The Welfare Effects of Gender-Inclusive Innovation: Evidence from Patents and Academic Publications

Matt Marx
Cornell University & NBER

Frank Müller-Langer
University of the Bundeswehr Munich & MPI-IC

Joel Waldfogel
University of Minnesota, NBER, & ZEW

March 26, 2026

PRELIMINARY DRAFT

Abstract

Female participation in science and innovation has long lagged male participation; but in the past half century, the shares of patents and academic articles with mostly-female inventors and authors has risen from 5 to 10 percent for patents and from 5 to 20 percent for articles. This paper explores the effect of these female influxes on downstream citing inventor and author welfare. Doing so requires gender attribution for authors and inventors, and we use name-based algorithms and an LLM to classify the genders of 66.7 million authors of 65.2 million academic articles published between 1965 and 2015; and we use USPTO data for patents. We find that female articles and patents are cited roughly as much as work by men and – especially for articles – substantially more by women. We see no evidence of displacement of male by female innovation, and additional work in both areas "expands the market," suggesting its value. The articles available in 2015 raise inventor surplus by 5.1 percent relative to an environment without the post-1965 female influx, with male authors gaining 3.6 percent and female authors gaining 12.8 percent. Results for patents are similar but muted by the smaller influx. Overall and male inventor surplus rise by 1 percent, and female surplus rises by 1.7 percent.

Introduction

Innovation is essential to growth, so it would be valuable to attract additional qualified people to participation in scientific and innovative activity. Women have traditionally participated in innovative endeavors at depressed rates relative to men. In the past half century women have entered the labor forces of industrialized countries in great numbers, growing from about 30 percent of the US labor force in 1950 to approximately 50 percent in 2020 (Toossi, 2002; U.S. Bureau of Labor Statistics, 2022). Yet, women have been slower to move into technical, innovative, and creative occupations, with substantial potential costs to society. There is reason to believe that the low female participation has caused valuable innovations to have been forgone (Bell et al., 2019).

Female-authored articles accounted for less than five percent of the academic articles published in 1965, with much lower shares in STEM fields. Half a century later—in 2015—female-authored articles accounted for 20 percent of new articles. Between 1965 and 2015 the number of new academic articles published per year rose from 324,000 to 3.24 million. Of these, the number by women rose from almost none (14,000) to 652,000, or 20 percent of the total. The pattern of female inventor entry into patenting is similar but less pronounced. The share of patents with mostly-female inventors rose from 5 percent in 1975 to 10 percent in 2020. The goal of this paper is to use the growth in female participation in patenting and academic authorship – and the resulting impacts on the pool of citeable articles and patents – to measure the benefit of inclusive scientific innovation.

Most patents and articles receive few citations, so it is not self-evident that the growth in female contributions is valuable. There are two reasons, however, why a female influx could be valuable. First, the women entering science between 1965 and 2015 reflect a social liberalization that may have allowed high-ability women to join the innovation workforce. Second, the quality of both research and patents is highly unpredictable, so additional schol-

ars, patents, and articles are like "draws from an urn." The greater the unpredictability, the more benefit that an influx could bring (Aguiar and Waldfogel, 2018; Waldfogel, 2025).

A natural way to measure the value of patents and academic contributions is through the citations they eventually receive. We make use of this idea to measure the value of the female influx into innovative activity. In particular, we adopt an industrial organization approach. We treat the citing behavior of new inventors and authors as demand, and we treat the extant patents and articles as the choice set. We quantify the effect of the female influx into authorship by comparing the "consumer surplus" that authors derive from the contemporary choice set compared with a counterfactual in which female-authored patents and articles are removed to mimic the female share prior to 1970. We estimate the surplus both overall and by gender of citing authors and inventors.

We undertake this exercise using two broad datasets. For academic articles, we use the Microsoft Academic Graph (MAG) on the citations received by all 65.2 million journal articles published between 1965 and 2015 in 40 distinct academic areas. We are able to classify each article according to its authors' genders, and we know how many of the citations it received each year came from male vs female-authored citing papers. We use forename-country-based algorithms and an LLM to classify the genders of 66.7 million authors. For patents we use the gender-identified data on 7 million patents granted between 1975 and 2020 created by the US Patent and Trademark Office.

Our paper proceeds in three steps. First, we provide background information, based on education and achievement and standardized test scores, on the growing volume of women with preparation for careers in innovative occupation. Second, we explore a series of descriptive questions about each of the two – academic and patenting – domains. (i) We document the growing shares of female contributions; (ii) we explore the relationship between the female shares of patents and articles, by area, and the tendency for these works to receive citations; (iii) we look for evidence that female entry displaces male entry; and (iv) we ex-

plore whether additional patents or articles attract citations. Third, we use nested logit demand models to quantify the welfare impact of female entry into innovation; we compare the value of the end-of-sample choice sets that include the female influx with counterfactual alternatives that remove female works to bring the female shares to their values at the start of the sample.

We have five descriptive findings, and we provide measurements of the proportionate impact of the influx on the welfare of citing authors. First, there is a substantial female influx into academic authorship in the last half century. The share of articles with mostly-female authors rises from 5 percent in the 1965 vintage to 20 percent in 2015. Second, articles by women attract roughly as many citations, per article, as articles by men although women cite work by women substantially more. Third, increases in the female-authored share of articles, by field and vintage, have been accompanied by roughly proportional increases in the share of citations garnered by female-authored articles, which is consistent with the female influx delivering high-quality research. Fourth, we see no evidence of displacement of male-authored articles by female-authored articles across fields and vintages. Fifth, fields and years with more extant articles attract more citations, which is consistent with the additional articles being valuable. The female influx into academic research has raised overall author CS by 5.1 percent; female author CS has increased by 12.8 percent while male author CS has risen by 3.6 percent. The smaller female influx into patenting has had a commensurately smaller effect: Overall inventor surplus has risen by 1.02, male surplus by 0.96 percent, and female inventor surplus by 1.67 percent.

The impact of the influx depends, in part, on how male innovative activity might have differently proceeded in its absence. Our baseline approach assumes no displacement. We also entertain extreme alternatives in which the female influxes are replaced, one-for-one, with additional male articles and patents, drawn at random from the existing male distributions. In these extreme cases overall – and male – surplus falls slightly while female surpluses

rise.

The paper proceeds in five sections after the introduction. Section 1 provides background, including historical context and literature. Section 2 describes a theoretical setup that implies empirical questions and provides a framework for welfare analysis. Section 3 describes the data used in the study. Section 4 presents evidence on the female influx and its impact on citations. Section 5 presents estimates of the demand model and the counterfactual exercises showing the proportionate gain to authors, overall and by gender, from the female influx.

1 Background

1.1 Historical context: the supply of female innovators

Women's participation in science and engineering was limited in the early and mid-20th century; women's paid work was concentrated in a narrow set of low-skill occupations with little participation in technical careers (Costa, 2000; Goldin, 2014). For instance, women constituted between 0.3% and 1.2% of engineers from 1940 to 1950 with similarly small shares in related technical occupations (U.S. Women's Bureau and Pidgeon, 1954). In 1940, women accounted for only 0.8% of employed chemists and engineers (Blank and Stigler, 1957).

In 1940, less than one percent of US engineers were women, and the share remained at one percent for the following two decades. Perhaps not surprisingly, women were also absent from PhD programs in technical fields. In 1960, 6 of 959 US PhDs in engineering were awarded to women. In 1949, women earned 175 of 52,246 engineering bachelor's degrees, 15 of 4,496 engineering master's degrees, and just 1 of 417 engineering doctorates (Pidgeon, 1954).

Relatedly, from at least the 1960s onward, U.S. women represented a large latent pool of scientific potential: Altonji et al. (2016) suggests that ability distributions by themselves

do not explain women's under-representation in STEM occupations. Jiang (2021) finds that gender differences in ability and ability sorting explain only a small portion of the gender gap in STEM majors, implying that many women with STEM-relevant skills do not select into STEM fields. Finally, evidence from standardized math and science assessments shows only modest differences in quantitative performance, implying a substantial overlap in male and female quantitative aptitude distributions (Hyde and Linn, 2006; Hyde et al., 2008).

The absence of women in technical fields suggests a failure to exploit some innovative potential, and performance on standardized tests for graduate school provides some evidence. By the early 1980s, women were taking the Graduate Record Exam (GRE) in greater numbers than men. In 1982, 71,845 of the US test takers were men, while 95,235 were women. After 2000, the majority – over 60 percent – of test takers were women.

