting findings, he finds little difference in citation rates for scholars with early versus late names, especially among junior faculty.

We apply the alphabetical ordering of names in economics differently, not to investigate how last names affect career outcomes, but to identify the effect of fame on citations.

Our paper also relates to two recent studies following citations of papers whose authors become more famous. Azoulay, Stuart, and Wang (2014) and McCabe and Babutsidze (2020) select authors who have won the Howard Hughes Medical Investigator Award and the Nobel Prize in Economics, respectively, and follow citations of their papers before and after they win the award. In both cases, using matched samples, they find a substantial increase in citations post-award even though their papers were already well-known and well-cited pre-award. We follow the same approach in some of our tests, but at a broader and more granular level. We measure the fame of all authors, award winning or not, over time and measure citations to their papers as their fame rises and falls. It is perhaps not surprising that, in the extreme example of Nobel Prize winners, fame causes citations. It may be more surprising that it does so for more mundane examples.

3. Data

Our data include all papers published in the set of journals outlined in Table 1. Most journals appear in our dataset in 1974, though some appear later. Each journal's date of first appearance is listed in Table 1. The journal list comes from Brogaard *et al*. (2014), and the data for each paper, including citations, were downloaded from Web of Science in July 2017.

Our analysis is simple. An observation is a published paper, which may have two or three authors. All analyses are performed separately for those two groups. There are two relevant variables that are not in the original data and must be constructed, and we construct these variables in three steps. First, we calculate, for each paper, its *citation percentile*. For each year in the data, we select all papers published in that year across all journals in the sample for that year, and we rank those papers by citations as of July 2017. Each paper's percentile in that ranking is defined as its citation percentile.

Second, we define each paper as a *home run* or not based on its citation percentile.[11] Depending on the regression, we may define a paper to be a home run if its citation percentile is >95, >90, >75, >50, or >0. This last category simply defines all papers to be home runs. We use citation percentile as our measure of a paper's citations because papers are cited more as they age. A four-year-old paper with 10 citations recorded in Web of Science has been fairly successful. A 20-year-old paper with 10 citations has not. Citation percentile also benefits from being uniform, whereas raw citations are highly skewed.

Third, for each paper, we calculate the number of home runs that each author has in the sample, *not including the paper in question*. Suppose that paper A is a home run, for example, and paper B is not, and suppose that both papers share an author. If the author's home run count associated with paper A were X, then her home run count associated with paper B would be X+1. If both or neither were home runs, then her home run count for both papers would be the same. Throughout the rest of the manuscript, we shorten *home run count* to *fame*.

Table 2 provides summary statistics for our measures of fame. Statistics are calculated separately for two- and three-author papers. The average number of papers written by the first author on a two-author paper is 16.05, and the average number written by the second author is 16.10. These are higher than for three-author papers which are, given the rise in co-authoring over the last few decades, written by younger scholars.

Continuing down to increasingly well-cited papers, the average number written drops. The average number of papers written by the first (second) author of a two-author paper in the top 5% of all papers published in its year is 1.59 (1.53). For two- and three-author papers, and for all definitions

[11] The term is taken from Brogaard, Engelberg, and Van Wesep (2018).

of fame, the average publication rates for authors in different positions in the author order are similar.

There are naturally many authors who have only one publication in our journal list, and we assign a value of 0 for fame. Of interest may be the authors with the most highly cited publications. Andrei Schleifer, who has been a first and second author on two-author papers, has 64 publications in the top 5% of papers published in the same year. He was never first author on a three-author paper in the top 5%. The most prolific first author on a top-5% three-author paper is James Heckman.

4. Results

4.1 Baseline effects of author order on citation percentiles

Our research design is variants of the following baseline regression:

$$y_{ijt} = \alpha + \beta_1 \times Fame_{ijt1} + \beta_2 \times Fame_{ijt2} + \beta_3 \times Fame_{ijt3} + \gamma_{jt} + \epsilon_{ijt,}$$

where y_{ijt} is the outcome of interest, either *citation percentile* or $\log{(1 + citations)}$, for paper i, published in journal j in year t; $Fame_{ijtk}$ is the *fame* of the k^{th} author of paper i; γ_{jt} is a fixed effect for journal j published in year t, and ϵ_{ijt} is an error. [12]

We believe this to be the simplest design that can deliver causal claims regarding our research question. Each β should be positive if we believe some authors tend to write more highly cited papers than other authors. We are not, therefore, interested in the null hypothesis that the true coefficients are zero. Instead, we provide results of F-tests for restrictions that $\beta_1 = \beta_2$, etc. If fame causes citations and if citers tend to notice earlier authors more, then our alternative hypothesis is that $\beta_1 > \beta_2 > \beta_3$.

