Dissemination, Publication, and Impact of Finance Research: When Novelty Meets Conventionality*

Rui Dai, Lawrence Donohue, Qingyi (Freda) Drechsler, and Wei Jiang †

Current Version: March 15, 2022

*We thank Alex Edmans, Itay Goldstein, Cam Harvey, David Hirshleifer, Andrew Karolyi, Song Ma, Matt Spiegel, Laura Starks, and Frank Yu for their suggestions and encouragement. We are particularly grateful for the inputs from an anonymous referee, an Associate Editor and Marcin Kacperczyk (Editor). The study also benefits from feedback from participants at seminars at SUFE, NFA, and the AFA/AFFECT mentoring event.

†Rui Dai, Lawrence Donohue, and Qingyi (Freda) Drechsler are from Wharton Research Data Services (WRDS), University of Pennsylvania. Wei Jiang is affiliated with Columbia Business School, NBER, and ECGI. Contact information: Dai: rdai@wharton.upenn.edu, Donohue: donohuel@wharton.upenn.edu, Drechsler: qsong@wharton.upenn.edu, and Jiang: wj2006@columbia.edu.

Dissemination, Publication, and Impact of Finance Research: When Novelty Meets Conventionality

ABSTRACT

Using numeric and textual data extracted from over 50,000 finance articles in SSRN during 2001–2019, we examine the relationship between measured qualities and a paper's readership, eventual outlet, and impact. Conventionality (semantic similarity with existent research) helps boost readership and publication prospects. However, novelty in the forms of emerging topics and databases are associated with better publishing outcomes. Studies that do not easily map into established finance subfields or that introduce non-finance elements face a higher hurdle. Finally, papers whose research questions span multiple fields are a hard sell, but those building on prior knowledge from multiple fields are valued.

Keywords: Finance Research; Impact; Publication; Innovation, Machine Learning, Textual Analysis
JEL Classification Number: A14, D85, G00

1 Introduction

Finance research has shaped many aspects of markets, businesses, and society at large (see a summary by Zingales (2015)). While financial economists devote most of their research resources into analyzing the operational and financial efficiency of business models and entities, they have not, until quite recently, put their own modus operandi—the production, dissemination, and recognition of research—under the same level of scrutiny. While there is a growing literature within the academic community on the "science of scientific research," most such studies are based on theory (e.g., Ellison, 2002), editorial and bibliographic data usually confined to published papers (e.g., Spiegel, 2012; Card and DellaVigna, 2013; Hirshleifer, 2014),[1] or surveys (e.g., Card et al., 2020). This paper aims to be the first large-scale empirical study that uncovers the determinants of research outcomes, including publication outlet, impact, and readership in academic finance.

The objective of our study is to reconcile—confirming, contrasting, and uncovering nuances—the empirical data with the anecdotal evidence and common wisdom regarding how research is evaluated and recognized by the community and its premier outlets. Such information should help the profession in reflecting on its prevailing practices and in seeking improvement so that finance academia can continue to incubate creativity and innovation with real and long-lasting impact. We hope to accomplish this goal by overcoming two empirical challenges. First, unlike prior research that mostly focuses on the sample of papers accepted by journals and conferences, we build our study on the near universe of finance research from working papers to published work. Such an unconditional analysis not only overcomes data censoring but also allows an examination of the publication filtering process per se, which is arguably the profession's most important machinery given the "publish or perish" environment. Second, we leverage current state-of-art machine-learning techniques to characterize research papers along dimensions, such as conventionality and novelty, that have not previously been quantified, and gain objective inferences about the relationship of these qualities with publication and impact.

Our sample starts with the universe of finance papers posted in the Financial Economics Network (FEN) on SSRN, an online depository of scholarly research in social sciences since 1994. After filtering down to papers with traits that are usually considered viable for submission to peer-reviewed journals by academically oriented authors, we end up with a sample of 52,497 research articles. About 22.5% of this sample have eventually been published in one of the 28 leading journals in finance and related disciplines. In reverse, these articles represent 75.0% of the papers published in the *Journal of Finance*, *Journal of Financial Economics*, and the *Review of Financial Studies* (the "top three finance journals"). We explore the paper and author characteristics with research outcomes related to publication (especially in the top three finance journals), readership (measured by downloads), impact (measured by Google Scholar citation), and conference acceptances at the two leading general-interest finance conferences (the annual meetings of the American Finance Association (AFA) and the Western Finance Association (WFA)). The

[1] The literature based on data related to published articles also includes studies on the research productivity of schools (Borokhovich, Bricker, Brunarski, and Simkins, 1995; Kim, Morse, and Zingales, 2009), journal ranking (Alexander and Mabry, 1994; Arnold, Butler, Crack, and Altintig, 2003), the editing and review processes (Brogaard, Engelberg, and Parsons, 2014; Welch, 2014), and coauthorship networks (Chen and Huang, 2007; Ductor et al., 2014)

results, not all of which conform to the common wisdom, could be summarized as follows.

The first and perhaps most important finding is that familiarity breeds publication and readership. We measure a given paper's semantic similarity with the existing stock of finance research using natural language processing tools including both bag-of-words (BOW) and the more advanced Universal Sentence Encoding model (USE) (Cer et al., 2018). A higher similarity measure is sign of conventionality in the work's language. A one standard deviation increase in conventionality (using the USE measure) is associated with an incremental probability of 12.5% to 15.3% hitting a top three finance journal, or 0.9 to 1.1 percentage points over the unconditional probability of 7.2%. High conventionality papers also invite significantly more readership, though lose in longer-term citations. Such a result might seem puzzling, as it suggests that a paper's measured differentiation from previous work is not as prized by our profession's top outlets despite the academic community's stated goal of pursuing original research. However, there is also a twist to this relation in that if the focal paper is too close to an existing paper (with a USE similarity score above 0.8), then the prospect of publication plummets by 13.9% (1.0 percentage points).

The combined results reveal a nuanced attitude toward semantic conventionality within the research community. On the one hand, both readers and journals welcome papers that have a large footprint on the existing knowledge, possibly because readers and reviewers are better able to connect and resonate with such research (Spiegel, 2012) or because referees and editors are risk averse and are thus reluctant to endorse research whose quality is difficult to assess (Shepherd, 1995). On the other hand, readers do value, and journals do require, new papers to be clearly differentiated from their intellectual next-of-kin. Top journals are particularly unforgiving when a paper's writing fails to distance from its "highly similar precursor." Admittedly, qualities key to research success, notably topic importance, being of interest to a wide audience, as well as execution quality are hard to quantify directly. Such a challenge is familiar to researchers in the innovation literature, e.g., in assessing the value and quality of patents.

We conduct several tests to rule out the possibility that our conventionality measure is a proxy for the omitted variable for quality. The first test involves the addition of *PrePubCites*, the number of citations received by an article before publication (or the total number of citations for unpublished papers to our data vintage date), in the research outcome regressions as a control for research quality. The effect of conventionality retains its significance after such an adjustment. In the second test, we focus on a selected sample of the 500, 1,000, and 2,000 papers (published or not) with the highest quality-adjusted citation metric. Within such subsamples of undeniably high-quality work, it is still the case that conventionality eases publication. Finally, we show that unconventional studies enjoy significantly higher impact since two to three years post to publication among published papers, supporting the hypothesis that they were subject to a higher hurdle in the publication process.

Second, "novel elements," in the form of an emerging topic or a novel dataset, are associated with better outcomes throughout. However, the race is also tight in that novel topics become saturated and reshuffle quickly,[2] and papers deploying novel datasets have been far from rarity since the mid-2000s. On the other hand, novelty in terms of deviating from the standard finance subfields, or being in an "Other" category[3]

[2]For example, in one year after the eruption of COVID-19 pandemic, the NBER has hosted over 400 papers on the topic.

[3]Based on a data-driven categorization, we focused on the following five main subfields: Asset Pricing, Corporate Finance, Financial

(as is this paper), or introducing non-finance elements deters readership and dims publication prospects.[4] Relatedly, field concentration, measured as the Herfindahl-Hirschman Index (HHI) of the probabilities that a paper falls into each of six outcome fields, is associated with better outcomes. However, diversity (i.e., a lack of concentration) in the research elements from which the focal paper's references are drawn—analogous to the "originality" measure for patents (Hirshleifer, Hsu, and Li, 2018)—turns out to be a plus. Thus, there seems to be a contrast in our profession's attitude toward papers whose research questions span multiple fields versus papers that build on prior knowledge from multiple fields. Readers and journals in finance value authors' effort and ability in bridging knowledge from different research elements but are nevertheless skeptical of research that tackles questions without a well-defined habitat.

Third, the correlation between author characteristics and research outcome turns out to be less surprising. However, the magnitude of the association should still be interesting. Authors from well-endowed schools (as measured by the number of WRDS databases that the institution subscribes to) enjoy an advantage. Having at least one author from a top 20 research school (per UT Dallas research ranking) is associated with a 68.4%, 17.5%, and 59.7% (or 4.3 percentage point) increase in citation, downloads, and probability of publication in the top three finance journals, respectively. Finally, papers by authors that are "central" in the profession (measured by the eigenvalue centrality in the coauthor network) also enjoy a boost. Determinants pertaining to author characteristics are likely a combination of selection and treatment effect. Affiliation with a high-reputation/resource institution and centrality within the network of authors are successful outcomes by themselves; in the meantime, resources and centrality also facilitate dissemination and recognition of the resulting research product.[5]

People are often under the impression that authors' experience could be quite different based on their seniority, reputation, and affiliations. We find that the general patterns discussed above hold for all groups of authors sorted along these dimensions, but there are a few secondary-order differences. Finance papers with a nontraditional focus that are authored by researchers from the top 20 schools receive even more citations, but do not invite more audience. Readers, as well as peer reviewers, are relatively more welcoming when highly "central" authors present atypical and unconventional work. On the other side, early stage researchers' articles have a greater likelihood of being published in the top three finance journals but receive fewer citations when they introduce non-finance elements. Additionally, new data sources give them a bigger boost in attracting readership, compared to senior scholars.

Finally, we confirm that admissions into top conferences such as AFA and WFA are significant precursors to publication and impact. Being on the program of either conference predicts a 48.1% increase in the probability of publishing in the top three finance journals relative to propensity-matched control papers, and an 87.3% increase in citation. The two conferences, however, exhibit differences in preferences. A paper outside the usual topic areas is more likely to be accepted at the WFA but less so at the AFA.

Intermediation, Investment, and Market Microstructure. A residual category is a catch-all "Diverse Field." We apply a convolutional neural network (CNN) model to calibrate the probability that a paper falls into any of the six categories using sessions at the top conferences as the training sample.

[4]The most famous example is probably the quick rejections Fischer Black and Myron Scholes received from two journals for their seminal option pricing paper, which built on the knowledge of particle movements from physics.

[5]See Brogaard, Engelberg, Eswar, and Van Wesep (2020) for a carefully identified study on the causal effect of author reputation on citations.

Moreover, author reputation appears to be a stronger predictor for inclusion in the AFA than the WFA. Such a comparison is indicative of the differences in the selection mechanism between a peer-review model and a session-chair-decision model.

The main findings of our study probably confirm the common understanding that while exploratory projects have the potential for high recognition, researchers need to first overcome the hurdles of the publication process, especially the completeness and robustness requirements expected at top outlets (Ellison, 2002) and the uncertainty in receiving proper valuations from referees from the more conventional domains (Hirshleifer, 2014). While the objective of this research is not to provide a "how to" guide for publication, we do not mind if some findings are to be taken as "careerist" lessons for people in the finance academia, especially junior scholars with limited capacity to tolerate publication failures. For example, one can make the inference that the prospect of publication is maximized, all else equal, when authors write their papers in semantics that are familiar to the literature and hence to the review team,[6] but at the same time provide a clear distinction against the closest earlier work. Our research may inform profession leaders on directing the disciplines through encouraging researchers to maintain a proper mix of creativity and familiarity given the limited resources in editorial process relative to the substantial growth of research outputs, especially taking into account junior scholars' career concerns that are similar to early-stage professionals in other sectors such as stock analysis and portfolio management (e.g., Hong, Kubik, and Solomon, 2000; Chevalier and Ellison, 1999).

Our ultimate aim is to provide comprehensive and objective evidence that could help the profession, especially its leaders, to reflect on the prevailing patterns and to think of creative ways to encourage and promote innovations in research. While a large and growing literature aims to measure innovation, explain successes, and promote best practices in firm innovation policies (e.g., Manso, 2011; Cohen, Diether, and Malloy, 2013), finance scholars should apply the same approach to our own production. Many of our findings echo or validate the comments made by various professional leaders (e.g., Ellison, 2002; Spiegel, 2012; Hirshleifer, 2014; Welch, 2014). We also note that there have already been targeted efforts by leading journals to promote research in "unfamiliar" topic areas via novel procedures such as registered reports (e.g., Goldstein, Jiang, and Karolyi (2019) on FinTech[7] and Hong, Karolyi, and Scheinkman (2020) on climate finance) and the launching of "perspective"-oriented journals. We expect to see more efforts along these lines aiming at a balance between encouraging true breakthroughs (often with some loose ends) and holding high execution standards on incremental refinements.

Finally, this study hopes to contribute to the general innovation literature. The trade-off between exploration (deviating from the established knowledge domain) and exploitation (refinement of existing findings) is not at all unique to the process of academic research. Neither are the challenges in motivating innovation and in measuring the quality of innovation outputs (such as patents). They are recurring themes in the research and practice of decision making in corporate innovation strategies and government

[6] This was, presumably, the strategy by Charles Darwin who devoted the first part of *On the Origin of Species* to well-accepted knowledge at the time about the selective breeding of dogs and cattle. This was also, according to Schaefer (1998), how Black and Scholes (1973) was eventually published in the *Journal of Political Economy* after the authors incorporated the comments from the more experienced Merton Miller and Eugene Fama.

[7] The editors and authors explained that a main motivation behind finance's first registered report process is to overcome potential authors' concern that a FinTech paper did not fit into any established literature and that there might be too few qualified referees (at the time).

industrial policies.[8] Some of the findings from our study are applicable as general lessons for spurring and disseminating innovations in that we cannot improve on what we cannot measure; moreover, we need to think hard how to *measure what we want* because we will *get what we measure*.

The rest of the paper is organized as follows. Section 2 describes the sample construction and provides a sample and data overview. Section 3 introduces the variable construction methodology. Section 4 presents empirical results. Finally, Section 5 concludes.

2 Related Literature and Hypotheses

Our paper falls into the broad literature on innovation because dissemination and consumption of research (via publication, downloads, and citations) could be viewed as a demand for innovation. The existent innovation research has primarily focused on innovation supplies and has not systematically examined the relationship between novelty/conventionality and demand. A growing strand of the literature has zoomed into the academic review and publication process but has neither composed a comprehensive sample that includes unpublished papers nor has the existent research modeled a comprehensive set of factors. Closely related to our theme are studies using articles from scientific disciplines (e.g, Uzzi et al., 2013) which find that papers with conventional frames are more likely to become be well-cited in the long run (eight years down the road), and atypical bibliographic combination is a riskier strategy for authors to seek high-impact outlets (e.g, Stephan, Veugelers, and Wang, 2017).

In addition to working on an unconditional sample, we leverage the state-of-art textual analysis toolkit to quantify semantic similarity (e.g., Hoberg and Phillips, 2016). We further supplement the textual measures with various indicated for "novel elements" in research, including emerging topics and databases (Hanley and Hoberg, 2019; Hope, Hu, and Lu, 2016) and atypical fields (e.g., Dyer, Lang, and Stice-Lawrence, 2017).

True breakthroughs, by definition, are rare. Schilling and Green (2011) show that introducing new ideas or variations away from the conventional domain tends to be uncertain, complicated, and futile on average; but successful efforts would receive outsized attention from sequential research in the form of citation in the longer run. On the other hand, consumers of research rely on the established research findings for decision-making (e.g., whether to recommend a paper for publication) and learning (e.g., whether to read or reference an article). As a result, readers and gatekeepers of research, while aspiring new knowledge, are more receptive to absorbing the expected findings from familiar settings. Therefore, we hypothesize that research articles deviating from the conventional domain will eventually acquire more citations but enjoy less readership, while articles introducing novel concepts or data would receive more attention in both citation and readership. Finally, conducting research in a well-inhabited and well-defined field is an easier path to gaining readership and citations compared to research in atypical topic areas.

While long-term recognition comes to successful studies that deviate from conventional paths, such studies initially need to overcome a higher hurdle of the publication process. It can be challenging for

[8]In addition to the studies already referenced, additional representative work includes Holmstrom (1989), Bena and Li (2014), Bernstein (2015), Kerr and Nanda (2015), Kogan, Papanikolaou, Seru, and Stoffman (2017), and Lerner and Seru (2017).

projects of this nature to satisfy the definitiveness and robustness requirements of a top journal defined mainly by existing research in the conventional domain (Ellison, 2002) and to receive a proper evaluation from referees (Hirshleifer, 2014). Identifying true gems from a set of unfamiliar original work is a riskier endeavor and more complex, given that editorial and refereeing capacity has been outpaced by research production and submission flows. To maximize overall welfare (i.e., filtration and recognition of high-quality research) given the limited resources, the peer-review process of journals rationally sets a higher bar for "unfamiliar" work due to the higher cost and risk in assessing true quality. Such a practice inevitably delays the certification and dissemination of a small number of atypical but potentially impactful research and thus deters their production.

From an author's perspective, career concerns respond to the demand side, influencing the author's choice of research agendas and topics. In various professions such as stock analysts and portfolio managers, previous literature has documented a tendency for individuals to herd to consensus or common practice, especially among people in their early career or under performance pressure (Hong, Kubik, and Solomon, 2000; Hong and Kubik, 2003; Chevalier and Ellison, 1999). The innovation literature, e.g., Ferreira, Manso, and Silva (2014), also suggests that inventors under career pressure are more likely to conduct less exploratory projects, especially when they anticipate a lower tolerance for failure. Applying to the setting of academic research, we expect a divergence between top journals' craving for originality and their inconsistency in evaluating research in unfamiliar settings or on atypical topic areas. Such a misalignment funnels researchers, especially junior scholars, into herding, creating an excess supply of research of derivative nature, i.e., of marginal contribution to the scientific discovery process.

3 Data and Overview

Our analyses build on the universe of finance papers posted on SSRN, which was formed in 1994 as an archive (especially for preprints) devoted to the wide and timely dissemination of scholarly research in social sciences.[9] By 2020, SSRN hosted nearly one million research articles by around 500,000 researchers from more than 50 disciplines. For the purpose of this study, we focus on articles submitted to one or more "subject matter e-journals" under the Financial Economics Network (FEN), which contains roughly 21% of all SSRN articles and counts for more than 34% of total downloads.[10]

It is not uncommon for financial economists to post their works on SSRN as early as a few years before journal submission so as to promote and time-log their studies in the public domain. SSRN does not require peer reviews, but it ensures articles are classified, usually by authors themselves, to relevant e-journals. In addition, author disambiguation and article version tracking are two ongoing processes at SSRN.[11] While readers can view article abstracts anonymously, they must log into their account to download an article, allowing SSRN to count downloads by unique readers. This also imposes a higher hurdle for web crawlers

[9] In a public letter from SSRN cofounder Michael C. Jensen (Dec, 2004), the vision of SSRN is described as to "enable scholars to share and distribute their research worldwide, long before their articles work their way through the multi-year journal refereeing and publication process, at the lowest cost possible for authors and readers."

[10] For complete description of the SSRN FEN, please refer to: https://www.ssrn.com/index.cfm/en/fen/

[11] For clearer author information, SSRN merges accounts when an author is associated with more than one account. Similarly, SSRN endeavors to combine different versions of the same article submitted by different authors or through event venues such as conference submission.

scraping the site for articles.

Start with the universe of 100,056 working papers posted on SSRN FEN from 2001 to 2019, we would like to narrow down to a sample of serious research papers aiming for peer-reviewed publication in leading journals in finance and related disciplines.[12] Such a process entails five steps. First, we remove all articles that are shorter than 25 pages or longer than 100 pages. Second, we remove articles that report more than five coauthors.[13] Third, eligible articles need to have at least one author that is affiliated with an academic institution, defined as one that subscribes to Wharton Research Data Services (WRDS) or is accredited by AACSB. Fourth, we remove presentation slides, book chapters, and law articles deposited in HeinOnline (an internet database service for legal materials).[14] Finally, we exclude articles written in a language other than English and those that have been downloaded fewer than ten times. The combined filters, after removing duplicate versions of the same article, reduce our final sample to 52,497 research articles.

From this finalized sample of SSRN articles, we collect article and author information, including article title, abstract, author identification and affiliation, posted date, written date (if available), article length, number of downloads, number of news and social media mentions, as well as its latest PDF posted (if available) as of May 2020. The news and social media mentions directed to the SSRN, obtained from PlumX, capture attention from outside academia, such as investors, think tanks, and news media. Importantly, these mentions refer to the preprint version of an article and link to SSRN, allowing us to disentangle the external impact of a working paper independent from impact from journal acceptance and in-print.

As one of the goals of this research is to analyze the determinants that lead to publication and impact, we need to trace the SSRN papers' publication status and track their citations. The set of publication outlets considered in our analyses consists of 28 journals (see Table 1): (1) "leading finance journals," the 14 finance journals that are rated B+ or higher by Currie and Pandher (2020) plus a new journal, *Review of Corporate Finance Studies*; (2) "top five economics journals" and "top five accounting journals" as defined by the "*50 Journals used in Financial Times Research Rank*";[15] and (3) a selection of interdisciplinary journals, namely *Management Science*, *Journal of Business Ethics*, and *Journal of International Business Studies*.[16] We then obtain information about the articles published in these journals from the Web of Science (WoS), and we resort to Google Scholar to locate the records of publication status for the articles outside the WoS data coverage.[17] About 22.5% of our sample, or 11,809 FEN articles, have eventually been published in one of the 28 journals. In reverse, these articles represent 75.0% of the papers published

[12] Before 2001, most FEN articles do not include downloadable or text-based PDF manuscripts required for database name extraction.

[13] We select the upper limit of five coauthors as it is the maximum number of coauthors in our published sample from Web of Science (WoS). In recent years, papers with over five authors have started to emerge, but their numbers remain small.

[14] Law research overall has a different layout and different research designs from those of finance articles and is subject to a different journal review process.

[15] *Financial Times* ranks MBA and EMBA programs annually based on the academic and placement performance of business schools. One of the crucial metrics is a research ranking calculated by the number of faculty publications in 50 academic and practitioner journals. The current 50 journals used in *Financial Times Research Rank* are selected according to the consensus among 200-odd business schools in 2016.

[16] Depending on research focus and topics, some finance papers end up in these three interdisciplinary journals. *Management Science* and *Journal of Business Ethics* maintain editorial board members that specialize in finance research. Similarly, *Journal of International Business Studies* regards some traditional finance topics, such as cross-border finance, M&A, and corporate governance, as essential components within its knowledge domain.

[17] WoS does not contain article records of all the journals since their inception date. For example, the *Journal of Financial Econometrics* is established in 1974, but the earliest articles from the *Journal of Financial Econometrics* in WoS are dated back to 2007.

in the top three finance journals and 32.7% of all papers published in the full list of journals.

Furthermore, we obtain reference list and specific topics of referenced articles from Microsoft Academic Graph (MAG), a knowledge graph of scholarly works collected around publication, conference, and author posting events across all disciplines (Sinha, Shen, Song, Ma, Eide, Hsu, and Wang, 2015; Wang, Shen, Huang, Wu, Eide, Dong, Qian, Kanakia, Chen, and Rogahn, 2019). Innovation and bibliometric research, e.g., Hirshleifer, Hsu, and Li (2018), primarily relies on citation analysis. However, unlike for published papers, it is challenging to accurately obtain a complete list of references from SSRN PDF files, especially for references not part of the FEN universe. MAG deploys AI-powered machine readers to process all documents discovered by Bing's crawler. For our FEN articles, MAG can retrieve reference records for nearly 80% PDF files. More importantly, MAG applies reinforcement learning algorithms to build a proprietary multiple-layer topic system across all the disciplines, which allows us to measure the bibliographic quality of a FEN paper.

Table 1 tabulates the entire list of journals considered for this study. The earliest and latest publication year of articles from each journal included in our FEN sample is reported in Column 1, and the total number of FEN articles published in a journal is reported in Column 2. The selection of articles from all leading finance journals is densely populated, while the other journals also host a large number of FEN articles. The time series of the numbers of FEN articles plotted in Figure 1 shows an annualized growth of 8.7% from 2001 to 2019. Also plotted are the numbers of FEN articles that are eventually published in the top three finance, top five economics, or top five accounting journals, as well as the numbers of articles in the top three finance journals each year. Figure 1 confirms the common impression that the vast majority of top three finance journal publications had been posted on FEN beforehand, and that the growth of research production outpaces journal space. Less expected is that close to half of the publications in top accounting journals also appear in FEN, suggesting a sizable overlap of topics and dissemination channels between finance and accounting.

In addition to publication outlets, citation is another important metric for the success and impact of research. One challenge in analyzing citation of SSRN articles is that many popular sources for citations, such as WoS and Scopus, focus on the references to the published version of a research paper, while the citation numbers from SSRN are primarily limited to the papers posted on SSRN platform. To reconcile the two, we use Google Scholar's citation record as it tracks and aggregates such metrics among different versions of an article, starting in the working paper stage all the way to the published version. Martín-Martín et al. (2018) find that Google Scholar consistently covers more citations across all disciplines than other citation sources, and more than half of unique Google Scholar-only citations are from theses, books, conference articles, and unpublished materials. Our citation data is collected from Google Scholar over a four-week window from May 2020 to June 2020, and we supplement automated scraping programs with a manual check to ensure accurate attribution to the right papers.

Table 1 further reports the distribution of normalized citations in Column 3 to 7 as well as normalized download counts in Column 8 to 12 among the articles by publishing journal. We group all papers that are not published in the 28 listed journals in one "Other" category. A paper's citations (downloads) are normalized by the average number of citations (downloads) of papers posted in FEN in the same year. Such a vintage-adjusted metric is often referred to in the innovation literature as relative citation strength. The mean (and median) of normalized citation and download counts validate the dominant impact of the top three finance journals in their field. Comparing to papers published in top finance and accounting journals, papers in the top five economics journals have, on average, higher citation counts but receive less readership, reflecting the varying acceptance preferences of journals in these different but related disciplines. While the skewness of research impact is well documented, it is notable that many well-disseminated papers remain unpublished. For example, 10% (4,068) of those papers enjoy 2 to 3 times more vintage-adjusted Google Scholar citations than the 10 percentile articles in top three finance journals.

Figure 2 plots the average number of years from when a paper first appeared in SSRN till the time it is published. The mean (standard deviation) of publication duration across all 28 journals is 3.0 (1.8) years, while the top three finance journals have a mean (standard deviation) of 2.9 (1.5) years. It is taking increasingly longer over the last twenty years for newly posted working papers to be published in one of the selected 28 journals (Panel A) or in top three finance journals (Panel B). The duration doubled from 1.7 years in 2001 to over 3.5 years in 2020, likely due to the combination of two forces: First, research articles tend to stay as circulated working papers longer, and as a result, preprint dissemination has become an increasingly important venue where research ideas receive early feedback and gain recognition. The process then feeds back to encourage researchers to share their work (by, e.g., posting on SSRN) at a relatively early stage. Second, it is also consistent with the theoretical models (Ellison, 2002; Hirshleifer, 2014) and evidence (as shown in journals' annual editors' reports) that the journey to publication has lengthened through the years, even for those that eventually get to the destination.

Most published papers were first submitted to and presented at seminars and conferences before their eventual appearance in journals. During the process, certain conferences have established a reputation for being a precursor to publication, in terms of both quality filtering and pre-journal-review feedback for authors. Given their roles, we extend the analyses to the two most competitive general-interest finance conferences: The American Finance Association Annual Meeting (AFA) and the Western Finance Association Annual Meeting (WFA). Their statistics are reported in Table 2. About 40.7% (36.6%) of the conference papers from the WFA (AFA) eventually make their way to the top three finance journals, and another 3.5% (2.6%) landed in the top five economics journals.

4 Methodology and Variable Construction

This section constructs and describes variables that capture paper and author characteristics. With recent developments in machine learning, especially regarding natural language processing (NLP) techniques, we are able to calibrate and quantify characteristics, such as a paper's novelty or its fit to a research field, which have been generally considered "soft" information or subjective. The methodologies are not the creation or the focus of this study, but we nevertheless discuss the general structure and properties of the three models we tailored to our setting that help construct the key inputs to our analyses.