Data on Graduate Record Examination (GRE) scores are available by gender for US test takers back to 1982 (Bleske-Rechek and Browne, 2014). Although average scores for women are lower than for men, there is substantial dispersion; and both means and standard deviations are reported by gender for the math and verbal tests in some years. Assuming that the math and verbal scores are distributed normally, one can calculate the numbers of men and women with scores above any absolute level. Suppose that G_m and G_f are the numbers of men and women taking the exam, μ_m^v and σ_m^v are the average and standard deviation of the men's verbal score in a year, and Φ is the normal cumulative distribution function. Then the number of women scoring above, say, 700 on the verbal component in a year is given by

$$G_f \left(1 - \Phi(\frac{700 - \mu_f^v}{\sigma_g^v})\right),$$

and the numbers of men or women scoring above any particular threshold in a year is calculated analogously. As Figure 1 shows, given the rather substantial dispersion in scores – along with the large entry of women into test-taking – the influx of women into test taking

roughly doubled the number of US test takers receiving a quantitative score over 700. This score is high enough to gain entry to most top engineering programs.[1]

Beginning in the late 1960s and through the late 20th century, women's educational patterns changed substantially (Mammen and Paxson, 2000). Goldin (2014) finds that during the late 20th century U.S. women increased their years of education relative to men and shifted the content of secondary and college education toward more remunerative and career-oriented fields. This lead to rising professional and graduate enrollment underscoring changing educational patterns. A comparison of degree attainment in the U.S. over the late 20th century reveals substantial changes in women's representation. $In 1959-60, women earned 35.3\% of Bachelor's degrees, 31.6\% of Master's degrees, and 10.5\% of Doctor's degrees, resp$

Figure 2 shows how the female shares of new PhDs have evolved over the past half century, as reported by the US Department of Education.[2] The female share of PhD grantees varies across fields but has risen in all of them since 1960. At the high end, the female shares are over 75 percent in psychology, nearly 70 percent in education, and roughly 55 percent in the arts. Among the fields more directly relevant to innovative activity, the female share exceeds 50 percent in health and biology and by 2018 had reached roughly 20-25 percent in math, computer science, physical science, and engineering. It is clear that the cadre of doctorate holders, including those in STEM fields, is growing more female over time. This raises the possibility of harnessing the innovative potential of a previously less engaged group.

1.2 Relevant literature

A burgeoning literature has documented both the gaps and growth in women's participation in the scientific enterprise. Larivière et al. (2013) documented persistent gender disparities

[1] According to `https://www.studentprogress.org/gre/score/engineering/`, the average GRE scores for most kinds of engineering average between 152 and 161 on the new scale. This maps to 670-770 on the old scale. See `https://www.prepscholar.com/gre/blog/old-gre-to-new-gre-score-conversion-charts/`.

[2] See Table 352, `https://nces.ed.gov/programs/digest/2021menu_tables.asp`.

in scientific publishing, with women comprising less than 30% of fractionalized authorships on papers from 2008–2012. Women were moreover underrepresented as first authors, and received fewer citations overall when occupying key authorship positions (first, last (i.e., PI), or solo). Shen (2013) quantified ongoing inequalities in U.S. academia, highlighting that despite gains in entry-level representation, women remain underrepresented in senior faculty positions, suggesting of a "leaky pipeline" despite efforts to close the gap. Related work explored underlying reasons for the gap. National Academy of Sciences (2006) report concluded that barriers arise largely from implicit bias and legacy institutional structure as opposed to intentional discrimination or innate ability differences. Ceci and Williams (2011) for example claimed that gender imbalance in math-intensive fields stems not primarily from overt discrimination in peer review, grants, or hiring; rather, individual choices (e.g., regarding time allocation to family).

At the same time, the gender gap in scientific production seems to be shrinking, albeit not according to every metric. Huang et al. (2020) note that even as women's annualized participation in science increased substantially, gender gaps in total career productivity and impact paradoxically widened (e.g., from a 10% productivity gap in the 1950s to 35% favoring men in the 2000s). Iaria et al. (2022) notes similarly that only one percent of academics were women in 1900, and the share of women increased slowly until 1969 but faster in the last three decades of the 20th century. Still, by the end of the 20th century women remain underrepresented in the sciences, particularly in prestigious universities.

Similar patterns obtain not just in the publishing of scientific discoveries but their subsequent commercialization. Several studies suggest that women patent their scientific discoveries at lower rates (Thursby and Thursby, 2005; Whittington and Smith-Doerr, 2005; Ding et al., 2006; Fechner and Shapanka, 2018). Given that the commercialization of science is not limited to patenting, Tartari and Salter (2015) report less engagement in consulting and other translational activities. Fieldwork establishes that gender differences in institutional

and individual characteristics favoring the commercialization of academic research may contribute to this gender gap (Murray and Graham, 2007). Koffi and Marx (2025) report a double-digit gender gap in the commercialization of scientific discoveries as captured via patent-paper "pairs" (Ducor, 2000).

Despite the substantial progress to date in characterizing the magnitude and possible mechanisms underlying the gender gap in scientific production, comparatively less work has addressed the welfare effects of such. In other words, even if the raw percentage of papers published by women has increased, what has been the impact of this narrowing of the gap on science more generally? One could imagine several outcomes. For example, perhaps the increase in participation by women has displaced contributions by men, with unpredictable consequences for consumer surplus. On the other hand, it may be that contributions by women have been primarily additive—and although to our knowledge this has not been explicitly measured, it remains a general consensus among policy and academic work (see for example Huang et al. (2020)). Moreover, recent work on innovations is suggestive of welfare-enhancing outcomes where participation by women yields inventions (and products) that are targeted at female consumers (Einiö et al., 2019; Koning et al., 2021). But again, we are unaware of any quantification of consumer surplus for female participation in academia.

2 Theory

2.1 Analogy to demand and entry

We are interested in measuring the value of the female influx. We do so by means of an analogy to the IO approach to analyzing product markets.

Demand. We view the academic articles $j \in J$ extant – and available to be cited – at time t as the supply side of the market, or as the product "choice set." We view the authors of

new articles appearing in calendar year t as the "consumers" of extant citations and therefore the demand side. The author, or authors, of each new article are making a series of choices of which of the existing articles to cite. For each of the instances in which an author i might make a citation to article j, there is a utility of doing so:

$$u_{ij} = \delta_j + \zeta_{ij} + (1-\sigma)\epsilon_{ij},$$

where $\delta_j = x_j\beta - \alpha p_j + \xi_j$ is the mean utility of article j, and the utility of the outside good, not citing a paper, is $u_{i0} = 0$. There is no literal price of citing a paper, so there is no p_j. Yet, there is an implicit price, as papers cite a finite number of papers, but it is not observed. Hence, we cannot include a variable p_j, nor can we estimate the money metric parameter α. Each article j has a vintage in which it appeared (v) as well as an academic area (a).

We are interested in the effect of the influx of female-authored papers on the benefit that authors derive from the available choice set. We make an analogy between consumer surplus and the "author surplus" that authors derive from the available choice set. Author surplus is given, by analogy to consumer surplus, by

$$CS(D) = \frac{M}{\alpha}\ln(1 + D^{1-\sigma}),$$

where $D = \Sigma_{j\in J} e^{\delta_j/(1-\sigma)}$, and J is the set of extant articles at the time that citation decisions are being made. Define D_{sq} as the result of the actual choice set of articles at the end of our sample period, when the female influx has occurred. Define D_0 as the analog if the influx had not occurred, e.g. removing mostly-female-authored papers at random to bring the female share to its relative share prior to 1970.

Our main welfare object of interest is given by

$$\%\Delta CS = 100 \times \left(\frac{CS(D_{sq})}{CS(D_0)} - 1\right),$$

which does not depend on the price parameter α. We can estimate this object overall as well as by gender of the citing author.

Supply/entry. Authors' decisions to write/publish a paper are entry decisions. They do so when they have an idea/finished project whose expected influence exceeds the standards of the publication process. We can think of this as an article having sufficient promise so that the expected number of citations it will receive exceeds a threshold K.

The growth in female participation reflects a reduction in the participation cost to women. Changed social roles has increased the numbers of women pursuing education, which has in turn delivered more women to research pursuits. How might the quality of the additional scholarship from women relate to the pre-existing quality. There are two possible reasons for the marginal quality to be high. First, the growing participation of women might reflect changed social attitudes. Women with substantial ability might have been prevented from participating until recently, so that the increased participation makes use of high-ability women, rather than arising from drawing lower-ability women into scholarship. Second, the realized number of citations – indicating the article's quality – is a random variable. The quality of creative and scientific innovations is generally unpredictable (Caves, 2000; Scherer and Harhoff, 2000). If quality were entirely unpredictable, then additional entry, like that induced by the female influx into academic science, would bring additional articles as good, on average, as pre-existing articles. Consequently, the welfare benefit of the female influx would be large (Aguiar and Waldfogel, 2018; Waldfogel, 2025).

One observable consequence of unpredictability is manifested in the relationship between the share of articles by women (n_f) and the share of citations garnered by articles by women (q_f). If a growth in the female-authored share brought low-quality publications (relative to existing female-authored articles), then the change in the female-authored share ($\frac{\partial q_f}{\partial n_f}$) would be low. On the other hand, if the additional articles were as good, on average, as existing

articles, then the derivative would equal 1.

2.1.1 Empirical questions

These considerations lead directly to our four descriptive empirical questions as well as our welfare measurement exercise. First, how much are female articles and patents cited, relative to male-authored analogues; and how does this vary according to the citation behavior of the citing authors? Second, how does the share of citations garnered by female articles and patents (q_f) evolve as the share of works by women (n_f) rises; and how does this vary according to the genders of the citing authors and inventors? Third, does the entry of female innovators – and the ensuing female-authored articles and patents – displace male author activity? Fourth, do additional articles and patents expand the pie of citations received? Or do additional articles and patents simply "steal business" from existing works? Finally, what proportionate benefit does the influx of female-driven innovation bring to citing authors and inventors, overall and by citer gender?