In Table 3, Panel A, we present estimates in which y_{ijt} is citation percentile, and papers are two-authored. Each regression uses a different definition of home run when generating the fame variable.

[12] We force the author order to be alphabetical. That is, even if authors choose a non-alphabetical order in practice, we assume that they chose to list names alphabetically. Alternate orderings are not particularly common, and this procedure works against our results, relative to simply dropping those observations.

In column (1), every paper is defined to be a home run, so we are simply comparing how each author's total publications (except for the paper in question) correlates with citations. Surprisingly, the first author's total publications matter substantially more, and the difference is highly statistically significant: the p-value for the F-test of $\beta_1 = \beta_2$ is 0.000. To get a sense of the difference, suppose that the first author has 40 publications and the second 15. The paper in question, holding journal-year constant, would be expected to have a $(0.153 \times 40 + 0.116 \times 15) - (0.153 \times 15 + 0.116 \times 40) = 0.925$ higher citation percentile than if the author order were reversed. This might not seem large, but recall that this is for two-authored papers, and we are defining fame to simply be publication count.

As we move to columns (2), (3), (4), and (5), we raise the threshold for a paper to be defined as a home run, and therefore reduce the number of home-run papers. The coefficients on first- and second-author fame increase monotonically as we define fame more strictly. This should not be surprising: authors whose papers X and Y are more highly cited will also tend to receive more citations for paper Z. In each column, the first author's fame is more important than the second author's fame, and the differences are always highly significant.

Consider the same author pair as before, but now the first author has 40 95^{th} percentile papers and the second has 15 95^{th} percentile papers. The paper in question, holding journal-year constant, would be expected to have a 4.05 higher citation percentile than if the author order were reversed. In this example, the author with 40 home runs is very famous – known (by name at least) to nearly everyone in economics. The author with 15 is very well known within her field but may not be known (yet) to everyone. The additional citations that follow, if the more famous author is first, are large.

Table 3, Panel B, displays results from the same analysis for three-authored papers. The number of observations drops by more than half, as two-authored papers are much more common in the profession, but the first-author effect should rise, as the use of *et al* to hide second and third authors usually begins at three. We would expect larger coefficient differences and larger standard errors. This is indeed what we see.

Beginning again with the case where fame is measured simply as the number of papers that a person has published, X=0, the effect of a first author's fame is double that of a third author's. The differences in the coefficients are again highly significant, with p-values of the tests that they are equal less than 0.01. Returning to our example of authors of varying fame, consider three authors, now with 40, 20, and 10 papers. The additional citation percentile points if the most famous is first and the least famous is last, versus the opposite, is 1.89. This is not a large magnitude, but this is also a weak definition of fame.

As we increase the threshold for a paper to be considered a home run, coefficients once again monotonically increase, consistent with fame more closely matching how we think of it intuitively. In all cases, the coefficient for first author fame exceeds that for second author fame, which exceeds that for third author fame, suggesting that the effect is not entirely related to names being subsumed by the term *et al*. The differences between the coefficients on the second and third author are, however, much smaller, and not always statistically significant at standard levels, consistent with *et al* being a primary driver of the effect.

Focusing on the strictest definition of a home run, X=95, we again consider a hypothetical paper whose authors have 40, 20, and 10 home run papers. This paper's citation percentile would be 17.16 higher if the author order were by fame rather than in reverse. This is a large gap.[13]

4.2 Results using citations rather than citation percentiles

In order to make interpreting magnitudes easier, in Table 4 we re-do all analyses, replacing citation percentile with $log(1 + citations)$. In order to be consistent, we perform all analyses precisely the same way as in Table 3, except with logged citations as the dependent variable. We use the logarithm because citations are highly skewed and, more importantly, they accumulate over time. The additional citations accruing to papers with more famous first authors are likely to grow as a paper becomes more cited. Using a log allows us to measure the percentage increase in citations from changing author order.

[13] Due to space constraints, we do not report results for papers with four or more authors. The number of four-authored papers in the sample is 3,327, so standard errors are large. The overall pattern is similar to what we observe in Panels A and B. As the definition of a home run is made more stringent, the association between author fame and a paper's citations grows monotonically for all four author locations. The association between author order and a paper's citations is not monotonic in this sample, but in three of fifteen comparisons between the coefficient on the first author and the coefficient on another author, the difference is positive and statistically significant at the 10% level. The difference is never negative and statistically significant at the 10% level. As the definition of home run narrows, the association is more pronounced and the p-values for the restrictions that the coefficient on the first and later authors are equal falls to as low as 0.021.