4.1 Quantifying Conventionality Based on Semantic Similarity

Non-conventionality is the most coveted quality in scholarly research. We assess this quality by quantifying an article's similarity with the existing literature using two leading methods. The first model is commonly referred to as the *bag-of-words* (BOW) similarity, which has been widely used in finance and accounting literature (e.g., Hoberg and Phillips, 2016). Applying BOW to the FEN articles in our sample, we first take the text of each article's title and abstract (article corpus hereafter) and clean it following conventional textual analysis procedures.[18] We then construct a vector summarizing a vocabulary of the most representative 1,000 finance research-related unigrams and bigrams.[19] The corpus of each article is thus condensed into a 1,000 by 1 vector, in which each entry records the number of times a given unigram/bigram appears in the article.

For illustration, consider a pair of two articles: paper i with title "The Effect of Litigation Risk on Management Earnings Forecasts" and paper j titled "Management Forecasts and Litigation Risk." The titles and abstracts of the two papers are shown in Figure 3. If we measure the bag-of-words similarity based on article title only, there are six-word roots in the union of both documents: effect, litigation, risk, management, earning, and forecast. In our example, we have:

$$BOW_{paper_i} = [1,1,1,1,1,1], BOW_{paper_j} = [0,1,1,1,0,1].$$

The cosine similarity between the two articles is then defined as

$$sim_{BOW} = \frac{BOW_{paper_i} \cdot BOW_{paper_j}}{\|BOW_{paper_i}\| \times \|BOW_{paper_j}\|},$$

where the dot, $\cdot$, is the scalar or inner product and the norm, $\|\square\|$, is the Euclidean norm. The resulting similarity is $sim_{BOW} = 0.82$.[20] Two main drawbacks of the bag-of-words methodology are that it does not extract information from the ordering of words, which is often material to the underlying semantic meaning, and that the high dimensionality of the vector space imposes a computational burden that grows linearly with the size of the entire corpus's vocabulary.

Those drawbacks motivated our adoption of the second similarity model, which is based on sentence embeddings, a language modeling technique from NLP that relies on word and sentence co-occurrence to create a representation in a relatively low-dimensional Euclidean space (Mikolov et al., 2013). More specifically, we use a Universal Sentence Encoding model (USE) (Cer et al., 2018).[21] The paragraph representations from the USE model are low-dimensional dense vectors which can accommodate large

[18] Such procedures involve (i) removing non-textual content such as numbers; (ii) removing "stop words," which are commonly occurring words that provide little or no unique information, as defined by the NLTK package and supplemented by a few stop words specific to research papers, such as *find*, *show*, *et al*, *paper* and *research*; and (iii) lemmatizing (grouping together different forms of a word into a single item) the remaining words to their root.

[19] We arbitrarily select a vector size of 1,000 for this research. We have tested the robustness of vector size using a subsample of over 3,000 papers with different vector sizes including 500, 750, 1,000, 1,250, and 1,500. The correlation among the BOW similarity outputs using different vector sizes is over 98%. Examples of vocabularies in the 1,000 vector-size specification include *abnormal return*, *compensation*, *cross-sectional*, *dividend*, *earnings announcement*, and *institutional investor*.

[20] In this case, $sim_{BOW} = \frac{1\times0+1\times1+1\times1+1\times1+1\times0+1\times1}{\sqrt{1^2+1^2+1^2+1^2+1^2+1^2}\times\sqrt{1^2+1^2+1^2+1^2}} = 0.82$

[21] We use a pretrained USE model publicly available from Google's TensorFlow Hub. We choose the encoding model trained with a deep averaging network since it is less demanding on computing resources.

vocabularies and corpora. Mathematically, a paragraph vector indicates a position in a 512-dimension vector space[22] in which the distance between any two paragraph vectors can be measured.

Using the same example given above, the titles of two articles are encoded into:

$$\overrightarrow{USE}_{paper_i} = [0.067, 0.061, 0.005, -0.070, \dots], \overrightarrow{USE}_{paper_j} = [-0.001, 0.005, -0.015, -0.043, \dots]$$

where $\overrightarrow{USE}_{paper_i}$ and $\overrightarrow{USE}_{paper_j}$ are 512-dimension vectors. The cosine similarity between the two article titles is then

$$sim_{USE} = \frac{\overrightarrow{USE}_{paper_i} \cdot \overrightarrow{USE}_{paper_j}}{\|\overrightarrow{USE}_{paper_i}\| \times \|\overrightarrow{USE}_{paper_j}\|},$$

which results in $sim_{USE} = 0.68$, notably lower than sim_{BOW} (0.82). If we bring the texts of abstracts into the similarity calculation, sim_{USE} increases to 0.91 while sim_{BOW} decreases to 0.48.

The potency of USE over BOW similarity could be illustrated by a few examples. First, Figure 3 reports another pair of articles for which sim_{BOW} is much higher than sim_{USE}. A reading by an educated reader of the pair of abstracts is likely to confirm the judgment of sim_{USE} more than that of sim_{BOW}. Second, consider two phrases: "*Activists tend to affect a firm's executives.*" vs "*Active investors can influence a company's management.*" sim_{USE} measures the similarity between the two phrases at 0.53, in contrast to 0 by sim_{BOW} measure, as they do not share any common words. Third, we link pairs of abstracts of the same articles in their SSRN and published versions in the top three finance journals. The two versions should, with rare exceptions, correspond to the same research questions and major inferences. Out of 1,670 article pairs, we find that the average sim_{USE} is higher than that of sim_{BOW} (0.88 vs. 0.78). For the subset of 160 papers with major revisions in their abstracts, the average sim_{USE} becomes significantly higher than that of sim_{BOW} (0.82 vs. 0.63).[23]

There have been quite a few anecdotal stories where an unconventional but ground-breaking piece of work eventually attained widespread acceptance but only after a substantial delay (Schaefer, 1998). Our sample enables us to quantify the dynamics of the integration of nonconventional work into finance literature. Focusing on FEN papers published in peer-reviewed journals from 2005 to 2014, we compute the Sim_{USE} metric of a given article separately with the FEN papers posted during the five years before and those after the year of publication. We find that the papers that were in the lowest quintile of Sim_{USE} before publication experience significantly greater increases in their semantic similarity with the literature following their publication, relative to those that were most conventional to start with (in the highest quintile of Sim_USE).[24] In addition, the time series of overall semantic similarity offers an encouraging

[22]The USE method first fits individual words (unigrams) or pairs of consecutive words (bigrams) into a 512-dimension vector space based on the words surrounding each unigram or bigram. As a result of the computationally intensive loss-minimizing processes, each word vector within the 512-dimensional space represents a computer-understandable semantic meaning defined by a sizable English corpus, such as Wikipedia. For example, $\overrightarrow{Queen}$ can be approximated by $\overrightarrow{King} - \overrightarrow{Man} + \overrightarrow{Woman}$. The word vectors of all the words in a sentence are then combined into a 512-dimension sentence vector to fit the surrounding vectorized sentences through a similar loss-minimizing process.

[23]For this comparison, we start with 3,734 articles and restrict our sample to the articles with different wording in the abstracts. In particular, we only keep the article in our comparing sample when its generalized edit distance (a generalization of the Levenshtein edit distance) between the SSRN abstract and the journal abstract is greater than 5,000 (or 60,000 for "major revisions.") For more details about the procedure, summary statistics, and illustrative cases, please see Table IA.1 in the Internet Appendix.

[24]Table IA.6 in the Internet Appendix describes the procedure and reports the results.

sign about our profession's growing acceptance of unfamiliar work. When we trace the mean similarity of all FEN articles posted (or all papers published in the Top Three finance journals) in a given year, relatively to articles posted in the previous five years, the time series exhibits a smooth downward trend.[25] Such a trend reveals a collective endeavor to integrate more novel and atypical knowledge elements into financial research, which might be attributed to increasingly more diverse editorial boards at journals and referee pools.

4.2 Field Classification

Much of our analysis requires classifying articles into topical fields, not only to account for field-specific effects, but also to build variables that capture characteristics associated with atypical topics and interfield research. We classify FEN articles into different research fields through a convolutional neural network (CNN) model that takes into account both word order and semantic meaning.[26]

The CNN model adopted in this study is a supervised learning model that involves extra knowledge of the environment through examples of desirable behavior, and is trained to learn the appropriate field classification in finance.

A natural candidate for model training is an academic conference, where papers are classified into different sessions based on research fields. As before, we focus on the two leading general-interest conferences for finance researchers. For AFA, we collect the title and abstract of 758 articles from the conference website from 2016 to 2019. Similarly, we collect paper information for WFA annual meetings and obtain the corresponding abstract records from SSRN for 1,038 selected articles from 2008 to 2019. We first label all articles primarily based on their conference session titles (e.g., "Asset Pricing Anomalies"). For sessions with a title that has no clear indication of a research field (e.g., "The Role of Media in Finance"), we read through the title and abstract of the articles in these sessions and label those articles independently by at least two researchers to ensure our classifier reflects the consensus of multiple reviewers.

Finally, we aggregate all session topics into six categories: asset pricing (AP), corporate finance (CF), financial intermediation (FI), investments (IV), market microstructure (MM), and the residual diverse field (DV). The output of our classification model is a vector of probabilities for each of the six fields. Though there is no sum-up constraint in the algorithm, the summation of all those probabilities is close to one. We define a probability of 60% or above being matched to one field to be the criterion for a paper to have a dominant field. About 85.7% of all articles fall into this category of clear resolution.

Through our analysis, we use the field winning the highest probability as the primary field of an article. We verify that the trained algorithm entails low errors rates through a set of human-labeled out-of-sample articles accepted by WFA 2020.[27]

Panel A of Figure 4 plots the time series of the number of FEN articles based on their primary fields.

[25]See Panels A and of Figure IA.1 in Internet Online appendix.

[26]Convolutional neural networks make up a class of deep neural networks commonly applied in image and video recognition, image classification, and natural language processing. The successes and innovations in the application of CNN-based classification tasks are often considered a cornerstone for the rise in popularity of deep learning neural network techniques over the past decade.

[27]Table IA.2 in the Internet Appendix shows the top and bottom articles in terms of assigned field probability for each field in the out-of-sample test. Based on a group of Ph.D. students and our own review, there were two articles whose dominant field assignments were incorrect, and four articles whose second-highest probability fields should have been the primary (but not dominant) field, out of the 144 WFA 2020 articles.

Rising from rough parity with AP, the number of papers in CF increases over time, peaking as the most popular field toward the end of the sample period. FI and DV outgrow IV and MM from 2008 onwards. Panel B of Figure 4 repeats the time series among the articles for which dominant field is classified with a probability of 80% and higher. The pattern remains largely the same except for the noticeable leveling off of DV, suggesting that articles combining multiple topics have driven the diverse field's growth. It is worth noting that the field classification is topic-based rather than methodology- or approach-based. Some clusters of papers, such as structural estimation or behavioral finance, are forming their own distinct literature. In our setting, we match these papers to their respective fields based on the research topic.

In Panel A of Figure 5, we plot the annual share of Google Scholar citations for a given field compared to all FEN articles in that year. As the number of papers in each field can vary dramatically, we normalize the citations by the fraction of papers in that field relative to the entire FEN universe. For example, the CF articles in 2006 account for 40% of all citations to FEN articles, but the CF field accounts for only 33% of FEN articles in that same year; hence, the CF field has a relative citation share of 1.2, or 20% above the level of proportional influence for 2006. Panel B of Figure 5 repeats the same exercise using download data. Both panels of Figure 5 reveal the impact of timely research on emerging capital market phenomena. For example, both citations and downloads of the FI articles jumped right after the Subprime Crisis, as did those of the MM articles around the 2010 Flash Crash. The two figures also expose differences between citation and download as proxies for research impact. Notably, downloads lean towards research fields that are more relevant to practitioners, such as portfolio managers and regulators.

We cross-check the CNN classifier's output with the textual similarity introduced in the previous section by computing average sim_{BOW} and average sim_{USE} across the 6 different fields. Each row in Panel A of Table 3 represents the average of sim_{BOW} for articles in the primary and other fields.

Panel B of Table 3 reports the same figures for sim_{USE}. The higher diagonal values indicate that the articles from the same field on average have the highest similarity in terms of both sim_{BOW} and sim_{USE}, except for those articles that are categorized as DV. The off-diagonal values are also significantly positive, and the relative ranking of cross-field similarity measures seems to be intuitive, as that between AP and MM stands out, followed by the AP and IV pairing. One noteworthy difference between the two measures is that sim_{BOW} disregards semantically related words, resulting in a magnitude that is substantially less than that of sim_{USE}, particularly for off-diagonal or cross-field terms.

Finally, Figure 6 displays the trend of growing paper length in terms of pages (Panel A) and number of coauthors (Panel B). From 2001 to 2019, the articles' average length in our sample increases from 42 pages to over 50 pages, and the number of coauthors also increases from an average of 2.1 to 2.6 during the same period. There is no clear cross-field difference, suggesting converging norms across different FEN fields to demand more work and higher robustness standards out of a single paper. The result confirms Spiegel (2012)'s discussion based on articles published in 2010 versus in 1980 in two top finance journals. Ellison (2002) attributed such a trend to the journal reviewing process in which researchers spend more and more time on "padding and polishing" papers relative to developing insights due to mutual reinforcement between the two roles (authors and reviewers) researchers serve.

4.3 Database Entity Recognition

The exponential growth of data available for academic research is one major engine of empirical research in the past two decades. To characterize papers in terms of data intensity as well as novelty, we collect information about the databases used in each SSRN article by processing the full paper text. In preparation, we first convert article PDF files into text files.[28] We then go through three steps to identify the databases mentioned. First, we build up a database of database entities collected by a group of research assistants, based on databases mentioned in top three finance journals from 2014 to 2016, supplemented by the names of databases available through WRDS if not already included. Finally, we detect the databases on the list in the text through lemmatized unigrams and bigrams.

One drawback of this keyword approach is that it could miss new databases that have not been encountered, in other words, it fails to "know the unknown." To exhaust all unlisted databases as well as unlisted name variations of known databases, we trained a CNN name entity recognition (NER) model to detect such additional databases. To make the task even more challenging, authors can sometimes refer to a database-related entity name in a context outside of data discussion. For example, NBER or the World Bank may be referred to as an institution or a data source, depending on the context. For this reason, we trained a CNN classifier to detect the probability of a sentence referring to data discussion.[29] Our final sample only keeps the database entities identified by both NER or the database lists in a sentence with greater than 0.9 probability mentioning a database entity.

After the multilayered effort to exhaustively identify databases used by research papers, we then study the trend of databases in empirical finance studies. Panel A of Figure 7 illustrates the trend over time of the proportion of FEN papers that are empirically oriented, defined by the inclusion of at least one database and the presence of at least one of the three basic elements: mentions of "summary statistics," regressions, and numerical tables. The proportion of empirical articles has increased across the board, from just below 70% in 2001 to over 80% in 2019. A similar trend has been under way in economics: Hamermesh (2018) finds that research in economics has become more empirical, and, among papers published in top five economics journals, those empirical studies generated more citations than did articles categorized as theory or econometric theory. Furthermore, Berninger, Kiesel, and Schnitzler (2021) reveal that database selection also affects citations among published articles in 16 financial economics journals.

Panel B of Figure 7 shows that the average number of databases per empirical article has also increased. At the beginning of our sample, except for Market Microstructure (MM), the number of databases used in a given paper is clustered around three. Toward the end of our sample period, the number rose to above five databases per empirical IV or CF article and that number also surpassed four among AP and FI articles. The average statistics have been built into expectations, leading to growing efforts to consolidate or connect databases from different sources or providers and to secure novel or proprietary data. Examples of such efforts including consolidation of five or more major commercial hedge fund databases, and merging census data with standard sources such as CRSP and Compustat.

[28]When the PDF files are in images, we use optical character recognition (OCR) to convert PDF files into text files.

[29]Logically, our NLP models for database detection determine an entity and the probability of the entity to be mentioned in a database-mentioning sentence based on the frequency of semantically similar syntax in which databases are mentioned. Please see Internet Appendix Figure IA.4 for more details.

4.4 Elements of Novelty: Emerging Topics, New Databases, and Diverse Field

Given the central importance of measured novelty to our study, we construct three additional variables capturing the novelty components of a research paper: the number of novel databases in a research project, the number of emerging topics, and the originality of topics. The first two variables are constructed using textual analysis tools in a similar spirit to Hoberg and Phillips (2016). Admittedly, these measures do not directly reflect the novelty in insights, but they serve as quantifiable metrics for studies' contribution to expanding topical domains and data knowledge. Moreover, new data are often instrumental for creative research designs and exploration of new ideas. A database is deemed novel if its current use in the focal article is within three years of its first appearance in any FEN article. We count the number of *new* databases used in a paper as the first measure of novelty.

In defining an emerging topic in a given year, we start with the lemmatized unigrams and bigrams from titles, keywords (if available), and abstracts of all the articles in the year, and narrow these down to those containing at least one noun detected by *part of speech* tagging.[30] A topical keyword is considered new if it appears fewer than three times in total in the full past history of article textual records among articles in the top three finance journals plus JFQA going back to 1980 and all FEN papers going back to SSRN's inception year (1994). There are 3,650 unigrams and bigrams to start with. After manually excluding ones that are not meaningful as research keywords (e.g., "Value Add," "Family Member"), long-standing topics (e.g., "Corporate Lawsuit," "ROA"), as well as database, person, and country names, we maintain a list of 439 topical keywords which were novel at some point of time during our sample period. We report three topics each year with the highest number of appearances during that year in Figure 8.[31] For our second measure of novelty, an article is deemed to be on an emerging topic if, on its first SSRN posting date, it covers a topic on the topical keyword list within three years of its first appearance.

An innovation that draws upon knowledge from a wide range of knowledge areas indicates originality that deviates more from current knowledge trajectories (Balsmeier, Fleming, and Manso, 2017; Hirshleifer, Hsu, and Li, 2018). In the innovation literature, a patent is considered to be more original if it cites previous patents spanning a wider range of technology classes. Analogously, drawing upon diverse topics in a scholarly work also reflects the authors' ability to integrate research ideas in a novel way. Following Trajtenberg, Henderson, and Jaffe (1997), we define the *CiteBreadth* of an article as one minus the Herfindahl-Hirschman Index for a fraction of citations made by the focal article to articles from different MAG research topics. MAG employs reinforcement learning methods to construct a multilayer topic system applicable to all disciplines. The majority of these articles are categorized as Business or Economics at the root level and as Financial Economics, Monetary Economics, or Accounting at the first sublevel. For our *CiteBreadth* measure, we focused on second sublevel (level 2).[32]

[30] *Part of speech* is a supervised learning solution that uses features like the previous word, next word, capitalized first letter, etc. to break a sentence into parts of speech (unigrams) such as nouns, pronouns, etc. Such a method does not require a predefined list of nouns (as most existent research has done in a similar setting) so as to give novel phrases the best chance to emerge.

[31] Internet Appendix Figure IA.5 provides the full list.

[32] Our final sample includes 19 root level topics, 288 level 1 topics, and 8,396 level 2 topics among all cited articles available through MAG. Table IA.6 in the Internet Appendix provides a word cloud for the most frequent level-2 MAG topics for the referenced papers of FEN articles, including *Market Liquidity*, *Rational Expectation*, *Unit Root*, and *Civil Law*.

4.5 Other Variables: School Ranking and Coauthor Centrality

Arguably, research resources and school ranking influence affiliates' capacity to conduct and disseminate research, as well as immersion to emerging topics, a combination of selection and treatment effects (Swanson, 2004). For this reason, we construct a few additional control variables. The first is the number of databases subscribed to collectively by the superset of coauthors' institutions through WRDS (based on WRDS records). This variable proxies for the total research supports the author team has access to; it captures the affiliation effect rather than the database resources actually used in the paper. Second, we assign a dummy variable for any author of a given article to be affiliated with a top 20 research school, according to the University of Texas at Dallas (UTD) Top 100 Business School Research Rankings (based on publications in the top three finance journals). Third, we construct a measure for author network reputation at the article level. Using the two-step reach centrality, as described in Bajo et al. (2016), we first calculate the author-specific eigenvector centrality measure based on the author-level coauthorship network within the author universe of FEN. The measure takes into account indirect and direct connections an author has in their network in that an author boasts a high centrality if they are working with coauthors who are themselves central. Averaging the centrality measure across all coauthors, we obtain an author centrality measure for each FEN article.

4.6 Summary Statistics

We report in Panel A of Table 4 the summary statistics of the final sample of 52,497 articles posted on SSRN FEN from 2001 through 2019. The average (median) citations in our sample is 60.86 (11), and the average (median) downloads is 404.98 (189). Various statistics remind us of the scarcity of novelty even among academic research: the 90th percentile values of both new databases used and emerging topics discussed are zero. The variable sim_{BOW} (sim_{USE}) has a mean of 0.078 (0.365) with a standard deviation of 0.024 (0.051). Panel B of Table 4 reports the pairwise correlations among our independent variables. While none of the individual correlations are particularly striking, we note that the similarity measures have a low correlation with any of the other independent variables. It is perhaps a little surprising that research resources, school ranking, and author centrality measures are not highly correlated with one another at the article level.

5 Dissemination and Recognition of Finance Research: Empirical Analyses

This section empirically examines how measured research quality, especially in terms of conventionality and novelty, is associated with success in dissemination (measured by downloads), publication (in leading journals), and impact (measured by citation). Given that originality is the focal quality valued by the academic research community, one would expect that, all else being equal, a piece of research that is more differentiated from the current stock of work or less conventional (e.g., one that is semantically dissimilar to previous work or one that introduces a new topic or new database) should be more prized by journals and also generate higher impact due to follow-up research. On the other hand, research with novel topics or unconventional designs does not enjoy a captive audience and might also incur a higher hurdle going through a rigorous peer review process in which editors and reviewers may still apply the well-developed standards applicable to mature topics; they may even require more evidence in order to convinced of something that lacks precedence (Lee et al., 2013).

5.1 Conventionality and Research Outcomes

We start with the following regression to examine the empirical relation between conventionality and research outcome:

$$Outcome_i = \alpha\ Conventionality_i + \sum_{k=1}^{K} \beta_k X_{i,k} + \delta_t + \theta_j + \epsilon_i, \tag{1}$$

where i indexes a given article. $Outcome_i$ is one of our research outcomes, including citations ($Citation$), downloads ($Downloads$), and a dummy variable for the appearance in a top three finance journal ($Top3Fin$). $Conventionality_i$ is semantic similarity, sim_{USE} or sim_{BOW}, defined as the average similarity between the corpus (title and abstract) of the focal article against that of each of the FEN papers posted in previous years.[33] $X_{i,k}$ is a vector of additional article and author characteristics. δ_t and θ_j are year and field fixed effects respectively, and ϵ_i is the error residual. The covariates vector $X_{i,k}$ includes the number of pages in logarithm ($LogPage$), number of coauthors ($\#Authors$), and number of databases used in the article in logarithm ($LogData$). The vector also includes author affiliation information capturing research resources underlying the production of the article: number of WRDS databases (in logarithm) the author group as a whole has access to via their institutions ($LogWRDS$), whether any of the authors is affiliated with a top 20 school as defined by the UTD annual school research ranking ($Top20School$), and average eigenvector centrality measure of authors ($AuthorCentrality$).[34] We use logit regression for publication analysis ($Top3Fin$).

Table 5 reports the results based on Equation (1). Each outcome variable takes up four columns, which alternate between the two similarity measures and the on/off state of field fixed effects. Yearly fixed effects are included in all columns. The α coefficient associated with conventionality is uniformly significantly negatively in Columns 1–4, which report citation outcome. Based on the specifications with field fixed effects, a one standard deviation increase in *Conventionality* is associated with 5.53% to 7.69% decrease in citation. Such a relation may be somewhat surprising as it is easier for researchers to see the connection between their own and others' work, and hence cite the latter, when they are more similar. Our result uncovers a desirable element of the citation metric—the most natural measure of research impact—in that it rewards novelty.

Given the results regarding citations, Column 5–8 of Table 5 show an intriguing variation when we measure research impact using download count, which shows an overall positive correlation with conventionality measures. The coefficients are significant in three out of four specifications, with stronger results using the

[33] Restricting similarity comparison to papers posted during the past five years yields qualitatively similar results.
[34] Sensitivity check using the maximum *AuthorCentrality* of an author group yields very similar results.

USE similarity. The correlation between downloads and conventionality is notably weakened when field fixed effects are introduced. To some extent, the results also confirm that the USE similarity measure is more refined and thus sharper at capturing reader experience, in that papers that are connected with the current literature attract more reader attention. The results also suggest that a broader readership, as measured by the number of SSRN downloads, does not always translate into more citations.[35]

An academic career is often summarized as "publish or perish"; as such, qualities that are predictive of a publication in a coveted outlet are perhaps of greatest interest. The last four columns of Table 5 report that both similarity measures are strongly and positively correlated with the chance of an article getting published in a top three finance journal. Using the specification with USE similarity and field fixed effects, a one standard deviation increase in conventionality is associated with an incremental 16.4% to 19.4% increase in the probability of hitting a top three, or 0.9 to 1.2 percentage points over the unconditional probability of 7.2%.[36] The result might seem puzzling, as it suggests that a paper's measured differentiation from previous work is not as prized by our profession's top outlets despite the academic community's stated goal of pursuing original research. It could be that conventionality as measured by semantic similarity with existing research stock does not do justice to truly creative and novel work; but the empirical evidence seems to strongly support that authors tend to have a harder time with journals if they do not write their papers in semantics that are familiar to the literature, and hence to the review team. It is worth emphasizing that this positive coefficient does not necessarily indicate that research published in leading finance journals do not constantly move into new territories, but instead suggests that papers written less conventionally at a point in time face a higher hurdle to publish.[37]

A positive correlation between research outcomes and *Conventionality* could be justified if the measure happens to be a proxy for desirable qualities that are hard to quantify, for example, the importance of the topic, its attraction to the wide audience, and execution quality. A priori, however, research projects that are intrinsically important or interesting should not appear conventional. Moreover, important and innovative issues would be quickly "arbitraged" away if they could be expressed and resolved with semantic framework that is similar to the exiting work. In fact, if we relate ex post extreme research success, defined as top 500 to 2,000 papers from the SSRN FEN by MAG saliency following Wang et al. (2019), to *Conventionality*, we find the relation to be significantly negative,[38] making *Conventionality* an unlikely proxy for the omitted qualities.

Table 5 reveals several interesting relations between research outcomes and the covariates, which are all statistically significant in impacting research outcomes. Longer papers (*LogPage*) and more coauthors

[35] Given that downloads and citations are highly right-skewed, we conduct a median regression analysis for the same empirical model to mitigate the influence of outliers. Results are reported in Table IA.3 of the Internet Appendix. The result remains qualitatively similar.

[36] In a sensitivity analysis in which we build the *Conventionality* variable using papers posted during the past five years, a one standard deviation increase in *Conventionality* defined by the USE similarity with articles in the past five years is associated with a 5.28% decrease in citation as well as a 2.63% increase in downloads and a 1.02 percentage point increase in hitting a top three finance publication. All results have comparable magnitude with those in the main specification.

[37] In Panel B of Figure IA.1 in the Internet Appendix, we traces the evolution of average semantic similarity of accepted papers in a given year and all accepted articles in prior 5 years in the universe of articles published by leading finance journals through the years. It indicates that journals have became more welcoming to unconventional studies during the last two decades.

[38] The regression with control variables is presented in Table IA.4 in the Internet Appendix. MAG saliency is an eigenvalue centrality-based measure with citation weights proportional to the saliencies of the authors, their affiliations, the dissemination venue, and the recency of the last citations. Benchmarking against traditional research impact measures, such as citations and *H-Index*, Wang et al. (2019) shows the MAG saliency consitutes a better filtering of quality from quantitative metrics.

(*#Authors*) are overall associated with better outcomes, except that increasing number of authors actually decreases the odds of publishing in the top three finance journals (the magnitude of the effect is, however, quite negligible). Card and DellaVigna (2013) show that, conditional on publication in top economics journals, citation counts are significantly higher for longer papers and those written by more coauthors. We thus confirm a similar finding at the working paper stage. More databases utilized in a given paper (*LogData*) is also a straight plus: a one standard deviation increase in databases used in the paper is associated with a 26% increase in the probability of publication. All these results suggest that comprehensive data and exhaustive analyses (which lead to lengthier manuscripts) have become part of the standards for quality research.