3 Data

We construct two different basic datasets, one for academic articles and another for patents. We use each of these to create four analysis datasets.

3.1 Academic dataset construction

We obtain the Microsoft Academic Graph (MAG) dataset (Sinha et al., 2015) from Färber (2021) (see also Färber (2019)). Our article-level sample consists of all 65.2 million journal articles in 40 distinct academic areas published between 1965 and 2015.[3] We define academic

[3]We focus on journal articles as they constitute the most standardized and consistently indexed unit of scholarly output within and across scientific fields. We exclude working papers, books, book chapters, and other non-journal scholarly outputs as they exhibit heterogeneous citation practices and less reliable

areas following the OECD Field of Science and Technology subfield-classification.[4] For this purpose, we use the crosswalk dataset from Koffi and Marx (2025) that maps scientific papers to specific scientific areas using a combination of MAG scientific paper fields, OECD fields, and Web of Science (WoS) subject categories.

For each article, we observe author names and the countries of their affiliated institutions, the journal in which the article was published, and the number of citations the article had received in a given citation year and in total as of 2020.[5] The dataset comprises 66.7 million distinct authors and 48,016 distinct journals.

We use the article as the unit of observation for many exercises below, and we need to classify articles according to whether they are female authored. We do this according to whether half or more of its total authors are female (see Section 3.1.1 for a more detailed description on how we determine author gender). Define $\delta_j^{f,all} = 1$ if half or more of the authors of article j are female. Define f_j as the female share of authors on article j. We also create a measure of female authorship for each area and vintage based on the continuous share (average of or sum).

Table 1 provides summary statistics at the article level. 13.4% of articles in our sample are mostly-female-authored articles. The average female share of authors is 13.8% while the average share of male authors and unidentified-gender authors is 49.4% and 36.7%, respectively. For 23.7% of articles in our sample the gender of authors is unknown. We control for these unidentified-gender articles in our regressions. The average article is published in

meta-data and coverage in MAG (Moed, 2005; Visser et al., 2021).

[4]The academic areas are Agricultural and Biological Sciences, Arts and Humanities, Biochemistry, Genetics and Molecular Biology, Business, Management and Accounting, Chemical Engineering, Chemistry, Computer Science, Decision Sciences, Dentistry, Earth and Planetary Science, Economics, Econometrics and Finance, Energy, Engineering, Environmental Science, General, Health Professions, Immunology and Microbiology, Materials Science, Mathematics, Medicine, Multidisciplinary, Neuroscience, Nursing, Pharmacology, Toxicology and Pharmaceutics, Physics and Astronomy, Psychology, Social Sciences, Unknown, and Veterinary.

[5]We restrict the sample to articles published before 2016 to have at least five post-publication citation years.

2000 and receives a total of about 20 cites, of which 3.3 cites can be assigned to female authors, 12.3 cites to male authors, and 4.2 cites to unknown-gender authors, respectively. 36.5% of the studied journal articles (23.8 million) do not receive any citation. We also observe the yearly citations for the citation years 1965 to 2020, in total and assigned by gender. While these variables are not shown in Table 1 for readability, we report summary statistics for total citations aggregated into five-year intervals. This indicates that most citations occur between 2000 and 2020. The average share of female authors by area and vintage is 14.3% while the average total number of female authors by area and vintage is 113,188, respectively.

We have so far considered the construction of our *cited*-article variables. We obtain our *citing*-article variables using MAG reference-level data consisting of 780 million citing-article-cited-article pairs. For consistency, we also restrict the citing-article sample to journal articles published between 1965 and 2015, i.e., we consider the citing behavior of authors of journal articles. In this respect, the citing-article sample is a subset of the cited-article sample, e.g., not all of the 65.2 million journal articles under study are citing articles. Merging via the unique MAG paper ID, for each cited article, we obtain the full set of article-level characteristics for its citing articles including the aforementioned gender-related variables. This approach allows us to obtain, for a given cited article, the citations from mostly-female citing articles and male citing articles, by citing year and in total. It also allows us to distinguish between the following citation patterns: (i) mostly-female article citing mostly-female article, (ii) mostly-female article citing male article, (iii) male article citing mostly-female article, and (iv) male article citing male article, respectively.

3.1.1 Determining author gender

It is common practice in such studies to infer gender from given names ("forenames") using algorithms that detect the likelihood of a given forename being male vs. female (Lerchen-

mueller and Sorenson, 2018). Even when incorporating information such as location into probability tables, however, name-prediction algorithms can make substantial errors (Marx, 2024).[6] Our approach is to combine high-confidence prediction from names with Large Language Models (LLMs) for authors where forename-based prediction yields lower confidence.

The intuition behind this approach is that information available publicly on the internet, and subsequently incorporated into the training of LLMs, could usefully disambiguate ambiguous forenames. Consider Chris Forman, faculty member at Cornell University. Chris' forename is highly ambiguous and will likely remain unclassified via algorithm. However, Chris' faculty page at Cornell contains gender-identifying pronouns that make it clear he is a male. Therefore, an LLM-based approach can help fill in unclassified names. Moreover, LLMs can help to avoid classification errors. Consider former Harvard president Drew Gilpin Faust, whose forename is given to males far more often. President Faust might well be misclassified as male by an algorithm, even though she is well known as the first female president of Harvard and therefore easily disambiguated by the LLM.

We begin with 66.7 million authors in MAG. For names that can be confidently classified (i.e., with at least 90% confidence), we use forename-country-based algorithms.[7] This corresponds to 34.3% of distinct authors in our dataset. For the remaining 44.5 million authors, we proceed as follows. Given that the LLM has a better chance of retrieving the correct gender if the author's affiliation is known, as per in the examples above, we look up the author using his or her most recent affiilation in MAG. This can be done for 22.9 million authors.

We then "interpolate" possible affiliations for authors by a) taking the most frequent affiliation on a paper where the focal author's affiliation is missing, and b) crosswalking

[6] Koffi, Marx, and Yimfor (2026) report that the gender flag in the USPTO PatentsView dataset has about 8% error, based on their own merge of inventors to LinkedIn and subsequently identifying gender via profile pictures. As many academics do not have LinkedIn pages, this method is less attractive for us.

[7] We use `gender.api` for this purpose.

the paper (by DOI) to the Clarivate Web of Science and using the affiliation for the paper. Interpolation covers another 13.3 million authors, leaving 7.8 million for whom we lack any affiliation.

We use the 8-billion parameter Llama 3.1 model to retrieve gender using the following prompt: *Determine the gender of these researchers using only internal knowledge: . For each researcher, assign exactly one gender: Male, Female, or Unknown (use Unknown for ambiguous names). Respond strictly with a comma-separated list of exactly genders in the same order as the input names. also include a confidence score from 0-100 for how sure your classification is. Do not include explanations, notes, or any other text. Example: Male/100,Female/50,Unknown/30.* Importantly, the LLM is asked to provide a confidence score for each classification. We retained only LLM classifications with a 90% confidence score.

3.2 Patent data

The patent data for this study consist of 7,626,120 patents granted in the US between 1976 and 2020, along with their 106.0 million citations to US patents granted over the same period. For each citing and cited patent, we know its WIPO technology category, its grant year, and the number of its inventors who are men, women, or persons whose genders cannot be ascertained from their names. We classify patents into 35 technology categories according to their first listed WIPO category.[8] The data are from the USPTO.[9]

[8]The resulting categories are analysis of biological materials; audio-visual technology; basic communication processes; basic materials chemistry; biotechnology; chemical engineering; civil engineering; computer technology; control digital communication; electrical machinery, apparatus, energy; engines, pumps, turbines; environmental technology; food chemistry; furniture, games; handling; IT methods for management; machine tools; macromolecular chemistry, polymers; materials, metallurgy; measurement; mechanical elements; medical technology; micro-structural and nano-technology; optics; organic fine chemistry; other consumer goods; other special machines; pharmaceuticals; semiconductors; surface technology, coating; telecommunications; textile and paper machines; thermal processes and apparatus; and transport.

[9]See `https://patentsview.org/`.

3.3 Analysis datasets

We use the underlying datasets above to create two kinds of datasets for analysis. First, we create aggregate data showing the numbers of articles or patents by academic or technological area and by vintage. We also distinguish articles and patents according to whether they are predominantly (half or more) female-authored or invented. These aggregate data also include the numbers of citations received by male and female articles or patents in each vintage and topical area. In some versions of these dataset we also disaggregate to citing year. We use these aggregate datasets for the descriptive analysis.

The second sort of broad dataset is at the patent or article level. We observe characteristics of each article and patent, including its topical area and appearance vintage, as well as whether it is predominantly female (half or more of its authors or inventors are female). We also observe the number of citations that each article or patent receives from male or female authors or inventors in each calendar year. We use these datasets to estimate citation demand models.

4 Descriptive evidence

4.1 The female influx

For simplicity of exposition, we refer to articles whose authors are mostly female as "female-authored;" and we refer to the complement, the non-majority-female-authored articles, as "male-authored."