For two-authored papers, the coefficient on first-author fame on citations is consistently about 25% to 30% larger than the coefficient on second-author fame, and the difference is always highly statistically significant. When X=0, and author fame is simply her publication count (minus the paper in question), a paper whose first author's count is 40 and whose second author's count is 15 will receive $(0.010 - 0.007) \times (40 - 15) = 0.075$ more logged citations, or a bit more than 7.5% more citations, relative to an identical paper for which the author order is switched.

Moving to the case where a paper is considered a home run if it is in the top 5% of papers published that year, and again assuming authors with 40 and 15 home runs, the paper will receive $(0.053 - 0.041) \times (40 - 15) = 0.30$ more log citations, which equates to approximately 35% more citations, if the prolific author is first. Depending on one's frame of reference, this may be large or small, but these are effects coming only from author order, which is one of many ways that an author can be more visible.

If we consider the most extreme case, in which Andrei Schleifer of Harvard University publishes a paper with a research assistant with no top 5% publications, the difference in predicted citations if professor Shleifer were listed first would be more than double the count if he were listed second. We perform this calculation with tongue firmly in cheek: it is not clear that one can linearly extrapolate to such an outlier.

For three-author papers, our results are much starker. The coefficients on author fame are monotonically decreasing for later authors and the coefficients are monotonically growing as we make the definition of a home run more stringent. F-tests for equality of the coefficients on first and third authors always reject the null that true coefficients are equal at the 0.1% level. F-tests for equality of the coefficients on first and second authors reject the null that the true coefficients are equal at the 5% level in four cases, and reject the null at the 1% level in two. Importantly, we can more strongly reject the null as we raise the threshold for a paper to be defined to be a home run. At the highest threshold, we can reject the null at the 0.1% level for both second and third authors.

We also compare the coefficients for second- and third-author fame, and have more mixed results. When the threshold for a home run is weak, we cannot reject that the true coefficients are equal, but as the threshold increases, the p-values associated with these tests hover around 5%.

Returning to our example of authors of varying fame, consider again three authors with 40, 20, and 10 papers. If we order authors from most cited to least, rather than the other way around, the log citations are predicted to increase by 0.75 points. As is the case with Professor Shleifer and his RA, predicted citations would be more than twice as high if the most prolific author is listed first rather than last.

These results are generally consistent with our claim that fame affects citations, and that the effects are large. As when we measure citations with citation percentile, the effects are much larger for three-author papers than two-author papers. This is consistent with *et al* burying later authors' names, as is the fact that the difference in coefficients for second and third authors is always smaller than the difference from first to second.

4.3 Results restricting the sample to pre-publication fame

Our baseline analysis considers citations and fame as of 2017 to all papers in our set of 48 journals. The advantage of choosing this single year is that we can use data from Web of Science, which identifies citations from all publications, including those outside this set. There are two disadvantages. First, we cannot rule out the alternative explanation for our results that, for some reason, papers in which the first author is famous are simply better than those for which the second or third author is famous. Second, many citations appear in journals that are not widely read. Our results that fame causes citations may be more limited: perhaps fame causes citations that appear in journals with limited impact. Perhaps the citations that truly matter – those appearing in widely-read journals – reflect only the true impact of the paper, not the fame of the authors.

We therefore restrict attention to citations that appear in our set of 48 journals, to papers published in that same set. The number of citations for each paper is much smaller than we observe in the Web of Science data because the set of journals in which the citations appear is much smaller, but we are able to observe the year of each citation. Furthermore, by restricting attention to only these citations, we are also restricting attention to only high-quality citations.

We perform regressions of the form:

$$y_{its} = \alpha + \beta_1 \times Fame_{is1} + \beta_2 \times Fame_{is2} + \beta_3 \times Fame_{is3} + \gamma_i + \epsilon_{is,}$$

where y_{its} is the outcome of interest, either *citation percentile* or $\log(1 + citations)$, in year $s \geq t$, for paper i, published in year t; $Fame_{isk}$ is the *fame* of the k^{th} author of paper i in year *s*; γ_i is a fixed effect for paper i and ϵ_{is} is an error.

That is to say, each observation is a paper-year. A paper published in 1998, for example, will be associated with observations in 1998, 1999,…,2017. Each paper published before 2017 is associated with multiple observations and, for papers published before 1998, we retain only the first 20 years after publication of the paper. For each paper-year, the dependent variable is a measure of the paper's impact, either using citation percentile or $\log(1 + citations)$, as of that year. The independent variables are measures of each author's fame as of that year. We include a paper fixed effect so the variation in the paper's impact is driven by changes in each author's fame over time.

