Is the USPTO Gender Neutral?*

Hadar Avivi†

November 2, 2023
Newest version here

Abstract

This paper studies the prevalence and evolution of gender bias in the USPTO examination process and assesses the consequences of this bias on economic outcomes. Applying Natural Language Processing tools to patent applications submitted between 2001 and 2013, I estimate gender gaps conditional on the content of the patent application, thereby comparing allowance probabilities between teams of inventors with different gender compositions but similar inventions. Despite a substantial raw gender gap in the probability of initial allowance, I document that there is no average difference in initial allowance rates between mixed-gender and all-male teams. This average masks important heterogeneity. Allowance rates for mixed-gender teams were significantly lower between 2001 and 2003, a gap that shrank to zero by 2005. Gender gaps also vary substantially across examiners, with bias against mixed-gender patents concentrated among senior examiners and bias in favor of women concentrated among young examiners. A mean zero gender gap with positive variance generates economic loss due to the misallocation of granting rights. Building on the methodology of Kogan et al. (2017), I estimate that these biases depressed the value of initially approved patents by at least $1.4 million per year.

*I am thankful to Patrick Kline and Christopher Walters for their invaluable guidance and support on this project. For very helpful comments, I also would like to thank Tslil Aloni, Livia Alfonsi, Amir Bar, David Card, Sydnee Caldwell, Luisa Cefala, Nick Flamang, Ingrid Haegele, Conrad Miller, Enrico Moretti, Sendhil Mullainathan, Tatiana Reyes, Jesse Rothstein, Sophie Sun, Ben Scuderi, Yotam Shem-Tov, Nick Swanson, Damián Vergara, Heidi Williams, Alice Wu, and the participants in the UC Berkeley Labor Lunch Seminar, NBER productivity seminar, SOLE conference, and the GSB Causal Panel Data Conference. Outstanding research assistance was provided by Joy Xie and Hao Wang. I gratefully acknowledge financial support from the Institute for Research on Labor and Employment at UC Berkeley.

†Economics Department, U.C. Berkeley. Email: havivi@berkeley.edu

1 Introduction

Over 200 years have passed since Hannah Wilkinson Slater became the first woman to be granted a patent in the United States.[1] Although female participation rates in the education system and labor market have risen dramatically since then, women nonetheless remain grossly underrepresented in the patenting system.[2] While various explanations have been offered for this persistent gender gap, including differences in occupation choice (Hunt et al., 2013), rejection aversion (Subramani et al., 2021), and lack of exposure to innovation (Bell et al., 2019), the concern that gender bias causes patent examiners to over- or under-value contributions based on inventors' gender persists. Such a concern is particularly pronounced in the US patenting systems in which examiners' discretion has been documented to be consequential in allowance decisions (Frakes and Wasserman, 2014; Sampat and Williams, 2019). Bias in patenting generates not only equity but also efficiency concerns. Misallocation of patent rights could result in long-run real economic consequences as patents and innovative activities drive economic growth (Akcigit et al., 2017; Acemoglu et al., 2018), and are meaningful for inventors' careers (Toivanen and Vaananen, 2012; Kline et al., 2019) and firm profits (Hall et al., 2005; Kogan et al., 2017).

Previous attempts to estimate the importance of gender bias in the patent system suffer from several fundamental problems. Jensen et al. (2018) have documented a gender gap in the likelihood a patent is granted. However, a gender gap in granting doesn't necessarily imply gender bias. The patent-granting process comprises multiple rounds of revisions before a patent is actually granted, and Subramani et al. (2021) show that women are less likely to persist and resubmit their patent applications after early rejections. Efforts to investigate gender differences in the first-round examination decisions have estimated a smaller and insignificant gender gap (Choi et al., 2019). Even so, all previous estimates comparing allowance rates by gender are susceptible to omitted variable bias, as gender gaps could result from a correlation between inventors' gender and patent quality.

[1] Her patent was registered in 1793 and introduced a new method of producing cotton-sewing thread. Interestingly, the inventor's name on the patent was Mrs. Samuel Slater, her husband's name (Khan, 1996; Government, 1888).

[2] In 2016, only 12% of US patent inventors were women, only 21% of granted patents had at least one woman in the inventors' team (Lissoni et al., 2018), and patent granting rates for women are significantly lower than for men (Jensen et al., 2018).

In this paper, I address this concern by exploiting state-of-the-art tools from the Natural Language Processing (NLP) literature to control for the patent application text and content. I analyze whether the inventors' gender, implied by their name on the patent application, affects the outcome of the application process conditional on rich measures of patent quality and content, as captured by the text of the application. I estimate the overall and examiner-level gender disparities in initial allowance decisions and study the characteristics of biased examiners between the years 2001 and 2014.

I encode the patent text by transforming it into *text-embeddings*, a moderate-size vector of text features generated by a pre-trained BERT (Devlin et al., 2018) neural network model that was trained exclusively on nearly the entire corpus of patents by Google (Srebrovic and Yonamine, 2020). I show that the patent text embeddings are highly predictive of both USPTO decisions and inventor gender. Moreover, I provide evidence that conditional on the text embeddings, the gender mix of the inventors and other non-text characteristics are *not* systematically correlated with examiner characteristics and patent classification.

The measure of bias I estimate – the gender gap in allowance conditional on the text – encompasses both disparate treatment and disparate impact. Disparate treatment can arise from models of taste-based discrimination (Becker, 1957), inaccurate stereotypes (Bordalo et al., 2016), or statistical discrimination (Aigner and Cain, 1977). Disparate impact in which examiners discriminate over non-text non-gender application characteristics that are correlated with gender and affect decisions are more closely related to the measure of "indirect discrimination" formalized in Bohren et al. (2022) and studied in the legal system in Arnold et al. (2022). In either case, the bias I measure is policy-relevant as USPTO examiners are legally required to rely primarily on the patent application's text and claims.