As expected, author reputation (*Top20School*) and research resources (*LogWRDS*) of authors' institutions are significantly positive determinants. Having an author from a top 20 school is associated with a 75% increase from the unconditional probability of publication, and the same effect from a one standard deviation increase in author school resources (as proxied by WRDS data subscriptions) is 42%. Finally, the coefficients associated with *AuthorCentrality* are always positive and significant, implying a 31% increase in publication probability and 27% increase in citations for a one standard deviation change. Well-connected authors have disproportionate impact, consistent with Chung, Cox, and Mitchell (2001) finding that a few prominent researchers dominate in the leading finance journals, despite the fact that the productivity gap has shrunk between researchers affiliated with top- and non-top-ranked schools (Kim, Morse, and Zingales, 2009).[39] We acknowledge that such a relation is likely a combination of selection and treatment effects. Affiliation with a high-reputation/resource institution and centrality within the network of authors are successful outcomes by themselves and correlated with the focal author's own scholarly quality; at the same time, resources and centrality also facilitate dissemination and recognition of the research (West et al., 2013).[40]

The positive coefficients of *Conventionality* for download and publication outcomes in Table 5 suggest the readership and journal review teams favor the works similar to previous ones because they closely build on the existing knowledge as measured by the combined semantic space spanned by previous knowledge. On the other hand, a paper's innovativeness is often judged by its incremental contribution relative to the most closely related current art. We define a "highly similar precursor" to be an earlier paper that has a sim_{USE} to be at least 0.8 or sim_{BOW} of at least 0.7.[41] If a paper has at least one highly similar precursor (about 5% of our sample), the dummy variable *HighSim* is coded as one. We then repeat the regression in Table 5 except adding to the conventionality measure the additional variable of *HighSim*. Table 6 reports the results.

Because all specifications in the Table 5 see fixed effects capturing a noticeable degree of heterogeneity

[39] In fact, we find that the coefficient on *Top20School* is significantly larger for the post-2010 subperiod for publication, suggesting a widening advantage in favor of papers with authors from top schools. The shrinking gap in research productivity is thus likely driven by the growing cross-school coauthoring made possible by communication, data-, and code-sharing technologies as well as profession-wide events that reduce the importance of immediate colleagues, a key mechanism in Kim, Morse, and Zingales (2009).

[40] To the extent that this study focuses on article-level analyses, disentangling the selection from treatment effect of author and school reputation goes beyond the scope of the current article. However, we refer the readers to Brogaard et al. (2020) for a carefully identified study on the causal effect of author reputation on citations.

[41] These cutoff values are selected based on the observation of similarity values of different versions of the same paper. One such example is reported in Internet Appendix Table IA.1. We also use 0.75 for semantic similarity and 0.65 for bag-of-words similarity, and results are qualitatively similar.

among FEN papers, in all tables onward we only report the results with time and field fixed effects in regression analyses. Table 6 shows that having a highly similar precursor does not impact citation, but significantly reduces downloads, possibly due to less new content as perceived by prospective readers. Moreover, the presence of a highly similar precursor as measured by USE semantic similarity (but not by BOW similarity) also significantly dampens publication prospect. The sharper results with sim_{USE} vs sim_{BOW} again demonstrate the strength of the USE algorithm in accurately calibrating the originality of the focal paper to its closest predecessor. In an unreported exercise, we also discover that well-cited highly similar precursors, as measured by an interaction variable of $HighSim$ and high citation (in the top 20 percentile of a year), further dump the citations of subsequent works.

Tables 5 and 6 reveal a nuanced attitude toward semantic innovativeness within the research community, especially as manifested in the publication process. On the one hand, both readers and journals welcome papers that have a large footprint on the existing knowledge (i.e., high semantic similarity with the universe of earlier papers). This could be due to the fact that readers and reviewers are better able to connect and resonate with such research (Spiegel, 2012), and could also be explained by the risk aversion of referees and editors who feel more confident in judging "familiar" research and are also reluctant to endorse research whose quality is difficult to assess, a recurring theme in the book by Shepherd (1995) in which leading economists pondered on the publication process. On the other hand, readers do value, and journals do require, new papers to be clearly differentiated from the outstanding work that are closest to them. Top journals are particularly unforgiving when the writing of a paper fails to distance from its "highly similar precursor." A "careerist" lesson is thus that the prospect of publication is maximized when authors work on research projects with well-populated prior knowledge but at the same time provide a clear distinction against the closest predecessors. In other words, incremental contribution in a mature research area seems to be the easiest path toward publication (but not necessarily knowledge creation).

5.2 Novel Elements and Research Outcomes

This section expands the measures for conventionality beyond textual similarity to more tangible dimensions. More specifically, we consider three elements: the presence of a new topic ($NewTopic$); the log of the number of new datasets used ($LogInnovData$), and the ease with which an article could be matched to an established field, as measured by the highest probability that a paper falls into a research field ($Prob_{Field}$). All these variables are defined and discussed in detail in Section 4. These three sets of measures constitute the $Novelty$ variable in the following regression:

$$Outcome_i = \alpha_1 Conventionality_i + \alpha_2 Novelty_i + \sum_{k=1}^{K} \beta_k X_{i,k} + \delta_t + \theta_j + \epsilon_i \tag{2}$$

where all other variables are defined identically to those in Equation (1). Once again, we use logit regression for publication analysis.

Results involving novel elements are reported in Table 7. Both new variables are uniformly associated with significantly better research outcomes in terms of citations, downloads, and publication in the top three finance journals. The effects are economically significant as well. A paper delving on a new topic enjoys 58% higher citation, 55% more downloads, and stands 1.7 percentage points higher chance of landing into a top three finance journal (relative to the unconditional probability of 7.2%). One additional novel database is associated with an increase of citation (downloads) number by roughly 31.0% (25.3%), and the probability to be accepted by the top three finance journals by approximately 2.2 percentage points.

We then turn to our next group of tests in which we look at the atypical knowledge elements added from the fields other than the well-established five. We first add the probability of a paper belonging to the diverse field, $Prob_{DV}$ into Equation (1) and report the regression results in Column 1, 4, and 7 in Table 8. The results reveal that $Prob_{DV}$ is positively correlated with citation counts, but negatively correlated with readership and chance of top tier publication, and all three coefficients are significant at the 1% level.[42] The results are consistent with those in Stephan et al. (2017)— the novelty from an atypical combination, on average, attracts more citations but faces a higher hurdle for publication in top journals, possibly due to the reduced chance of encountering a referee who is both sympathetic to the research question and is willing to exercise judgment with less-than-clear standards. Next, we add $Prob_{DV}$ into Equation (2) and report the results in Column 2, 5, and 8. The results largely remain qualitatively similar. Interestingly, $Conventionality_i$'s coefficients become insignificant in 2 out of 3 download analyses (Columns 4 and 6). This may suggest that the research from traditional finance focus fields drives the positive correlation between $Conventionality_i$ and downloads.

We take the opportunity to investigate the research impact of papers from different finance fields by adding individual field probability measures into Equation (2) and report results in Column 3, 6, and 9 in Table 8.[43] While coefficients of $Prob_{DV}$ once again remain qualitatively unchanged, there are a few fresh results worth discussing. Comparing to other fields' probabilities, the probability of a paper being in the asset pricing field is significantly and negatively associated with citation counts over our sample period. However, its coefficient in the publication outcome is still marginally significant and positive. The results also indicate that the papers related to the corporate finance field and the financial market fields, such as asset pricing, investment, and market microstructure, receive more readership, confirming our descriptive results reported in Figure 5. Papers belonging to the field of financial intermediation tend to receive more citations from sequential research but no more downloads. Finally, it seems that the close match to any field, other than DV, is not significantly associated with publication probability, though, numerically, AP-focused papers stand the highest chance.

5.3 Dynamics of Research Demand and Productions

It is fair to argue that our findings, particularly those concerning publication outlets, are influenced by hard-to-measure qualities such as the importance and sharpness of a piece of research. We make a best-effort attempt at controling such quality in two ways. First, we add among control variables the number of citations pre-publication – which avoids the reverse causality that citations are boosted by top outlets

[42] In an unreported table, we also find articles with a high proportion of atypical topics receive more citations but enjoy less readership and top finance journal publications. In particular, we consider a MAG level-2 topic "typical" if it hosts more than 1,000 FEN articles, or if its parent-level topics (level 1) include Financial Economics or Finance.

[43] Because the field classifications are not mutually exclusive and because all probabilities do not sum up to a constant (e.g., one), there is no strict multi-collinearity when the regression includes probabilities associated with all six fields.

post publication, even conditional on genuine quality. Specifically, we resort to MAG's reference data, which has slightly lower coverage than Google Scholar yet includes references to SSRN papers while they are working papers. Additionally, the pre-publication citation has become more informative over time as research papers have prolonged their life in working paper stage in the past two decades. Second, we investigate the relationship between journal outlets and conventionality and novelty measures among *ex-post* unambiguously impactful, high-quality studies based on MAG's saliency metric, an H-index equivalent assessing research impact based on the quantity and quality of citations in research network. The top 500, 1,000, and 2,000 papers in our sample, representing nearly 1%, 2%, and 4% of our sample, form our sample of truly high quality, impactful papers.

Column 1 of Table 9 demonstrates that the semantic conventionality coefficient remains positive and significant after controlling for pre-publication citations. Most of our findings also remain qualitatively similar. It is worth noting that the pre-publication citation absorbs the significance of the *NewTopic* covariate, suggesting that the *NewTopic* metric is a vital component to attract citations before publication. Columns 2 to 4 report the results for the subsample analysis among those most impactful research. There is a significant reduction in sample size, and some coefficients, e.g., those associated with *NewTopic* and *LogInnovData*, become insignificant. In contrast, the coefficients for our semantic conventionality and atypical field proxies remain qualitatively similar, suggesting that these two metrics influence the publication chance in the top three finance journals, even after controlling for potential research impact or among the subsample of unambiguously impactful research.

When certain type of research is subject to a higher hurdle for publication (because it requires more effort and risk-taking in quality assessment), then the accepted papers from this group should enjoy better outcomes (i.e., higher impact) *ex-post*. Figure 9 confirms such dynamics. In the chart, we sort articles published in the 28 journals listed in Table 1 into quintiles of their *Conventionality* (sim_{USE}) at the year of publication ("year 0"), and trace out their median cumulative citations from MAG up to each year from -3 years to $+10$ years.[44] The median citations across quintiles evolve in lock-steps up to two years post publication, from which point the lowest conventionality quintile starts to diverge from the pack. Such a trend affirms a publication bias against non-conventional research, which is held to higher hurdle for publication (in terms of quality and potential as judged by long-term citation). On the positive side, unusual research becomes more mainstream and popular over time, generating follow-up research.

The relation between research fields and outcomes could also be assessed from the angle of "field concentration" of a paper. A research paper could be a "pure play," focusing on one specific field and with little matching qualities to any other fields, or could be "interdisciplinary" (within finance) by combining the knowledge from multiple fields. Such an attribute could be measured by the variable $FieldFocus$, defined as the Herfindahl-Hirschman Index (HHI) over the values of field-specific probabilities. The average article has a $FieldFocus$ value of 0.806. Results, reported in Table 10, show that concentrated research (high $FieldFocus$) generates significantly more citations and downloads; presumably, they are more likely to match prospective readers' interests. Such papers are also significantly more welcome by top

[44] To have a balanced sample, for those articles with sparse or uneven citations across years or shorter than 10-year history since publication, we consider a paper referenced zero times in the year without MAG citation, instead of dropping it from the sample in that year.

finance journals. A one standard deviation increase in $FieldFocus$ is associated with a 30.4% increase in the probability of publication in top finance journals. As a comparison, top five economics and top three accounting journals (shown in the last two columns of the table) demonstrate no "diversification discount" for papers that strand multiple finance fields. On the contrary, the discount turns into a significant premium in accounting journals. In other words, finance papers published in top accounting journals are more likely to be combinations of different fields instead of a pure play in one well-defined field.

A related HHI measure could be applied to the references a paper cites. If a paper builds on prior work from a diverse set of fields and topics, its broad scope may uncover unusual linkages in knowledge networks that generate new ideas.[45] It is a priori unclear as how the atypical combinations of knowledge are received by readers and journals. We define $CiteBreadth$ as one minus the Herfindahl-Hirschman Index for citations made by the focal article to articles across MAG level-2 topics,[46] resulting in an average value of 0.88 across all FEN articles. Results, reported in Columns 2–3 of Table 10, demonstrate that the quality proxied by $CiteBreadth$ generates significantly more citations, corroborated by prior citation-based works, such as Wang, Veugelers, and Stephan (2017). Papers building on a broader base on knowledge are also significantly more welcome by top finance journals. A one standard deviation increase in $CiteBreadth$ (24.8 percentage points) is associated with an increase in the probability of being accepted by the top three finance journals by approximately 2.4 percentage points (relative to the unconditional probability of 8.6% for this regression sample). On the other hand, $CiteBreadth$ leads to a lower readership, suggesting that a broader scope of prerequisite knowledge may intimidate the readers.

Table 10 exposes an intriguing contrast in our profession's attitude toward papers whose research questions span multiple fields and papers that build on prior knowledge from multiple fields. A lack of field focus of the research per se incurs a discount in publication and citation, but the same lack of focus, if applied to the references (or prerequisite knowledge to the current research), is prized in similar ways as *high-breadth* patents. The combined results suggest that readers and journals in finance value authors' effort and ability in bridging knowledge from different subfields but still appreciate more research that tackles questions with a well-defined habitat.

5.4 Career Status and Research Outcomes

This section examines the effect of conventionality and novelty-related metrics on research outcomes, conditional on the seniority and reputation of the authors. We classify coauthor teams by whether they contain 1) an author from one of the top 20 schools, 2) an author with centrality in the top 5 percentile, or 3) all early stage researchers with fewer than ten years of SSRN (and MAG) history. We are particularly keen on the coefficient of the interaction terms because they provide estimate differences in the conventionality and novelty measures between articles drafted by a specific group of authors and those written by the rest. Table 11 summarizes the results.

The coefficients of interaction terms in Columns 1–3 in Table 11 reveal a few absorbing estimate

[45] The analogous measured applied to patents is usually terms as "originality" in the innovation literature, e.g., in Hall, Jaffe, and Trajtenberg (2001). In our context, such a measure has little correlation with conventionality (or lack of originality). Instead, it captures the breadth of knowledge the research builds on.

[46] We also calculate an $CiteBreadth$ measure with MAG level-1 topics, and the results are qualitatively similar.

differences between researchers from the top 20 schools and the rest. Nontraditional finance focus-field papers authored by academics at the top 20 schools receive even more citations, despite their lower audience. On the other hand, subsequent studies cite conventional research by the top 20 school researchers more frequently, though they would receive less readership. As suggested in Columns 4–6, articles written by high-centrality researchers tend to obtain more citations when they investigate new ideas or use new data, and readers as well as peer reviewers prefer atypical and unconventional works from this group of researchers. Finally, we noticed some intriguing trends in the final three columns of Table 11. Early stage researchers' articles have a greater likelihood of being published in the top three finance journals but receive fewer citations when they cover nontraditional finance issues. Additionally, their work with new data sources brings in additional readers via SSRN.[47]

5.5 Dissemination of Finance Research

5.5.1 Conferences and Research Outcomes

Dissemination of research precedes publication and builds up research impact beyond the publication channel. SSRN, seminars, and conferences are the top academic venues for dissemination. In this section, we highlight the dissemination channel of the AFA and WFA, because they are the two top general-interest academic finance conferences and have a reputation of promoting paper visibility as well as providing extensive feedback from the discussants and the broad audience. Table 12 analyzes the determinants of as well as outcomes from the conference inclusion. The first three columns report the predictive regressions based on the superset of the main variables using in the previous tables. The outcome variables are inclusion of the paper in the AFA and WFA.

Not surprisingly, longer articles with innovative data by authors from well-resourced high-reputation schools and by authors who are well-connected are significantly more likely to be accepted at both conferences. Papers with more coauthors stand a lower chance, and a new topic is a positive input with only marginal significance. An interesting but perhaps not surprising contrast between the two conferences emerges in that a paper outside the usual topic areas is more likely to be accepted at the WFA but less so at the AFA. Moreover, author reputation appears to be a stronger predictor for inclusion in the AFA than the WFA. While the WFA program selection is peer review-based, the decision makers at the AFA are session chairs who are more likely to pick papers and authors in areas that they themselves are more familiar with.

Next, we define a paper that was accepted at either of the two conferences as "treated" and examine whether such a treatment is associated with differential research outcomes on a one-to-one matched sample based on the propensity score using the predictive model in Column 3. In other words, the sample in Columns 4 to 8 includes all papers that were ever accepted to any AFA or WFA conference, plus their close matches (based on the propensity score) in the same year vintage that never made into the conferences. The latter columns of the table show that papers included in the programs of AFA and WFA saw a boost in

[47] Additionally, in unreported tests, we find that *CiteBreadth* increases the likelihood of an article being published and the number of citations for articles written by early stage researchers. Similarly, *FieldFocus* also increases the number of citations for these articles.

citations, downloads, and prospects in publication in the top finance and economics journals. Such papers are significantly less likely to publish in accounting journals. Being included in the two top conferences is associated with a 48.1% and 13.5% increase in the probability of publishing in the top finance and economics journals, and -68.8% decrease in the probability of publishing in the top accounting journals, relative to the regression-sample probabilities of 24.20%, 2.46%, and 2.63%.

5.5.2 Dissemination of Research via Media

Many finance researchers value impact of their research outside the ivory tower. Leading business schools also pride in serving thought leadership for the business world and society at large, and provide extensive support for faculty to disseminate their research outward.[48] We thus relate non-academic attention to research quality measures, especially novelty and conventionality. More specifically, we consider two measures of research dissemination via media as dependent variables for Equation 2. The first is $News\ Mentions$, defined as the log number of news mentions, and the second is $Social\ Media\ Mentions$, defined analogously. Results are reported in Table 13.

As expected, the news media has a penchant for "newsworthy" research. Articles delving into a timely topic ($NewTopic$) and those with low semantic similarity to the existing research (sim_{USE}) receive significantly more media mentioning, consistent with the conventional wisdom that content selection in the news media is primarily driven by timeliness and novelty (Kennedy, 1988) to retain a broad, dominantly non-academic, audience base. A new topic alone increases news mentioning by 2% and social media mentioning by 14%. Empirical intensity (as measured by $LogData$) is also a helping factor, but novel data ($LogInnovData$) does not have a notable effect, suggesting that novel databases perhaps mainly serve a purpose for researchers to conquer previously untestable hypotheses, which is important to researchers seeking breakthroughs but not in its own of interest to news media. Not surprisingly, the success of researchers with high reputation themselves (measured by $AuthorCentrality$) or affiliated with high-reputation institutions ($Top20School$) in disseminating their research outside academic matches their success inside academia.

5.6 Comparisons of Publication in Multiple Related Disciplines

This section extends our analyses over a broader selection of journals to uncover more insights into the influence of novelty and conventionality components on various publication outcomes in multiple related disciplines. Ellison (2002) predicts that the review process's insular nature should result in different novelty and conventionality standards across different disciplines. The breadth of journals across finance, accounting, economics, and others in which FEN articles are published provides a testing ground to calibrate this cross-disciplinary difference in social norms. On the other hand, another implication of the model is that all peer-reviewed disciplines would show increasing demands for robustness, rigor, and polish as a literature matures, which may imply a low variation in standards, especially with regard to requirements for robustness quality and incremental contribution.

[48] Examples of such endeavors include Wharton@Work of Wharton, *Ideas at Work* of Columbia Business School, and *Insights* of Stanford Business School.

We report the multinomial logit regression analysis results in Table 14. This table includes all the control variables described in Table 6 and 8, except we replace the number of authors with three dummy variables for coauthorship size analysis: papers with two, three, and four coauthors respectively. In Model 1 (Columns 1–3), the dependent variable includes four outcomes: publication in the top three finance journals ($Top3Fin$), publication in the top five economics journals ($Top5Econ$), publication in the top three accounting journals ($Top3Acc$), and others (the reference and omitted outcome). The multinomial logit model nests the set of parallel and exhaustive outcomes. We repeat a similar analysis in Model 2 (Columns 4–7), except we include an additional category of publication in the three interdisciplinary journals ($Top3Int$).

The results of $Top3Fin$ (Column 1) are similar to those reported in previous tables, except that the articles with two coauthors have enjoyed a better chance of being published in a top three finance journal. The results for $Top5Fin$ (Column 4) are largely identical to $Top3Fin$, with the only exception that the coefficient of three coauthor dummy becomes significant. In addition, the coefficient of the emerging-topic and highly similar precursor dummies seemingly become weaker.[49] The results of the logit regressions reveal the correlations of the highly similar precursor and emerging-topic dummies with $Top10Fin$ are no longer significant (see Internet Appendix Table IA.5), suggesting the next-tier journals are more open to follow-up studies and competing articles of existing papers. Also, it appears that the articles with four coauthors tend to have a higher likelihood of publishing in the less prestigious finance journals. The impacts from research resources and school ranking lose their significance in $Top10Fin$ analysis, suggesting that these factors are more instrumental for publication in the most coveted outlets.

Columns 2 and 5 of the table demonstrate the pattern in economics journals ($Top5Econ$). The coefficients of our novelty measures, $NewTopic$ and $LogInnovData$, and conventionality, approximated by semantic similarity, as well as the highly similar precursor dummy ($HighSim$), become insignificant. These results are consistent with the prediction by Ellison (2002) in the sense that the independently evolving peer-review norms would value novelty differently among different disciplines. Next, the number of databases used in the article is negatively and significantly correlated with $Top5Econ$, suggesting that many FEN articles may have appeared in top economics journals due to their theoretical or conceptual innovation instead of intensive data exploration. Furthermore, the close connection to the center of the FEN coauthorship network ($AuthorCentrality$), is not significantly correlated with $Top5Econ$, suggesting a rather separate author network between finance and economics. It is interesting to note that the coefficient of Prob_{DV} becomes significantly positive. The two results suggest that the authors of papers published in the top economics journals are more likely to have a research focus outside the traditional finance fields.

In Columns 3 and 6, we repeat the same investigation on top accounting journals. Similar to articles published in top economics journals, we find that our novel databases' coefficients and highly similar precursor dummy are no longer positively significant. However, emerging topics and conventionality remain significant and positive. Also, similar to those of the authors published in top economics journal, the centrality measures are not correlated with research outcomes in terms of accounting publications. Finally,

[49] In Table IA.5 of Internet Appendix, we further add in JCF, JBF, FM, JFM, and JFI to construct a binary dummy of the top 10 finance journals ($Top10Fin$). Journal abbreviations are provided in Table 1.

Column 7 reports the results of the interdisciplinary journal ($Top3Int$) analysis. We find that the coefficient of $LogInnovData$ is positively and significantly correlated with publication. The correlation between conventionality and $Top3Int$ becomes insignificant, and so are the coefficients of Prob_{DV} and $NewTopic$. As expected, the top interdisciplinary journals do not penalize papers from a nontraditional finance field.

6 Conclusion

Academics devote a large portion of their working hours to research. However, the question of what affects their contribution to society (or just to the scholarly community) has received surprisingly little attention. The existing literature focuses on published papers in prestigious journals, ignoring many research projects that end up in less prestigious journals or are discontinued due to various considerations. Motivated by pieces of countless anecdotal evidence and research from other disciplines, we adopt a few proprietary databases and newly developed textual analysis tools to examine the extent to which the disciplinary norm of finance values novelty and conventionality.

Our results suggest that novelty through investigating emerging topics and integrating new data sources significantly increases an article's citation, download, and publication likelihood in a top journal. We also find that novelty through combining atypical knowledge outside the finance focus fields would, on average, increase citation counts while reducing download counts and publication prospects, which is consistent with previous research using science literature (Uzzi et al., 2013; Wang et al., 2017). Furthermore, conventionality increases a paper's download counts and the chance of getting published in top journals while reducing citations from the subsequent research. These results remain after controlling for other aspects of article and author characteristics, such as research resources, top business school affiliation, and author network. Our study contributes to both textual analysis literature in finance and the study of finance literature, and offers new insights into the impactful research of financial studies.