It is not obvious how to measure the fame of each author or the impact of each paper in each year of its life, so we present eight analyses, each occupying a space in a 2x2x2 matrix. A paper's impact in a given year could be measured by the citations that it receives *in that year* or by the citations that it has received cumulatively *up to that year*. In either case, its impact might be best measured by citation percentile or by $\log(1 + citations)$. Similarly, an author's fame could be measured by the citations that she receives in that year or by the citations that she has received cumulatively up to that year.

Our first specification measures a paper's impact as its citation percentile in each year, using contemporaneous citations. That is, for a paper published in 1998, we measure its impact in 2002 using its percentile ranking of citations *received* in 2002 among all papers *published* in 1998. This specification measures each author's fame in a given year using her total citations to her other papers in that year. We present results of this regression in column (1) of Table 5 in Panel A, for two-authored papers, and Panel B, for three-authored papers. We include paper fixed effects so, whether a paper is typically 85th or 10th percentile in our sample, coefficients on each author's fame are not affected. Instead, variation in each author's fame drives variation in citations over time.

The advantage of using contemporaneous citations as the measure of a paper's and an author's impact in a given year is that it is "real time". As a paper ages, its citations in each year vary. Similarly, the aggregate citations that each author receives on all of her other papers in that year varies. The disadvantage is that the number of citations that a particular paper receives in a given year is often zero, and variation in citation percentile is therefore noisy.

Column (2) presents results from similar regressions, where each author's fame is measured using her cumulative citations on her other papers as of the year of the observation. Cumulative citations might be a better measure of fame, as they capture the longevity of an author's career as well as her current impact. Column (3) measures each paper's impact in a given year using its cumulative citations from its publication in year t to the observation year s. Column (4) measures both paper impact and author fame using cumulative citations.

This last specification is most similar to our baseline analysis, as it measures both paper impact and author fame using cumulative citations. The difference is that it does so in each year of a paper's life rather than just in 2017, and considers only citations appearing in our list of 48 journals.

In Table 6, we repeat the analyses using $\log(1 + citations)$ rather than citation percentile. These results are somewhat easier to interpret, as coefficients represent the impact of an increase in author fame on a paper's citations, in percentage terms. They are inferior, however, because paper fixed effects are less able to control for the average impact of a paper over its life. Citations grow over time, whereas paper fixed effects are fixed.

Across all eight regressions for two-authored papers, presented in Panels A in Tables 5 and 6, the coefficient on the first author's fame is statistically significantly larger than the coefficient on second author's fame in two regressions, smaller in two, and not significantly different in four. Because it is not obvious that one specification is better than the others, our interpretation is that there is no evidence that author order matters for two-authored papers. These results stand out as different from those resulting from our baseline analysis considering only cumulative impact as of 2017, which showed an effect of author order. Theory does not clearly suggest whether there should be an effect for two-authored papers, and our mixed results are consistent with this ambiguity.

Panels B in Tables 5 and 6 perform the same analyses for three-authored papers. In each case, as in prior tables, we present results from linear restriction F-tests for the equality of coefficients on first and later authors. Across 16 comparisons, equality is rejected at the 1% level in all but three tests and at the 5% level in all but two tests. It is fair to say that, as in our baseline analysis, the evidence that author order affects citations is clear.

5. Conclusion

A paper in economics, a field in which author order is typically alphabetical, is cited more if its most famous author appears first. Given that people are more attentive to the first item in a list, and that author alphabetization is unrelated to quality, we therefore provide causal evidence of an effect of fame on citations.

Because we find large effects on a paper's citations simply by changing the ordering of authors within a paper, the effect in the ideal experiment, in which known authors are replaced with unknown authors, would likely be larger. In short, our estimates probably underestimate the effect of fame on citations.

Perhaps fame causes citations because people tend to cite a paper when they recognize the name associated with it. If so, this does not imply that citers are making mistakes: familiar names probably write more interesting and well-executed papers. Given limited attention, it is individually rational to read, and therefore cite, papers written by more familiar names.

While it may be rational for an individual not to read a manuscript written by unfamiliar names, this behavior likely results in many high-quality manuscripts failing to receive the recognition that they deserve. This cautions against using citations in promotion and tenure decisions, and instead reading the papers themselves.