I document three novel facts about the distribution of gender bias in the first round of the patent application process. First, while the raw gender gap in initial allowance is substantial, I find that after controlling for the text, the average gender gap entirely disappears. This result is robust to different gender definitions, including thousands of art-unit,[3] year, and class fixed-effects, and alternative matching estimates. Second, I document the evolution of gender bias in patent applications from 2001 to 2013. I find that between 2001 and 2003, patents with mixed-gender authors were 0.8 percentage points less likely to be initially allowed compared to all-male or gender-unknown patents. However, that gap shrank over

[3]An art unit is an examination unit, a group of examiners specializing in a particular technology.

time, and since 2005, it has converged to zero. This finding mirrors the recent evidence from academia indicating that the gender gap shrank over time and turned in favor of women in recent years (Card et al., 2021, 2022). Lastly, I show that, although the mean gender gap is zero, there is substantial variation in gender bias across examiners. After adjusting for sampling error using the leave-out estimator of Kline et al. (2020), the standard deviation of gender bias across examiners is 2.3 percentage points, more than 25 percent of the initial allowance rate for no-female teams. Such two-sided gender discrimination has also been found in audit experiments in the labor market (Arceo-Gomez and Campos-Vazquez, 2014; Kline and Walters, 2020; Kline et al., 2022).

Next, we ask if some groups of examiners are more discriminatory than others. We find that examiner seniority and patent art-unit partially explain the variation in gender bias across examiners. Examiners' start-year at the USPTO explains roughly 25% of the variation in gender bias, with senior examiners more likely to exhibit bias toward mixed-gender patents and younger examiners more likely to be biased against all-male patents. This result is not explained by time effects, ruling out the possibility that examiners modified their behavior over time. The art unit of examiners explains approximately 25% of the variation in gender bias, suggesting a wide across-field variation. Interestingly, while the gender bias against women is negatively correlated with the share of mixed-gender applications in a sub-class, it is positively correlated with that share within art units. Such a pattern is consistent with the finding from the entrance exam for French higher education in Breda and Ly (2015), where male-dominated fields were more biased in favor of women. Exploring the legal ground behind initial rejections reveals that although most rejections are based on a lack of novelty or obviousness, the gender gap in initial allowance is disproportionally correlated with technical rejections based on writing. Previous research argues that such rejections are "simpler" and less time-consuming as they do not require prior-art search (Frakes and Wasserman, 2014).

In line with previous research (Frakes and Wasserman, 2014; Sampat and Williams, 2019), I find substantial variation in examiner discretion when controlling for the patent text embeddings. The standard deviation of examiners' base-level leniency is nine percentage points, slightly higher than the sample's mean initial allowance rate.

Next, I study the variability of gender bias within examiners' characteristics and over time. I find that the standard deviation of male examiners' gender bias is twice the size of the standard deviation of female examiners, suggesting that male examiners pose a higher risk in the system. Similarly, I find that lower levels of variability of discretion and bias also characterize younger cohorts of examiners that joined the USPTO later. Finally, combining the facts that the mean preferences of younger examiners differ from those of senior examiners, I find that over time, the risk of encountering an abnormally biased examiner has increased. This result suggests an increase in polarization within the examiners' community.

Most of the discrimination literature fixates primarily on average gaps, which map only partially to fairness and inefficiencies. First, even though there is no ex-ante bias, heterogeneity in bias undermines ex-post horizontal equity. Second, an exclusive focus on mean bias overlooks the detrimental ramifications of misallocation, which may manifest even if the mean bias is zero. Utilizing Kogan et al. (2017)'s stock market return model for patents, I estimate the effect of bias on the stock market return of publicly traded firms. Since patents' stock market return is observed only for granted patents, I estimate a selection model, exploiting the examiners as instruments, conditional on the patent content. My analysis estimates the lower bound on the annual cost of having a positive variance of bias for patent applications assigned to publicly traded firms to be approximately $1.4 million. This facilitates that the misallocation of patent grants due to gender bias leads to inefficiencies and generates considerable economic costs per year.

This paper contributes to several strands of the literature. First, this paper is the first systemic evaluation of gender bias in the USPTO patent application process that accommodates major identification concerns. Closely related work by Coluccia et al. (2023) provides evidence for racial discrimination in patenting during the 20th century in the US, and Li and Liu (2023) estimates gender gaps in citations. More broadly, this paper contributes to the literature on the lack of recognition and under-representation of female scientists (Rossiter, 1993; Silver et al., 2018; Ellinas et al., 2019; Koffi, 2021), and aligns with literature in economics studying discrimination and how it varies across gate-keepers (Abrams et al., 2012; Arnold et al., 2018, 2020; Feigenberg and Miller, 2022; Kline et al., 2022).

Methodologically, this paper exploits recent advancements in the NLP literature and builds a matching estimator that conditions on the patent application text (for a review on the use of text in economics, see Gentzkow et al., 2019). Recent years have seen growing efforts to exploit the patent data text, where most efforts use the text embeddings to generate other intermediate variables (Acikalin et al., 2022; Suzgun et al., 2022; Li and Liu, 2023).

To my knowledge, this is the first paper that offers a concrete method for controlling for the patent text. Using moderate dimensional embeddings, I demonstrate that these embeddings can balance observables of patent applications across inventing teams with different gender mixes, suggesting that these comparisons are unconfounded. The approach taken in this paper could serve as a useful framework to study questions in other settings where evaluations are mandated to depend upon textual content.

The rest of the paper is organized as follows. Section 2 provides the institutional background, and Section 3 describes how we use the text data for analysis, Section 4 describes the patent data and the text embedding features, and Section 5 outlines the conceptual framework and identification assumptions. Section 6 reports the average gender gap overall and by years and examiners' years of experience. Section 7 documents the variance components of gender bias and its variation across groups. Section 9 provides robustness tests and mechanisms. Section 10 estimates the impacts of bias on market return, and Section 11 concludes.