Our first question is simply, how large is the female influx into academic science? Panel (a) of Figure 3 shows the share of articles with mostly-female authors. Under 5 percent of the articles published in 1965 have mostly-female authors. This rises steadily and reaches 20 percent by 2015. Panel (b) of Figure 3 provides a different perspective on the influx. Both

female – and male – authorship are growing substantially over the half-century. Female gains appear not, on their face, to come at the expense of male participation. Panels (c) and (d) repeat the exercises for patents. Female-inventor patents rise from under 5 percent in 1975 to over 9 percent in 2020. As with articles, both male and female-inventor patents granted rise over time.

The level, and increase, varies across academic fields. Panel (a) of Figure 4 shows the mostly-female authorship share in the 1960s and the 2010s by field, ordered by the mostly-female share in the 2010s. The mostly-female share rises in all areas, but the eventual share varies sharply across fields. At the high end, it reaches roughly a third in psychology, sociology, education, and art. By contrast, the mostly-female share remains below 10 percent in physics and various engineering fields. Panel (b) repeats the exercise for patents. Female inventors are most prevalent in biology, biotechnology, and pharmaceuticals, with over 15 percent of inventors. Women are least prevalent in civil and mechanical engineering.

4.2 Citations per article

How well do female-authored articles get cited, relative to male-authored articles; and how does this change with the influx of female-authored articles? For female-authored, vintage-v articles in area a, the average (eventual, as of the end of the sample period) citations per article are given by $\frac{Q_{avf}}{N_{avf}}$. The analogous ratio for male-authored articles is $\frac{Q_{avn}}{N_{avn}}$. Hence, we can characterize the citation influence of female-authored articles, relative to male-authored articles, by

$$r_{va} = \left(\frac{Q_{avf}}{N_{avf}} / \frac{Q_{avn}}{N_{avn}} \right).$$

The left panel of Figure 5 shows this ratio for academic articles. The heavy line shows that female articles receive more citations per article until about 2010. Moreover – as the dashed line shows – women cite women at elevated rates relative to men. The right panel of Figure 5 repeats the exercise for patents. Overall – as the solid line shows – women receive

roughly the same number of citations per article as men, Women cite women more, initially at a ratio of 1.5 and later rising above 3.

4.3 The value of the influx

If the influxes of female articles and patents is valuable, then female innovations should attract citations. We explore this via regressions of the form

$$q_{avf} = \phi n_{avf} + \mu_a + \mu_v + \varepsilon_{av}. \tag{1}$$

where q_{avf} is the share of citations attracted by female innovations, among those in topic area a and vintage v, n_{avf} is the female share of innovations, and μ_a and μ_v are area and vintage fixed effects.

Table 4 reports variants on this regression for academic articles. Column (1) includes neither area nor vintage fixed effects, and the coefficient is 1.15 (se = 0.011). Column (2) adds area and vintage fixed effects, and the coefficient falls to 1.06 (0.031). Columns (3) and (4) repeat the exercises of columns (1) and (2) with the entire sample and not just the citing papers whose author genders are identifiable, with very similar results. Essentially, as the share of articles by female author rises, the share of citations that these articles garner rises roughly proportionally. Table 5 repeats the exercise in Table 4 for patents, with similar qualitative patterns.

The remainder of the table repeats the exercise by citing author gender. Columns (5) and (6) use the share of citations from female-authored citing papers to female-authored cited papers, or q^f_{avf}. Columns (7) and (8) use the citations from male-authored citing articles q^n_{avf}. The female-citing coefficients are larger. As the female-authored share of articles rises, what happens to the share of citations from women to women, compared to the coefficient for men? Increases in the female-authored share of articles raise the share of citations to female

authors more for female than for male citing articles. This reinforces the finding above that female authors derive more benefit from female-authored articles than male authors derive.

The evidence in this section suggests that the female influx is valuable. Patents and articles by women attract roughly as many citations as articles by men; and as the share of patents and articles by women rises, the share of citations to those works by women rises roughly proportionally. Moreover, female authors cite patents and articles by women more than do male authors.

That the contributions from women are valuable does not, by itself, demonstrate that the influx raises the value of the scientific corpus. It is possible, for example, that contributions by women displace contributions from men that would have been similarly valuable. To address these concerns, we look for evidence of displacement, as well as whether growth in articles draws more citations.

Tables 6 and 7 and revisit the exercises in Table 4 and 5 using data by area, vintage, and calendar times. The regressions augment Equation (1) with a calendar year fixed effect. Results are nearly identical. The effects that we document across vintage remain when we allow for citing behavior that varies across calendar time.

4.4 Displacement

Does the entry of articles or patents by female authors and inventors displace male entry? We study this by estimating the following regression model relating the number of articles or patents in a topical area by males appearing at each vintage on the number of female articles or patent in the area appearing in the vintage, along with vintage and area fixed effects.

$$N_{avn} = \psi^f N_{avf} + \mu_a + \mu_v + \varepsilon_{av},$$

where N_{avn} is the number of male works in area a and vintage v and N_{avf} is the analogous measure for works by women. If female-authored articles simply displaced male-authored articles, we would see that $\psi^f = -1$.

Table 8 reports results for academic articles, in both levels and logs, with and without fixed effects. All of the ψ^f coefficients are positive. There is no evidence that male entry is displaced by female-authored article entry. We interpret these results, along with panel (b) of Figure 3, to mean that there is substantial growth in demand for male and female scholars and, moreover, that the influx of women has allowed an increase in the amount of research created. Table 9 repeats the displacement exercise for patents. All coefficients are positive, so we see no direct evidence of displacement.

4.5 Business stealing vs. market expansion

We explore possible market expansion by asking whether growth in the number of articles in an area raises the number of citations that those articles receive.

The choice set of articles to cite evolves over time. One simple summary of that choice set is N_{at}, the number of extant articles in field a published prior to year t. The question is whether the number of citations from year-t articles or patents (Q_a^t) to the extant articles or patents in field a (N_{at}) rises faster in fields in which the number of articles or patents grows more. We examine this via the following regression

$$Q_a^t = \nu N_{at} + \mu_a + \mu_t + \varepsilon_{at}$$

We face the complication that we only observe articles published since 1965. Hence, $N_{at} = \Sigma_{v=1965}^{t-1} N_{av}$. Consequently, we experiment with different start dates for the regression. Table 10 shows results for academic articles, in levels and logs. Academic areas whose extant articles grow attract more citations, after accounting for the general time pattern. This

market expansion supports the notion that additional articles are valuable. Table 11 repeats the exercise for patents, with similar results.

5 Welfare estimation

The descriptive results above suggest that the female influx into academic publishing is valuable. Female-authored articles attract roughly as many citations per article as male-authored articles (at least until recently). Moreover, as the share of articles by women rises, across vintages within areas, the share of citations garnered by those articles rise. This increase is proportional overall, although the increase is lower for citations from male-authored articles than from female-authored articles.

This section quantifies the value of the female influx using simple industrial organization tools. First, we present a model of the demand for articles to cite. We treat the authors of each year's entering articles as the "buyers" of existing articles, which they buy via citation. We treat the extant articles at each point in time as the choice set. Our goal is to measure the value of the evolving choice set, and we do this both from the standpoint of citing authors overall, as well as from the respective perspectives of female and male authors.

The exercise has two components. First, we have a nested logit demand model, which we estimate separately for three groups (overall authors, male authors, and female authors). Second, we have counterfactual assumptions about how many of the female-authored articles existing after the female influx need to be removed to characterize an environment in which the influx had not occurred.

5.1 Demand

We estimate various nested logit models of demand for cited articles and patents, using different sets of citing authors. First, we estimate models using all articles or patents.

Second, we estimate separate variants for citations from patents or articles with female (majority female) vs male (other) inventors or authors. Following Berry (1994), we estimate nested logit demand models via

$$\ln(s_j^t) - \ln(s_0^t) = \gamma\delta_j^f + \sigma \ln(s_j^t/(1 - s_0^t)) + \mu_v + \mu_a + \xi_j,$$

where s_j^t is the number of citations received by work j (which was published in vintage v and is in field a) during year t, divided by "market size" M^t (which we operationalize as the number of articles or patents appearing in year t times 100). The term δ_j^f is an indicator that is 1 if article or patent j is female-authored or invented.

The parameter σ shows the extent to which articles are substitutes for one another. If $\sigma = 1$, then works are close substitutes for one another; if $\sigma = 0$, then this simplifies to the plain logit model; and works are not close substitutes for one another so that additional articles or patents bring large welfare benefits. It is identified by the extent to which total citations rise across years with the growth in the number of articles or patents to cite. We use the log of the number of extant articles or patents in each year N_t as an instrument for the log of the inside share, $\ln(s_j^t/(1 - s_0^t))$.

Table 12 reports first stage regressions for academic articles. The log number of products is strongly related to the inside shares.