References

Alexander, J. C. and Mabry, R. H. (1994) Relative significance of journals, authors, and articles cited in financial research, *Journal of Finance* 49:697–712

Arnold, T., Butler, A. W., Crack, T. F., and Altintig, A. (2003) Impact: What influences finance research?, *Journal of Business* 76:343–361

Bajo, E., Chemmanur, T. J., Simonyan, K., and Tehranian, H. (2016) Underwriter networks, investor attention, and initial public offerings, *Journal of Financial Economics* 122:376–408

Balsmeier, B., Fleming, L., and Manso, G. (2017) Independent boards and innovation, *Journal of Financial Economics* 123:536–557

Bena, J. and Li, K. (2014) Corporate Innovations and Mergers and Acquisitions, *The Journal of Finance* 69:1923–1960

Berninger, M., Kiesel, F., and Schnitzler, J. (2021) Commercial data in financial research, *Available at SSRN 3943132*

Bernstein, S. (2015) Does Going Public Affect Innovation?, *The Journal of Finance* 70:1365–1403

Black, F. and Scholes, M. (1973) The pricing of options and corporate liabilities, *The Journal of Political Economy* 81:637–654

Borokhovich, K. A., Bricker, R. J., Brunarski, K. R., and Simkins, B. J. (1995) Finance research productivity and influence, *Journal of Finance* 50:1691–1717

Brogaard, J., Engelberg, J., Eswar, S., and Van Wesep, E. D. (2020) On the Causal Effect of Fame on Citations, *Available at SSRN 3565487*

Brogaard, J., Engelberg, J., and Parsons, C. A. (Jan. 2014), Networks and productivity: Causal evidence from editor rotations, *Journal of Financial Economics* 111:251–270

Card, D. and DellaVigna, S. (2013) Nine facts about top journals in economics, *Journal of Economic Literature* 51:144–61

Card, D., DellaVigna, S., Funk, P., and Iriberri, N. (2020) Are Referees and Editors in Economics Gender Neutral?, *The Quarterly Journal of Economics* 135:269–327

Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., John, R. S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., et al. (2018) Universal sentence encoder, arXiv preprint arXiv:1803.11175

Chen, C. R. and Huang, Y. (Dec. 2007), Author Affiliation Index, finance journal ranking, and the pattern of authorship, *Journal of Corporate Finance* 13:1008–1026

Chevalier, J. and Ellison, G. (1999) Career concerns of mutual fund managers, *The Quarterly Journal of Economics* 114:389–432

Chung, K. H., Cox, R. A., and Mitchell, J. B. (2001) Citation patterns in the finance literature, *Financial Management* pp. 99–118

Cohen, L., Diether, K., and Malloy, C. (2013) Misvaluing Innovation, *The Review of Financial Studies* 26:635—-666

Currie, R. R. and Pandher, G. S. (2020) Finance journal rankings: Active scholar assessment revisited, *Journal of Banking and Finance* 111:105717

Ductor, L., Fafchamps, M., Goyal, S., and van der Leij, M. J. (2014) Social networks and research output, *Review of Economics and Statistics* 96:936–948

Dyer, T., Lang, M., and Stice-Lawrence, L. (2017) The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation, *Journal of Accounting and Economics* 64:221–245

Ellison, G. (2002) Evolving standards for academic publishing: A q-r theory, *Journal of Political Economy* 110:994–1034

Ferreira, D., Manso, G., and Silva, A. C. (2014) Incentives to innovate and the decision to go public or private, *Review of Financial Studies* 27:256–300

Goldstein, I., Jiang, W., and Karolyi, G. A. (2019) To FinTech and beyond, *The Review of Financial Studies* 32:1647–1661

Hall, B. H., Jaffe, A. B., and Trajtenberg, M. (2001) The NBER patent citation data file: Lessons, insights and methodological tools, *NBER working paper*

Hamermesh, D. S. (2018) Citations in economics: Measurement, uses, and impacts, *Journal of Economic Literature* 56:115–56

Hanley, K. W. and Hoberg, G. (2019) Dynamic interpretation of emerging risks in the financial sector, *Review of Financial Studies* 32:4543–4603

Hirshleifer, D. (12 2014), Editorial: Cosmetic Surgery in the Academic Review Process, *Review of Financial Studies* 28:637–649

Hirshleifer, D., Hsu, P.-H., and Li, D. (2018) Innovative originality, profitability, and stock returns, *The Review of Financial Studies* 31:2553–2605

Hoberg, G. and Phillips, G. (2016) Text-based network industries and endogenous product differentiation, *Journal of Political Economy* 124:1423–1465

Holmstrom, B. (1989) Agency costs and innovation, *Journal of Economic Behavior & Organization* 12:305–327

Hong, H., Karolyi, A., and Scheinkman, J. (2020) Climate Finance, *The Review of Financial Studies* 33:1011—1023

Hong, H. and Kubik, J. D. (2003) Analyzing the analysts: Career concerns and biased earnings forecasts, *The Journal of Finance* 58:313–351

Hong, H., Kubik, J. D., and Solomon, A. (2000) Security analysts' career concerns and herding of earnings forecasts, *The Rand journal of economics* pp. 121–144

Hope, O.-K., Hu, D., and Lu, H. (2016) The benefits of specific risk-factor disclosures, *Review of Accounting Studies* 21:1005–1045

Kennedy, G. (1988) How journalists select "news", *ETC: A Review of General Semantics* pp. 164–173

Kerr, W. R. and Nanda, R. (2015) Financing Innovation, *Annual Review of Financial Economics* 7:445–462

Kim, E. H., Morse, A., and Zingales, L. (2009) Are elite universities losing their competitive edge?, *Journal of Financial Economics* 93:353–381

Kogan, L., Papanikolaou, D., Seru, A., and Stoffman, N. (2017) Technological Innovation, Resource Allocation, and Growth, *The Quarterly Journal of Economics* 132:665–712

Lee, C. J., Sugimoto, C. R., Zhang, G., and Cronin, B. (2013) Bias in peer review, *Journal of the American Society for Information Science and Technology* 64:2–17, ISBN: 1532-2882 Publisher: Wiley Online Library

Lerner, J. and Seru, A. (November 2017), The Use and Misuse of Patent Data: Issues for Corporate Finance and Beyond, Working Paper 24053 National Bureau of Economic Research

Manso, G. (2011) Motivating Innovation, *The Journal of Finance* 66:1823–1860

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., and López-Cózar, E. D. (2018) Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories, *Journal of Informetrics* 12:1160–1177

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013) Distributed representations of words and phrases and their compositionality, In *Advances in Neural Information Processing Systems* pp. 3111–3119

Schaefer, S. M. (1998) Robert Merton, Myron Scholes and the development of derivative pricing, *The Scandinavian Journal of Economics* 100:425–445

Schilling, M. A. and Green, E. (2011) Recombinant search and breakthrough idea generation: An analysis of high impact papers in the social sciences, *Research Policy* 40:1321–1331

Shepherd, G. (1995) *Rejected: Leading Economists Ponder the Publication Process*, Thomas Horton and Daughters

Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.-J., and Wang, K. (2015) An overview of Microsoft Academic Service (MAS) and applications, In *Proceedings of the 24th international conference on world wide web* pp. 243–246

Spiegel, M. (03 2012), Reviewing Less—Progressing More, *Review of Financial Studies* 25:1331–1338

Stephan, P., Veugelers, R., and Wang, J. (2017) Reviewers are blinkered by bibliometrics, *Nature* 544:411–412

Swanson, E. P. (2004) Publishing in the majors: A comparison of accounting, finance, management, and marketing, *Contemporary Accounting Research* 21:223–255

Trajtenberg, M., Henderson, R., and Jaffe, A. (1997) University versus corporate patents: A window on the basicness of invention, *Economics of Innovation and new technology* 5:19–50

Uzzi, B., Mukherjee, S., Stringer, M., and Jones, B. (2013) Atypical combinations and scientific impact, *Science* 342:468–472

Wang, J., Veugelers, R., and Stephan, P. (2017) Bias against novelty in science: A cautionary tale for users of bibliometric indicators, *Research Policy* 46:1416–1436

Wang, K., Shen, Z., Huang, C., Wu, C.-H., Eide, D., Dong, Y., Qian, J., Kanakia, A., Chen, A., and Rogahn, R. (2019) A review of Microsoft Academic Services for science of science studies, *Frontiers in Big Data* 2:45

Welch, I. (2014) Referee recommendations, *Review of Financial Studies* 27:2773–2804

West, J. D., Jensen, M. C., Dandrea, R. J., Gordon, G. J., and Bergstrom, C. T. (2013) Author-level Eigenfactor metrics: Evaluating the influence of authors, institutions, and countries within the social science research network community, *Journal of the American Society for Information Science and Technology* 64:787–801

Zingales, L. (2015) Presidential address: Does finance benefit society?, *Journal of Finance* 70:1327–1363

Appendix Table A.1

Variable Definition and Data Source

Variable	Definition and Data Source
Research Outcome Variables	
LogCitations	The log of number of citations (Google Scholar)
LogDownloads	The log of number of downloads (SSRN)
Fin3Top	A dummy variable equal to one if a paper is accepted by a top three finance publication and zero otherwise (WoS and Google Scholar)
LogNews	The log of the sum of total news mentions (SSRN)
LogSocial	The log of the sum of total social media mentions (SSRN)
Novelty and Conventionality Variables	
NewTopic	A dummy variable equal to one if the paper covers an emerging topic that first appeared within 3 years prior to the article and zero otherwise (SSRN)
LogInnovData	The log of the number of databases that first appeared in SSRN FEN within 3 years prior to an article (SSRN)
Prob_{DV}	The probability for a paper to be outside the five traditional focus fields of finance. The probability is assigned by a CNN classifier trained by WFA and AFA sessions as well as manual labeling
sim_{USE}	Average of semantic similarity with previous papers (SSRN)
sim_{BOW}	Average of bag-of-word similarity with previous papers (SSRN)
HighSIM_{USE}	A dummy variable equal to one if there is at least one previous paper with SIM_{USE} greater than 0.8 and zero otherwise (SSRN)
HighSIM_{BOW}	A dummy variable equal to one if there is at least one previous paper with SIM_{BOW} greater than 0.6 and zero otherwise (SSRN)
CiteBreadth	One minus the Herfindahl-Hirschman Index of papers cited by a FEN article across different MAG level-2 topics. (SSRN and MAG)
Career Stage/Status	A dummy variable equal to one when all authors of an article posted their first article on SSRN, published it in a journal, or had it accepted by conferences within 10 years prior to the article or when one of its authors has a coauthorship network centrality that is more than the 95th percentile of all FEN authors in the previous year, and zero otherwise.
Covariate Variables	
LogPage	The log of the reported number of pages (SSRN)
#Authors	Number of coauthors (SSRN)
LogData	The log of the total number of databases used (SSRN)
LogWRDS	The log of the total number of distinct databases the coauthors' institutions subscribed to (WRDS)
Top20School	A dummy variable equal to one if an author is affiliated with a top 20 business school as defined by top 3 finance journal publications and zero otherwise (UTD)
AuthorCentrality	Average 2-step eigenvector of all the authors, where the eigenvectors measure coauthorship connections weighted by each connection's relative importance in the network (SSRN)
Prob_{AP}	The probability for a paper to be in the asset pricing field. The probability is assigned by a CNN classifier trained by WFA and AFA sessions as well as manual labeling (SSRN).
Prob_{CF}	The probability for a paper to be in the corporate finance field (SSRN)
Prob_{FI}	The probability for a paper to be in the financial intermediation field (SSRN)
Prob_{IV}	The probability for a paper to be in the investment field (SSRN)
Prob_{MM}	The probability for a paper to be in the market microstructure field (SSRN)
HHI_{Fields}	The sum of squared probabilities of a paper belonging to one of six fields (SSRN)

Figure 1: Finance Working Papers and Top Finance, Economics, and Accounting Journal Publications

This figure reports time series of the numbers of FEN articles; those FEN articles that are eventually published in top finance, economics, or accounting journals; and the numbers of articles in top finance journals in each year as benchmarks for illustration purposes. Articles eventually in one of top journals in a given year indicates the number of articles that become available in SSRN that year and that end up in a top 3 finance, top 5 economics, or top 5 accounting journal. "# Top Fin" reflects the total number of articles published in top finance journals in a given year.

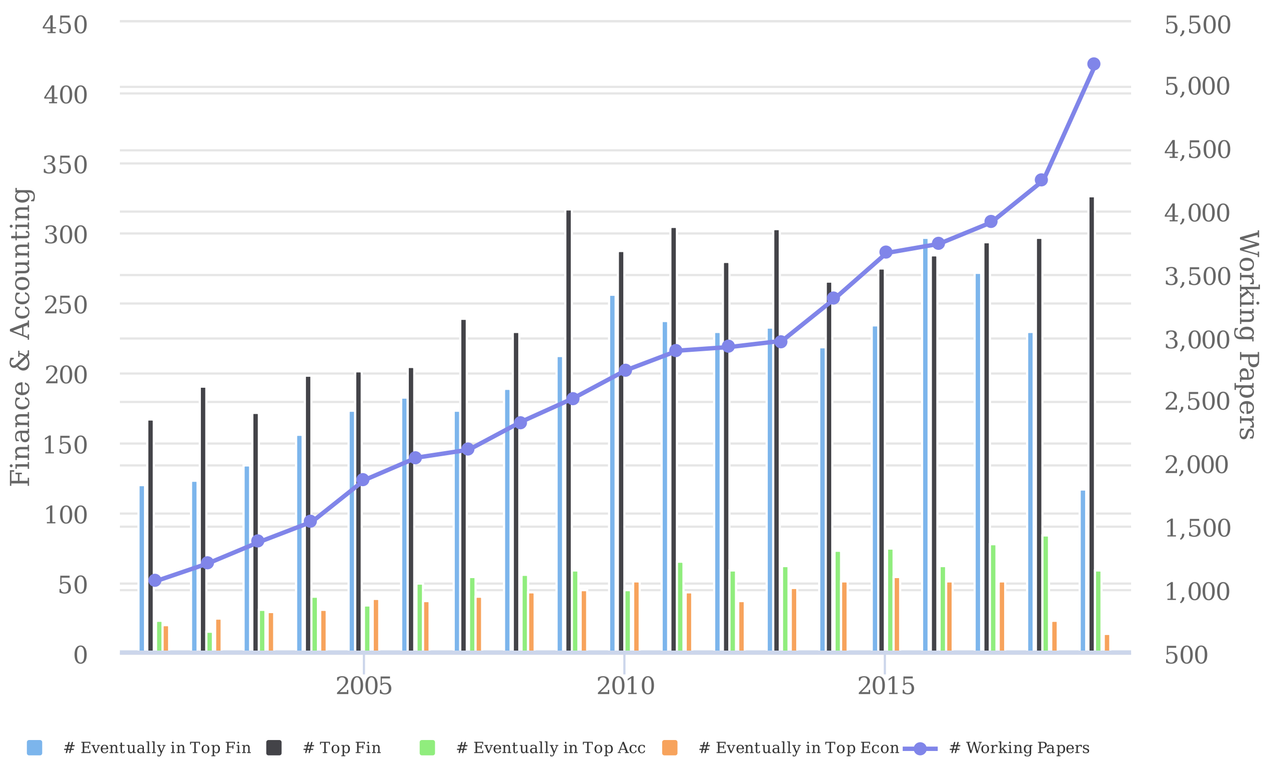


Figure 2: Years from SSRN Posting to Publication

Panel A of this figure reports the average number and one standard deviation band of years that span from SSRN posting of FEN articles to their publication in 1 of 28 journals. Panel B plots the same figures for the FEN articles published in the top 3 finance journals. We remove articles posted in and after their publishing year. The posting year is the earliest year of posting, being written, or first conference acceptance of a FEN article.

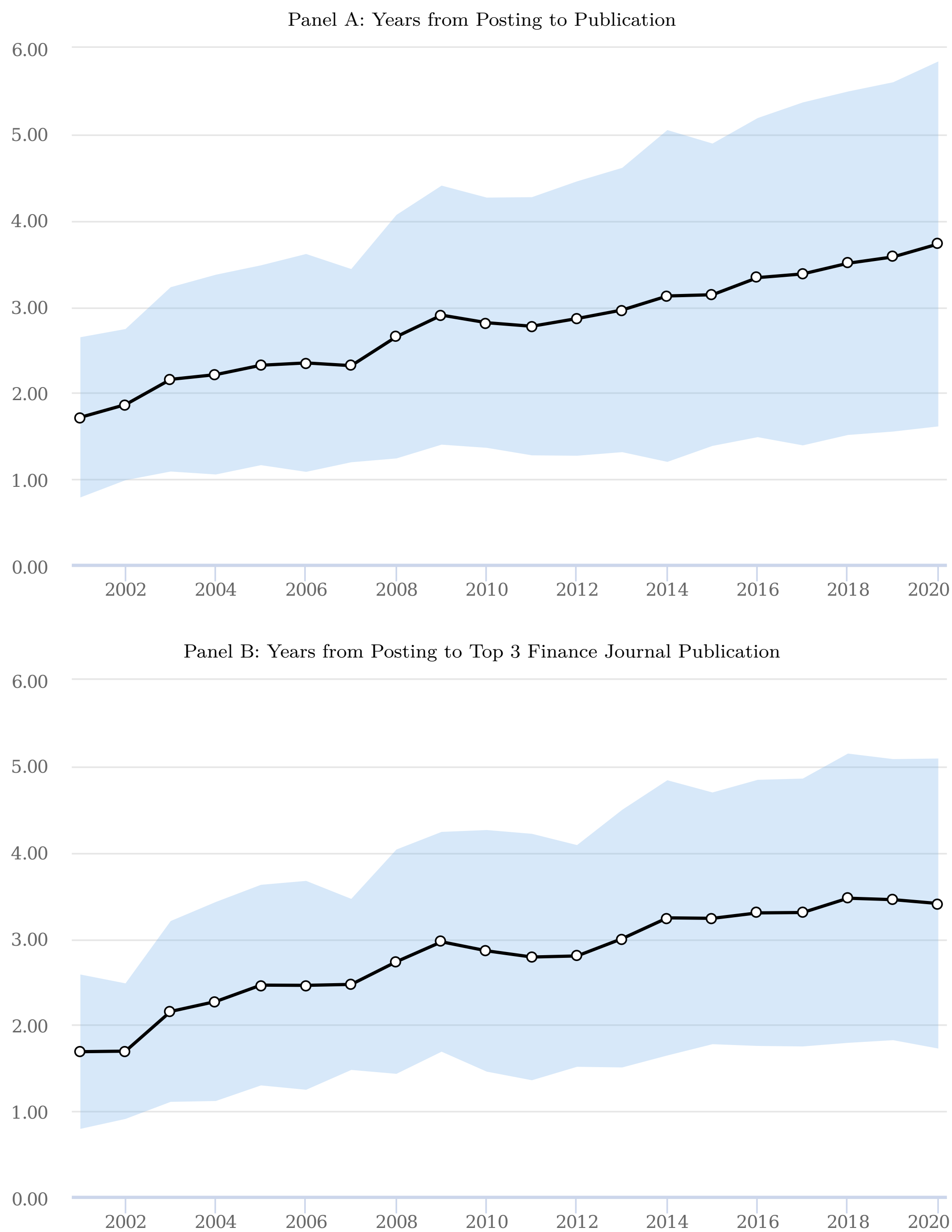


Figure 3: Illustration of sim_{BOW} and sim_{USE} Disagreement

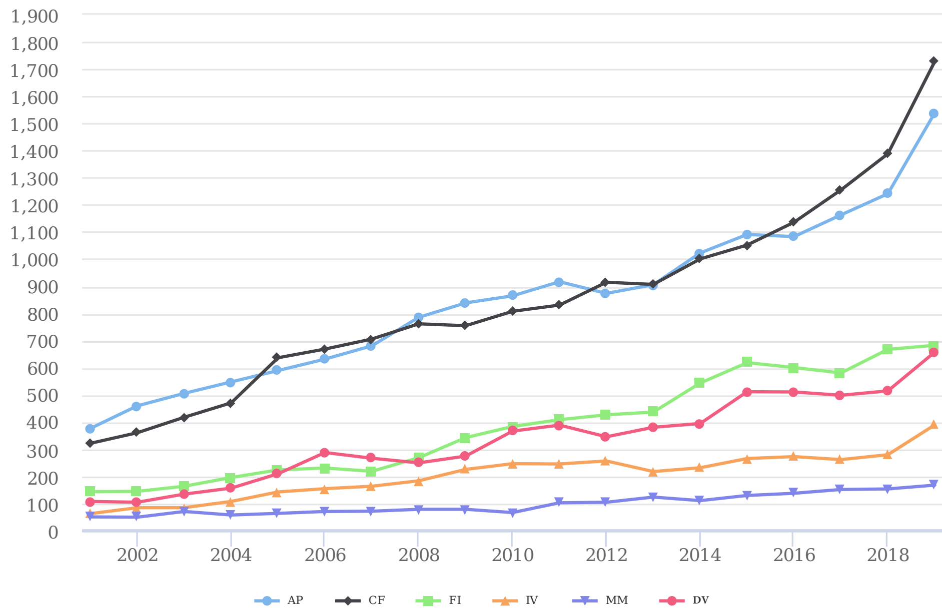
This figure illustrates two cases in which sim_{USE} and sim_{BOW} disagree each other. sim_{USE} is measured by the standard model of Google's Universal Sentence Encoder, while sim_{BOW} is calculated by unigram and bigram word vector cosine similarity.

Paper A	Paper B	sim_{USE}	sim_{BOW}
The Effect of Litigation Risk on Management Earnings Forecasts	*Management Forecasts and Litigation Risk*	0.91	0.48
We examine the effect of litigation risk on management's decision to issue earnings forecasts. We use a new ex ante measure of litigation risk, namely, the Directors and Officers liability insurance premium. This measure bypasses significant problems associated with the estimation of ex ante litigation risk in prior studies. By using this measure of litigation risk, our results are more intuitive. We find that, when faced with ex ante litigation risk, managers with bad news are more likely to issue an earnings warning. For good news firms, we do not see this effect. We also examine three forecast characteristics: forecast horizon, extent of news revealed, and forecast precision. Firms with higher litigation risk tend to issue earnings forecasts earlier if they have bad news, but this is not so when they have good news. They also reveal less news in the forecasts if they have good news. As litigation risk increases, bad news earnings forecasts tend to become more precise while good news earnings forecasts tend to become less precise. This differential effect of litigation risk on management earnings forecasts, based on the direction of the news, has not been documented by previous studies.	We examine the influence of the ex-ante risk of class action securities litigation on firms' decisions to issue management earnings forecasts as well as the characteristics of those forecasts. We find that litigation risk is positively associated with the likelihood of issuing a forecast for both good- and bad-news firms. While the association is marginally stronger for firms with bad earnings news, our results suggest that litigation risk is unlikely to explain the observed preponderance of bad-news forecasts. We examine the effect of litigation risk on the amount of the total earnings news released in the forecast, on forecast horizon, and on forecast precision. These results indicate that higher litigation risk is associated with a higher proportion of news being released when firms have bad news. Finally, higher litigation risk is associated with forecasts being released earlier and being more precise.		
Tail Risk and Pk-Tail Risk	*Regulatory Arbitrage of Risk Measures*	0.52	0.87
This paper discusses the notion of tail risk, and the ability of a tail risk measure to reflect this kind of risk. In particular, Yamai and Yoshiba's (2001, 2002) notion of strict risk measure tail risk is discussed and linked with a different notion of tail risk, the pK-tail risk, which is the risk associated with the probability measure conditional on the event that the losses are at least as large as K. A subset of pK-tail risk measures that are free of strict risk measure tail risk is introduced. These notions are then extended to Yaari's (1987) dual theory and the distorted risk measures framework.	We introduce the regulatory arbitrage of risk measures, one of the key considerations in choosing a suitable risk measure to use in banking regulation. A regulatory arbitrage is the amount of capital requirement reduced by splitting a financial risk into several fragments, regulated via a risk measure separately. Coherent risk measures by definition are free of regulatory arbitrage; dividing risks will not reduce the total capital requirement under a coherent risk measure. However, risk measures in practical use, such as the Value-at-Risk (VaR), are often not coherent and the magnitude of their regulatory arbitrage is then of significant importance. We quantify the regulatory arbitrage of risk measures in a rigorous mathematical framework, and categorize risk measures into three classes: free of regulatory arbitrage, of limited regulatory arbitrage, and of infinite regulatory arbitrage. We provide explicit results to characterize the regulatory arbitrage for general classes of risk measures, including distortion risk measures and convex risk measures. Several examples of risk measures of limited regulatory arbitrage are illustrated, as possible alternatives for coherent risk measures.		

Figure 4: Total Paper Count by Field and Year

Panel A illustrates the number of articles in each research field through the years. The probability of being in a research field is determined by a supervised CNN classifier which is trained by the articles that are accepted in American Finance Association and Western Finance Association annual meetings. A field of an article is the one with highest probability among all the fields. Panel B demonstrates the number of articles determined to be in a field with probability greater than 0.8.

Panel A: Number of Papers in Each Field



Panel B: Number of Papers in Each Field with Probability > 0.8

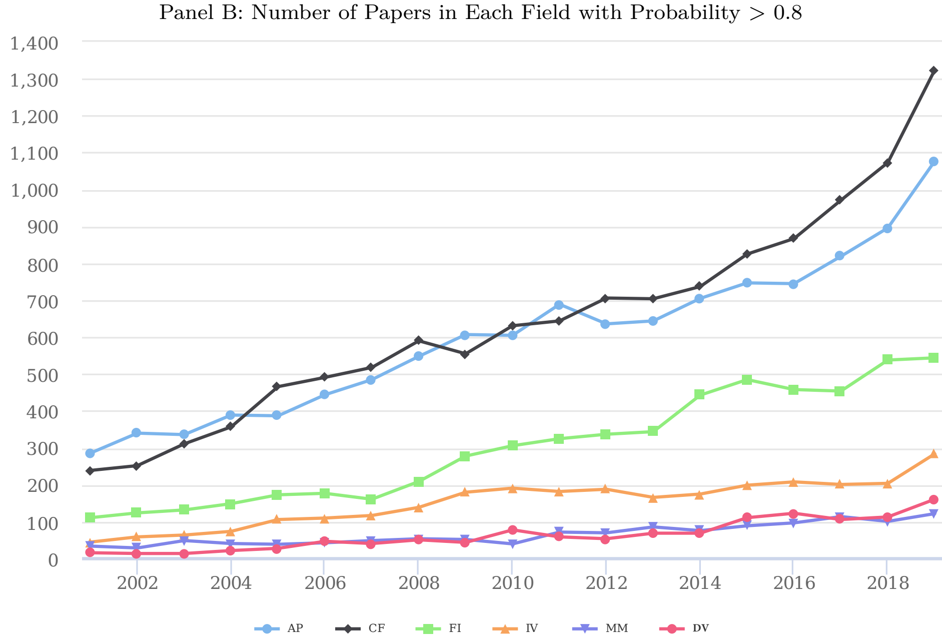
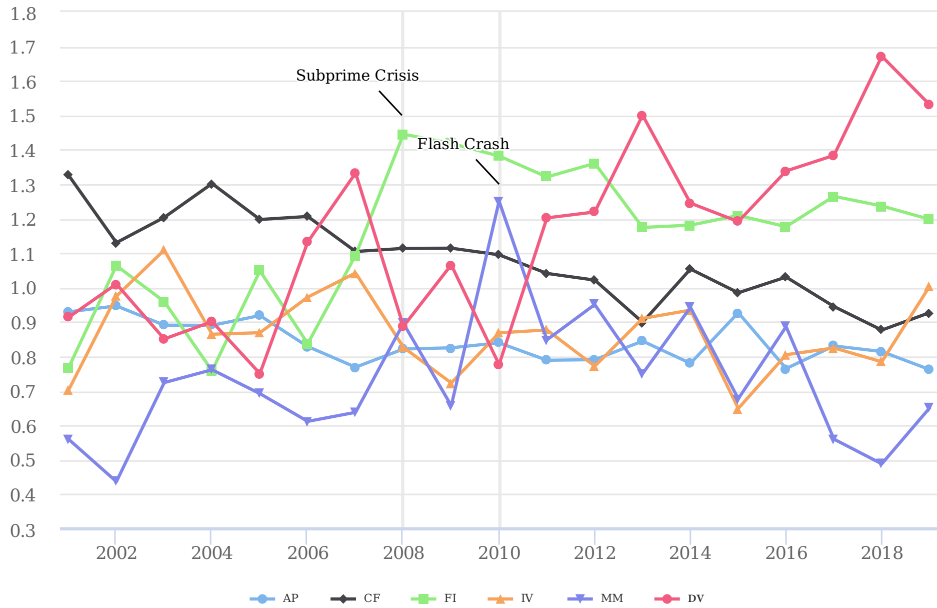


Figure 5: Number of Citations and Downloads by Field and Year

This figure reports the year-by-year relative share of Google Scholar citations (Panel A) and SSRN downloads (Panel B) for the articles in a given field compared to the citations to all the articles in a year, normalized by the fraction of the number of articles for that same field out of all FEN articles in that same year.

Panel A: Relative Share of Google Scholar Citations



Panel B: Relative Share of SSRN Downloads

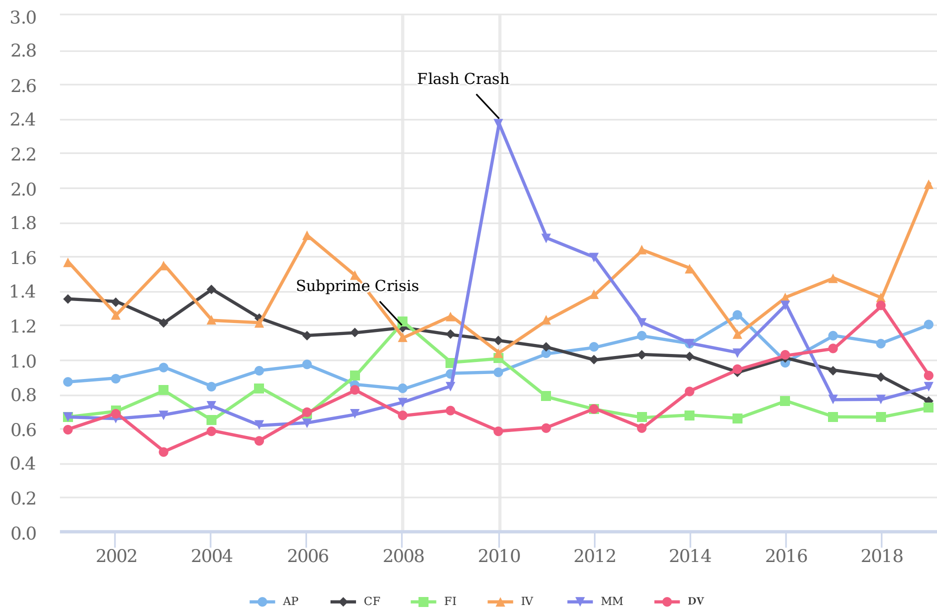


Figure 6: Paper Length and Author Count

Panel A of this figure reports the year-by-year average number of pages per article within each finance field, while Panel B reports the year-by-year average number of authors per article within each finance field. All values are smoothed using a 3-year centered moving average.