2 Institutional Background

In this section, I present the institutional background of the U.S. Patent and Trademark Office (USPTO).[4] I describe the process through which patent applications are allocated to examiners and the important role the patent application text plays in the examination process.

2.1 Patent Examination Process

A patent application is a form describing an invention that is being requested for a grant at the patent office. It includes a title, written description, abstract, at least one claim, and, if necessary, some drawings and is usually written by patent attorneys using legal language. Once the USPTO receives the application, it undergoes a pre-examination review to ensure that all necessary forms have been completed and all fees paid. The application claims are then classified and forwarded to the relevant USPTO technology center and art unit[5] for examination.

[4] For a more thorough introduction of the patent examination process see Marco et al. (2017)

[5] Technology Centers are groups of examination units, called art units, divided by a broad technology.

Within an art unit, a supervisory examiner (SPE) assigns the application to a specific examiner, who oversees the application for the remainder of its existence. Previous research argues that, at least in some art units, applications are assigned to examiners randomly (Frakes and Wasserman, 2014; Sampat and Williams, 2019). However, some evidence also indicates within art-unit specialization by class and subclass (Righi and Simcoe, 2019). Importantly for this work, and regardless of the extent to which applications are being randomly assigned within art units, all existing evidence points to the central role patent application content plays in the assignment of applications to examiners.

The assigned examiner then evaluates the application content for compliance with law and regulation. She makes sure that the patent claims include only a single invention,[6] that the claims clearly define the invention, and that the description describes the invention adequately. To determine whether the claimed invention is novel and nonobvious, the examiner also conducts a prior art search by looking for related previous patents or non-patent literature. Based on this examination, the examiner may either allow all claims, an event I refer to as Initial Allowance (IA) or issue an Office action indicating a Non-Final Rejection rejecting or objecting to one or more of the claims made in the application. A typical Non-Final Rejection Office action identifies the specific claims and the grounds on which each is being objected to and/or rejected.

Upon receiving a Non-Final Rejection, the applicant is generally given three months in which to respond. The applicant's response is some combination of arguments and amendments to the claims, usually narrowing their scope. The applicant may also ask for a telephone or in-person interview with the examiner.[7]

2.2 The Role of the Patent Application Text

The patent application text has a central and significant role in the application process. Based on it, patents are classified into classes and subject matters, which determine the

[6] If multiple inventions appear in the claims, the examiner issues a restriction, and the applicant is then required to choose claims drawn to a single invention.

[7] An interview with the applicant or the applicant's attorney for the discussion of the patentability of a pending application will not occur before the first Office action unless the application is a continuing or substitute application. Therefore, in my analysis, I exclude continuation applications and analyze decisions only at the first stage of the application process.

specific technology centers, art units, and examiners to which they are assigned. Upon assignment, the Patent Act defines the criteria for examiners when assessing the patentability of an invention to be based primarily on the information in the patent applications.

An examiner can deny an invention based on several grounds. One is lack of novelty and obviousness, which requires the examiner to make a comparison of the claimed invention with the prior art. Other grounds for denial are missing statutory subject matter, non-usefulness of the proposed invention, and failure of the application to satisfy disclosure requirements, all of which the examiner determines based on the patent text and its content. Although examiners are allowed to use non-patent-text information such as prior art search, non-patent literature, and personal knowledge when making allowance decisions, all such information should be used in conjunction with the patent text.

While each patent application is required to identify the inventors in the applications properly, the inventors' names are not expected to be used for assessing patentability and definitely cannot serve as a ground for rejection. Moreover, the role of uncertainty is also limited in the patent application process. Unlike in academic papers, whose content could change dramatically during the review process, the invention technology is not allowed to be changed after a patent application has been submitted for it. Instead, a patent application should represent a final output. Should inventors seek to change their invention, they must file a new application with a new set of claims. Amendments are allowed only to the scope of the claims. Examiners, of course, can still face uncertainty or have varying skills in identifying high-quality patents. However, any variation attributable to such sources that is systematically associated with inventor gender would be classified in my analysis as a measure of bias, as I outline formally in section 5.

The unique setting of the patent examination process and the important role played by the patent application text lead to the following question: can we perform matching based on the patent application text? In the following sections, I provide a framework to do so and further evidence for its validity.

3 Text as Data

"You shall know a word by the company it keeps!"

- John Firth, "A synopsis of linguistic theory 1930-1955", 1962.

Text is a form of unstructured, high-dimensional data that encodes rich information. This property has led generations of scientists and linguistics to wonder how to properly represent it numerically. In this paper, I follow the state-of-the-art NLP solution of text representation and assume that words can be represented as *word embeddings*, dense real-valued vectors of a finite dimension dwelling in some predefined vector space. Each vector dimension is associated with a feature that represents different aspects of the word, and each word is represented by a point in the vector space such that words with the same meaning are closer to each other.

This approach to text representation is based on the "*Distributional Hypothesis*" from the literature on Distributional Semantics in linguistics, an idea that was first popularized by Firth (1957). It is derived from the semantic theory of language usage, in which lexical items (words, sentences, or paragraphs) that are used and that occur in similar contexts tend to convey similar meanings. This representation, which has led to unprecedented success in multiple AI tasks, is thus considered one of the key breakthroughs in NLP research in recent years.

In this section, I describe in detail how I use the neural network language model BERT (Devlin et al., 2018) to convert the high-dimensional patent application text inputs into text embeddings, thereby creating a list of p continuous covariates $x_{i1}, ..., x_{ip}$ of moderate size.

3.1 Contextualized Word Embeddings Using BERT Model

The Bidirectional Encoder Representations from Transformers, known as the BERT model (Devlin et al., 2018), is a deep neural network model trained simultaneously to predict two unsupervised tasks. The first is to randomly predict a certain percentage of the words in each paragraph, and, given a pair of sentences, the second is to predict whether one sentence proceeds another. In this section, I briefly describe the BERT model's main structure and how it generates word embeddings. See the original paper (Devlin et al., 2018) for more detailed information.