1. **References**

- Aad, G. *et al*. (ATLAS Collaboration, CMS Collaboration) (2015) Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{\boxed{s}}$=7 and 8 TeV. *Phys. Rev. Lett.* **114**, 191803 (2015)
- Arsenault, C. and V. Larivière (2015). Is paper uncitedness a function of the alphabet? *Proceedings of the 15th International Conference of the International Society for Scientometrics and Informetrics*: 286–287.
- Azoulay, P., Stuart, T., & Wang, Y. (2014). "*Matthew: Effect or Fable?" Management Science*, *60*(1), 92-109.
- Baldi, S. (1998) "Normative versus social constructivist processes in the allocation of citations: A network-analytic model", *American Sociological Review 63* (6), 829-846.
- Beaver, D. (2004) "Does collaborative research have greater epistemic authority?" *Scientometrics* 60, 399-408.
- Brogaard, J., J. Engelberg, and E. Van Wesep (2018), "Do economists swing for the fences after tenure?" *Journal of Economic Perspectives* 32(1): 179-194.
- Brogaard, J., J. Engelberg, and C.A. Parsons (2014), "Networks and Productivity: Causal Evidence from Editor Rotations", *Journal of Financial Economics* 111(1): 251-270.
- Card, D. and S. DellaVigna (2019), "What do editors maximize? Evidence from four economics journals", *The Review of Economics and Statistics* 102(1): 195-217.
- Cole, J.R. and B. Singer (1991), "A theory of limited differences: Explaining the productivity puzzle in science", in H. Zuckerman, J. Cole & J. Bruer (eds), "The Outer Circle: Women in the Scientific Community", New York: WW Norton: 277-10.
- Efthyvoulou, G. (2008), "Alphabet economics: The link between names and reputation", *The Journal of Socio-Economics* 37(3): 1266–1285.
- Feenberg, D., I. Gagnuli, P. Gaule, and J. Gruber (2017), "It's Good to Be First: Order Bias in Reading and Citing NBER Working Papers", *Review of Economics and Statistics* 9(1): 32-39.
- Hamermesh, DS. (2013), "Six decades of top economics publishing: Who and how?" *Journal of Economic Literature* 51(1): 162–172.
- Hamermesh, DS. (2018), "Citations in economics: Measurement, uses, and impacts", *Journal of Economic Literature* 56(1): 115-156
- Hargens, Lowell L. (2000), "Using the Literature: Reference Networks, Reference Contexts, and the Social Structure of Scholarship", *American Sociological Review*, 65 (6), 846–865.
- Haque, A., and P. Ginsparg, (2009), "Positional Effects on Citation and Readership in arXiv," *Journal of the American Society for Information Science and Technology* 60(11): 2203–2218.
- Huang, W. (2015), "Do ABCs Get More Citations than XYZs?" *Economic Inquiry* 53(1): 773–789.
- Kamenica, E., S. Mullainathan and R. Thaler (2011), "Helping consumers know themselves", *American Economic Review* 101(3): 417-422.
- MacRoberts, M. and B. MacRoberts, (1996) "Problems of Citation Analysis", *Scientometrics* 36, 435-444.

- Mahlck, P., and O. Persson (2000), “Socio-bibliometric mapping of intra-departmental networks”, *Scientometrics* 49 (1), 81-91.
- McCabe, M. J., & Babutsidze, Z. (2020). Scientific Prizes And Post-Award Attention: Evidence From The Nobel Prize In Economics. *Working Paper.*
- Merton, R. K. (1968). “The Matthew effect in science: The reward and communication systems of science are considered”, *Science*, *159*(3810), 56-63.
- Ray, D. ⓡ A. Robson (2018), "Certified Random: A New Order for Coauthorship", *American Economic Review*, 108(2): 489-520.
- Simcoe, T. S., & Waguespack, D. M. (2011). “Status, quality, and attention: What's in a (missing) name?” *Management Science*, *57*(2), 274-290.
- Van Praag, C. M. and B. van Praag, B (2008), “The benefits of being economics professor A (rather than Z)”, *Economica* 75(300): 782–796.
- Weber, M. (2018), “The effects of listing authors in alphabetical order: A review of the empirical evidence”, *Research Evaluation*, Volume 27, Issue 3, July 2018, Pages 238–245.

Table 1: List of Journals

This table reports all included journals, and the first year in which the journal appears in our dataset. Our dataset begins in the year 1974.