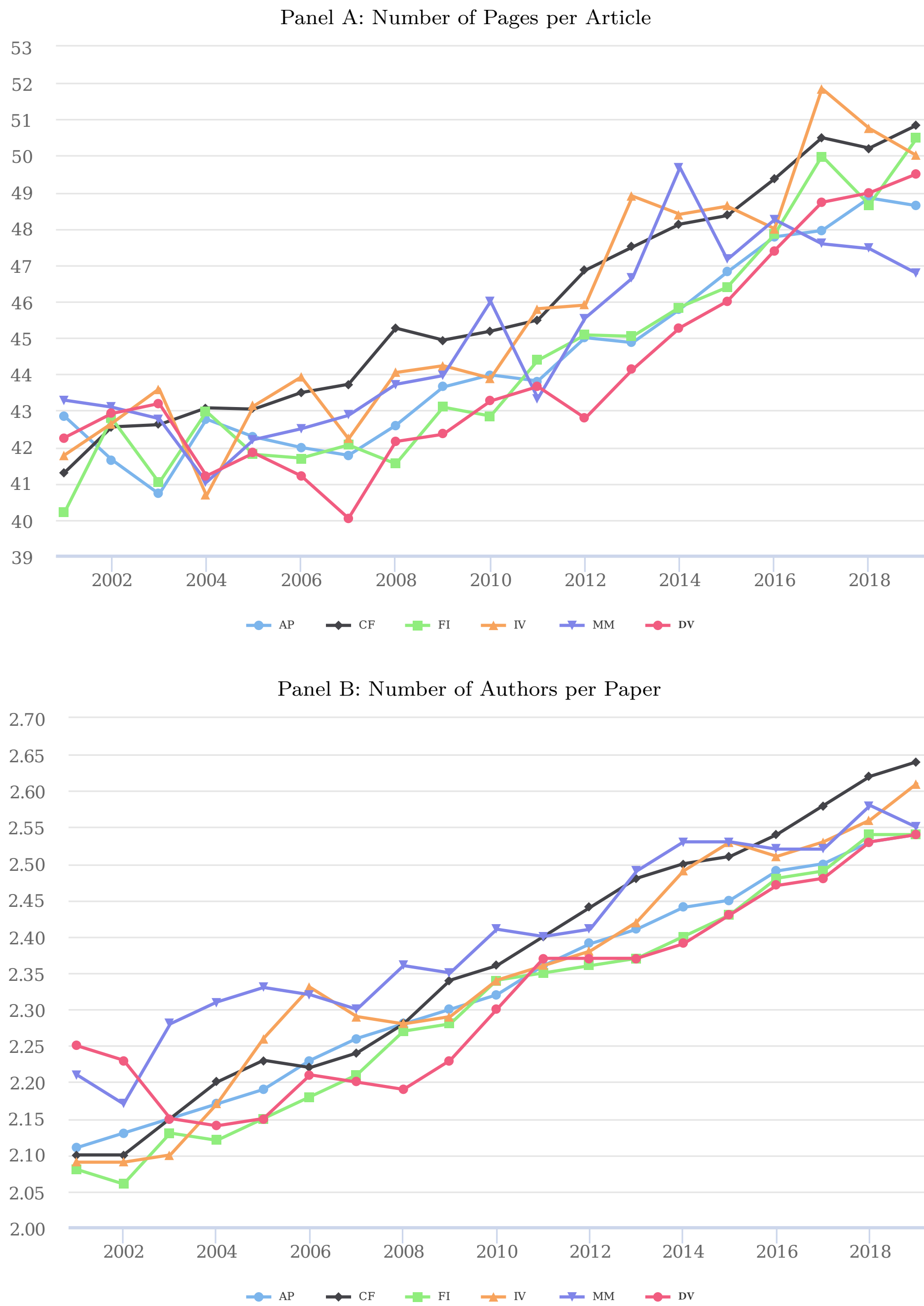
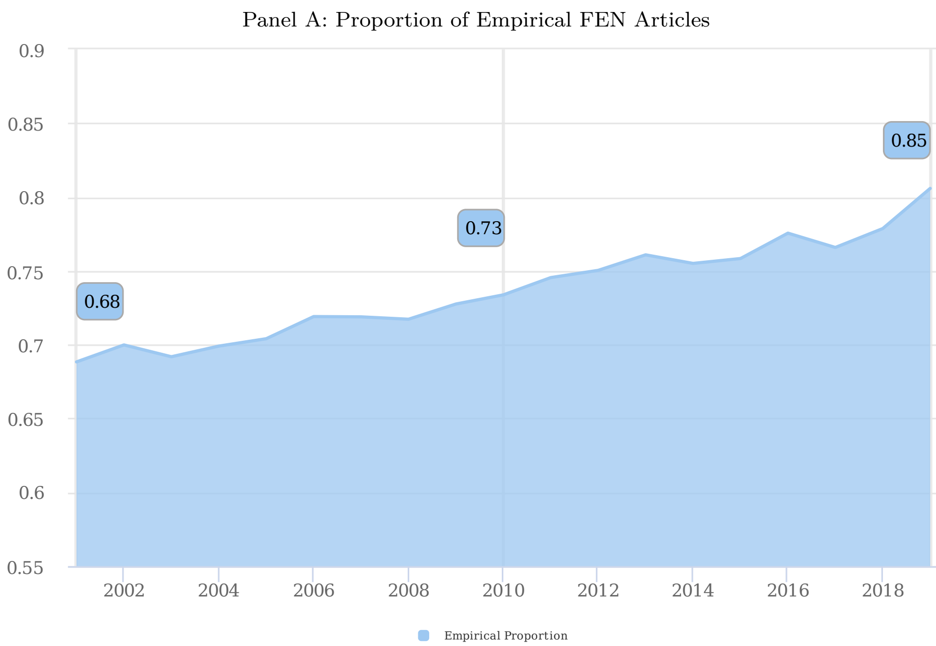


Figure 7: Empirical Papers and Database Count

Panel A of this figure reports the proportion of empirical FEN articles, defined as those which include at least one database. Panel B reports the number of databases used per empirical article. Each database entity is collected through a keyword matching method and/or a supervised name entity recognition model and is assigned a probability that the potential entity is mentioned in a context referring to databases. We include records with a context that refers to a database with at least a probability of 0.9 and higher.



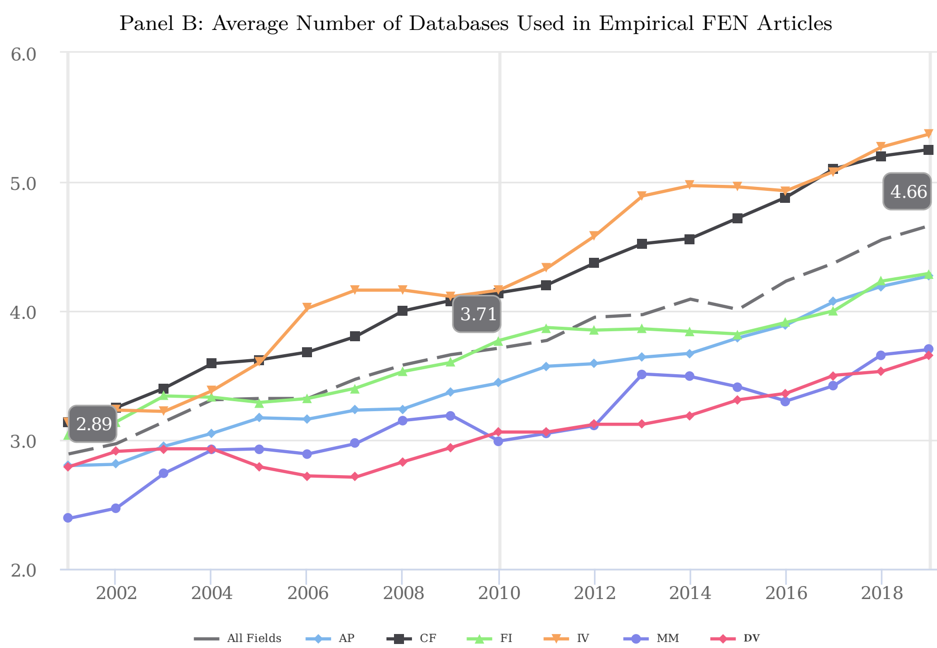


Figure 8: Randomly Selected Emerging Topics Through Years

This figure illustrates three emerging topics per year through our sample period. An emerging topic is a topic that appears the first time in a year over our FEN sample (2001–2019), but has not been mentioned more than 3 times among FEN papers since FEN inception (1994) or among papers published in JF, JFE, RFS, and JFQA back to 1980. A topic is selected from the lemmatized unigrams and bigrams from titles, keywords, and abstracts among all the articles in the year and must contain at least one noun detected by *part of speech* tagging. All meaningless and ambiguous grams, variations of existing ideas, and database, human, firm, and region names are removed.

		Emerging Topics	
2019	GDPR (General Data Protection Reg.)	Cyberattacks	Autoencoder
2018	Stable Coins	Judge Ideology	Carbon Pricing
2017	Bitcoin Market	Regulatory Sandbox	Demonetization
2016	Brexit	Panama Papers	Wind Power
2015	Cyber Risk	Blockchains	Cloud Computing
2014	Smart Beta	Sentiment Analysis	Bunching
2013	Bitcoin	Banking Union	Crowdfunding Platform
2012	JOBS Act	Forward Guidance	LASSO
2011	ESG	Clawback Provision	CSR Performance
2010	Dodd-Frank Act	LTG Forecast	Flash Crash
2009	Dark Pool	Securitized Bond	Asset Relief
2008	Sovereign Wealth (Fund)	Credit Crisis	Housing Bubble
2007	Funding Liquidity	Mandatory IFRS	Forecast Guidance
2006	Emission	Crosslisting Premium	Egalitarianism
2005	Predatory Lending	Material Weakness	Managerial Overconfidence
2004	Tax Haven	Money Illusion	CDS Market
2003	Agglomeration	Accrual Anomaly	SRI (Socially Responsible Investing)
2002	Regulation FD	Temperature	Price Synchronicity
2001	Target Price	Interchange Fee	Religion

Figure 9: **Conventionality, Publication and Citations**

This figure presents year-by-year median cumulative citations within each quintile ranked by USE abstract similarity of FEN papers published in 28 selected journals, measured against all prior FEN works in our sample, from three years prior to up to ten years after the publication year.

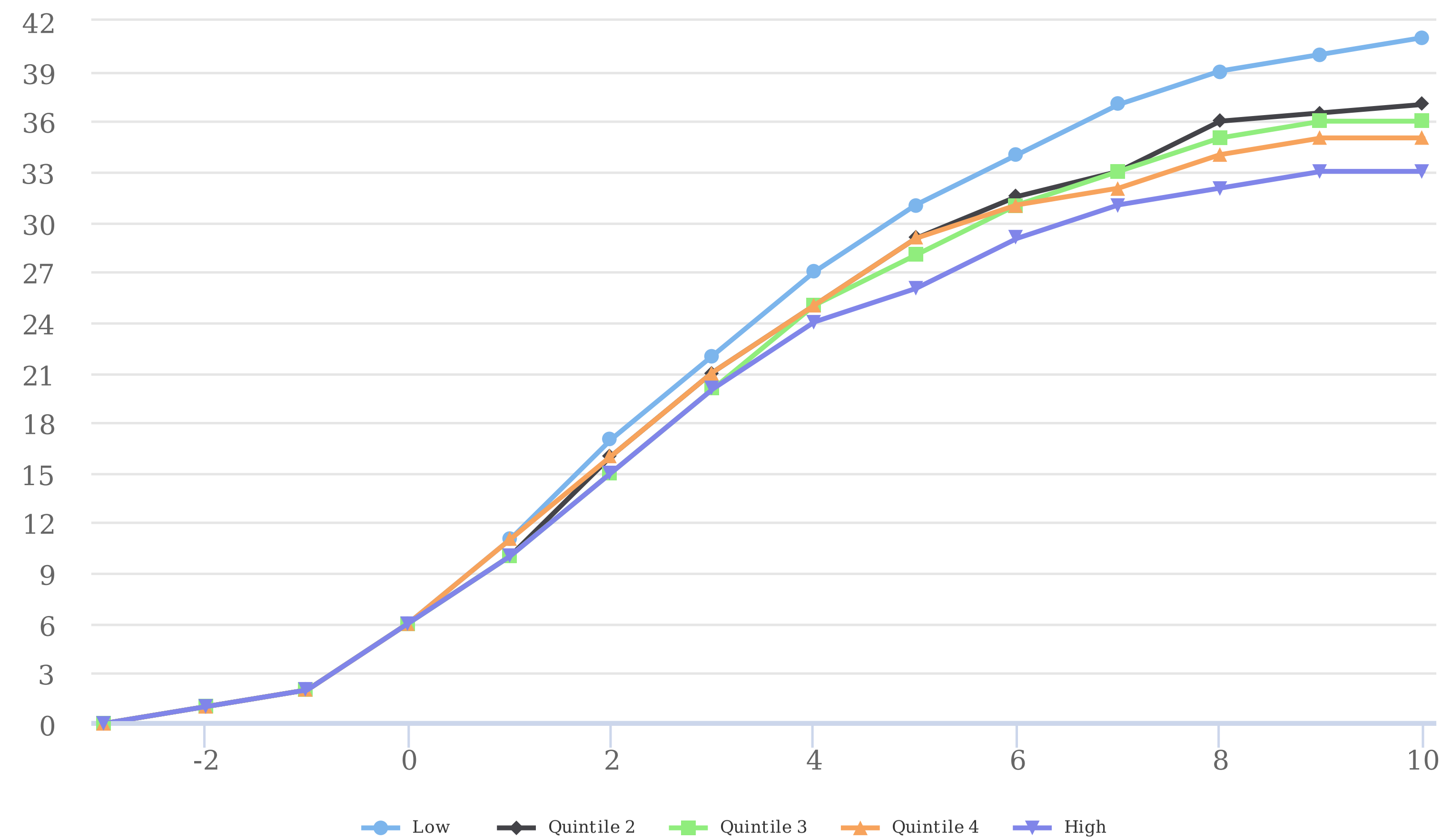


Table 1: **Destination Journals, Citations, and Downloads of FEN Articles**

This table reports number of destination journal articles, citations, and downloads of finance articles posted to SSRN from 2001 to 2019. The destination journals include 5 finance journals and 13 journals from economics, accounting, and interdisciplinary fields that are closely related to finance from the *Financial Times* 50 Journal list. In addition, we select 9 finance journals that are ranked as B+ and above by Currie and Pandher (2018) as well as a new journal, *Review of Corporate Finance Studies*. The finance articles in the Others group are the ones that are published in other journals or that remain as working articles upon our data collection. The number of downloads is collected from SSRN in May 2020, while the citation data is collected from Google Scholar over a four-week window from May 2020 to Jun 2020. The citations (downloads) of an article are normalized by the average number of citations (downloads) of articles posted on FEN in the same year, which is often referred to in innovation literature as relative citation strength.

				Number of Citations					Number of Downloads				
Journal Name	Abb.	Year	Articles	Mean	Std	P10	Median	P90	Mean	Std	P10	Median	P90
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Top Finance Journals													
Journal of Finance	JF	2001-2020	1,022	3.44	3.60	0.62	2.33	7.07	2.12	3.45	0.18	1.20	4.47
Journal of Financial Economics	JFE	2001-2020	1,557	3.11	4.08	0.48	1.95	6.67	2.27	5.78	0.28	1.26	4.34
Review of Financial Studies	RFS	2001-2020	1,212	2.70	3.97	0.41	1.64	5.52	2.09	4.48	0.22	1.11	3.81
Journal of Financial Intermediation	JFI	2001-2020	272	1.60	2.25	0.14	0.80	4.05	0.70	0.64	0.13	0.49	1.47
Journal of Money Credit And Banking	JMCB	2001-2020	251	1.33	1.73	0.11	0.74	3.40	0.44	0.53	0.07	0.27	0.97
Review of Asset Pricing Studies	RAPS	2011-2020	59	1.10	1.35	0.24	0.71	2.19	1.71	1.98	0.32	1.25	3.28
Journal of Corporate Finance	JCF	2001-2020	745	1.23	1.88	0.13	0.70	2.76	0.88	1.35	0.18	0.54	1.72
Journal of Banking & Finance	JBF	2001-2020	1,473	1.19	1.68	0.14	0.67	2.76	0.75	0.94	0.15	0.48	1.57
Journal of Financial & Quant. Analysis	JFQA	2001-2020	714	1.08	1.26	0.14	0.65	2.43	1.45	1.90	0.33	0.99	2.89
Review of Finance	ROF	2001-2020	417	1.25	2.07	0.15	0.64	2.67	1.31	2.00	0.19	0.74	2.76
Journal of Financial Markets	JFM	2002-2020	236	0.84	1.30	0.08	0.51	1.73	1.13	1.96	0.21	0.59	2.14
Financial Management	FM	2001-2020	277	0.80	1.43	0.10	0.43	1.77	0.86	1.09	0.17	0.53	1.81
Journal of Financial Econometrics	JFEC	2003-2020	93	1.05	1.71	0.10	0.41	2.79	0.90	1.36	0.13	0.49	1.61
Review of Corporate Finance Studies	RCFS	2012-2020	39	0.69	0.80	0.03	0.37	2.14	1.08	1.22	0.24	0.77	2.12
Journal of Empirical Finance	JEF	2001-2020	355	0.66	0.94	0.05	0.36	1.57	0.68	0.75	0.14	0.45	1.39
Panel B: Top Economics Journals													
Quarterly Journal of Economics	QJE	2001-2020	147	6.70	9.25	1.32	3.77	15.64	1.81	3.86	0.05	0.40	5.10
Journal of Political Economy	JPE	2002-2020	94	5.04	5.55	0.80	3.12	10.72	1.44	3.83	0.04	0.23	3.93
American Economic Review	AER	2001-2020	292	5.36	8.70	0.72	2.88	11.05	0.83	2.02	0.05	0.25	1.95
Econometrica	ETCA	2003-2019	96	3.42	3.05	0.63	2.23	8.86	0.88	1.55	0.05	0.32	2.29
Review of Economic Studies	RES	2003-2020	109	2.95	3.32	0.45	1.81	7.26	0.51	0.60	0.05	0.30	1.36
Panel C: Top Accounting Journals													
Journal Of Accounting & Economics	JAE	2001-2020	343	3.15	4.45	0.46	1.84	7.18	1.99	1.96	0.56	1.39	4.25
Journal of Accounting Research	JAR	2002-2020	331	2.42	2.92	0.41	1.52	5.53	1.85	2.54	0.40	1.08	3.67
Accounting Review	TAR	2001-2020	367	2.22	2.97	0.29	1.34	4.90	2.10	3.04	0.57	1.44	3.94
Review of Accounting Studies	RAS	2003-2020	286	1.29	1.63	0.17	0.82	2.97	1.70	2.85	0.48	1.05	3.26
Contemporary Accounting Research	CAR	2002-2020	286	1.04	1.28	0.13	0.63	2.38	1.48	1.81	0.33	1.06	2.81
Panel D: Interdisciplinary Journals													
Journal of Inter. Business Studies	JIBS	2002-2020	59	1.72	1.81	0.21	1.17	4.08	0.76	0.73	0.16	0.49	1.50
Journal of Business Ethics	JBE	2005-2020	119	1.70	2.27	0.22	0.95	4.08	0.88	1.50	0.13	0.53	1.85
Management Science	MS	2003-2020	558	1.34	2.28	0.14	0.70	3.14	1.42	1.88	0.26	0.87	2.93
Panel E: Other Articles													
Others	OT		40,688	0.67	2.40	0.00	0.19	1.56	0.86	2.10	0.11	0.42	1.73

Table 2: **Conference Acceptance, Citations, and Downloads of FEN Articles**

This table reports the summary statistics of FEN articles accepted by the two most competitive general-interest finance conferences and their citations and downloads from 2001 to 2019. The conferences are the annual meetings of the American Finance Association (AFA) and the Western Finance Association (WFA). Panel A reports the distribution of citations and downloads, and Panel B reports the number of articles published in top journals of finance, economics, and accounting as shown in Table 1. The number of downloads is from SSRN in May 2020, while the citation data is collected from Google Scholar over a four-week window from May 2020 to June 2020. The citations (downloads) of an article are normalized by the average number of citations (downloads) of articles posted on FEN in the same year, which is often referred to as relative citation strength in innovation literature.

				Number of Citations					Number of Downloads				
Conference Name	Abb.	Year	Articles	Mean	Std	P10	Median	P90	Mean	Std	P10	Median	P90
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Citation and Downloads													
Western Finance Association Meetings	WFA	2001-2020	1,591	2.14	3.07	0.11	1.13	5.14	1.84	3.56	0.22	0.97	4.00
American Finance Association Meetings	AFA	2001-2020	2,260	2.04	3.25	0.10	1.04	4.76	1.91	3.15	0.31	1.15	4.07
Panel B: Top Journal Acceptance													
			Top 3 Fin.		Top 5 Econ.		Top 3 Acct.						
			No.	%	No.	%	No.	%					
Western Finance Association Meetings	WFA	2001-2020	647	40.67	56	3.52	10	0.63					
American Finance Association Meetings	AFA	2001-2020	827	36.59	60	2.65	23	1.02					

Table 3: **Textual Similarity with Previous Works**

This table documents average cross-field similarity among all articles posted on SSRN FEN over the period of 2001 to 2019. The similarity is the average similarity between the title and abstract of an article with those of all priorly available articles in each of six fields. An article's available year is the year when the article is posted on SSRN or its reported written year, whichever comes first. Panel A reports the similarity score using the *bag-of-words* (BOW) approach, and Panel B reports the similarity scores based on the standard version of Google's Universal Sentence Encoder (USE).

	AP	CF	FI	IV	MM	DV
Panel A: *Bag-of-Words* (BOW)						
Asset Pricing (AP)	0.104	0.067	0.065	0.071	0.094	0.072
Corporate Finance (CF)	0.066	0.112	0.067	0.068	0.065	0.063
Financial Intermediation (FI)	0.064	0.066	0.135	0.057	0.065	0.065
Investment (IV)	0.069	0.070	0.058	0.110	0.069	0.060
Market Microstructure (MM)	0.096	0.067	0.069	0.070	0.144	0.078
Diverse Field (DV)	0.072	0.064	0.066	0.060	0.076	0.072
Panel B: *Universal Sentence Encoder* (USE)						
Asset Pricing (AP)	0.431	0.350	0.354	0.382	0.404	0.349
Corporate Finance (CF)	0.344	0.381	0.340	0.361	0.334	0.322
Financial Intermediation (FI)	0.356	0.343	0.433	0.354	0.345	0.343
Investment (IV)	0.376	0.363	0.349	0.425	0.359	0.330
Market Microstructure (MM)	0.405	0.339	0.343	0.364	0.456	0.334
Diverse Field (DV)	0.350	0.325	0.343	0.334	0.335	0.331

Table 4: **Summary Statistics**

Panel A reports summary statistics for finance articles posted on SSRN over the period of 2001 to 2019. The summary statistics include research outcome characteristics, number of citations, number of downloads, as well as article-level covariates, including number of news mentions, number of news, blogs, and social media mentions, innovative databases, number of pages, number of coauthors, number of the databases used in the article, number of total databases accessible by coauthors through WRDS, top 20 research school dummy variable, average centrality of all coauthors, $Prob_i$ is the probability for an article belonging to the field i, HHI_{Field} is the summation of the squared probability that an article belongs to one of the five traditional finance focus fields and the diverse field, and *CiteBreadth* is 1 - the summation of the squared shares of cited topics. Panel B reports the correlations between the control variables used in this article. All the variables are defined in Appendix Table A.1.

Panel A: Summary Stats

Variable	N Obs	Mean	Std Dev	1st Pctl	5th Pctl	50th Pctl	95th Pctl	99th Pctl
Citations	52,497	60.857	143.082	0.000	0.000	11.000	307.000	949.000
Downloads	52,497	404.978	617.912	13.000	24.000	189.000	1,549.000	3,895.000
Top3Fin	52,497	0.072	0.259	0.000	0.000	0.000	1.000	1.000
NewTopic	52,497	0.060	0.238	0.000	0.000	0.000	1.000	1.000
InnovData	52,497	0.027	0.163	0.000	0.000	0.000	0.000	1.000
sim_{USE}	52,497	0.365	0.051	0.222	0.273	0.370	0.439	0.460
sim_{BOW}	52,438	0.078	0.024	0.028	0.040	0.077	0.120	0.136
#Page	52,497	45.611	13.035	26.000	28.000	44.000	70.000	86.000
#Authors	52,497	2.382	0.945	1.000	1.000	2.000	4.000	5.000
Data	52,497	3.021	2.840	0.000	0.000	2.000	9.000	12.000
WRDS	52,497	58.323	48.783	0.000	0.000	53.000	149.000	219.000
Top20School	52,497	0.307	0.461	0.000	0.000	0.000	1.000	1.000
AuthorCentrality	52,497	0.008	0.010	0.000	0.000	0.004	0.028	0.047
HHI_{Field}	52,497	0.806	0.228	0.191	0.331	0.907	0.999	1.091
Prob_{DV}	52,497	0.171	0.224	0.000	0.002	0.063	0.695	0.871
Prob_{AP}	52,497	0.315	0.385	0.000	0.001	0.072	0.990	0.997
Prob_{CF}	52,497	0.325	0.397	0.000	0.001	0.069	0.996	0.999
Prob_{FI}	52,497	0.156	0.313	0.000	0.000	0.007	0.990	0.999
Prob_{IV}	52,497	0.092	0.237	0.000	0.000	0.006	0.896	0.999
Prob_{MM}	52,497	0.043	0.164	0.000	0.000	0.001	0.198	0.996
CiteBreadth	41,291	0.874	0.119	0.375	0.673	0.908	0.961	0.972

Panel B: Correlation

	NewTopic	LogInnovData	$Prob_{DV}$	sim_{USE}	sim_{BOW}	LogData	#Authors	LogWRDS	Top20School	AuthorCent.	HHI_{Field}
LogInnovData	0.063										
$Prob_{DV}$	0.011	-0.029									
sim_{USE}	-0.010	0.018	-0.216								
sim_{BOW}	0.018	0.024	-0.162	0.500							
LogData	0.031	0.157	-0.158	0.172	0.145						
#Authors	-0.016	0.014	-0.018	0.031	0.021	0.090					
LogWRDS	-0.040	-0.005	0.058	-0.054	-0.070	0.166	0.200				
Top20School	0.045	0.024	0.071	-0.023	-0.021	0.063	0.082	0.333			
AuthorCentrality	0.059	0.070	-0.109	0.110	0.084	0.190	0.233	0.062	0.215		
HHI_{Field}	0.009	0.030	-0.445	0.127	0.113	0.157	0.006	0.017	-0.013	0.094	
CiteBreadth	0.007	0.009	0.011	0.020	0.063	0.041	0.045	0.053	0.032	0.047	0.022

Table 5: **Conventionality and Research Outcomes**

This table reports results from regressing research outcome, measured by the number of citations, the number of downloads, and a dummy variable indicating whether an article is eventually published in one of the top 3 finance journals on the measures of conventionality, author-, and article-specific covariates as follows:

$$Outcome_i = \alpha Conventionality_i + \sum_{k=1}^{K} \beta_k X_{i,k} + \delta_t + \theta_j + \epsilon_i$$

where $Conventionality_i$ is either the average of semantic similarity with previous FEN articles (sim_{USE}), or the average of *bag-of-word* similarity with previous FEN articles (sim_{BOW}), and $X_{i,k,t}$ is a vector of covariates, including logarithm of number of pages ($LogPage$), number of coauthors ($\#Authors$), log of number of databases used in the article ($LogData$), log of number of total databases accessible for coauthors through WRDS ($LogWRDS$), top 20 research school dummy variable ($Top20School$), and average centrality of all coauthors ($AuthorCentrality$), as well as year and field fixed effects (FE). t-statistics (z-statistics) of the OLS (logit) regression coefficients are shown in parentheses and are computed based on standard errors clustered at the year level. The number of observations (N Obs), adjusted R^2 (Pseudo R^2) for OLS (logit) regression are reported. The variable construction is presented in Appendix A.1.