Columns (1) and (2) of Table 14 report results for academic articles, estimated on calendar years 1980-2015.[10] Odd-numbered columns in the Table are OLS, while even-numbered columns report IV. The IV estimate of σ is 0.611, and female-authored articles have less (overall) utility. Columns (3) and (4) repeat the exercise of columns (1) and (2) using the citations from female-authored citing papers. The utility function parameters here reflect the "preferences" of female authors. These authors find female-authored articles to be more

[10]Starting the sample at 1980 means we observe most of the articles that the demand side uses. That is, we miss only the articles over 15 years old in the first year of the sample.

useful; and the substitution parameter is 0.475. The last two columns report the utility function parameters for male authors. These authors attach lower value to female-authored articles, and the substitution parameter is 0.647. Standard errors are clustered on citing year.

Tables 13 and 15 repeat the exercises of Tables 12 and 14 for patents. Table 13 shows that the instruments "work." Table 15 presents the demand model estimates. As with articles the IV substitution parameters are lower.

5.2 Counterfactuals

We measure the impact of the female influx into academic publishing and patenting by comparing end-of-sample choice sets (2015 for articles and 2020 for patents) to counterfactual choice sets in which female-authored articles or patents are eliminated, at random, to bring articles to their share in each field at the start of the sample (1965-1969 for articles, 1975-1979 for patents). We rely on the IV estimates in Tables 14 and 15.

The left panel of Figures 6 reports the overall $\%\Delta CS$ for the female influx into each field for academic articles. Taking all fields together, the gain to citing authors is a 5.1 percent increase in surplus. The gain is largest — 10 to 15 percent — in the fields with large eventual female author shares. These include law, education, health science, psychology, and sociology. The overall gains are lowest – under 3 percent – for engineering, computer science, and math.

The overall effects in patents are smaller. Overall inventor surplus rises by 1.02 percent with the influx, while the gains to male and female inventors are 0.96 percent and 1.67 percent, respectively. The right panel of Figure 6 does the exercise across patent categories. Gains to inventors are largest in nanotechnology, biology, and pharmaceuticals (2-3 percent) and substantially smaller in civial and mechanical engineering (under half a percent).

Given the different utility function parameters for male and female authors in Table 14,

the effects of the female influx can be different across author gender. The overall gain for female citing authors is a 12.8 percent increase in surplus, while the overall gain for male authors is a 3.6 percent gain. The left panel of Figure 7 juxtaposes the $\%\Delta CS$ for male and female authors of academic articles, ordered from largest to smallest according to the benefit to female authors. A few results emerge. First, the benefit of the female influx is larger, regardless of field, for female citing authors. Even in the low-benefit engineering fields, CS rises by about 5 percent (compared with one or less for men). The benefit to women as citers exceeds 20 percent in heavily female fields such as other humanities, health sciences, education, law, and sociology. In these fields, the benefits to women citing authors are roughly double the proportionate benefits to male authors.

The right panel of 7 repeats the exercise for patents. Given the smaller female influx into patenting, the effects are smaller; but as with academic articles, the female influx has a larger impact on the surplus of female authors.

6 Conclusion

The past half century has seen a substantial female influx into patenting and, especially, into academic research. The female-authored share of new academic articles quadrupled from five to 20 percent between 1965 and 2015; and the female share of patent inventors rose from 5 to 9 percent ove a similar period. We offer descriptive findings suggestive that the female influx is valuable. The growth in female-authored work is accompanied by roughly proportionate growth in the share of eventual citations garnered by female-authored work, and female-authored work attracts roughly as many citations, per patent or article, as male-authored work. Women cite patents and articles by fellow women at elevated rates. Moreover, additional articles and patents appear to "expand the market" by attracting additional citations and not just diverting citations from other works. Finally,

we see no evidence that the female influx displaces research activity by men. We measure the proportionate impact of the female influx on the welfare of researchers. All researchers benefit from the influx of work by women. Women benefit substantially more.

References

Aguiar, L. and J. Waldfogel (2018): "Quality predictability and the welfare benefits from new products: Evidence from the digitization of recorded music," *Journal of Political Economy*, 126, 492–524.

Altonji, J., P. Arcidiacono, and A. Maurel (2016): "Chapter 7 - The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects," Elsevier, vol. 5 of *Handbook of the Economics of Education*, 305–396.

Bell, A., R. Chetty, X. Jaravel, N. Petkova, and J. Van Reenen (2019): "Who becomes an inventor in America? The importance of exposure to innovation," *The Quarterly Journal of Economics*, 134, 647–713.

Berry, S. T. (1994): "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 242–262.

Blank, D. M. and G. J. Stigler (1957): "Appendix B: Census Data on Number of Engineers and Chemists, 1890–1950," in *The Demand and Supply of Scientific Personnel*, Princeton, NJ: National Bureau of Economic Research, 143–155.

Bleske-Rechek, A. and K. Browne (2014): "Trends in GRE scores and graduate enrollments by gender and ethnicity," *Intelligence*, 46, 25–34.

Caves, R. E. (2000): "Creative industries: Contracts between art and commerce," *Harvard University*.

Ceci, S. J. and W. M. Williams (2011): "Understanding current causes of women's underrepresentation in science," *Proceedings of the National Academy of Sciences*, 108, 3157–3162.

Costa, D. L. (2000): "From Mill Town to Board Room: The Rise of Women's Paid Labor," *The Journal of Economic Perspectives*, 14, 101–122.

Ding, W. W., F. Murray, and T. E. Stuart (2006): "Gender differences in patenting in the academic life sciences," *Science*, 313, 665–667.

Ducor, P. (2000): "Coauthorship and coinventorship," *Science*, 289, 873–875.

Einiö, E., J. Feng, and X. Jaravel (2019): "Social push and the direction of innovation," *Available at SSRN 3383703*.

Fechner, H. and M. S. Shapanka (2018): "Closing diversity gaps in innovation: Gender, race, and income disparities in patenting and commercialization of inventions," *Technology & Innovation*, 19, 727–734.

Färber, M. (2019): "The Microsoft Academic Knowledge Graph: A Linked Data Source with 8 Billion Triples of Scholarly Data," in *Proceedings of the 18th International Semantic Web Conference*, ISWC'19, 113–129.

——— (2021): "The Microsoft Academic Graph in RDF: A Linked Data Source with 8 Billion Triples of Scholarly Data," Zenodo, version 2020-06-19.

Goldin, C. (2014): "A Grand Gender Convergence: Its Last Chapter," *American Economic Review*, 104, 1091–1119.

Huang, J., A. J. Gates, R. Sinatra, and A.-L. Barabási (2020): "Historical comparison of gender inequality in scientific careers across countries and disciplines," *Proceedings of the National Academy of Sciences*, 117, 4609–4616.

Hyde, J. S., S. M. Lindberg, M. C. Linn, A. B. Ellis, and C. C. Williams (2008): "Gender Similarities Characterize Math Performance," *Science*, 321, 494–495.

Hyde, J. S. and M. C. Linn (2006): "Gender Similarities in Mathematics and Science," *Science*, 314, 599–600.

Iaria, A., C. Schwarz, and F. Waldinger (2022): *Gender gaps in academia: Global evidence over the twentieth century*.

Jiang, X. (2021): "Women in STEM: Ability, preference, and value," *Labour Economics*, 70, 101991.

Koffi, M. and M. Marx (2025): "Cassatts in the Attic: Is There a Gender Gap in the Commercialization of Science?" *American Economic Journal: Applied Economics*, forthcoming.

Koning, R., S. Samila, and J.-P. Ferguson (2021): "Who do we invent for? Patents by women focus more on women's health, but few women get to invent," *Science*, 372.

Larivière, V., C. Ni, Y. Gingras, B. Cronin, and C. Sugimoto (2013): "Bibliometrics: Global gender disparities in science," *Nature*, 504, 211–213.

Lerchenmueller, M. J. and O. Sorenson (2018): "The gender gap in early career transitions in the life sciences," *Research Policy*, 47, 1007–1017.

Mammen, K. and C. Paxson (2000): "Women's Work and Economic Development," *Journal of Economic Perspectives*, 14, 141–164.

Marx, M. (2024): "Can we win the "names game"? Tactics for large-scale demographic research in entrepreneurship and innovation," *Industry and Innovation*, 1–9.

Moed, H. F. (2005): *Citation Analysis in Research Evaluation*, Dordrecht: Springer.

Murray, F. and L. Graham (2007): "Buying science and selling science: Gender differences in the market for commercial science," *Industrial and Corporate Change*, 16, 657–689.

National Academy of Sciences (2006): *Beyond bias and barriers: Fulfilling the potential of women in academic science and engineering*, National Academies Press.

National Center for Education Statistics (2023): "Table 318.10. Degrees conferred by degree-granting institutions, by sex of student and level of degree: Selected years, 1869–70 through 2020–21," `https://nces.ed.gov/programs/digest/d23/tables/dt23_318.10.asp`, accessed: 2026-02-04.

Pidgeon, M. E. (1954): *Changes in women's occupations, 1940-1950*, 253-256, US Department of Labor, Employment Standards Administration, Women's Bureau.

Scherer, F. M. and D. Harhoff (2000): "Technology policy for a world of skew-distributed outcomes," *Research policy*, 29, 559–566.

Shen, H. (2013): "Mind the gender gap," *Nature*, 495, 22–24.