The BERT model is a neural network model, a statistical model inspired by neuroscience, which allows for high levels of dependencies and nonlinearities between inputs. Illustrated in appendix Figure A.1, any neural network model can be represented as a graph comprising input, hidden, and output layers. Each node in the graph is called a neuron, and each edge represents an estimated parameter such that each neuron is a function of the linear combination of the neurons from the previous layer. In the BERT model, the input layer is a piece of text, represented by a dummy for each word, and each word dummy is then mapped to several hidden layers of a moderate dimension. Between layers, different words are connected using a special attention layer architecture (Vaswani et al., 2017). Each model's hidden layer is considered to be an embedding vector, representing a word in a lower dimensional space.

To prepare the input of the BERT model, each word in the input text is broken into subwords and tokenized, and a special token, termed the [CLS] token, is added to the beginning of the text. Then, in the last layer, the model predicts the two tasks described above: predicting the masked word using its embedding representation and predicting the next sentence using the last layer embedding of the special [CLS] token.

The outputs of a pretrained model are the estimated parameters of each neural network layer. As such, there are several possible options to generate word embeddings. Devlin et al. (2018) suggest using BERT embeddings in various ways, including the last, second-to-last, or concatenating the last four layers. Moreover, to represent a document comprised of many paragraphs, it is customized to use either the average word embeddings in the paragraph or the embeddings of the first special [CLS] token as an aggregate representation of the entire paragraph. In this work, I use the CLS token as a text representation. In section 4.2, I compare different combinations of embedding layers and proceed with those that attain the highest predictive power subject to my machine memory constraint.

To generate the word embeddings, I utilize a pretrained large BERT model that was trained exclusively on over 100 million patents by Google (Srebrovic and Yonamine, 2020) and on all the parts of the text, including the abstract, description, and claims. The model has 24 hidden layers, each embedding vector is of a size of 1024, and each input text could have up to 512 tokens. Since the language model was estimated exclusively on patents, the estimated parameters represent the language structure, word distribution, and unique semantics of the specific domain of patents.

The BERT word embeddings, among other deep neural network NLP models, are known for their high predictive power. Artificial Intelligence (AI) and data science applications have long used the word embeddings generated using BERT or similar models. In Section 4.2, I show that also in this setting, the patent application text embeddings have high predictive power for the contents, allowance probability, and the decisions of the examiners.

4 Data

The primary data source employed in this research is the USPTO Patent Examination Research Dataset (Graham et al., 2015) which includes the universe of all public patent applications available online in the Public Patent Application Information Retrieval system (Public PAIR).[8] I restrict the sample to utility[9] patent applications filed after November 29th, 2000[10] and before January 1st, 2014. For every patent application, the Public PAIR data includes information on inventors' first and last names and additional variables such as country, application number, publication number, the grant date if granted, and examiners, art-unit, and technological classes and sub-classes identifiers. For gender name coding, I relied on the gender distributions of first names provided by the U.S. Social Security Administration. To avoid detecting differential behavior to non-US inventors, I include only patent applications submitted by US inventors.

I merged this dataset with several other datasets: 1) The USPTO Patents View data, which includes detailed information on both granted patents and patent applications. Specifically, this source includes the list of all the patents it cites, including granted and non-granted patents and a predicted identifier for inventors. 2) The Patent Claims Research Dataset (Marco et al., 2019) from which I obtain detailed information on the number of claims per patent, claim text, and the change in the claims between application to granting for granted patents. 3) "Google Patents Research Data" from which I merged the abstract and description text of each patent application. 4) Examiners' roster, pay scale, and education levels from Frakes and Wasserman (2014) Freedom of Information Act request. 5) Kogan et al. (2017) patent market value data, which run event studies to estimate the excess stock market return realized on the grant date of patents assigned to publicly traded firms. 6) USPTO Office Action Rejection, which documents the grounds of rejections for all rejected patent applications from 2008 to 2017. For additional information about the patent data, see Appendix section B.

[8] The data can be found here:

[9] Utility patents are granted for the "invention of a new and useful process, machine, manufacture, or composition of matter" (USPTO 2010).

[10] Since the American Inventors Protection Act of 1999, almost all the USPTO patent applications filed after November 29th, 2000 were published online, regardless of whether they are granted or not.

Applicants and examiner characteristics: I assume that examiners imply the gender of inventors based on their names. Therefore, to classify inventors' gender, my procedure relies on the assumption that examiners located in the US are likely to infer the gender of names common in the US. Hence, my gender coding procedure assigns gender based on the distribution of the first name by gender provided by the US Security Administration (SSA).[11] Using that method, I could assign gender to 75% of the names in my final sample. Names not available in the SSA tables are considered foreign names with unknown gender.

Unlike with inventors' gender coding, my goal is to assign gender to all possible examiners. Therefore, I use a collection of sources to identify the maximum amount of names, including foreign-sounding names. Specifically, in addition to the US SSA administrative tables, I use the name gender published by the United Kingdom Intellectual Property Organization, the name gender dataset published by the World Intellectual Property Organization (WIPO), the gender-guesser Python package, and the Genderize.io program. Using all these sources, I successfully identified 85% of the examiners' names.

Examiners' years of experience and education level are based on the FOIA roster tables provided by Frakes and Wasserman (2014), which date back to 1992 and end in 2012. To account for the accurate years of experience of examiners who joined the USPTO before 1992 or after 2012, I combine the information of the first office action and validate my approach using the administrative records. For more information, see appendix section B.