Journal Name		Journal Name	
ACCOUNTING REVIEW	1974	JOURNAL OF FINANCIAL ECONOMETRICS	1998
AMERICAN ECONOMIC JOURNAL	1976	JOURNAL OF FINANCIAL ECONOMICS	1974
AMERICAN ECONOMIC REVIEW	1974	JOURNAL OF FINANCIAL INTERMEDIATION	1990
ECONOMETRICA	1974	JOURNAL OF FINANCIAL MARKETS	1974
ECONOMIC JOURNAL	1974	JOURNAL OF FINANCIAL RESEARCH	1974
ECONOMIC THEORY	1974	JOURNAL OF HUMAN RESOURCES	1974
FINANCIAL ANALYSTS JOURNAL	1974	JOURNAL OF INDUSTRIAL ECONOMICS	1974
FINANCIAL MANAGEMENT	1974	JOURNAL OF INTERNATIONAL ECONOMICS	1974
GAMES AND ECONOMIC BEHAVIOR	1974	JOURNAL OF LABOR ECONOMICS	1976
INTERNATIONAL ECONOMIC REVIEW	1974	JOURNAL OF LAW & ECONOMICS	1974
JOURNAL OF ACCOUNTING & ECONOMICS	1974	JOURNAL OF MONETARY ECONOMICS	1974
JOURNAL OF ACCOUNTING RESEARCH	1974	JOURNAL OF MONEY CREDIT AND BANKING	1974
JOURNAL OF APPLIED ECONOMETRICS	1974	JOURNAL OF POLITICAL ECONOMY	1974
JOURNAL OF APPLIED ECONOMICS	1974	JOURNAL OF PUBLIC ECONOMICS	1974
JOURNAL OF BANKING & FINANCE	1974	MANAGEMENT SCIENCE	1974
JOURNAL OF BUSINESS	1974	MATHEMATICAL FINANCE	1974
JOURNAL OF BUSINESS & ECONOMIC STATISTIC	1974	QUARTERLY JOURNAL OF ECONOMICS	1974
JOURNAL OF CORPORATE FINANCE	1974	RAND JOURNAL OF ECONOMICS	1974
JOURNAL OF ECONOMETRICS	1974	REVIEW OF ECONOMIC DYNAMICS	1974
JOURNAL OF ECONOMIC GROWTH	1989	REVIEW OF ECONOMIC STUDIES	1974
JOURNAL OF ECONOMIC LITERATURE	1974	REVIEW OF ECONOMICS AND STATISTICS	1974
JOURNAL OF ECONOMIC PERSPECTIVES	1974	REVIEW OF FINANCE	1990
JOURNAL OF ECONOMIC THEORY	1974	REVIEW OF FINANCIAL STUDIES	1974
JOURNAL OF FINANCE	1974		
JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS	1974		

Table 2: Summary Statistics

This table reports the summary statistics for each variable used in the regression specifications. For two-author papers, we report summary statistics for a count of the number of papers excluding the current paper which are above the 0^{th} (50^{th}, 75^{th}, 90^{th}, 95^{th}) percentile by the first author (Author 1), and the second author (Author 2). We report the same statistics for three-author papers in our sample.

		Two-Author Papers		Three-Author Papers		
		Author 1	Author 2	Author 1	Author 2	Author 3
		N = 42,358		N=17,341		
Papers above 0th Percentile	Mean	16.05	16.10	14.03	13.68	13.52
	Std Dev	19.03	20.10	17.25	17.44	17.61
	Min	0	0	0	0	0
	Max	160	160	138	160	160
Papers above 50th Percentile	Mean	10.09	10.09	8.95	8.64	8.57
	Std Dev	13.87	14.62	12.52	12.57	12.82
	Min	0	0	0	0	0
	Max	132	132	104	131	132
Papers above 75th Percentile	Mean	6.09	5.98	5.39	5.10	5.06
	Std Dev	10.02	10.31	8.93	8.91	9.08
	Min	0	0	0	0	0
	Max	102	102	77	101	102
Papers above 90th Percentile	Mean	2.88	2.79	2.51	2.41	2.43
	Std Dev	6.12	6.10	5.21	5.66	5.74
	Min	0	0	0	0	0
	Max	83	83	52	83	83
Papers above 95th Percentile	Mean	1.59	1.53	1.36	1.32	1.33
	Std Dev	4.02	4.04	3.35	3.87	3.89
	Min	0	0	0	0	0
	Max	64	64	35	64	64

Table 3: Effect of author citation percentiles on paper citation percentile

This table reports the impact of each author's position and a count of the number of papers by the author with citations above a certain percentile on the paper citation percentile. The dependent variable in all the specifications is the citation percentile of the paper. In Panel A, we include two-authored papers, and in Panel B, three-authored papers. In each column, we count the number of papers by each author which contain citations which are above the Xth percentile. For columns 1, 2, 3, 4 and 5, X is equal to 0, 50 ,75, 90, and 95 respectively. Our sample includes all papers published after the year 1974. Robust standard errors are reported in parentheses, only for the tests of difference in coefficients. *, **, and *** indicate significance at 5%, 1%, and 0.1% respectively.