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

Tartari, V. and A. Salter (2015): "The engagement gap: Exploring gender differences in University–Industry collaboration activities," *Research Policy*, 44, 1176–1191.

Thursby, J. G. and M. C. Thursby (2005): "Gender patterns of research and licensing activity of science and engineering faculty," *The Journal of Technology Transfer*, 30, 343–353.

Toossi, M. (2002): "A Century of Change: The U.S. Labor Force, 1950–2050," *Monthly Labor Review*, 125, 15–28.

U.S. Bureau of Labor Statistics (2022): "Women in the Labor Force: A Databook," Report 1097, U.S. Bureau of Labor Statistics, data from the Current Population Survey (CPS). Accessed 2026-02-04.

U.S. Women's Bureau and M. E. Pidgeon (1954): "Changes in Women's Occupations, 1940–1950," Women's Bureau Bulletin 253, United States Department of Labor, Women's Bureau, Washington, DC.

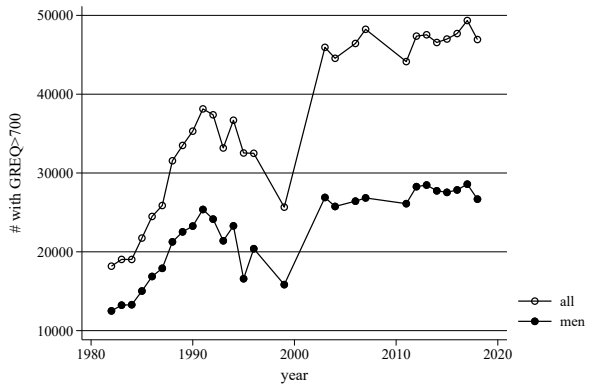
Visser, M., N. J. van Eck, and L. Waltman (2021): "Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic," *Quantitative Science Studies*, 2, 20–41.

Waldfogel, J. (2025): "The Welfare Effects of Gender-Inclusive Intellectual Property Creation: Evidence from Books," *Journal of Political Economy*, 133, 2229–2264.

Whittington, K. B. and L. Smith-Doerr (2005): "Gender and commercial science: Women's patenting in the life sciences," *The Journal of Technology Transfer*, 30, 355–370.

7 Tables and figures

Figure 1: Male and total persons with GREQ 700



Notes: The number of quantitative GRE test takers with scores above 700 using the old scoring standard. Computed from reported means and standard deviations of scores, assuming normality.

Figure 2: Female share of PhDs earned

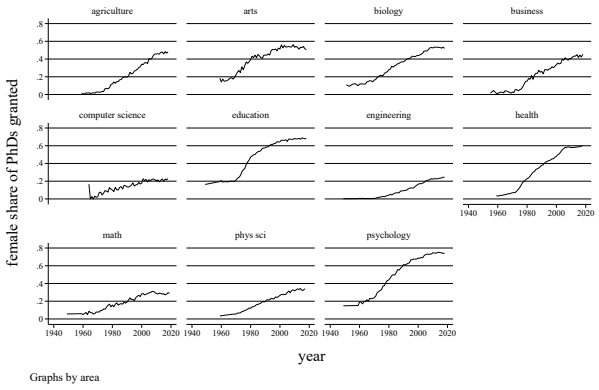
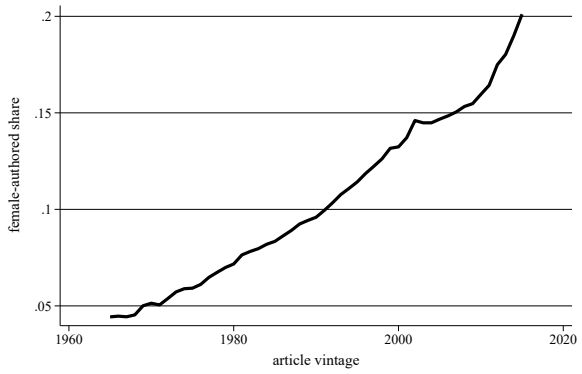
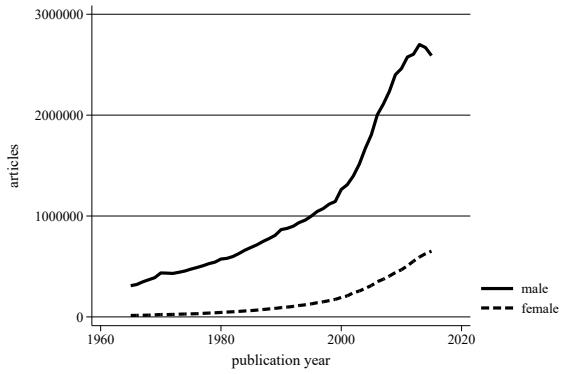


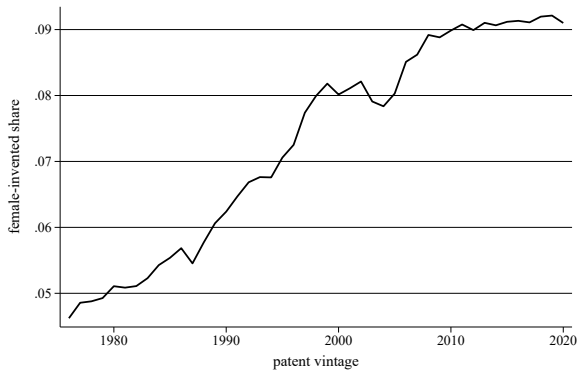
Figure 3: Male and female articles and patents



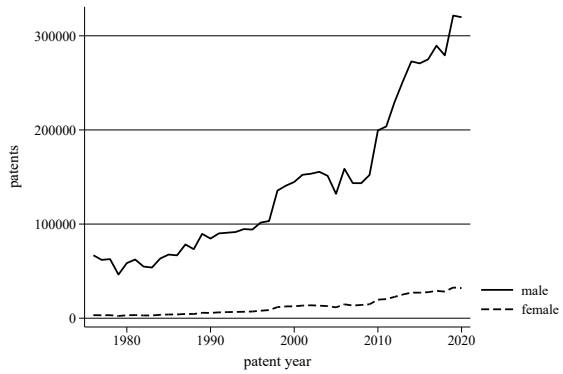
((a)) female authorship share



((b)) male and female articles



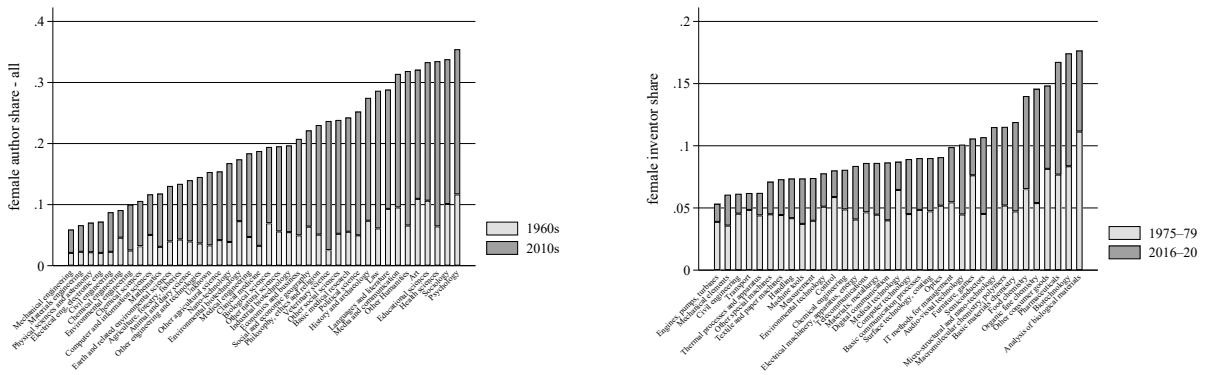
((c)) female patent inventor share



((d)) male and female patents

Notes: Panel (a) shows the share of new articles by publication year with mostly-female authors. Panel (b) shows the absolute number of new articles by year from majority-female author teams and the remainder (male). Panel (c) shows the share of new patents by year with mostly-female inventors. Panel (d) shows the absolute number of new patents by year from majority-female inventor and the remainder (male).

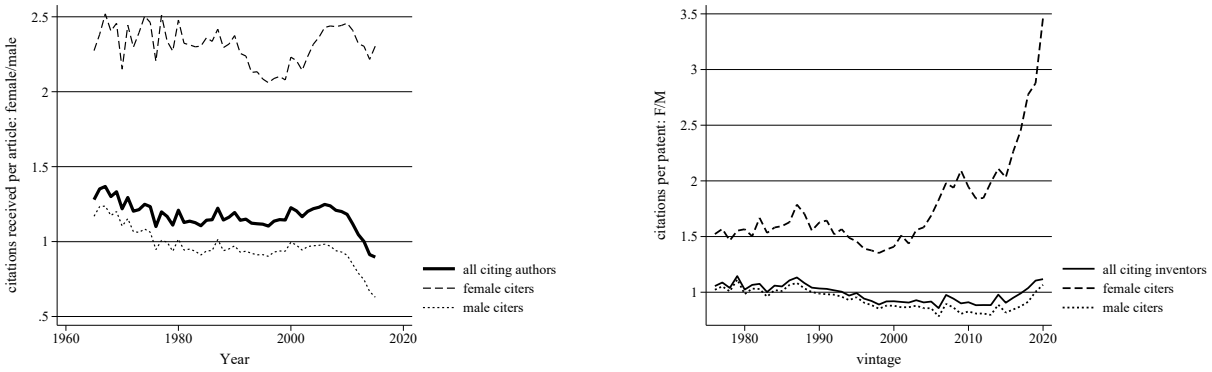
Figure 4: Female author and inventor shares by topic, 1960/70s vs 2010s



((a)) articles ((b)) patents

Notes: Shares of articles (a) and patents (b) in each field that are mostly-female-authored in the 1960/70s (light portion of bar, 1965-1969 for articles, 75-79 for patents) vs 2010s (full bar, 2010-2019).