4.1 Descriptive Statistics

Table 1 provides summary statistics of the main sample of patent applications. The baseline sample includes over 1.2 million patent applications; only 16% includes at least one female inventor, and in only 11% of which there is a female inventor as a first or second author.[12] The average team has 2.5 inventors, whereas teams with at least one female inventor are bigger. 35% of the patent applications are of sole inventors, most of them male inventors. The raw gender gaps in initial allowance and grants are presented in the last two rows of table 1. An initial allowance is a rare event: only 8.3% of the patent applications are allowed in the first round of examination. On the other hand, 64.6% of the patent applications

[11] https://www.ssa.gov/oact/babynames/limits.html

[12] The inventors' rank in the inventors' list doesn't necessarily indicate the contribution to the invention. Nevertheless, I use this measure to account for the visibility of females in the inventors' team.

are eventually granted. Patent applications with at least one female inventor are almost 2 percentage points less likely to find their patent initially allowed (23% compared to all male or unknown patent applications) and 6.1 percentage points less likely to eventually grant their patents (10% compared to all male of unknown grant rate).

Table 2 reports the main descriptive statistics of the patent examiners in my analysis. I have a total of 8,550 examiners; 2,055 of them are classified as female, 5,128 as male, and for 1,367, I could not code their gender based on their name. The mean examiner-level initial allowance rate is 6.7 percentage points, with male examiners having a 1.9 higher allowance probability. In addition, 20 percent of the examiners in my sample joined the USPTO before 1995, 31 percent joined between 1996-2001, and the rest joined after the year 2001.

4.2 Selection and Visualization of the Patent Text Embeddings

The BERT model has several layers, each potentially serving as text embeddings. To choose my final list of embeddings, I adhere to the following protocol: For every part of the patent application text, description, and claims, I generate a unique embedding vector using the CLS token. Since the BERT text input is constrained to at most 512 tokens, I parse each piece of text into paragraphs and generate the CLS embeddings of each paragraph. The final embedding representation of each part of the text is the average across all the paragraphs. Then, following the recommendations in Devlin et al. (2018), I generate the embeddings vector for every part of the text using the last, second-to-last, and third-to-last layers. This procedure generates (3×2) six possible embedding vectors, each of a size of 1023 features.[13]

The preferred set of embeddings is selected to be the ones that achieve the biggest improvement in adjusted R^2 when predicting Initial Allowance (IA) and any female (F) in linear regression, subject to my machine memory constraint.[14] Results of the adjusted R^2 are presented in Figure 1.[15] Darker circles indicate that more embedding vectors were included, and the "X" point represents the adjusted R^2 from regressing IA and any female on the set of art-unit-year and class fixed-effects usually used in the literature to control for

[13]Each original embedding vector has 1024 features that are mapped into the probability space and sum to one. Therefore, in our analysis, we drop one of the features to avoid multicollinearity.

[14]The analysis in this paper runs on a high-performance computing cluster with 768 GB of memory.

[15]For a detailed description of the adjusted R^2 values of each combination embeddings vector see appendix table A.1

confounding (e.g, Jensen et al., 2018; Choi et al., 2019).

Figure 1 shows that all the sets of embedding vectors dominate the art-unit-year and class fixed effects in their predictive power of mixed-gender teams. Moreover, embeddings that include at least two vectors are more likely to dominate the fixed effects in predicting IA. This result is even more striking when realizing the fixed-effects control for around 8,000 covariates. Second, it seems that the improvement in the predictive power is marginal beyond the combination of two embeddings, one from each part of the text, description and claims. Therefore, as my preferred set of embeddings, I use the second layer of the claims and the description text. To further explore the superiority of embeddings on the art-unit-year fixed effect, Figure 2 presents the partial adjusted R^2 within art-unit-year and class fixed effects and examiner fixed effect.[16] This Figure emphasizes that the embedding vector introduces new information beyond the text's classification to fine categories of classes and sub-classes.

Predicting quality: To assess whether the embedding vector encodes information about the content and the quality of the text assumption, Figure 3 presents the split sample binned scatter plot of the predictions of rejections based on novelty, rejections based on obviousness, log number of citations, and log Kogan et al. (2017)'s market return. The Figures provide clear evidence that the embeddings I use have strong predictive power to quality proxy variables.

Describing embeddings: To describe the embeddings vector, I perform two exercises. First, I use the Uniform Manifold Approximation and Projection (UMAP) algorithm (McInnes et al., 2018) to reduce the dimensionality of the embeddings vector.[17] Figure 4 displays the 2-dimension UMAP reduction on a random sample of 50% of the patent applications where different colors encode different technology centers. It shows that although the embeddings were not trained directly to predict technology centers and fields, they are able to predict

[16] Partial R^2 measures the proportion of variation explained by the embeddings in a model with both fixed effects and embeddings that cannot be explained by the fixed effects alone. It is measured as $R^2_{partial} = \frac{SSE(FE)-SSE(X_i,FE)}{SSE(FE)} \times \frac{N-K_{FE}}{N-K_{FE}-K_{X_i}}$ where $SSE(X)$ is the sum of squared errors for a model that controls for X.

[17] UMAP is a non-linear dimension reduction algorithm that assumes the data is drawn from Riemannian manifold. Compared to other techniques, it has a more efficient estimation algorithm and provides a better representation of the underlying structure.

these dimensions. As a continuous index, it is evident that the embeddings can provide richer information than just field indicators. They inform us, for example, which patent applications from the mechanical engineering technology center are more similar to those from the computers and communication technology centers and which others are more similar to those from the chemical engineering technology center. Such similarity, which speaks to how we grasp ideas as a continuous domain, could not be captured with dichotomous indicators of art units or classes.[18]

Second, I describe the distribution of the text embeddings across gender groups by running the following regression:

$$IA_i = \alpha + \beta F_i + C_i'\gamma + \epsilon_i,$$

where IA_i is an indicator for initial allowance, F_i is an indicator for having at least one woman in the inventors' team, and C_i is the embedding vector. We can think of the embeddings index, $C_i'\hat{\gamma}$, as a quality index of the patent application. Figure 5 presents the distribution of the embedding index by the inventors' team gender for the entire sample of patent applications, and Figure 5 the distribution by technology centers. The distribution of the embedding index of mixed-gender applications is shifted to the left, which could suggest that there are more low-quality mixed-gender patents than all-male patents. Interestingly, when exploring the distribution of $x'\hat{\gamma}$ by technology center, we find that in the technology centers with the highest mixed-gender applications share, biotechnology, the two distributions are identical. Such a pattern could be explained by either a Roy-type model where women sort into technologies based on comparative advantage or a model with spillovers where women's productivity increases with the share of women.