	Citations				Downloads				Top3Fin			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
sim_{USE}	-1.555***	-1.108***			0.719***	0.441**			3.457***	2.966***		
	(-7.93)	(-4.78)			(4.26)	(2.35)			(8.96)	(7.27)		
sim_{BOW}			-2.451***	-1.823***			0.736**	0.007			4.583***	4.626***
			(-9.33)	(-6.33)			(2.50)	(0.02)			(6.13)	(6.02)
LogPage	0.618***	0.620***	0.610***	0.613***	0.525***	0.511***	0.529***	0.515***	2.824***	2.829***	2.828***	2.832***
	(9.72)	(9.66)	(9.65)	(9.55)	(15.56)	(15.81)	(15.90)	(16.21)	(36.37)	(36.39)	(36.51)	(36.51)
#Authors	0.141***	0.142***	0.140***	0.140***	0.018***	0.017**	0.019***	0.018**	-0.082***	-0.080***	-0.081***	-0.078***
	(11.09)	(11.11)	(11.12)	(11.09)	(3.09)	(2.80)	(3.12)	(2.85)	(-3.89)	(-3.79)	(-3.85)	(-3.70)
LogData	0.135***	0.133***	0.129***	0.128***	0.195***	0.177***	0.200***	0.182***	0.322***	0.315***	0.332***	0.322***
	(4.13)	(4.46)	(3.91)	(4.17)	(20.47)	(18.38)	(20.49)	(18.89)	(12.00)	(11.55)	(12.39)	(11.86)
LogWRDS	0.047***	0.049***	0.047***	0.049***	0.025***	0.023***	0.024***	0.022***	0.205***	0.207***	0.204***	0.208***
	(5.77)	(6.17)	(5.81)	(6.16)	(3.40)	(3.21)	(3.29)	(3.11)	(10.69)	(10.74)	(10.65)	(10.78)
Top20School	0.534***	0.527***	0.536***	0.527***	0.156***	0.172***	0.154***	0.170***	0.599***	0.609***	0.594***	0.605***
	(18.43)	(18.04)	(18.59)	(18.17)	(4.68)	(5.32)	(4.61)	(5.24)	(14.59)	(14.79)	(14.48)	(14.70)
AuthorCentrality	21.775***	22.007***	21.463***	21.856***	18.264***	17.729***	18.415***	17.826***	29.369***	28.924***	29.597***	28.993***
	(19.15)	(19.35)	(19.24)	(19.44)	(10.30)	(10.16)	(10.37)	(10.22)	(20.75)	(20.33)	(20.98)	(20.41)
N Obs	52,497	52,497	52,438	52,438	52,497	52,497	52,438	52,438	52,497	52,497	52,438	52,438
$\bar{R}^2$	0.371	0.375	0.370	0.375	0.273	0.288	0.272	0.288				
Pseudo R^2									0.167	0.170	0.166	0.169
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Table 6: **Highly Similar Precursor and Research Outcomes**

This table reports results from regressing research outcomes, measured by the number of citations, numbers of downloads, and a dummy variable indicating whether an article eventually published in one of the top 3 finance journals, on the dummy of *Highly Similar Precursor*, as well as article-specific covariates including the logarithm of innovative databases (*LogInnovData*), log number of pages (*LogPage*), number of coauthors (*#Authors*), log number of databases used in the article (*LogData*), log number of total databases accessible for coauthors through WRDS ($LogWRDS$), top 20 research school dummy variable ($Top20School$), and average centrality of all coauthors (*AuthorCentrality*), as well as year and field fixed effects (FE). t-statistics (z-statistics) of the OLS (logit) regression coefficients are shown in parentheses and are computed based on standard errors clustered at the year level. The number of observations (N Obs), adjusted R^2 (Pseudo R^2) for OLS (logit) regression are reported. The construction of the variables is presented in Appendix A.1.

	Citations		Downloads		Top3Fin	
Variable	(1)	(2)	(3)	(4)	(5)	(6)
HighSim$_{USE}$	0.016		-0.079***		-0.193**	
	(0.77)		(-4.15)		(-2.43)	
sim$_{USE}$	-1.113***		0.466**		3.032***	
	(-4.76)		(2.50)		(7.42)	
HighSim$_{BOW}$		0.025		-0.035**		-0.130
		(0.83)		(-2.16)		(-1.49)
sim$_{BOW}$		-1.844***		0.036		4.717***
		(-6.15)		(0.12)		(6.11)
LogPage	0.620***	0.613***	0.511***	0.515***	2.826***	2.831***
	(9.66)	(9.56)	(15.89)	(16.24)	(36.34)	(36.48)
#Authors	0.141***	0.140***	0.018**	0.018**	-0.079***	-0.078***
	(11.10)	(11.08)	(2.84)	(2.85)	(-3.76)	(-3.70)
LogData	0.133***	0.128***	0.177***	0.182***	0.315***	0.322***
	(4.45)	(4.17)	(18.41)	(18.94)	(11.56)	(11.85)
LogWRDS	0.049***	0.049***	0.023***	0.022***	0.207***	0.207***
	(6.17)	(6.18)	(3.21)	(3.10)	(10.74)	(10.76)
Top20School	0.527***	0.527***	0.171***	0.170***	0.608***	0.605***
	(18.03)	(18.18)	(5.31)	(5.25)	(14.78)	(14.70)
AuthorCentrality	21.995***	21.838***	17.787***	17.853***	29.036***	29.088***
	(19.34)	(19.36)	(10.16)	(10.22)	(20.39)	(20.45)
N Obs	52,497	52,438	52,497	52,438	52,497	52,438
$\bar{R}^2$	0.375	0.375	0.288	0.288		
Pseudo R^2					0.170	0.169
Field FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 7: **Emerging Topics, Novel Databases, and Research Outcome**

This table reports results from regressing research outcomes, measured by the number of citations, the numbers of downloads, and a dummy variable indicating whether an article eventually published in one of the top 3 finance journals, on the measures of novelty, conventionality, author-, and article-specific covariates as follows:

$$Outcome_i = \alpha_1 Conventionality_i + \alpha_2 Novelty_i + \sum_{k=1}^{K} \beta_k X_{i,k} + \delta_t + \theta_j + \epsilon_i$$

where $Novelty_i$ is either a dummy variable for emerging topics, number of novel databases used in an article, or both; $Conventionality_i$ is the average of semantic similarity with previous FEN articles (sim_{USE}); and $X_{i,k,t}$ is a vector of covariates, including logarithm of number of pages ($LogPage$), number of coauthors ($\#Authors$), log number of databases used in the article ($LogData$), log number of total databases accessible for coauthors through WRDS ($LogWRDS$), top 20 research school dummy variable ($Top20School$), and average centrality of all coauthors ($AuthorCentrality$), as well as year and field fixed effects (FE). t-statistics (z-statistics) of the OLS (logit) regression coefficients are shown in parentheses. t-statistics are computed based on standard errors clustered at the year level. The number of observations (N Obs), adjusted R^2 (Pseudo R^2) for OLS (logit) regression are reported. The construction of the variables is presented in Appendix A.1.

	Citations			Downloads			Top3Fin		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
NewTopic	0.586***		0.579***	0.551***		0.545***	0.167**		0.155**
	(15.84)		(15.69)	(9.87)		(9.75)	(2.57)		(2.38)
LogInnovData		0.385***	0.338***		0.325***	0.281***		0.463***	0.452***
		(6.87)	(5.98)		(7.31)	(6.36)		(4.43)	(4.32)
sim_{USE}	-1.043***	-1.084***	-1.023***	0.502**	0.461**	0.518***	2.980***	3.023***	3.035***
	(-4.59)	(-4.60)	(-4.44)	(2.85)	(2.42)	(2.90)	(7.31)	(7.41)	(7.43)
LogPage	0.616***	0.618***	0.614***	0.508***	0.509***	0.506***	2.828***	2.824***	2.823***
	(9.84)	(9.71)	(9.89)	(15.61)	(15.57)	(15.40)	(36.36)	(36.31)	(36.29)
#Authors	0.143***	0.141***	0.142***	0.019***	0.017**	0.018***	-0.079***	-0.081***	-0.081***
	(11.01)	(11.07)	(10.97)	(3.07)	(2.73)	(3.00)	(-3.77)	(-3.86)	(-3.85)
LogData	0.125***	0.123***	0.116***	0.170***	0.169***	0.163***	0.312***	0.292***	0.290***
	(4.39)	(4.28)	(4.21)	(18.24)	(17.52)	(17.22)	(11.44)	(10.51)	(10.43)
LogWRDS	0.048***	0.049***	0.048***	0.022***	0.023***	0.022***	0.207***	0.206***	0.206***
	(6.01)	(6.16)	(6.01)	(3.06)	(3.18)	(3.04)	(10.74)	(10.71)	(10.70)
Top20School	0.519***	0.526***	0.519***	0.164***	0.171***	0.164***	0.606***	0.610***	0.607***
	(17.77)	(18.09)	(17.82)	(5.14)	(5.37)	(5.19)	(14.72)	(14.80)	(14.73)
AuthorCentrality	21.758***	21.942***	21.704***	17.495***	17.674***	17.449***	28.843***	28.897***	28.824***
	(18.92)	(19.41)	(18.96)	(9.78)	(10.20)	(9.80)	(20.27)	(20.29)	(20.24)
N Obs	52,497	52,497	52,497	52,497	52,497	52,497	52,497	52,497	52,497
$\bar{R}^2$	0.381	0.376	0.381	0.299	0.289	0.299			
Pseudo R^2							0.170	0.171	0.171
Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 8: **Novelty from Atypical Fields and Research Outcomes**

This table reports results from regressing research outcomes, measured by the number of citations, numbers of downloads, and a dummy variable indicating whether an article is eventually published in one of the top 3 finance journals, on the probabilities of being classified into a finance research field, as well as the measures of novelty, conventionality, author-, and article-specific covariates. $Prob_{DV}$ is the probability for an article to be outside the five traditional focus fields of finance, $Prob_{AP}$ is the probability for an article to be in the asset pricing field, $Prob_{CF}$ is the probability for an article to be in the corporate finance field, $Prob_{FI}$ is the probability for an article to be in the financial intermediation field, $Prob_{IV}$ is the probability for an article to be in the investment field, and $Prob_{MM}$ is the probability for an article to be in the market microstructure field. The novelty element measure is either a dummy variable for emerging topics ($NewTopic$) or the log number of novel databases used in an article ($LogInnovData$). The conventionality measure is the average of semantic similarity with previous FEN articles (sim_{USE}). The vector of covariates, including logarithm of number of pages ($LogPage$), number of coauthors ($\#Authors$), log number of databases used in the article ($LogData$), log number of total databases accessible for coauthors through WRDS ($LogWRDS$), top 20 research school dummy variable ($Top20School$), and average centrality of all coauthors ($AuthorCentrality$), as well as year and field fixed effects (FE). t-statistics (z-statistics) of the OLS (logit) regression coefficients are shown in parentheses. t-statistics are computed based on standard errors clustered at the year level. The number of observations (N Obs), adjusted R^2 (Pseudo R^2) for OLS (logit) regression are reported. The construction of the variables is presented in Appendix A.1.

	Citations			Downloads			Top3Fin		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Prob_{DV}	0.236***	0.222***	0.193***	-0.648***	-0.661***	-0.557***	-0.851***	-0.856***	-0.849***
	(3.88)	(3.92)	(3.27)	(-8.13)	(-8.87)	(-8.30)	(-5.66)	(-5.69)	(-5.36)
Prob_{AP}			-0.184**			0.241***			0.271
			(-2.73)			(3.76)			(1.52)
Prob_{CF}			0.010			0.372***			-0.133
			(0.17)			(5.10)			(-0.76)
Prob_{FI}			0.428***			-0.084			-0.033
			(3.88)			(-1.11)			(-0.13)
Prob_{IV}			-0.108			0.266***			-0.197
			(-1.60)			(4.27)			(-0.81)
Prob_{MM}			-0.291**			0.307**			-0.237
			(-2.49)			(2.58)			(-0.73)
NewTopic		0.578***	0.576***		0.550***	0.551***		0.158**	0.161**
		(15.69)	(15.64)		(9.50)	(9.47)		(2.41)	(2.46)
LogInnovData		0.337***	0.342***		0.281***	0.280***		0.453***	0.457***
		(5.99)	(6.10)		(6.18)	(6.02)		(4.33)	(4.36)
sim_{USE}	-1.049***	-0.968***	-0.868***	0.280	0.354**	0.279	2.795***	2.863***	2.697***
	(-4.55)	(-4.20)	(-3.61)	(1.63)	(2.16)	(1.59)	(6.83)	(6.98)	(6.45)
N Obs	52,497	52,497	52,497	52,497	52,497	52,497	52,497	52,497	52,497
$\bar{R}^2$	0.375	0.382	0.382	0.293	0.304	0.306			
Pseudo R^2							0.171	0.172	0.172
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 9: **Robustness for Conventionality and Top3 Finance Journals**

This table reports results from regressing a dummy variable indicating whether an article is eventually published in one of the top 3 finance journals on the logarithm number of pre-publication citations $PrePubCites$ over entire sample. We also implement a sub-sample analysis that excludes $PrePubCites$ as an independent variables among the papers ranked as Top 2000, 1000, or 500 base on MAG Saliency in our sample. We further exclude the Top5 Econ Journal articles to remove papers that may be otherwise accepted by the top 3 financial journals. The conventionality measure is the average of semantic similarity with previous FEN articles (sim_{USE}). The novelty element measure is either a dummy variable for emerging topics ($NewTopic$) or the log number of novel databases used in an article ($LogInnovData$). $Prob_{DV}$ is the probability for an article to be outside the five traditional focus fields of finance. The vector of covariates includes logarithm of number of pages ($LogPage$), number of coauthors ($\#Authors$), log number of databases used in the article ($LogData$), log number of total databases accessible for coauthors through WRDS ($LogWRDS$), top 20 research school dummy variable ($Top20School$), and average centrality of all coauthors ($AuthorCentrality$), as well as year and field fixed effects (FE). t-statistics (z-statistics) of the OLS (logit) regression coefficients are shown in parentheses. t-statistics are computed based on standard errors clustered at the year level. The number of observations (N Obs), adjusted R^2 (Pseudo R^2) for OLS (logit) regression are reported. The construction of the variables is presented in Appendix A.1.

	Full Sample	Impactful Sub-Sample		
		Top 2000	Top 1000	Top 500
	(1)	(2)	(3)	(4)
sim_{USE}	2.965***	5.100***	4.212**	4.971**
	(7.09)	(4.24)	(2.53)	(1.99)
PrePubCites	0.482***			
	(23.43)			
NewTopic	0.022	-0.302**	-0.059	0.001
	(0.33)	(-2.28)	(-0.34)	(0.00)
LogInnovData	0.334***	0.255	0.061	-1.155**
	(3.12)	(0.97)	(0.17)	(-2.20)
Prob_{DV}	-0.909***	-2.388***	-2.888***	-2.039**
	(-6.01)	(-5.38)	(-4.57)	(-2.47)
LogPage	2.606***	1.531***	0.775**	0.310
	(32.68)	(6.54)	(2.40)	(0.66)
#Authors	-0.106***	-0.220***	-0.248***	-0.261**
	(-4.92)	(-3.43)	(-2.67)	(-1.98)
LogData	0.248***	0.617***	0.680***	0.896***
	(8.80)	(7.06)	(5.37)	(4.86)
LogWRDS	0.187***	0.078	0.159**	0.102
	(9.68)	(1.51)	(1.98)	(0.80)
Top20School	0.471***	0.258**	0.099	0.027
	(11.18)	(2.10)	(0.55)	(0.10)
AuthorCentrality	24.914***	11.621***	9.530**	11.413*
	(17.09)	(3.44)	(2.23)	(1.88)
N Obs	52,497	1,838	892	440
Pseudo R^2	0.193	0.156	0.139	0.145

Table 10: **Field Focus, Citation Breadth and Research Outcome**

This table reports results from regressing various research outcome variables on the measures of $FieldFocus$ and $CiteBreadth$, as well as novelty, conventionality, author-, and article-specific covariates. $FieldFocus$ is the summation of the squared probability of an article belonging to one of the five traditional finance focus fields and the diverse field. $CiteBreadth$ is defined as the extent to which an article cites previous articles spanning a wide range of research topics. The novelty element measure is either a dummy variable for emerging topics ($NewTopic$) or the log number of novel databases used in an article ($LogInnovData$). $Prob_{DV}$ is the probability for an article belonging to the diverse field. The conventionality measure is the average semantic similarity with previous FEN articles (sim_{USE}). The vector of covariates, including logarithm of number of pages ($LogPage$), number of coauthors ($\#Authors$), log number of databases used in the article ($LogData$), log number of total databases accessible for coauthors through WRDS ($LogWRDS$), top 20 research school dummy variable ($Top20School$), and average centrality of all coauthors ($AuthorCentrality$), as well as year and field fixed effects (FE). t-statistics (z-statistics) of the OLS (logit) regression coefficients are shown in parentheses. t-statistics are computed based on standard errors clustered at the year level. The number of observations (N Obs), adjusted R^2 (Pseudo R^2) for OLS (logit) regression are reported. The construction of the variables is presented in Appendix A.1.

	Citations			Downloads			Top3Fin	Top5Econ	Top3Acc
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$FieldFocus$	0.262***		0.232***	0.268***		0.277***	1.109***	-0.261	-0.909***
	(5.26)		(4.02)	(11.16)		(10.58)	(10.09)	(-1.27)	(-5.46)
$CiteBreadth$		1.363***	1.357***		-0.149**	-0.156***	2.716***	7.882***	-0.872***
		(13.60)	(13.73)		(-2.84)	(-2.94)	(10.08)	(8.14)	(-3.45)
NewTopic	0.577***	0.611***	0.610***	0.549***	0.534***	0.532***	0.131*	-0.091	0.273**
	(15.52)	(16.19)	(15.99)	(9.49)	(9.11)	(9.06)	(1.92)	(-0.54)	(2.42)
LogInnovData	0.335***	0.326***	0.324***	0.279***	0.289***	0.287***	0.448***	0.197	-0.339
	(5.98)	(6.02)	(6.00)	(6.16)	(7.04)	(7.01)	(4.16)	(0.63)	(-1.55)
Prob_{DV}	0.314***	0.169**	0.258***	-0.567***	-0.690***	-0.583***	-0.327**	2.359***	-2.103***
	(6.24)	(2.62)	(4.09)	(-8.09)	(-10.95)	(-9.89)	(-2.04)	(8.81)	(-6.49)
sim_{USE}	-1.031***	-1.822***	-1.865***	0.289*	0.624***	0.572***	2.352***	-2.444***	3.129***
	(-4.51)	(-8.18)	(-8.33)	(1.77)	(3.30)	(3.04)	(5.37)	(-2.68)	(4.27)
N Obs	52,497	41,291	41,291	52,497	41,291	41,291	41,291	40,500	41,291
$\bar{R}^2$	0.383	0.310	0.310	0.306	0.268	0.271			
Pseudo R^2							0.183	0.175	0.154
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 11: **Coauthorship Characteristics and Research Outcome**

This table reports results from regressing various research outcome variables on interaction variables between *Sort Variable* and the measures of novelty and conventionality. *Sort Variable* is a top 20 research school dummy variable (*Top20School*); a dummy variable for "High Centrality Researcher", which equals one when one of its authors has a coauthorship network centrality that is more than the 95th percentile of all FEN authors in the previous year (Column 4–6) and zero otherwise; and a dummy variable for "Early Stage Researcher", which equals one when all authors of an article posted (or published) their first article within 10 years prior to the article (Column 7–9) and zero otherwise. The novelty element measure is either a dummy variable for emerging topics (*NewTopic*) or the log number of novel databases used in an article (*LogInnovData*). $Prob_{DV}$ is the probability for an article belonging to the diverse field. The conventionality measure is the average semantic similarity with previous FEN articles (sim_{USE}). The vector of covariates, including logarithm of number of pages (*LogPage*), number of coauthors (*#Authors*), log number of databases used in the article (*LogData*), log number of total databases accessible for coauthors through WRDS (*LogWRDS*), top 20 research school dummy variable (*Top20School*), and average centrality of all coauthors (*AuthorCentrality*), as well as year and field fixed effects (FE). t-statistics (z-statistics) of the OLS (logit) regression coefficients are shown in parentheses. t-statistics are computed based on standard errors clustered at the year level. The number of observations (N Obs), adjusted R^2 (Pseudo R^2) for OLS (logit) regression are reported. The construction of the variables is presented in Appendix A.1.

Sort Variable	Top 20 Schools			High Centrality Researcher			Early Stage Researcher		
	Citations	Downloads	Top3Fin	Citations	Downloads	Top3Fin	Citations	Downloads	Top3Fin
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
NewTopic × Sort Variable	0.085	-0.070	-0.067	0.262***	-0.015	0.006	-0.011	-0.016	0.171
	(1.33)	(-1.06)	(-0.51)	(4.86)	(-0.22)	(0.05)	(-0.18)	(-0.30)	(1.29)
NewTopic	0.545***	0.578***	0.196**	0.523***	0.562***	0.175**	0.585***	0.554***	0.087
	(12.01)	(9.70)	(1.98)	(11.89)	(9.33)	(2.07)	(15.43)	(9.23)	(1.04)
LogInnovData × Sort Variable	0.165	-0.012	-0.336*	0.212*	0.052	0.053	0.183	0.167**	-0.051
	(1.03)	(-0.10)	(-1.68)	(1.95)	(0.57)	(0.26)	(1.51)	(2.27)	(-0.25)
LogInnovData	0.277***	0.284***	0.622***	0.278***	0.270***	0.414***	0.276***	0.211***	0.458***
	(3.94)	(4.96)	(4.34)	(5.31)	(4.29)	(2.91)	(3.86)	(3.69)	(3.48)
Prob_{DV} × Sort Variable	0.168***	-0.298***	-0.271	-0.065	0.159**	0.501***	-0.173**	0.072	0.516***
	(2.94)	(-5.15)	(-1.41)	(-0.67)	(2.20)	(2.59)	(-2.53)	(1.62)	(2.62)
Prob_{DV}	0.167**	-0.560***	-0.700***	0.193***	-0.707***	-1.092***	0.269***	-0.676***	-1.011***
	(2.75)	(-8.85)	(-3.78)	(3.47)	(-9.55)	(-6.38)	(4.54)	(-9.40)	(-6.06)
sim_{USE} × Sort Variable	1.072***	-0.484**	-0.021	-0.104	-0.845***	-2.619***	0.226	-0.035	-0.046
	(4.39)	(-2.19)	(-0.03)	(-0.22)	(-3.02)	(-3.30)	(0.85)	(-0.15)	(-0.06)
sim_{USE}	-1.299***	0.502**	2.882***	-0.832***	0.580***	3.892***	-0.974***	0.310*	2.693***
	(-5.56)	(2.35)	(4.96)	(-2.98)	(3.30)	(7.51)	(-3.88)	(1.82)	(5.42)
N Obs	52,497	52,497	52,497	52,497	52,497	52,497	52,497	52,497	52,497
$\bar{R}^2$	0.382	0.305		0.373	0.302		0.383	0.306	
Pseudo R^2			0.172			0.169			0.175
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 12: **Conference Acceptance and Research Outcome**

This table reports results from regressing major finance conference acceptance on the measures of novelty, conventionality, author-, and article-specific covariates, and research outcome on conference treatment based on a 1-to-1 propensity-matched sample (based on the model in Column 3). The novelty element measure is either a dummy variable for emerging topics (*NewTopic*) or the log number of novel databases used in an article (*LogInnovData*). $Prob_{DV}$ is the probability of an article belonging to the diverse field. The conventionality measure is the average semantic similarity with previous FEN articles (sim_{USE}). The vector of covariates, including logarithm of number of pages (*LogPage*), number of coauthors (*#Authors*), log number of databases used in the article (*LogData*), log number of total databases accessible for coauthors through WRDS (*LogWRDS*), top 20 research school dummy variable (*Top20School*), and average centrality of all coauthors (*AuthorCentrality*), as well as year and field fixed effects (FE). t-statistics (z-statistics) of the OLS (logit) regression coefficients are shown in parentheses. t-statistics are computed based on standard errors clustered at the year level. The number of observations (N Obs), adjusted R^2 (Pseudo R^2) for OLS (logit) regression are reported. The construction of the variables is presented in Appendix A.1.

	AFA	WFA	Both	Citations	Downloads	Top3Fin	Top5Econ	Top3Acc
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment				0.624***	0.445***	1.479***	0.383**	-1.778***
				(17.13)	(14.39)	(23.82)	(2.50)	(-9.03)
NewTopic	0.090	0.132	0.126*	0.546***	0.501***	0.167*	-0.302	0.356
	(1.14)	(1.43)	(1.94)	(6.32)	(9.10)	(1.69)	(-1.03)	(1.60)
LogInnovData	0.259**	0.503***	0.407***	0.333***	0.264***	0.317**	0.272	-0.266
	(2.00)	(3.33)	(3.76)	(3.98)	(4.85)	(2.09)	(0.61)	(-0.69)
Prob_{DV}	-0.307*	1.064***	0.276**	0.194*	-0.527***	-0.268	1.416***	-1.527**
	(-1.78)	(5.94)	(2.08)	(1.79)	(-6.05)	(-1.23)	(3.08)	(-2.34)
sim_{USE}	3.338***	4.760***	4.262***	-1.420***	-0.683	0.167	-1.665	3.623**
	(6.75)	(8.15)	(10.60)	(-3.71)	(-1.61)	(0.25)	(-0.99)	(2.14)
LogPage	2.162***	2.499***	2.458***	0.646***	0.503***	1.824***	0.627*	0.876**
	(23.90)	(23.79)	(33.24)	(8.10)	(6.45)	(13.07)	(1.86)	(2.44)
#Authors	-0.055**	-0.131***	-0.101***	0.118***	0.048***	-0.013	0.002	0.341***
	(-2.20)	(-4.50)	(-5.00)	(5.81)	(3.14)	(-0.36)	(0.02)	(3.83)
LogData	0.284***	0.099***	0.228***	0.169***	0.201***	0.097**	-0.405***	0.417***
	(8.50)	(2.61)	(8.55)	(5.73)	(9.37)	(2.24)	(-3.84)	(3.26)
LogWRDS	0.237***	0.204***	0.230***	0.092***	0.053***	0.084**	0.804***	0.674***
	(9.49)	(7.08)	(11.71)	(3.61)	(3.03)	(2.41)	(4.39)	(3.98)
Top20School	0.555***	0.597***	0.606***	0.441***	0.054*	0.369***	0.888***	0.234
	(11.25)	(10.18)	(15.17)	(11.24)	(1.98)	(5.52)	(4.19)	(1.33)
AuthorCentrality	25.336***	19.927***	26.842***	17.362***	10.986***	15.506***	-10.138*	-3.383
	(15.51)	(10.29)	(18.98)	(16.13)	(9.80)	(7.29)	(-1.68)	(-0.64)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N Obs	52,497	52,497	52,497	7,798	7,798	7,798	7,798	7,798
$\bar{R}^2$				0.424	0.264			
Pseudo R^2	0.125	0.114	0.147			0.122	0.130	0.190

Table 13: **Novelty, Conventionality, and Media Attention**

This table reports results from regressing news (*LogNews*) and social media (*LogSocial*) mentions on the measures of novelty, conventionality, author-, and article-specific covariates. $Prob_{DV}$ is the probability for an article to be outside the five traditional finance focus fields. The novelty element measure is either a dummy variable for emerging topics (*NewTopic*) or the log number of novel databases used in an article (*LogInnovData*); the conventionality measure is the average semantic similarity with previous FEN articles (sim_{USE}). The vector of covariates, including logarithm number of pages (*LogPage*), number of coauthors (*#Authors*), log number of databases used in the article (*LogData*), log number of total databases accessible for coauthors through WRDS (*LogWRDS*), top 20 research school dummy variable (*Top20School*), and average centrality of all coauthors (*AuthorCentrality*), as well as year and field fixed effects (FE). t-statistics of the regression coefficients are shown in parentheses and are computed based on standard errors clustered at the year level. The number of observations (N Obs) and adjusted R^2 are reported. The construction of the variables is presented in Appendix A.1.