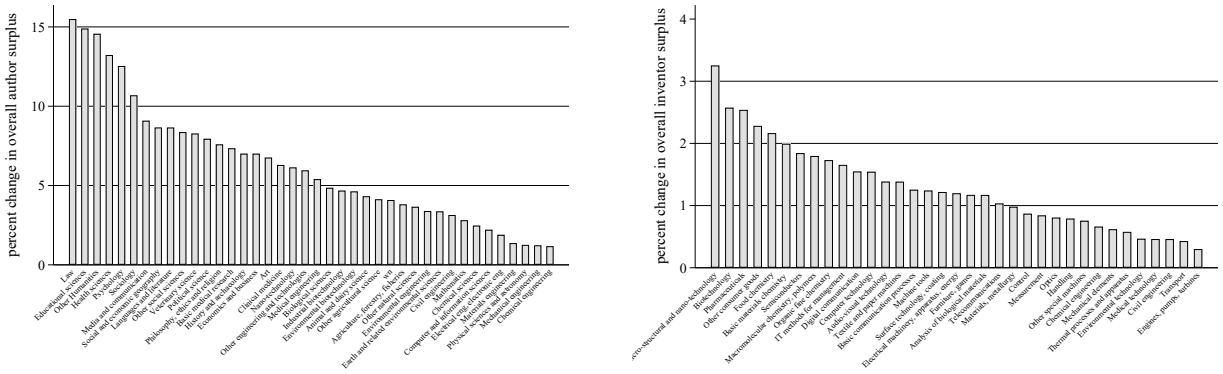
Figure 5: Relative citations per article and patent: F/M



((a)) articles ((b)) patents

Notes: The ratio of citations per female-authored articles to the citations per non-female-authored articles (a) or patent (b).

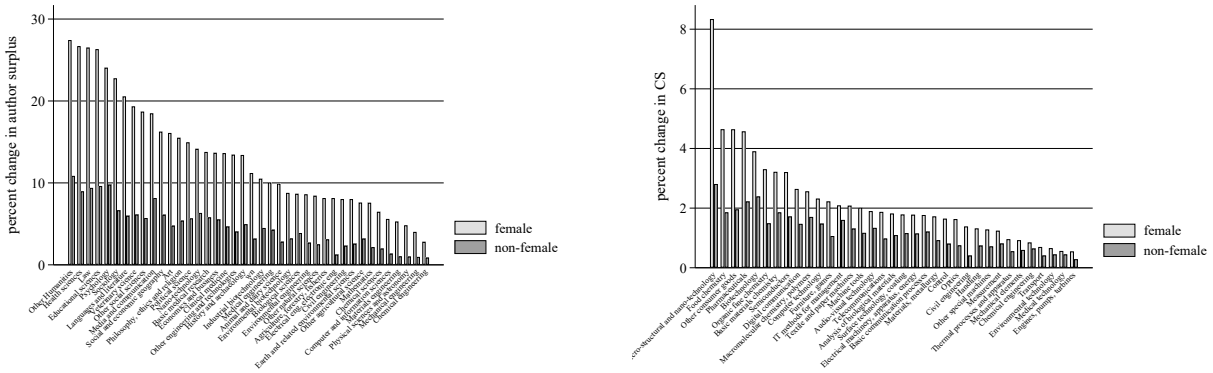
Figure 6: Percentage gain in overall citer surplus



((a)) articles ((b)) patents

Notes: Percent increase in author (a) or inventor (b) surplus from the growth in female-authored articles and patents. Based on the nested logit demand model using all authors, and based on a comparison of 2015 to a counterfactual environment in which female-authored papers are removed at random to bring the female-authored share to its level, by field, during 1965-1969.

Figure 7: Percentage gain in overall citer surplus



((a)) articles ((b)) patents

Notes: Percent increase in consumer (inventor) surplus from the growth in female-authored articles. Based on the nested logit demand model using mostly female – and non-mostly-female authors, respectively – and based on a comparison of 2015 to a counterfactual environment in which female-authored papers are removed at random to bring the female-authored share to its level, by field, during 1965-1969.

Table 1: Summary statistics, article level (full sample)

	mean	sd
Mostly-female-authored article	0.134	0.341
Share female authors	0.138	0.272
Share male authors	0.494	0.405
Share unidentified-gender authors	0.367	0.414
Article where gender is unknown	0.237	0.425
Year	1999.817	13.071
No. of citations	19.870	110.832
Female share of total cites received	3.323	27.119
Male share of total cites received	12.317	86.905
Unknown-gender authors' share of total cites received	4.222	42.944
Total citations, 1965–1969	1.704	4.963
Total citations, 1970–1974	2.079	6.968
Total citations, 1975–1979	2.113	8.025
Total citations, 1980–1984	2.220	10.456
Total citations, 1985–1989	2.277	14.546
Total citations, 1990–1994	2.426	16.515
Total citations, 1995–1999	2.715	16.551
Total citations, 2000–2004	3.455	18.272
Total citations, 2005–2009	4.594	23.981
Total citations, 2010–2014	5.655	32.240
Total citations, 2015–2020	6.276	43.235
Share female authors, by area and vintage	0.143	0.078
No. of female authors, by area and vintage	113187.658	193092.344

Notes: Summary statistics for article-level variables. The sample covers journal articles published between 1965–2015, $N = 65{,}219{,}500$.

Table 2: Summary statistics, academic area–vintage level (estimation sample)

	mean	sd
mostly- female-authored share of articles	0.121	0.080
share of citations to mostly female-authored (all)	0.144	0.104
share of citations to female-authored (subset)	0.143	0.100
share: female to female	0.222	0.129
share: male to female	0.119	0.079
Year	1990.000	14.723
number of male-authored	27685.292	53646.252
number of mostly female-authored	4285.051	10662.338
citations to male-authored	539841.550	1194006.242
citations to mostly female-authored	95399.900	252733.381

Notes: Data, covering articles published 1965–2015, are aggregated to academic area by vintage, $N = 2{,}040$.

Table 3: Summary statistics, demand estimation sample

	mean	sd
dv	-18.415	0.908
dv_female	-18.871	0.689
dv_male	-18.470	0.885
female-authored article	0.132	0.338
year	2003.544	10.396
vintage	1994.014	12.261
log extant articles	14.479	0.549
log inside share	-13.180	1.170
log inside share (female)	-12.060	1.092
log inside share (male)	-13.024	1.130

Notes: $N = 10{,}916{,}395$.

Table 4: Regressions of q_{avf} on n_{avf} for articles

	(1) both	(2) both	(3) both-all	(4) both-all	(5) female citers	(6) female citers	(7) male citers	(8) male citers
mostly- female-authored share of articles	1.150***	1.062***	1.195***	1.141***	1.446***	1.055***	0.884***	0.803***
	(0.0113)	(0.0310)	(0.0118)	(0.0332)	(0.0164)	(0.0442)	(0.0100)	(0.0273)
Constant	0.00414*	0.0147***	-0.000290	0.00622	0.0475***	0.0947***	0.0122***	0.0220***
	(0.00164)	(0.00383)	(0.00171)	(0.00410)	(0.00238)	(0.00546)	(0.00145)	(0.00337)
Field & Year FE	No	Yes	No	Yes	No	Yes	No	Yes
Observations	2040	2040	2040	2040	2040	2040	2040	2040
R-squared	0.835	0.885	0.834	0.878	0.792	0.860	0.792	0.858

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Regression of the share of citations garnered by mostly-female-authored articles, by publication year and academic field (s^f_{av}, on the mostly-female share of author in each field and publication year (n^f_{av}). Even-number columns include field and article vintage fixed effects. Columns (1) and (2) use the citer-gender data, columns (3) and (4) use all citing articles (including those where author gender identification is not possible), columns (5) and (6) use citations *from* female-authored articles, and columns (7) and (8) use citations *from* non-female-authored articles.