As mentioned in section 5, this pattern could also reflect a bias based on the patent's content. For example, a case where examiners dislike specific topics and women are more likely to write patents on these topics could result in such a figure. In this paper, I analyze bias holding the content evaluation as constant. Studying biases with respect to writing styles and topics is an interesting topic for future research beyond the scope of this paper.

[18]In fact, using dummy variables to describe the text practically implies describing it with independent measures, enforcing the dummy variable of mechanical engineering to be independent of the dummy variable of computers.

5 Conceptual Framework

Consider a population of patent applications indexed by i. Each patent application is characterized by a patent application document T_i, describing textually the content of the patent, and a gender composition of the inventors' team F_i. To ease notation, I refer to F_i as an indicator that equals one if there is at least one woman in the inventors' team. In my analysis, I also explore other fractional versions of F_i, which do not have a meaningful effect on the definitions described below. For every patent application, I observe an indicator, IA_i, which equals 1 if the patent application was initially allowed in the first round of the examination process.

Each patent application is assigned to an examiner $j \in \{1, ..., J\}$ where the function $J(i)$ indicates the examiner to which application i was assigned, and $Z_{ij} = 1$ whenever $J(i) = j$. The goal of each examiner in the examination process is to decide the patentability of each application by evaluating the patent's text and assessing its quality. Following the exposition in section 3, I assume that the patent content can be represented numerically using a continuous vector of a finite dimension $C_i \in \mathbb{R}^p$.[19] My goal in this paper is to study the gender differences in allowance decisions conditional on C_i, which includes all the relevant information about the content of the patent that is required for the application process.

I assume that examiners' decision is a function of the patent content, $C_i \in \mathbb{R}^p$, the patent gender $F_i \in \{0, 1\}$, and $U_{ij} \in \mathbb{R}^l$ which captures all the other factors that affect examiner j decision other than the patent content and inventors' gender. Rather than modeling directly the examiners' decision, I assume that there exists a function $IA_j(c, f, u)$ that represents the examiner's j decision rule, such that $IA_i = IA_{J(i)}(C_i, F_i, U_{iJ(i)})$. Therefore, for every application i, the potential first-round examination outcome from assignment to examiner j is $IA_{ij} = IA_j(C_i, F_i, U_{ij})$.

For every examiner j, I measure the gender gaps at the examiner level as:

$$\beta_j = \int \omega(c)(\mathbb{E}[IA_{ij}|F_i = 0, C_i = c] - \mathbb{E}, [IA_{ij}|F_i = 1, C_i = c]])dG(c). \tag{1}$$

[19] Denote Σ^*, the set of all possible combinations of words (up to a certain number of words per piece of text). The real-valued vector $C(T_i) \equiv C_i \in \mathbb{R}^p$ is the *text embeddings* of T_i if, for every text T_i, T_j and T_l, T_i's content is more similar to T_j rather than T_l, if and only if the distance between C_i and C_j is smaller than the distance between C_i and C_l.

I am also interested in the overall gender gaps in the system, which I define as:

$$\beta = \int \omega(c)(\mathbb{E}[IA_i|F_i = 0, C_i = c] - \mathbb{E}[IA_i|F_i = 1, C_i = c])dG(c), \tag{2}$$

where integrals are taken over the distribution of the patent application content $G(c)$ in the overall population, and $\omega(c)$ are the weights of the estimand of interest. For example, if $\omega(c) = 1$, β is the well-known Average Treatment Effect (ATE). Other weights correspond to other possible weighted averages of the gender gap in the population. If $\mathbb{E}[IA_i|F_i = 0, C_i = c] - \mathbb{E}[IA_i|F_i = 1, C_i = c]$ does not change with the patent content $C_i = c$, then β is the same regardless of the chosen weights. β_j represents the tendency of examiner j decision to vary by the gender of the inventors' team, and β is the system-wide gender gap conditional on the patent application content.

Throughout the analysis, I will refer to the content-adjusted gender gaps, β_j and β, as gender bias. This measure of gender bias is motivated by the legal mandate that requires patentability decisions to be based on the content and meaning of the patent invention. Therefore, systematic variation in gender gaps among patents of the same content represents a deviation from the stated mandate of the Patent Act. Such deviations could originate from various sources, which I discuss below.

5.1 Sources of Gender Gaps

To rationalize examiners' behavior, suppose that each patent application is also characterized by latent quality $q_i^* \in \mathcal{Q}^* \subseteq \mathbb{R}$. The utility of each examiner, $u(d, q, f, u)$, depends on her allowance decision $d \in \{0, 1\}$, the patents' quality $q \in \mathcal{Q}^*$, and potentially the gender indicator $f \in \{0, 1\}$ or the other characteristics $u \in \mathbb{R}^l$. Since the true quality of each patent application is unobserved, the examiners form beliefs about the distribution of quality given the observed signals: the patent content C_i, inventors' gender $F_i = f$, and the other factors used by the examiner to form decision $U_{ij} = u$. Specifically, I denote by $\tilde{\mathcal{F}}_{f,u}(q)$ the examiner prior distribution. Assuming examiners are rational, they form decisions based on the respective posterior distribution $\tilde{\mathcal{F}}_{f,u}(q|c)$ for the quality of the patent after observing the application content c, and the other characteristics f and u. Therefore, the initial allowance decision maximizes the expected utility of examiners, where the expectation is taken over the posterior distribution:

$$IA(c,f,u) = arg\max_{d\in 0,1} {}_q u(d,q,f,u) d\tilde{\mathcal{F}}_{f,u}(q|c).$$

We allow the beliefs to be different than the true prior distribution of quality $\mathcal{F}_{f,u}(q)$, with the respectful posterior distribution $\mathcal{F}_{f,u}(q|c)$ given observables (c,f,u)

Disparate treatment - any decision rule that depends directly on the inventors' gender, either through preferences or through beliefs, represents a direct discriminatory behavior. It results in a decision rule such that for every $c \in \mathbb{R}^p$, and $u \in \mathbb{R}^l$, $IA(c,1,u) \neq IA(c,0,u)$, and would generate $\beta_j \neq 0$.