Variable: Citations percentile	(1)	(2)	(3)	(4)	(5)
Panel A: Two-authored papers	X=0	X=50	X=75	X=90	X=95
Author 1: Number of Publications above Xth Percentile	15.250	25.772	37.575	57.987	81.737
	(0.645)	(0.880)	(1.250)	(2.208)	(3.523)
Author 2: Number of Publications above Xth Percentile	11.613	20.085	30.402	48.542	65.200
	(0.595)	(0.827)	(1.223)	(2.097)	(3.195)
p-value for test of difference in coefficients of					
Author 1 and Author 2	0.0000***	0.0000***	0.0000***	0.0045**	0.0013**
Journal x Year FE	Yes	Yes	Yes	Yes	Yes
Observations	42,358	42,358	42,358	42,358	42,358
R-squared	0.284	0.296	0.301	0.298	0.293

Variable: Citations percentile	(1)	(2)	(3)	(4)	(5)
Panel B: Three-authored papers	X=0	X=50	X=75	X=90	X=95
Author 1: Number of Publications above Xth Percentile	12.675	22.499	34.817	60.208	92.528
	(1.172)	(1.601)	(2.254)	(3.906)	(6.261)
Author 2: Number of Publications above Xth Percentile	8.155	16.386	27.504	40.768	54.686
	(1.174)	(1.564)	(2.183)	(3.588)	(5.441)
Author 3: Number of Publications above Xth Percentile	6.430	13.018	21.035	31.475	38.731
	(1.099)	(1.495)	(2.118)	(3.272)	(4.724)
p-value for test of difference in coefficients of					
Author 1 and Author 2	0.0086**	0.0092**	0.0279*	0.0006***	0.0000***
Author 1 and Author 3	0.0000***	0.0000***	0.0000***	0.0000****	0.0000***
Journal x Year FE	Yes	Yes	Yes	Yes	Yes
Observations	17,341	17,341	17,341	17,341	17,341
R-squared	0.300	0.311	0.319	0.318	0.315

Table 4: Effect of author citation percentiles on paper citation

This table reports the impact of each author's position and a count of the number of papers by the author with citations above a certain percentile on the paper citation. The dependent variable in all the specifications is the natural logarithm of (1+ total number of citations received by the paper). In Panel A, we include two-authored papers, and in Panel B, three-authored papers. In each column, we count the number of papers by each author which contain citations which are above the Xth percentile. For columns 1, 2, 3, 4 and 5, X is equal to 0, 50 ,75, 90, and 95 respectively. Our sample includes all papers published after the year 1974. Robust standard errors are reported in parentheses, only for the tests of difference in coefficients. *, **, and *** indicate significance at 5%, 1%, and 0.1% respectively.

Variable: Log(1 + Citations)	(1)	(2)	(3)	(4)	(5)
Panel A:Two-authored papers	X=0	X=50	X=75	X=90	X=95
Author 1: Number of Publications above Xth Percentile	0.958	1.582	2.322	3.699	5.321
	(0.038)	(0.052)	(0.075)	(0.136)	(0.221)
Author 2: Number of Publications above Xth Percentile	0.725	1.209	1.835	2.971	4.115
	(0.035)	(0.049)	(0.073)	(0.128)	(0.201)
p-value for test of difference in coefficients of					
Author 1 and Author 2	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***
Journal x Year FE	Yes	Yes	Yes	Yes	Yes
Observations	42,358	42,358	42,358	42,358	42,358
R-squared	0.452	0.463	0.468	0.467	0.463

Variable: Log(1 + Citations)	(1)	(2)	(3)	(4)	(5)
Panel B: Three-authored papers	X=0	X=50	X=75	X=90	X=95
Author 1: Number of Publications above Xth Percentile	0.668	1.150	1.765	3.083	4.817
	(0.058)	(0.080)	(0.114)	(0.201)	(0.326)
Author 2: Number of Publications above Xth Percentile	0.455	0.878	1.495	2.291	3.130
	(0.060)	(0.082)	(0.116)	(0.193)	(0.297)
Author 3: Number of Publications above Xth Percentile	0.371	0.716	1.147	1.757	2.274
	(0.055)	(0.076)	(0.110)	(0.174)	(0.255)
p-value for test of difference in coefficients of					
Author 1 and Author 2	0.0131*	0.0226*	0.1144	0.0072**	0.0003***
Author 1 and Author 3	0.0003***	0.0001***	0.0002***	0.0000***	0.0000***
Journal x Year FE	Yes	Yes	Yes	Yes	Yes
Observations	17,341	17,341	17,341	17,341	17,341
R-squared	0.587	0.595	0.601	0.601	0.600

Table 5: Effect of author citations on paper citation percentile in the time-series

This table reports the impact of each author's position and a count of the number of citations by the author in all other papers published before or in the current year on the paper citation percentile. The dependent variable is the paper citation percentile. In Panel A, we include two-authored papers, and in Panel B, three-authored papers. In each column for the dependent and independent variables, we use either the count of the number of citations received by each paper in the current year, or the running total of the number of citations received by the paper up to and including the current year, i.e., cumulative citations. Robust standard errors are reported in parentheses, only for the tests of difference in coefficients. *, **, and *** indicate significance at 5%, 1%, and 0.1% respectively.