Table 5: Regressions of q_{avf} on n_{avf} for patents

	(1) all	(2) all	(3) female citers	(4) female citers	(5) male citers	(6) male citers
mostly-female-inventor share	0.879***	1.118***	2.052***	2.030***	0.755***	0.961***
	(0.0138)	(0.0297)	(0.0539)	(0.0888)	(0.0135)	(0.0293)
Constant	0.00488***	-0.0135***	-0.0304***	-0.0287***	0.0102***	-0.00562*
	(0.00115)	(0.00232)	(0.00449)	(0.00694)	(0.00112)	(0.00228)
Field	Year FE	No	yes	No	yes	No
yes						
Observations	1574	1574	1571	1571	1574	1574
R-squared	0.721	0.768	0.480	0.750	0.666	0.718

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Regression of the share of citations garnered by mostly-female-authored patents, by year and academic field (s^f_{av}, on the mostly-female share of inventors in each field and publication year (n^f_{av}). Even-number columns include field and patent vintage fixed effects. Columns (1) and (2) use the citer-gender data, columns (3) and (4) use citations *from* female-invented patents, and columns (5) and (6) use citations *from* non-female-invented articles.

Table 6: Regressions of q^t_{avf} on n_{avf} using vintage × field × citing year data, articles

	(1) q^t_{avf}	(2) q^t_{avf}	(3) q^{tf}_{avf}	(4) q^{tf}_{avf}	(5) q^{tn}_{avf}	(6) q^{tn}_{avf}
mostly- female-authored share of articles	1.125***	1.101***	1.447***	1.189***	0.875***	0.838***
	(0.00386)	(0.0616)	(0.00700)	(0.0862)	(0.00365)	(0.0574)
Constant	0.00374***	0.00609	0.0480***	0.0735***	0.00998***	0.0136*
	(0.000461)	(0.00589)	(0.000841)	(0.00832)	(0.000436)	(0.00548)
Field & Year FE	no	yes	no	yes	no	yes
Observations	58048	58048	56348	56348	57997	57997
R-squared	0.594	0.621	0.432	0.511	0.498	0.528

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Regression of the share of citations garnered by mostly-female-authored articles, by publication year, academic field, and calendar year (q^t_{avf}), on the mostly-female share of author in each field and publication year (n_{avf}). Even-numbered columns include field, article vintage, and citing year fixed effects. Columns (1) and (2) use the citer-gender data, columns (3) and (4) use citations *from* female-authored articles, and columns (5) and (6) use citations *from* non-female-authored articles.

Table 7: Regressions of q^t_{avf} on n_{avf} using vintage × field × citing year data, patents

	(1) all	(2) all	(3) female citers	(4) female citers	(5) male citers	(6) male citers
nfshare	0.895***	1.118***	1.684***	1.597***	0.780***	0.998***
	(0.0534)	(0.160)	(0.146)	(0.214)	(0.0589)	(0.193)
Constant	0.00490	-0.0108	-0.0122	-0.00615	0.00962*	-0.00587
	(0.00330)	(0.0113)	(0.00776)	(0.0149)	(0.00375)	(0.0136)
Field, citing, cited Year FE	no	yes	no	yes	no	yes
Observations	42351	40811	38478	38478	42230	40690
R-squared	0.155	0.162	0.204	0.313	0.125	0.132

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Regression of the share of citations garnered by mostly-female-authored patents, by publication year, area, and calendar year (q^t_{avf}), on the mostly-female share of inventors in each field and year (n_{avf}). Even-numbered columns include field, patent vintage, and citing year fixed effects. Columns (1) and (2) use the citer-gender data, columns (3) and (4) use citations *from* female-invented patents, and columns (5) and (6) use citations *from* non-female-invented patents.

Table 8: Displacement regressions, articles

	(1) levels	(2) levels	(3) logs	(4) logs
female-authored articles	4.491***	3.368***		
	(0.0503)	(0.0366)		
log female-authored articles			0.812***	0.549***
			(0.00787)	(0.0149)
Constant	8442.6***	13252.7***	3.490***	5.282***
	(577.5)	(306.6)	(0.0558)	(0.102)
Field & Year FE	no	yes	no	yes
Observations	2040	2040	2039	2039
R-squared	0.797	0.953	0.839	0.993

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Dependent variation is the number of articles produced by vintage and academic field. Columns (1) and (2) use levels, and columns (3) and (4) use logs. Independent variable is the number of articles, by vintage and field, with mostly-female authors.

Table 9: Displacement regressions, patents

	(1) levels	(2) levels	(3) logs	(4) logs
female-inventor patents	9.525***	9.675***		
	(0.0720)	(0.0689)		
log female-inventor patents			0.862***	0.888***
			(0.00812)	(0.00747)
Constant	651.2***	598.3***	3.280***	3.142***
	(42.69)	(32.17)	(0.0440)	(0.0397)
Field & Year FE	no	yes	no	yes
Observations	1574	1574	1559	1559
R-squared	0.918	0.971	0.878	0.981

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Dependent variation is the number of patents produced by vintage and area. Columns (1) and (2) use levels, and columns (3) and (4) use logs. Independent variable is the number of articles, by vintage and field, with mostly-female inventors.

Table 10: Market expansion regressions, articles

	(1) levels	(2) logs
extant articles in t	1.056***	
	(0.00975)	
log extant articles in t		1.119***
		(0.0319)
Constant	-190019.7***	-2.533***
	(7962.7)	(0.369)
Field & Year FE	yes	yes
Observations	2000	2000
R-squared	0.932	0.995

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Dependent variation is the number of citations to articles appearing in year t. The explanatory variable of interest is the number of articles available to be cited in year t. These include all articles published from 1965 until year $t-1$. Regressions include academic area and calendar year fixed effects.

Table 11: Market expansion regressions, patents

	(1) levels	(2) logs
N	1.399***	
	(0.0481)	
lN		1.112***
		(0.0142)
Constant	-38521.6***	-1.878***
	(3855.5)	(0.148)
Field & Year FE	yes	yes
Observations	1574	1572
R-squared	0.670	0.989

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Dependent variation is the number of citations to patents appearing in year t. The explanatory variable of interest is the number of patents available to be cited in year t. These include all patents from 1975 until year $t-1$. Regressions include academic area and calendar year fixed effects.

Table 12: First-stage results, articles

	(1) all	(2) female citers	(3) male citers
log extant articles	-1.649***	-1.964***	-1.578***
	(0.0365)	(0.0432)	(0.0351)
female-authored article	-0.00220	0.0627***	-0.0496***
	(0.00114)	(0.00215)	(0.00357)
Observations	10517480	3445325	9464784
R-squared	0.461	0.680	0.472

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Dependent variation is the log of the inside share. Column (1) is all, column (2) is female citing articles, and column (3) is male-authored citing articles. Standard errors are clustered on year.

Table 13: First-stage results, patents

	(1) all	(2) female citers	(3) male citers
log extant articles	-1.306***	-1.479***	-1.298***
	(0.0236)	(0.000443)	(0.000260)
female-authored article	-0.0233***	0.00988***	-0.0288***
	(0.00397)	(0.000735)	(0.000540)
Observations	31814328	4436391	30300665
R-squared	0.374	0.626	0.378

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Dependent variation is the log of the inside share. Column (1) is all, column (2) is female citing patents, and column (3) is male-authored citing patents. Standard errors are clustered on year.

Table 14: Nested logit demand models, articles

	(1) all	(2) all	(3) fem aut	(4) fem aut	(5) non-f aut	(6) non-f aut
female-authored article	0.000168	-0.000633	0.00643***	0.0310***	-0.00362***	-0.0167***
	(0.000172)	(0.000362)	(0.000811)	(0.00150)	(0.000665)	(0.00151)
log inside share	0.851***	0.611***				
	(0.0107)	(0.0119)				
log inside share (female)			0.691***	0.475***		
			(0.0174)	(0.0102)		
log inside share (male)					0.864***	0.647***
					(0.0103)	(0.0132)
IV	no	yes	no	yes	no	yes
Observations	10517480	10517480	3445325	3445325	9464784	9464784
R-squared	0.949	0.871	0.879	0.779	0.958	0.895

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Nested logit models of demand for articles. Odd-numbered columns are OLS, even-number columns instrument the inside share with the log of the number of articles. Columns (1) and (2) use all citing articles. Columns (3) and (4) use majority female-authored citing articles, and columns (5) and (6) use non-majority-female-authored citing articles. All specifications include field and article vintage fixed effects. Standard errors are clustered on calendar year.

Table 15: Nested logit demand models, patents

	(1) all	(2) all	(3) fem aut	(4) fem aut	(5) non-f aut	(6) non-f aut
lsina	0.868***	0.587***				
	(0.0103)	(0.0158)				
(mean) dmostfu_ed	-0.00534***	-0.0157***	-0.0117***	0.00259	-0.00431***	-0.0183***
	(0.000497)	(0.00184)	(0.00277)	(0.00292)	(0.000456)	(0.00178)
lsinf			0.755***	0.575***		
			(0.0148)	(0.0165)		
lsinn					0.868***	0.590***
					(0.0101)	(0.0159)
IV	no	yes	no	yes	no	yes
Observations	31814328	31814328	4436391	4436391	30300665	30300665
R-squared	0.953	0.848	0.922	0.860	0.953	0.851

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Nested logit models of demand for patents. Odd-numbered columns are OLS, even-number columns instrument the inside share with the log of the number of patents. Columns (1) and (2) use all citing articles. Columns (3) and (4) use majority female-authored citing patents, and columns (5) and (6) use non-majority-female-authored citing patents. All specifications include field and patent vintage fixed effects. Standard errors are clustered on calendar year.