The canonical model of disparate treatment is taste-base discrimination where examiners' utility $u(d,q,f,u)$ varies directly with inventors' gender (Becker, 1957). Another form of bias is statistical discrimination, which arises when examiners form beliefs based on the true prior distribution $\mathcal{F}_{f,u}(q)$, and this distribution varies by gender (Aigner and Cain, 1977). Lastly, bias could arise from inaccurate beliefs about the distribution of quality, which varies systematically by inventors' gender $\tilde{\mathcal{F}}_{f,u}(q) \neq \mathcal{F}_{f,u}(q)$ (Bordalo et al., 2016).

Disparate impact - gender gaps could arise additionally due to a relationship between the other unobserved factors that affect examiners' decision U_{ij} and the inventors' gender, as long as they influence allowance directly. For example, if examiners directly discriminate based on ethnicity, and ethnicity is correlated with the gender mix of the inventors, we then find $\beta \neq 0$. I cannot directly disentangle gaps based on this source vs. disparate treatment since inventors' genders are not randomly assigned in my design. Nevertheless, in Section 9.2, I find limited evidence that other non-text characteristics, such as ethnicity and country of origin, explain the observed β.

An additional example of disparate impact that could arise in my setting is disparities due to examiners' differential tastes in writing styles (Levitskaya et al., 2022). Since the mapping from text to content is not one-to-one, the same invention could be written using a different combination of words, disparities could also arise from differential tastes in writing styles. The extent to which the BERT text embeddings control for such features is an interesting avenue for future research.

5.2 Identification

The embedding representation of the patent content, $C(T_i) \equiv C_i \in \mathbb{R}^p$, is latent, where even the correct dimension, p, of the vector space is unknown. Instead, I observe the estimated BERT text embeddings $\hat{C}(T_i) \equiv \hat{C}_i \in \mathbb{R}^{p'}$, where I assume that $p' \leq p$. Identification of (2) is premised on the assumption that the BERT embeddings serve as reliable controls that sufficiently cover all the relevant information for patentability assessment:

Assumption A1 - $\hat{C}$ *is sufficient for* C. For every application variable $Y_i \in \{IA_i, F_i\}$

$$E[Y_i|C_i, \hat{C}_i] = E[Y_i|\hat{C}_i]$$

This assumption does not require the BERT embeddings to capture all the information encoded in the text, but the deviations from such information do not relate systematically to allowance decisions or inventors' gender. While this assumption cannot be directly tested, I provide evidence that the patent embeddings have a strong predictive power of examiners' decisions, patent quality proxies, and inventors' gender.

Under Assumption A1 and with a large number of observations, the overall gender bias estimand (2) is identified by:

$$\hat{\beta} = \int \omega(\hat{c})(\mathbb{E}[IA_i|F_i = 0, \hat{C}_i = \hat{c}] - \mathbb{E}[IA_i|F_i = 1, \hat{C}_i = \hat{c}])d\hat{G}(\hat{c}),$$

where $\hat{G}(\cdot)$ is the distribution of the BERT embbedings vector $\hat{C}_i$.

With infinite data, (2) is non-parametrically identified. Otherwise, additional parametric assumptions are required. In the analysis that follows, my preferred estimate is the Ordinary Least Square (OLS) coefficient of a female indicator, controlling linearly for the text representation, and I show in section 5.3 that the results are robust to other matching techniques.

Heterogeneity across examiners could either reflect selection patterns of applications to examiners or heterogeneity in examiners' preferences. Therefore, without further assumptions, β_j, which was defined based on examiners' allowance behavior, is not identified. We turn now to introduce an additional assumption on the nature of the assignment of patent applications to examiners conditional on the patent content that ensures identification of examiner-level behavior.

Assumption A2 *Conditional Independence:*

$$U_{ij}, F_i \perp Z_{ij}|C_i$$

This assumption requires that, conditional on the patent application content, there is no systemic correlation between examiners' assignment and non-text features that affect decisions, including inventors' gender. It does not require random assignment of examiners conditional on the text. For example, we know that some examiners do not work in the USPTO for some years; hence, the sample assignment probabilities will vary by year. However, this assumption requires that after controlling for the content of the patent application, the time of assignment has no meaningful effect on the decision process. This assumption relies on the nature of the assignment process depicted in Section 2, and in Section 5.3, I present several tests justifying its validity.

Combining assumptions A1 A2, and with large sample, each β_j is nonparamterically identified as:

$$\hat{\beta}_j = \int \omega(\hat{c})(\mathbb{E}[IA_i|F_i = 0, \hat{C}_i = \hat{c}, Z_{ij} = 1] - \mathbb{E}[IA_i|F_i = 1, \hat{C}_i = \hat{c}, Z_{ij} = 1])d\hat{G}(\hat{c}),$$

Similarly, my main analysis is restricted to the OLS coefficients on the interaction of gender and examiner indicators, and I show that the first two moments that describe the distribution of β_j are robust to other matching techniques.

5.3 Justifying Identification Assumptions

5.3.1 Balance Tests

Assumption A1 requires that sufficient information on patent content is encoded in the BERT embeddings $\hat{C}_i$, and that the BERT embeddings error is not systematically correlated with the gender of the inventors' team after controlling for $\hat{C}_i$. To test this assumption, I run an ordinary least squares (OLS) regression of patents' content proxies, such as technology center indicators and patent claims statistics, which are known to be correlated with patent scope and content (Lanjouw and Schankerman, 2001; Marco et al., 2019).