	News Mentions				Social Media Mentions			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
NewTopic	0.020**			0.020**	0.141***			0.139***
	(2.44)			(2.44)	(3.60)			(3.58)
LogInnovData		0.006		0.004		0.069*		0.058*
		(0.45)		(0.32)		(2.04)		(1.73)
Prob_{DV}			-0.004	-0.004			0.035	0.031
			(-0.50)	(-0.59)			(0.95)	(0.88)
sim_{USE}	-0.149***	-0.150***	-0.152***	-0.149***	-0.719***	-0.730***	-0.726***	-0.707***
	(-4.33)	(-4.29)	(-4.45)	(-4.45)	(-5.91)	(-5.75)	(-6.01)	(-6.08)
LogPage	0.026***	0.026***	0.026***	0.026***	0.108***	0.109***	0.109***	0.108***
	(5.94)	(5.97)	(5.97)	(5.94)	(6.44)	(6.44)	(6.48)	(6.41)
Author#	0.001	0.001	0.001	0.001	0.004*	0.004	0.004*	0.004*
	(0.86)	(0.81)	(0.82)	(0.85)	(1.94)	(1.72)	(1.76)	(1.91)
LogData	0.013***	0.014***	0.014***	0.013***	0.036***	0.036***	0.038***	0.035***
	(6.55)	(6.30)	(6.45)	(6.15)	(5.17)	(4.85)	(5.44)	(4.72)
LogWRDS	0.001	0.001	0.001	0.001	0.007**	0.007**	0.007**	0.007**
	(1.10)	(1.13)	(1.17)	(1.13)	(2.24)	(2.28)	(2.31)	(2.25)
Top20School	0.022***	0.022***	0.022***	0.022***	0.069***	0.071***	0.070***	0.068***
	(5.59)	(5.61)	(5.62)	(5.60)	(6.09)	(6.10)	(6.02)	(6.01)
AuthorCentrality	0.455***	0.463***	0.461***	0.451***	1.758***	1.806***	1.846***	1.775***
	(3.23)	(3.26)	(3.26)	(3.21)	(3.67)	(3.83)	(3.89)	(3.72)
Constant	-0.049***	-0.048***	-0.047***	-0.048***	-0.069	-0.061	-0.072	-0.077
	(-3.13)	(-3.15)	(-3.14)	(-3.10)	(-1.43)	(-1.30)	(-1.48)	(-1.53)
N Obs	52,497	52,497	52,497	52,497	52,497	52,497	52,497	52,497
$\bar{R}^2$	0.022	0.022	0.022	0.022	0.084	0.082	0.081	0.084
Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 14: **Novelty, Conventionality, and Publications**

This table shows coefficient estimates from multinomial logit regressions for the probabilities of publication in a top journal across different disciplines on measures of novelty, conventionality, author-, and article-specific covariates. The baseline outcome variable is the papers that are not published in the top 3 finance, top 5 economics, or top 3 accounting journals in Model 1, or in the top 5 finance, top 5 economics, top 5 accounting or top 3 interdisciplinary journals in Model 2. $Prob_{DV}$ is the probability for an article to be outside the five traditional finance focus fields. The novelty element measures are a dummy variable for emerging topics ($NewTopic$) and the log number of novel databases used in an article ($LogInnovData$); the conventionality measure is the average semantic similarity with previous FEN articles (sim_{USE}). The vector of covariates, including logarithm of number of pages ($LogPage$), number of coauthors ($\#Authors$), log number of databases used in the article ($LogData$), log number of total databases accessible for coauthors through WRDS ($LogWRDS$), top 20 research school dummy variable ($Top20School$), and average centrality of all coauthors ($AuthorCentrality$), as well as year and field fixed effects (FE). z-statistics of coefficients are shown in parentheses and are computed based on standard errors clustered at the year level. The number of observations (N Obs) and pseudo R^2 are reported. The construction of the variables is presented in Appendix A.1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Model 1			Model 2			
	Top3Fin	Top5Econ	Top3Acc	Top5Fin	Top5Econ	Top5Acc	Top3Int
NewTopic	0.189**	0.020	0.314**	0.131*	0.025	0.350***	0.314
	(2.49)	(0.11)	(2.30)	(1.76)	(0.12)	(3.37)	(1.57)
LogInnovData	0.441***	0.204	-0.159	0.386***	0.221	-0.208	0.655***
	(3.68)	(0.78)	(-0.78)	(3.55)	(0.86)	(-1.32)	(2.60)
Prob_{DV}	-0.851***	2.229***	-1.698***	-0.964***	2.145***	-1.802***	-1.298***
	(-4.60)	(9.58)	(-5.41)	(-6.07)	(9.33)	(-7.74)	(-4.24)
HighSim_{BOW}	-0.224**	-0.246	-0.121	-0.158*	-0.236	0.052	-0.008
	(-2.25)	(-1.03)	(-1.04)	(-1.94)	(-0.99)	(0.51)	(-0.05)
sim_{USE}	3.015***	-0.912	3.980***	2.890***	-0.805	4.473***	0.984
	(5.18)	(-1.17)	(5.78)	(5.11)	(-1.00)	(6.80)	(1.61)
Author2	0.216***	0.383***	-0.238**	0.257***	0.392***	-0.195*	0.174
	(2.96)	(3.27)	(-2.49)	(3.60)	(3.27)	(-1.86)	(1.39)
Author3	0.106	0.299***	0.153	0.214***	0.326***	0.331***	0.229*
	(1.34)	(3.44)	(1.36)	(2.84)	(3.62)	(2.90)	(1.83)
Author4	-0.125	0.025	0.067	0.006	0.057	0.387***	0.365**
	(-1.38)	(0.15)	(0.58)	(0.08)	(0.34)	(3.20)	(2.00)
LogPage	3.007***	2.558***	2.005***	2.919***	2.631***	1.921***	0.176
	(19.57)	(11.24)	(22.15)	(18.86)	(11.42)	(25.18)	(0.83)
LogData	0.287***	-0.243***	0.406***	0.326***	-0.224***	0.409***	-0.005
	(13.86)	(-3.92)	(5.06)	(14.99)	(-3.60)	(5.45)	(-0.14)
LogWRDS	0.214***	0.265***	0.506***	0.150***	0.267***	0.421***	0.161***
	(8.36)	(4.83)	(9.23)	(6.78)	(4.91)	(12.46)	(4.85)
Top20School	0.694***	1.429***	0.816***	0.533***	1.436***	0.744***	0.360***
	(13.81)	(11.05)	(10.76)	(10.94)	(11.19)	(10.54)	(3.44)
AuthorCentrality	27.924***	-4.509	3.381	30.520***	-2.648	-0.111	19.487***
	(9.78)	(-0.97)	(1.33)	(8.72)	(-0.56)	(-0.04)	(3.38)
N Obs		52,497			52,497		
Pseudo R^2		0.196			0.180		

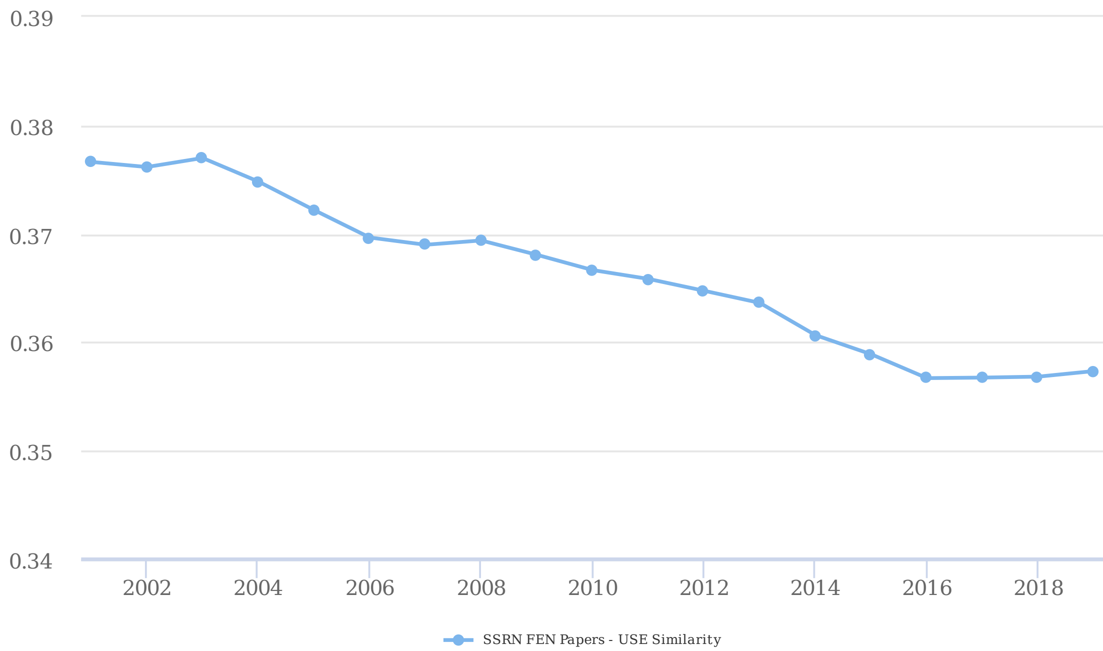
Internet Appendix (IA)

This Internet Appendix provides additional tables and figures for the paper, *Dissemination, Publication, and Impact of Finance Research: When Novelty Meets Conventionality.* Below we summarize the contents of this appendix.

Figure IA.1: **Average Similarity (USE), Conventionality, over Years**

Panel A reports year-by-year average USE abstract similarity among FEN articles with all FEN articles available 5 years prior. Panel B reports year-by-year average USE abstract similarity among the articles published in the Journal of Finance, Journal of Financial Economics, and Review of Financial Studies with all papers published in these 3 journals over 5 previous years. All the values in Figures A and B are smoothed using 3-year centered moving average.

A: Average Conventionality among FEN Papers



B: Average Conventionality among Papers Published in Top Three Finance Journals

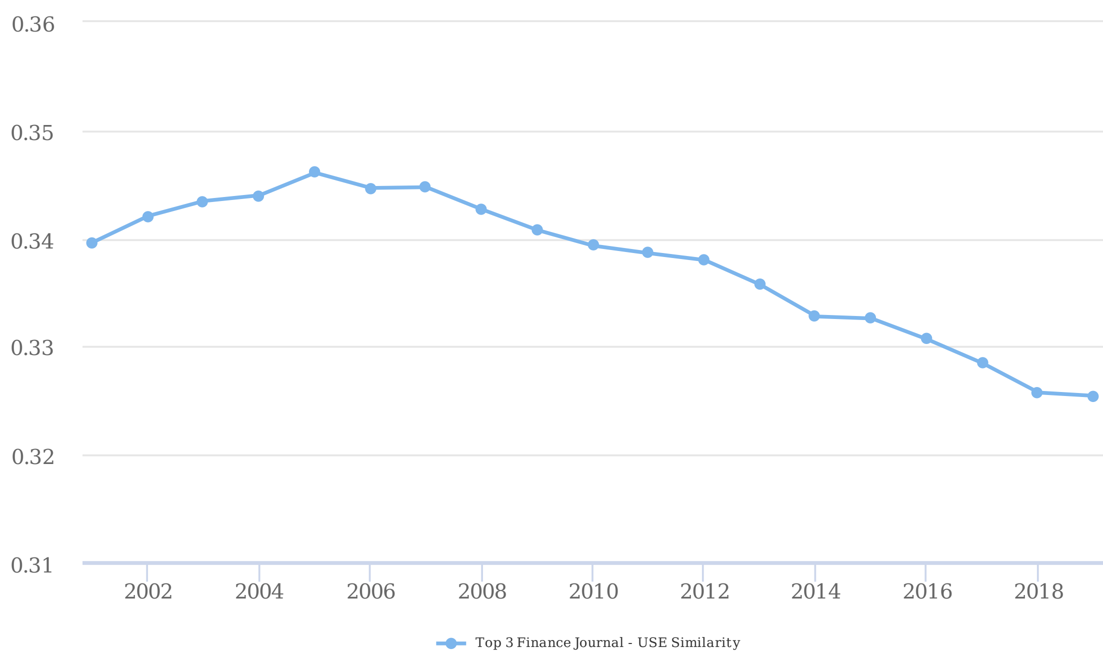


Figure IA.2: **Word Cloud of Each Field**

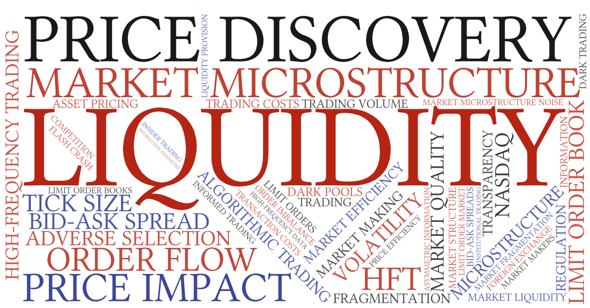
This figure reports the keywords reported in the articles that are posted on SSRN FEN. Each figure is constructed by the top 100 keywords in each field determined by a classifier described in the main text.

(a) Asset Pricing

(b) Corporate Finance

(c) Financial Intermediation

(d) Investment



(e) Market Microstructure

(f) Diverse Field

Figure IA.3: **Emerging Topics across Different Fields**

Panel A of this figure reports the number of articles categorized as Environmental, Social, and Corporate Governance (ESG) research by a classification model supervised by the title and abstracts of ESG papers suggested by the AFA/WFA session title and the top 3 journals' keywords. Panel B(C) plots the same figures for Macro Finance (Behavioral Finance) articles. Each year, we count the number of articles assigned at least 80% probability to belong to a topic while having at least 50% chance to be one of six research fields.

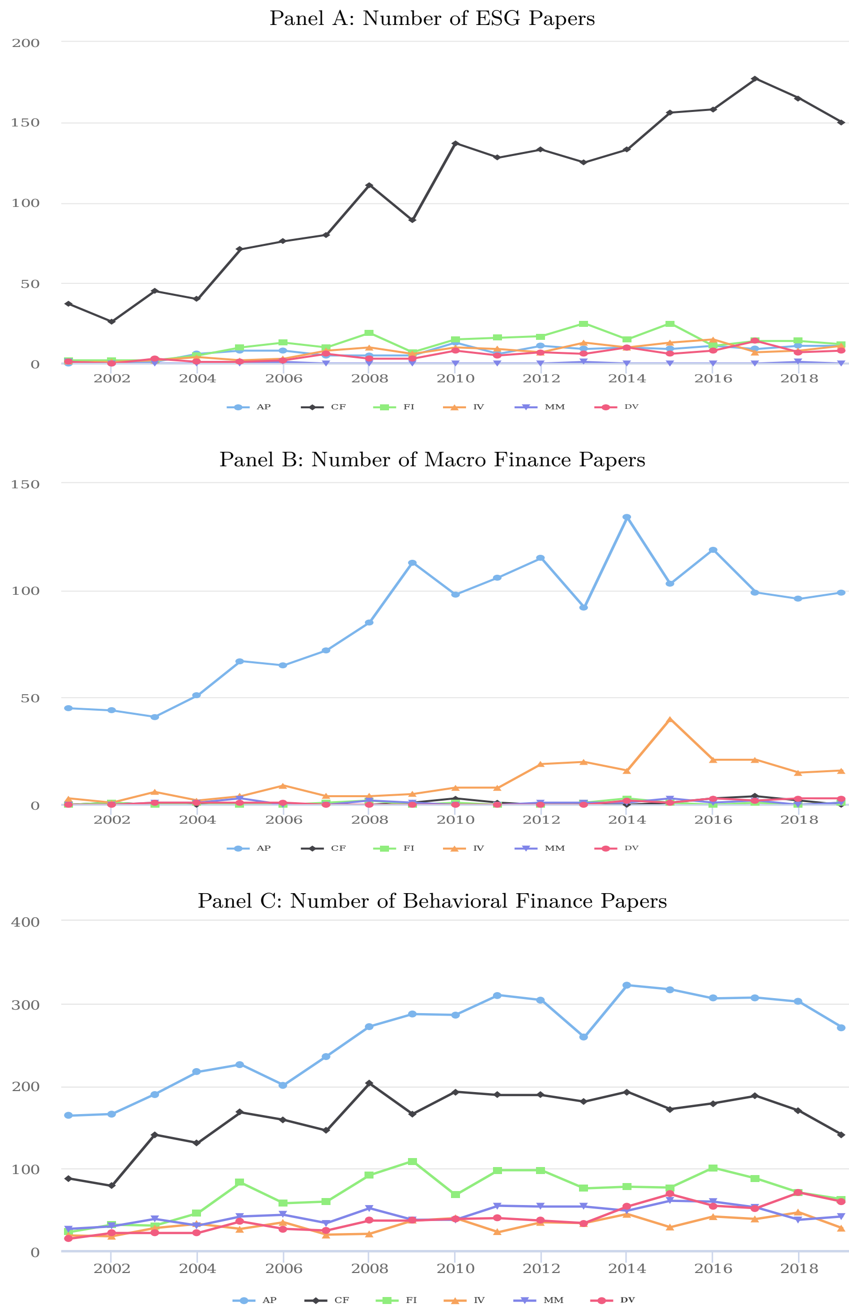


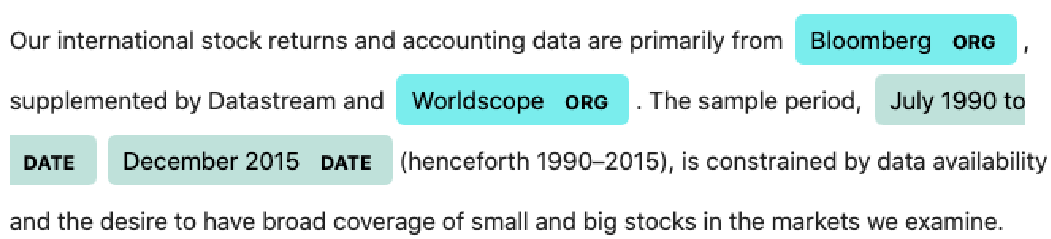
Figure IA.4: **Database Name Entity Recognition**

Panel A of this figure demonstrates databases that are identified through the text files of two Fama and French 5-factor articles, Fama and French (2016) and Fama and French (2017). Each database entity is collected through a keyword matching method and/or a supervised name entity recognition model, and is assigned a probability that the potential entity is mentioned in a context referring to databases. Panel B and Panel C illustrate the name entity recognition results from a pretrained model and a trained model respectively.

Panel A: Illustration of Database Entity Retrieval

Location: # Sentence	Possibility of a Data Sentence	Database
	Fama & French (2016)	
85	0.91	CRSP
85	0.91	Compustat
528	0.98	Compustat
	Fama & French (2017)	
49	0.98	Bloomberg
49	0.98	Datastream
49	0.98	Worldscope
150	0.50	Fama and French
472	0.76	Fama and French

Panel B: Pretrained Name Entity Recognition Model



Panel C: Trained Database Entity Recognition Model

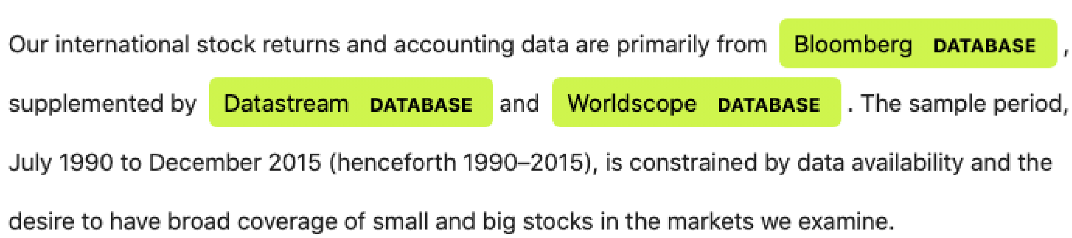


Figure IA.5: **Emerging Topics Through Years**

2001

Active Management
Basel
Behavioral Bias
Bond Risk
Common Knowledge
Corporate Credit
Crash Risk
Crossborder Merger
Crosslisted Firm
Data Snooping
Electronic Market
Enron
Equity Compensation
ETF
Ethic
Experimental Study
Female
Financial Advice
Financial Friction
Forecasting Performance
Foreign Affiliate
Gravity
Governance Index
Implied Cost
Information Diffusion
Information Leakage
Insolvency Risk
Interchange Fee
Limited Attention
Liquidity Measure
Microstructure Noise
Moneyness
Mood
Negative Externality
Newspaper
Option Trading
Pollution
Realized Volatility
Regulation FD
Religion
Rent Extraction
Riskshifting
Sanction
SNB
Social Interaction
Social Responsibility
Target Price
Tax Avoidance
Weather
WTO

2002

Analyst Behavior
Analyst Recommendations
Bubble Period
Cash Holdings
CEO Overconfidence
Collateral Constraint
Credit Score
Data Envelopment
Distancetodefault
Diversify Merger
DSGE
Electronic Payment
Enterprise Risk
Familiarity
Family Firm
Family Ownership
Frontend Load
Fund Flows
Governance Quality
Investment Vehicle
Liberalisation
Local Bias
Lookahead Bias
Managerial Power
Media
Misreporting
PIN
Price Synchronicity
Procyclicality
Quantile Regression
Real Earnings
REG FD
Risk Attitude
Sarbanes Oxley
Shortselling
Shorttermism
Sovereign Default
Stagger Board
Statistical Arbitrage
Strategic Default
Synchronicity
Temperature
Terrorist
TFP
Transportation
Women

2003

Accounting Conservatism
Accounting Quality
Accrual Anomaly
Agglomeration
Aggregate Volatility
Ambiguity Aversion
Asset Tangibility
Causal Effect
CEO Power
Content Analysis
Distress Risk
Dual-class Share
Entrepreneurial Finance
Financial Scandal
Immigration
Information Uncertainty
Institutional Quality
Intangible Capital
MBS
Medium Coverage
Misrepresentation
Natural Resource
Operational Risk
Overconfident CEOs
PEAD
Pension Scheme
Protection Law
Regulatory Arbitrage
Responsible Investing
Responsible Investment
Restatements
Shareholder Control
SIN
SOX
SRI
Stress Test
Tax Cut

2004

Bank Ownership
CDS Market
Climate Change
Contingent Capital
Corporate Fraud
Corporate Social
CSR
Donation
Endogeneity Concern
Environmental Factor
External Governance
Field Experiment
Imputed Equity
Inattention
Independent Directors
Investor Attention
Microfinance Institution
Money Illusion
News Disclosure
Non-GAAP Earnings
Nonperforming Loan
Overthecounter Market
Payout Decision
Peer Firm
Private Firms
Reporting Quality
Russian Crisis
Shareholder Litigation
Tax Haven
Tax Shelter
Unaffiliated Analyst
Unemployment Insurance

2005

ABS
Accrual Quality
Accrualbased Earnings
Accruals Quality
Backdating
Catering
CDS Spread
Control Weakness
Decision Theory
Discontinuity Design
Employee Share
FAS
Female CEO
Financial Covenant
Foreign Analyst
Gender Diversity
Identification Strategy
Individualism
Innovation Activity
Innovation Process
Inside Debt
Loss Recognition
Managerial Overconfidence
Material Weakness
Payday Lending
PIPE
Political Connection
Predatory Lending
Private Target
Regression Discontinuity
Scandals
SEO Discount
Shareholder Power
Social Connection
Sovereign CDS
Takeover Wave
Tax Enforcement
Treatment Effect
Variance Swap

2006

Carbon
Comment Letter
Crosslisting Premium
Disasters
Egalitarianism
Emission
Entry Deterrence
Goodwill Impairment
Household Finance
Housing Boom
IFRS Adoption
Labor Union
Lead Arranger
Loan Contracting
Loan Repayment
Option Backdating
Political Interest
Social Networks
VIX Future
Whistleblower

2007

Accrual Manipulation
Activist Hedge
Carry Trade
CAT Bond
CDS Trading
CMBS
Corrupt
CSR Disclosure
Derivative Litigation
Familiarity Bias
Financial Misconduct
Forecast Guidance
Fund Regulation
Funding Liquidity
Institutional Loan
Lifecycle Fund
Liquidity Regulation
Mandatory IFRS
MNES
Mortgage Crisis
Mortgage Securitization
Price Delay
Subprime Crisis
Tournament Incentive
Trademark
Uncertainty Avoidance
Volatility Smirk

2008

Aggressive Lending
Algorithmic Trading
Credit Crisis
Earnings Conference
Employee Relation
Female Director
Growth Anomaly
Housing Bubble
Implied Market
Lehman
Mortality Model
Negotiating
Network Analysis
P2P Lending
PCAOB
Peertopeer Lending
Placebo
Price Crash
Quasinatural Experiment
Regulation SHO
Relief Program
Religiosity
Renewable Energy
Rule 10b5
Sovereign Wealth (Fund)
SWF
TARP
Textual Analysis
Variance Premium

2009

Artificial Intelligence
Asset Relief
Central Clearing
Corporate Innovation
Counterparty Credit
Dark Pool
Gas Emission
Greenhouse Gas
Macroprudential Regulation
Quantitative Easing
Readability
Securitized Bond
Shadow Banking
Tax Risk
XBRL

2010

Anticorruption
Bribery
Common Ownership
Crowdfunding
Dodd-Frank Act
Ethanol
Eurozone Crisis
Fiscal Stimulus
Flash Crash
FMRI
Frequency Trading
HFT
High-Frequency Trader
Innovate
Linguistic Tone
LTG Forecast
Machine Learning
Newswires
Reversal Strategy
Social Medium
Violation Penalty

2011

ESG
Clawback Provision
Corporate Sustainability
CSR Performance
Customer Concentration
Dark Trading
Earningsbased Covenant
Fukushima
Government Guarantees
Millisecond
Mobile Money
Narcissism
Negative Interest
News Sentiment
Passive Investing
Shale
Text Mining
Timeseries Momentum
Uncertainty Shock

2012

Backer
Big Data
Dimension Reduction
EPU
Forward Guidance
Impact Invest
JOBS Act
LASSO
Societal Trust
Sustainability Disclosure

2013

Bank Levy
Banking Union
Bitcoin
Crowdfunding Platform
Cryptocurrency
Digital Currency
Factor Investing
Momentum Crash
Shareholder Engagement
Virtual Currency

2014

Bunching
Crypto
CSR Score
Deep Learning
Fintech
Political Corruption
Sentiment Analysis
Smart Beta

2015

Bitcoin Price
Blockchains
CEO Political
Cloud Computing
Cyber Insurance
Cyber Risk
Cybersecurity
ESG Rating
Fake News
Green Bond
Innovation Wave
Intraday Momentum
Language Processing
Press Conference
Reputational Damage
Staggered Implementation
Sustainable Investing

2016

Activist Shortselling
Brexit
Coin Offering
Cybersecurity Risk
Disclosure Doctrine
Distribute Ledger
ESG Incident
Gender Inequality
IDD
Inevitable Disclosure
Ledger Technology
Lending Club
Online Lending
Panama Papers
Pension Risk
Roboadvisors
Smart Contract
Synthetic Control
Wind Power

2017

Bitcoin Market
CEO-Employee Pay
Chinese Import
Demonetization
Donald Trump
Factor Zoo
Fintech Innovation
Fintech Lender
Import Penetration
Import Shock
Linguistic Distance
Medium Tone
Ownership Network
Regulatory Sandbox
Reporting Risk
Robo-advising
Sunset Provision

2018

Bitcoin Transaction
Carbon Pricing
Crypto Token
Direct Listing
Form AP
Judge Ideology
Populist
Redenomination Risk
Regulatory Ambiguity
Sexual Harassment
Stablecoins
Tokenization

2019

Autoencoder
Climate Sensitivity
Cyberattacks
Downside Variance
ESG Investment
GDPR
Green Asset
Libra
Loan Forgiveness
Patent Disclosure
Peer Disclosure
Platform Failure
Sustainability Practice
Token Offering
Unsupervised Machine

Figure IA.6: **Most Referenced MAG Topics**

This figure reports top topics cited by FEN articles. The topics are level-2 topics of referenced articles of FEN papers with the highest confidence score, both of which are determined by the proprietary NLP algorithm of MAG. The figure plots 999 topics with the most citations (out of 8,396).

Table IA.1: **A Comparison of USE and BOW Similarity**

This table reports USE and BOW similarity between two different versions of the abstract of the same article. The SSRN abstract is collected from SSRN at the beginning of May 2020 and journal abstract is obtained from WoS. The similarity measures are based on the title and abstracts of different version of same articles.