Variable: Citations percentile Panel A:Two-authored papers	(1)	(2)	(3)	(4)
Author 1: Fame	5.017	0.183	6.057	0.825
	(0.277)	(0.028)	(0.203)	(0.022)
Author 2: Fame	6.677	0.201	6.429	0.795
	(0.274)	(0.028)	(0.196)	(0.021)
p-value for test of difference in coefficients of				
Author 1 and Author 2	0.0002***	0.6895	0.2372	0.3871
Paper FE	Yes	Yes	Yes	Yes
Dependent variable cumulative	No	No	Yes	Yes
Independent variables cumulative	No	Yes	No	Yes
Observations	635,059	635,059	635,059	635,059
R-squared	0.468	0.467	0.831	0.832

Variable: Citations percentile Panel B: Three-authored papers	(1)	(2)	(3)	(4)
Author 1: Fame	8.853	0.273	7.713	1.025
	(0.533)	(0.058)	(0.373)	(0.042)
Author 2: Fame	3.384	0.106	3.320	0.375
	(0.389)	(0.043)	(0.285)	(0.034)
Author 3: Fame	5.525	0.292	4.992	0.594
	(0.469)	(0.054)	(0.349)	(0.041)
p-value for test of difference in coefficients of				
Author 1 and Author 2	0.0000***	0.034*	0.0000***	0.0000***
Author 1 and Author 3	0.0000***	0.8304	0.0000***	0.0000***
Paper FE	Yes	Yes	Yes	Yes
Dependent variable cumulative	No	No	Yes	Yes
Independent variables cumulative	No	Yes	No	Yes
Observations	202,247	202,247	202,247	202,247
R-squared	0.496	0.493	0.827	0.828

Table 6: Effect of author citations on paper log citations in the time-series

This table reports the impact of each author's position and a count of the number of citations by the author in all other papers published before or in the current year on the paper citations. The dependent variable is the natural logarithm(1 + paper citations). In Panel A, we include two-authored papers, and in Panel B, three-authored papers. In each column for the dependent and independent variables, we use either the count of the number of citations received by each paper in the current year, or the running total of the number of citations received by the paper up to and including the current year, i.e., cumulative citations. Robust standard errors are reported in parentheses, only for the tests of difference in coefficients. *, **, and *** indicate significance at 5%, 1%, and 0.1% respectively.

Variable: Log(1 + Citations) Panel A:Two-authored papers	(1)	(2)	(3)	(4)
Author 1: Fame	0.242	0.011	1.039	0.159
	(0.008)	(0.001)	(0.019)	(0.002)
Author 2: Fame	0.300	0.013	0.978	0.150
	(0.008)	(0.001)	(0.017)	(0.002)
p-value for test of difference in coefficients of				
Author 1 and Author 2	0.0000***	0.2494	0.0205*	0.0043**
Paper FE	Yes	Yes	Yes	Yes
Dependent variable cumulative	No	No	Yes	Yes
Independent variables cumulative	No	Yes	No	Yes
Observations	635,059	635,059	635,059	635,059
R-squared	0.593	0.588	0.730	0.755

Variable: Log(1 + Citations) Panel B: Three-authored papers	(1)	(2)	(3)	(4)
Author 1: Fame	0.416	0.017	1.134	0.183
	(0.015)	(0.002)	(0.027)	(0.003)
Author 2: Fame	0.160	0.007	0.597	0.082
	(0.012)	(0.001)	(0.024)	(0.004)
Author 3: Fame	0.283	0.017	0.753	0.117
	(0.014)	(0.002)	(0.030)	(0.004)
p-value for test of difference in coefficients of				
Author 1 and Author 2	0.0000***	0.0000***	0.0000***	0.0000***
Author 1 and Author 3	0.0000***	0.9416	0.0000***	0.0000***
Paper FE	Yes	Yes	Yes	Yes
Dependent variable cumulative	No	No	Yes	Yes
Independent variables cumulative	No	Yes	No	Yes
Observations	202,247	202,247	202,247	202,247
R-squared	0.609	0.601	0.715	0.748