Panel A: Comparison of Different Versions of Same Papers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(2)-(5)
		sim_{USE}			sim_{BOW}			
Edit Distance	N	Mean	Median	St.Dev	Mean	Median	St.Dev	Diff
>5,000	1,670	0.884	0.900	0.077	0.781	0.817	0.155	0.104
>30,000	854	0.842	0.075	0.075	0.695	0.725	0.155	0.147
>45,000	430	0.825	0.837	0.074	0.659	0.691	0.159	0.166
>60,000	160	0.821	0.830	0.064	0.634	0.663	0.163	0.187

Panel B: Different Versions of Abstracts

SSRN Abstract	Published Abstract	USE	BOW
Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy (1257858)			
While a host of economic variables have been identified in the literature with the apparent in-sample ability to predict the equity premium, Goyal and Welch (2008) find that these variables fail to deliver consistent out-of-sample forecasting gains relative to the historical average. Arguing that substantial model uncertainty and instability seriously impair the forecasting ability of individual predictive regression models, we recommend combining individual model forecasts to improve out-of-sample equity premium prediction. Combining delivers statistically and economically significant out-of-sample gains relative to the historical average on a consistent basis over time. We provide two empirical explanations for the benefits of the forecast combination approach: (i) combining forecasts incorporates information from numerous economic variables while substantially reducing forecast volatility; (ii) combination forecasts of the equity premium are linked to the real economy.	Welch and Goyal (2008) find that numerous economic variables with in-sample predictive ability for the equity premium fail to deliver consistent out-of-sample forecasting gains relative to the historical average. Arguing that model uncertainty and instability seriously impair the forecasting ability of individual predictive regression models, we recommend combining individual forecasts. Combining delivers statistically and economically significant out-of-sample gains relative to the historical average consistently over time. We provide two empirical explanations for the benefits of forecast combination: (i) combining forecasts incorporates information from numerous economic variables while substantially reducing forecast volatility; (ii) combination forecasts are linked to the real economy.	0.95	0.79
Is There Price Discovery in Equity Options? (1683587)			
This paper presents direct evidence that option price quotes do not contain any information about future stock prices beyond what is already reflected in current stock prices. We use trade and quote data for 39 liquid U.S. stocks and ETFs and options on them, and focus on events when the two markets disagree about the stock price in the sense that the option-implied stock price obtained from the put-call parity relation is inconsistent with the actual stock price. In these disagreement events the options market adjusts bid and ask prices to eliminate the disagreement, while the stock market behaves normally, as if there were no disagreement. The disagreement events are typically precipitated by stock price moves, and often exhibit signed option volume providing pressure to eliminate the mispricing. These results are consistent with the hypothesis that option price quotes do not participate in the price discovery process for the underlying stock price, and stand in contrast to much of the existing literature.	We use tick-by-tick quote data for 39 liquid US stocks and options on them, and we focus on events when the two markets disagree about the stock price in the sense that the option-implied stock price obtained from the put-call parity relation is inconsistent with the actual stock price. Option market quotes adjust to eliminate the disagreement, while the stock market quotes behave normally, as if there were no disagreement. The disagreement events are typically precipitated by stock price movements and display signed option volume in the direction that tends to eliminate the disagreements. These results show that option price quotes do not contain economically significant information about future stock prices beyond what is already reflected in current stock prices, i.e., no economically significant price discovery occurs in the option market. We also find no option market price discovery using a much larger sample of disagreement events based on a weaker definition of a disagreement, which verifies that the findings for the primary sample are not due to unusual or unrepresentative market behavior during the put-call parity violations.	0.90	0.73
Going Public to Acquire? The Acquisition Motive in IPOs (1153508)			
Using a sample of IPOs from 1994 to 2004, we show that newly public firms make acquisitions at a torrid pace. This acquisition activity is fueled not only through the initial IPO proceeds, but also through the creation of an acquisition currency that is used to raise capital for both cash and stock financed acquisitions and through debt issuance subsequent to the IPO. The IPO allows companies to use potentially overvalued stock to pay for acquisitions, but also facilitates M&A by resolving uncertainty about the true value of the acquiror. We show that acquisitions play as significant a role in the growth of newly public firms as do R&D and CAPEX outlays. The patterns of acquisition activity following an IPO are important in explaining the evolution of ownership structure of newly public firms.	Newly public firms make acquisitions at a torrid pace. Their large acquisition appetites reflect the concentration of initial public offerings (IPOs) in mergers and acquisitions (M&A) intensive industries, but acquisitions by IPO firms also outpace those by mature firms in the same industry. IPO firms' acquisition activity is fueled by the initial capital infusion at the IPO and through the creation of an acquisition currency used to raise capital for both cash- and stock-financed acquisitions along with debt issuance subsequent to the IPO. IPO firms play a bigger role in the M&A process by participating as acquirers than they do as takeover targets, and acquisitions are as important to their growth as research and development (R&D) and capital expenditures (CAPEX). The pattern of acquisitions following an IPO shapes the evolution of ownership structure of newly public firms.	0.88	0.60

Table IA.1 Comparison of USE and BOW Similarity – Continued

SSRN Abstract	Published Abstract	USE	BOW
Stock Option Vesting Conditions, CEO Turnover, and Myopic Investment (1707539)			
This paper analyzes the optimal design of stock option vesting conditions when the CEO faces a risk of being replaced at an interim date. First, I show that long vesting terms do not necessarily discourage but in fact can encourage short-termism. Second, the model demonstrates that the optimal vesting schedule involves balancing incentives for managerial effort with incentives for long-term investment. Due to this trade-off, overinvestment in myopic projects can arise from optimal contracting and is not necessarily an artifact of faulty pay arrangements. The study generates new empirical predictions regarding the determinants and impacts of stock option vesting terms in contract design.	Corporations have been criticized for providing executives with excessive incentives to focus on short-term performance. This paper shows that investment in short-term projects has beneficial effects in that it provides early feedback about Chief Executive Officer (CEO) talent, which leads to more efficient replacement decisions. Due to the threat of CEO turnover, the optimal design of stock option vesting conditions in executive compensation is more subtle than conventional views suggest. For example, I show that long vesting periods can backfire and induce excessive short-term investments. The study generates new empirical predictions regarding the determinants and impacts of stock option vesting terms in optimal contracting.	0.81	0.18
Banking Globalization and Monetary Transmission (1162253)			
The globalization of banking in the United States is influencing the monetary transmission mechanism both domestically and in foreign markets. Using quarterly information from all U.S. banks filing call reports between 1980 and 2006, we show that globalized banks activate internal capital markets with their overseas affiliates to insulate themselves partially from changes in domestic liquidity conditions. The existence of these internal capital markets directly contributes to an international propagation of domestic liquidity shocks to lending by affiliated banks abroad. While these results imply a substantially more active lending channel than documented in Kashyap and Stein (2000), they also imply that the lending channel within the United States is declining in strength as banking becomes more globalized and monetary transmission abroad likewise increases in strength.	Globalization of banking raises questions about banks liquidity management, their response to liquidity shocks, and the potential for international shock propagation. We conjecture that global banks manage liquidity on a global scale, actively using cross-border internal funding in response to local shocks. Having global operations insulates banks from changes in monetary policy, while banks without global operations are more affected by monetary policy than previously found. We provide direct evidence that internal capital markets are active in global banks and contribute to the international propagation of shocks. This feature was at play during the financial crisis of 20072009.	0.76	0.36
A Market-Clearing Role for Inefficiency on a Limit Order Book (894121)			
Using a stochastic sequential game, this paper models limit order book trading dynamics. It deduces ex ante surplus and some agents' strategies by using an intuitive stationarity property of equilibrium. This largely bypasses any need for direct analysis of agents' (traders') intricate forecasting problems. Surplus per agent, while decreasing in the bid-ask spread, is invariant to some interesting dynamic features of the model. One interpretation of this is that market inefficiency is fixed by the spread at a 'liquidity market-clearing' level. The model best describes cases where price-discreteness leads to a mainly-constant spread. Here, a smaller price tick size raises surplus, and is to be encouraged.	Limit order markets with stationary dynamics attract equal volumes of market orders and uncanceled limit orders, equalizing the supply and demand for liquidity and immediacy. To maintain this balance, market orders must share any benefit obtained by limit order traders from more efficient trading conditions, such as better order queuing policies. Therefore an efficient market places a low price on immediacy, producing small bid-ask spreads. Furthermore, when price-discreteness leads to a mainly constant spread, cutting the price tick raises surplus. This is modeled with a stochastic sequential game, using stationarity considerations to bypass direct analysis of traders' intricate market forecasts.	0.70	0.44
Funding Growth in Bank-Based and Market-Based Financial Systems: Evidence from Firm-Level Data (632503)			
How the relative development of a country's stock market and banking system affects firms' growth is closely tied to how well developed the country`s contracting environment is. How differences in the contracting environment affect the relative development of the stock market or banking system may have implications for which firms and which projects get financing.	We investigate whether firms' access to external financing to fund growth differs in market-based and bank-based financial systems. Using firm-level data for 40 countries, we compute the proportion of firms in each country relying on external finance and examine how that proportion differs across financial systems. We find that the development of a country's legal system predicts access to external finance, and stock markets and the banking system affect access to external finance differently. However, we find no evidence that firms' access to external financing is predicted by several proxies for relative development of stock markets to the development of the banking system.	0.65	0.60
Digesting Anomalies: An Investment Approach (2152674)			
Motivated from investment-based asset pricing, we propose a new factor model that consists of the market factor, a size factor, an investment factor, and a return-on-equity factor. The new model [i] outperforms the Carhart (1997) four-factor model in pricing portfolios formed on earnings surprise, idiosyncratic volatility, financial distress, equity issues, as well as on investment and return-on-equity; [ii] performs similarly as the Carhart model in pricing portfolios on momentum as well as on size and book-to-market; but [iii] underperforms in pricing the total accrual deciles. Our model's performance, combined with its clear economic intuition, suggests that it can serve as a new workhorse model for academic research and investment management practice.	An empirical -factor model consisting of the market factor, a size factor, an investment factor, and a profitability factor largely summarizes the cross section of average stock returns. A comprehensive examination of nearly 80 anomalies reveals that about one-half of the anomalies are insignificant in the broad cross section. More importantly, with a few exceptions, the q-factor model's performance is at least comparable to, and in many cases better than that of the Fama-French (1993) 3-factor model and the Carhart (1997) 4-factor model in capturing the remaining significant anomalies.	0.59	0.19

Table IA.2: **An Illustration of Paper Classifications**

This table demonstrates the two extreme cases (highest vs lowest probability) of each field from the out-of-sample classification of all 144 papers selected into the Western Finance Association 2020 annual meeting.

Rank	Paper	Category Possibility					
		AP	CF	FI	IV	MM	OT
	Asset Pricing (AP) – 38 papers						
Highest	**Asset Pricing with Heterogeneous Agents and Long-Run Risk** *Abstract: This paper shows that belief differences have strong effects on asset prices in consumption-based asset-pricing models with long-run risks…*	99.88%	0.06%	0.01%	0.01%	0.00%	0.18%
Lowest	**Attention and Biases: Evidence from Tax-Inattentive Investors** *Abstract: Using bunching induced by a policy notch for identification, we provide evidence of investor inattention to a very simple and well-known capital-gains tax exemption in the Brazilian stock market…*	22.12%	15.20%	0.39%	16.31%	6.51%	17.75%
	Corporate Finance (CF) – 35 papers						
Highest	**Technology Development and Corporate Mergers** *Abstract*: I examine the motives as well as consequences of merger-and-acquisition (M&A) transactions between companies with varying degrees of technological overlap…	0.02%	99.84%	0.00%	0.00%	0.03%	0.07%
Lowest	**Capital Composition and Productivity Under Uncertainty: Evidence from the Shipping Industry** *Abstract: We study how economic uncertainty affects corporate asset composition and productivity using near-universe data on shipping firms' new orders, secondary-market transactions, and demolition of ships...*	0.78%	45.08%	0.24%	2.17%	3.35%	30.43%
	Financial Intermediation (FI) – 18 papers						
Highest	**Financing Competitors** *Abstract: This paper studies banks' lending to shadow banks and its impact on mortgage market competition. I collect shadow bank call reports…*	0.26%	0.02%	99.94%	0.00%	0.00%	0.76%
Lowest	**A Fuzzy Bunching Estimator of Regulatory Costs** *Abstract: We propose a revealed preference approach to estimate the costs of financial regulation... Applying this estimator to banks' asset size distribution after the Dodd--Frank Act…*	31.35%	1.46%	71.19%	2.89%	0.76%	13.88%

Table IA.3 An Illustration of Paper Classifications – Continued

Rank	Paper	Category Possibility					
		AP	CF	FI	IV	MM	OT
	Investment (IV) – 20 papers						
Highest	**A Horizon Based Decomposition of Mutual Fund Performance using Transaction Data** *Abstract: This paper uses transaction-level data to decompose active mutual fund performance based on the past length of funds' holdings…*	0.26%	0.36%	0.02%	99.98%	0.02%	0.17%
Lowest	**Debt versus Equity in Liquidity Provision** *Abstract: We propose a unified framework to study liquidity provision by debt-issuing versus equity-issuing financial intermediaries... We find that, at the end of 2017, liquidity provision by a dollar of bond mutual fund shares amounts to a quarter of that by a dollar of uninsured bank deposits...*	11.10%	1.15%	36.96%	51.50%	0.31%	22.65%
	Market Microstructure (MM) – 5 papers						
Highest	**Effects of a Speed Bump on Market Quality and Exchange Competition** *Abstract: After a long period of facilitating faster trading, exchanges are now trying to slow down trading with speed bumps…*	0.05%	4.35%	0.19%	0.06%	99.75%	0.72%
Lowest	**Liquidity in the Cross Section of OTC Assets** *Abstract: We construct a dynamic model of a multi-asset over-the-counter (OTC) market that operates via search and bargaining and empirically test its implications using data from the US corporate bond market. ..*	40.23%	2.14%	0.81%	2.85%	76.14%	19.86%
	Diversified Field (DV) – 28 papers						
Highest	**Heterogeneous Real Estate Agents and the Housing Cycle** *Abstract: The real estate market is highly intermediated, with 90 percent of buyers and sellers hiring an agent to help them transact a house….*	9.89%	4.57%	8.16%	0.83%	2.35%	93.09%
Lowest	**Keeping Options Open: What Motivates Entrepreneurs?** *Abstract: Using French administrative data on job-creating entrepreneurs, I estimate a life-cycle model in which risk-averse individuals can start businesses and return to paid employment….*	28.72%	10.69%	2.77%	9.31%	0.46%	40.14%

Table IA.3: **Conventionality and Research Outcomes (Median Regression)**

This table reports quantile regression results from regressing research outcome, measured by the number of citations, and the number of downloads on the measures of conventionality, author-, and article-specific covariates. sim_{USE} is the average of semantic similarity with previous FEN articles, and sim_{BOW} is the average of *bag-of-word* similarity with previous FEN articles, and $X_{i,k,t}$ is a vector of covariates, including logarithm of number of pages ($LogPage$), number of coauthors ($\#Authors$), log of number of databases used in the article ($LogData$), log of number of total databases accessible for coauthors through WRDS ($LogWRDS$), top 20 research school dummy variable ($Top20School$), and average centrality of all coauthors ($AuthorCentrality$), as well as year and field fixed effects (FE). t-statistics of the quantile regression coefficients are shown in parentheses and are computed based on robust estimator of variance. The number of observations (N Obs), Pseudo R^2 for OLS qantile regression are reported. The variable construction is presented in Appendix A.1.

	Citations				Downloads			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
sim_{USE}	-1.684***	-1.220***			0.798***	0.484***		
	(-10.75)	(-7.45)			(7.29)	(4.31)		
sim_{BOW}			-2.500***	-2.023***			0.960***	0.048
			(-7.46)	(-6.03)			(4.51)	(0.23)
LogPage	0.603***	0.612***	0.602***	0.604***	0.654***	0.610***	0.658***	0.616***
	(19.78)	(20.10)	(19.44)	(20.00)	(31.39)	(29.33)	(32.12)	(30.15)
Author#	0.161***	0.163***	0.160***	0.160***	0.038***	0.037***	0.039***	0.036***
	(19.76)	(19.82)	(19.10)	(19.56)	(6.71)	(6.50)	(6.99)	(6.38)
LogData	0.113***	0.110***	0.101***	0.105***	0.210***	0.190***	0.214***	0.196***
	(10.30)	(9.90)	(9.03)	(9.50)	(27.97)	(25.41)	(28.94)	(26.56)
LogWRDS	0.049***	0.052***	0.051***	0.054***	0.029***	0.028***	0.028***	0.026***
	(8.10)	(8.67)	(7.91)	(8.77)	(7.59)	(7.30)	(7.63)	(7.15)
Top20School	0.545***	0.538***	0.541***	0.531***	0.213***	0.221***	0.209***	0.218***
	(28.61)	(27.98)	(28.10)	(28.15)	(16.19)	(16.67)	(16.00)	(16.46)
AuthorCentrality	25.017***	25.309***	24.634***	25.026***	17.938***	17.453***	17.939***	17.645***
	(27.88)	(28.73)	(26.15)	(27.55)	(35.74)	(31.93)	(38.34)	(31.27)
Nobs	52,497	52,497	52,438	52,438	52,497	52,497	52,438	52,438
Pseudo R^2	0.234	0.237	0.233	0.237	0.180	0.191	0.180	0.191
Field FE	No	Yes	No	Yes	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table IA.4: **Conventionality, Novelty, and Research Impact**

This table reports regression results from regressing logarithm value of MAG Saliency, an article level research impact measure, and dummy variables indicating whether an article are ranked as Top 2000, 1000, or 500 base on MAG Saliency in our sample. The sim_{USE} is the average of semantic similarity with previous FEN articles. NewTopic is a dummy variable for emerging topics, LogInnovData number of novel databases used in an article, and Prob_{DV} is the probability for an article belonging to the diverse field. $X_{i,k,t}$ is a vector of covariates, including logarithm of number of pages (*LogPage*), number of coauthors (*#Authors*), log of number of databases used in the article (*LogData*), log of number of total databases accessible for coauthors through WRDS (*LogWRDS*), top 20 research school dummy variable (*Top20School*), and average centrality of all coauthors (*AuthorCentrality*), as well as year and field fixed effects (FE). t-statistics (z-statistics) of the OLS (logit) regression coefficients are shown in parentheses. t-statistics are computed based on standard errors clustered at the year level. The number of observations (N Obs), adjusted R^2 (Pseudo R^2) for OLS (logit) regression are reported. The variable construction is presented in Appendix A.1.

	MAG Saliency	Research Impact Rank		
		Top 2000	Top 1000	Top 500
	(1)	(2)	(3)	(4)
sim_{USE}	-1.211***	-1.264**	-2.144***	-1.809*
	(-7.44)	(-2.51)	(-3.12)	(-1.89)
NewTopic	0.466***	0.990***	1.081***	1.120***
	(16.28)	(15.39)	(13.15)	(10.19)
LogInnovData	0.325***	0.337***	0.396**	0.325
	(8.75)	(2.62)	(2.38)	(1.41)
Prob_{DV}	0.236***	0.205	0.314	0.435
	(5.71)	(1.14)	(1.28)	(1.29)
LogPage	0.425***	1.278***	1.269***	1.479***
	(7.20)	(12.82)	(9.20)	(7.71)
Author#	0.159***	0.033	0.005	-0.001
	(26.41)	(1.19)	(0.14)	(-0.02)
LogData	0.033	0.267***	0.259***	0.221***
	(1.51)	(7.21)	(5.02)	(3.09)
LogWRDS	0.123***	0.136***	0.167***	0.243***
	(20.41)	(5.83)	(4.74)	(4.33)
Top20School	0.717***	0.811***	0.945***	1.015***
	(34.06)	(14.44)	(11.82)	(8.89)
AuthorCentrality	18.124***	20.387***	22.173***	22.108***
	(23.82)	(12.30)	(10.74)	(8.33)
N Obs	49,479	52,497	52,497	45,405
$\bar{R}^2$	0.241			
Pseudo R^2		0.112	0.125	0.132
Field FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Table IA.5: **Novelty, Conventionality, and Publications (Logit)**

This table reports results from logit regressing dummy variables indicating top journal publications across different disciplines on the measures of novelty, conventionality, author-, and article-specific covariates. Prob_{DV} is the probability for an article to be outside the five traditional finance focus fields. The novelty element measure is either a dummy variable for emerging topics ($NewTopic$) or the log number of novel databases used in an article ($LogInnovData$); the conventionality measure is the average semantic similarity with previous FEN articles (sim_{USE}). The vector of covariates, including logarithm of number of pages (LogPage), number of coauthors (#Authors), log number of databases used in the article (LogData), log number of total databases accessible for coauthors through WRDS (LogWRDS), top 20 research school dummy variable (Top20School), and average centrality of all coauthors (AuthorCentrality), as well as year and field fixed effects (FE). z-statistics of regression coefficients are shown in parentheses. The number of observations (N Obs) and pseudo R^2 are reported. The construction of the variables is presented in Appendix A.1.

	Top3Fin	Top5Fin	Top10Fin	Top5Econ	Top3Acc	Top5Acc
	(1)	(2)	(3)	(4)	(5)	(6)
NewTopic	0.162**	0.086	0.031	-0.030	0.265**	0.304***
	(2.48)	(1.42)	(0.60)	(-0.19)	(2.38)	(3.26)
LogInnovData	0.445***	0.376***	0.156*	0.075	-0.316	-0.399**
	(4.25)	(3.86)	(1.80)	(0.24)	(-1.47)	(-2.16)
Prob_{DV}	-0.875***	-0.918***	-1.330***	2.391***	-1.664***	-1.699***
	(-5.80)	(-6.78)	(-11.93)	(9.82)	(-5.73)	(-7.15)
HighSim_{USE}	-0.210***	-0.153**	-0.105*	-0.197	-0.068	0.098
	(-2.65)	(-2.21)	(-1.93)	(-0.90)	(-0.47)	(0.91)
sim_{USE}	2.865***	2.610***	2.895***	-1.323	3.594***	4.026***
	(6.96)	(7.12)	(9.83)	(-1.63)	(5.05)	(6.88)
Author2	0.214***	0.254***	0.286***	0.369***	-0.271***	-0.239***
	(3.84)	(5.10)	(7.27)	(3.03)	(-2.65)	(-2.80)
Author3	0.083	0.170***	0.219***	0.288**	0.144	0.299***
	(1.45)	(3.31)	(5.39)	(2.23)	(1.47)	(3.67)
Author4	-0.137*	-0.042	0.191***	0.055	0.101	0.398***
	(-1.73)	(-0.60)	(3.46)	(0.30)	(0.79)	(3.88)
LogPage	2.847***	2.728***	1.803***	2.139***	1.503***	1.368***
	(36.40)	(39.16)	(32.99)	(13.61)	(10.81)	(12.11)
LogData	0.277***	0.308***	0.380***	-0.290***	0.384***	0.374***
	(9.91)	(12.32)	(18.86)	(-4.97)	(7.51)	(9.02)
LogWRDS	0.202***	0.133***	-0.014	0.246***	0.492***	0.406***
	(10.47)	(8.71)	(-1.42)	(4.86)	(9.89)	(11.45)
Top20School	0.617***	0.427***	0.005	1.333***	0.692***	0.616***
	(14.92)	(11.52)	(0.17)	(12.84)	(9.38)	(10.36)
AuthorCentrality	27.815***	30.062***	30.966***	-11.093***	-4.007	-9.722***
	(19.49)	(22.49)	(26.04)	(-2.75)	(-1.37)	(-3.76)
N Obs	52,497	52,497	52,497	52,497	52,497	52,497
Pseudo R^2	0.173	0.161	0.110	0.148	0.150	0.153
Field FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Table IA.6: **Similarity before and after Publication**

This table reports mean and median of demeaned ratio of average SIM_{USE} of the abstract of each published FEN paper with other FEN papers posted during the five-year period before (-5) and after (+5) the publication year (year 0). We focus on the FEN papers published from 2005 to 2014 to ensure a balanced sample period. The publication information of FEN papers is obtained from MAG. The ratio of SIM_{USE} is defined as

$$ratio = \frac{SIM_{USE}(1,5)}{SIM_{USE}(-5,-1)} - 1.$$

We then demean the ratio by the annual average of ratios in order to have a zero mean distribution. The *** indicates the *p-value* is less than 1% for the null hypothesis that the mean (median) in the top and bottom quintile are equal, where the *p-value* is calculated by *Student-t* (*Wilcoxon Rank*) tests.

Quintile	Nobs	Mean	Median
	(1)	(2)	(3)
Low	2,115	0.62%	0.61%
2	2,122	0.19%	0.16%
3	2,122	-0.05%	-0.08%
4	2,122	-0.26%	-0.32%
High	2,115	-0.50%	-0.59%
Diff		1.12%***	1.20%***

Table IA.7: **Research Fields and Hosting Journals**

This table reports results from logit regression regressing dummy variables indicating top journal publications across different disciplines on the measures of novelty, conventionality, author-, and article-specific covariates. Prob_{DV} is the probability for an article to be outside the five traditional finance focus fields. The novelty element measure is either a dummy variable for emerging topics (*NewTopic*) or the log number of novel databases used in an article (*LogInnovData*); the conventionality measure is the average semantic similarity with previous FEN articles (sim_{USE}). The vector of covariates includes logarithm of number of pages (LogPage), number of coauthors (#Authors), log number of databases used in the article (LogData), log number of total databases accessible for coauthors through WRDS (LogWRDS), top 20 research school dummy variable(Top20School), and average centrality of all coauthors (AuthorCentrality), as well as year and field fixed effects (FE). z-statistics of regression coefficients are shown in parentheses. The number of observations (N Obs) and pseudo R^2 are reported. The construction of the variables is presented in Appendix A.1.

	Top3Fin	Top5Fin	Top10Fin	Top5Econ	Top3Acc	Top5Acc
	(1)	(2)	(3)	(4)	(5)	(6)
Prob_{AP}	0.485***	0.521***	0.358***	-0.158	0.110	0.419
	(2.79)	(3.35)	(2.81)	(-0.43)	(0.33)	(1.59)
Prob_{CF}	-0.086	-0.028	0.210*	-0.652*	1.525***	1.897***
	(-0.50)	(-0.18)	(1.68)	(-1.80)	(4.86)	(7.52)
Prob_{FI}	0.022	-0.030	0.645***	0.199	-0.087	0.112
	(0.09)	(-0.13)	(3.55)	(0.46)	(-0.17)	(0.27)
Prob_{IV}	-0.072	0.108	0.103	-2.114***	0.982**	1.110***
	(-0.30)	(0.50)	(0.59)	(-3.39)	(2.27)	(3.14)
Prob_{MM}	-0.075	0.143	0.233	-0.451	0.838	1.433**
	(-0.23)	(0.50)	(1.02)	(-0.63)	(1.18)	(2.57)
Prob_{DV}	-0.866***	-0.887***	-1.302***	2.270***	-1.230***	-1.184***
	(-5.49)	(-6.26)	(-11.16)	(8.47)	(-4.11)	(-4.83)
N Obs	52,497	52,497	52,497	52,497	52,497	52,497
Pseudo R^2	0.170	0.158	0.106	0.148	0.148	0.150
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Table IA.8: **Years from SSRN Posting to Publication**

This table reports the numbers of years a paper took to be published in one of the journals reported in Table 1 since it is posted to SSRN from 2001 to 2019. We remove the papers posted at and after its publishing year. The posting year is the earliest year of posting, being written, or the first conference acceptance of a FEN article.

			Years to Publication				
Journal Name	Abb.	Papers	Mean	Std	P10	Median	P90
Panel A: Finance Journals							
Journal of Finance	JF	980	3.00	1.42	1	3	5
Journal of Financial Economics	JFE	1,502	2.72	1.45	1	2	5
Review of Financial Studies	RFS	1,172	3.22	1.53	1	3	5
Journal of Fin. & Quant. Analysis	JFQA	685	3.75	1.93	2	4	6
Journal of Financial Intermediation	JFI	250	3.13	1.78	1	3	5
Journal of Money, Credit and Banking	JMCB	210	3.13	1.73	1	3	5
Review of Finance	ROF	398	3.98	2.14	2	4	7
Journal of Banking & Finance	JBF	1,244	2.48	1.54	1	2	5
Journal of Corporate Finance	JCF	621	2.85	1.96	1	2	5
Financial Management	FM	227	3.01	1.81	1	3	5
Journal of Financial Markets	JFM	215	3.40	1.94	1	3	6
Review of Asset Pricing Studies	RAPS	54	3.37	1.81	1	3	6
Journal of Empirical Finance	JEF	296	2.89	1.81	1	2	5
Review of Corporate Finance Studies	RCFS	36	4.50	2.44	2	4	8
Panel B: Economics Journals							
Quarterly Journal of Economics	QJE	132	2.22	1.34	1	2	4
Journal of Political Economy	JPE	75	3.07	1.80	1	3	5
American Economic Review	AER	252	2.55	1.37	1	2	4
Econometrica	ETCA	90	2.90	1.77	1	3	5
Review of Economic Studies	RES	106	3.62	2.07	1	3	6
Panel C: Accounting Journals							
Journal of Accounting & Economics	JAE	287	2.43	1.34	1	2	4
Journal of Accounting Research	JAR	258	2.20	1.22	1	2	4
Accounting Review	TAR	321	2.65	1.68	1	2	5
Review of Accounting Studies	RAS	252	2.93	1.83	1	3	5
Contemporary Accounting Research	CAR	261	3.52	2.17	1	3	6
Panel D: Cross-Disciplinary Journals							
Journal of Intl. Business Studies	JIBS	48	2.73	1.95	1	2	6
Management Science	MS	537	4.00	2.14	1	4	7
Journal of Business Ethics	JBE	96	2.29	1.21	1	2	4