Figure 7a displays the relationship between technology centers and whether there is at least one female inventor in the inventors' team with and without controlling for text embeddings. The gray dots display the uncontrolled relationship between mixed-gender patents and each technology center, and the blue ones display the relationship conditional on embeddings. Although mixed-gender applications are not equally distributed across technology centers, the text embeddings absorb that variation. Figures 7b, 7c, present the same exercise with other different measures for the "femaleness" of the patent application.

Accordingly, Figure 8 plots the relationship between the gender of the inventors' team with patent application claim statistics: the average and the minimum word count in all the claims and independent claims.[20] Similarly, we find that the BERT embeddings balance the relationship between claim counts and inventors' gender.

Assumption A2 states that examiners' assignments and characteristics should not be systematically correlated with any non-text characteristics. Figure 9 tests that assumption by displaying the relationship between a series of examiner characteristics with several non-text characteristics: having at least one female in the inventors' team, having at least one Asian name in the inventors' team, and the share of females in the attorney team. The first examiner characteristic in each Figure is the leave one out of initial allowance rate, which is often used in the literature to test random assignment of judges to cases (Arnold et al., 2018; Dobbie et al., 2018; Arnold et al., 2020). I also present the relationship with other observable examiner characteristics such as examiner education, gender, ethnicity, and years of experience. Besides a negligible relationship between examiners with 0-1 years of experience and female examiners, I do not find any statistically significant relationship between mixed gender patents and examiner characteristics after controlling for the text embedding, even though the uncontrolled relationship is not zero. Appendix Figure A.3 verifies that the results are robust to other definitions of the femaleness of the patent application and presents the same balance tests using the share of females in the inventors' team and whether the first or second inventors in the inventors' list is female.

Lastly, one might be worried that the distribution of patent text and content changes over time. If this is the case, the embeddings don't properly control all the information on the patents in different years. We repeat the same exercise of estimating the relationship between inventors' gender and the characteristics of the patent application and the examiner in Appendix Figure A.2 by a five-year time interval. In line with previous findings, we find that examiner gender is balanced across these characteristics within years.

[20] A claim may be written in independent or dependent form. An independent claim is a standalone claim that contains all the limitations necessary to define an invention. A dependent claim must refer to a claim previously set forth and must further limit that claim

5.3.2 Omitted Variable Bias - Long and Short Regressions

Under assumptions A1 and A2, conditional on the text embeddings, the examiner-level gender bias should be invariant to the inclusion of other non-text characteristics that are potentially correlated with examination outcome. To assess that hypothesis, I run the following "short" regression:

$$IA_i = \alpha_{j(i)} + \beta_{j(i)} F_i + C_i' \eta + \epsilon_i \tag{3}$$

and "long" regression

$$IA_i = \tilde{\alpha}_{j(i)} + \tilde{\beta}_{j(i)} F_i + C_i' \tilde{\eta} + w_i \gamma + \epsilon_i \tag{4}$$

where IA_i is an indicator for initial allowance, F_i is an indicator for mixed gender patent, w_i includes patent attorney gender, ethnicity, and experience, team size, and indicator for foreign priority, α_j and $\tilde{\alpha}_j$ are examiner fixed effects measuring examiners base leniency level, and β_j are examiner level tendency to prefer mixed-gender patents.

Under a constant effect assumption of α_j and β_j, the estimates of α_j and β_j from the short regression in equation 3 should not be sensitive to the inclusion of the additional controls w_i.[21] Figure 10 presents the main findings. Panel (a) displays the relationship between the examiner leniency levels $\hat{\alpha}_j$, and the controls index $w_i \hat{\gamma}$ estimated in the long regression of equation 4, and panel b presents the same relationship but with the examiner gender bias $\hat{\beta}_j$. $\hat{\alpha}_j$ and $\hat{\beta}_j$ were estimated in the "short" regression of equation 3, and Appendix Table **??** verifies that these omitted variables are predictive of initial allowance. The results suggest that examiner effects are uncorrelated with the controls index. The magnitude of each slope coefficient is tiny, where the slope coefficient for α_j is 0.0002, and the slope for β_j is -0.0002.[22]

[21] Sensitivity to the inclusion of additional controls could also reflect treatment effect heterogeneity as additional controls change the weights of each examiner effect and examiner bias.

[22] This implies that the impact of a one standard deviation increase in examiner leniency on initial allowance (std=9 percentage points, presented in table 5) may be biased by $0.09 \times 0.0002 = 0.000018$ due to omitted variables, and a one standard deviation increase in examiner gender bias on initial allowance (std=2.3, presented in table 5) may be biased by $0.023 \times 0.0002 = 0.0000046$.

6 Average Gender Gap in Initial Allowance

Table 3 investigates the average observed gender disparities in initial allowance rates in the USPTO. The main analysis includes estimating OLS regressions of the form:

$$IA_i = \alpha + \beta F_i + x_i' \gamma + \epsilon_i \tag{5}$$

where F_i is a variable characterising the "femaleness" of the inventors' team, and x_i is a vector of controls. Column 1 of Table 3 omits x_i, column 2 includes art-unit-year and class fixed effects, the commonly used controls in the literature to account for confounding, column 3 controls for the preferred set of 2,046 embeddings selected in section 4.2, and column 4 includes both the embeddings and the art-unit-year and class fixed-effects. In panel (i), F_i is an indicator for having at least one female in the inventors' team, in panel (ii), F_i measures the proportion of females, and in (iii), it is an indicator for having a female ranked first or second in the inventors' list.

The raw gender gap in initial allowance between mixed gender teams and no female teams is substantial and accounts for two percentage points with standard errors of 0.0006. The initial allowance gender gap is around 25 percent of the mean no-female initial allowance rate. After controlling for the fixed effects, the gap falls to 0.35 percentage points (SE= 0.0006). Controlling for the text embeddings shrinks the gender bias to 0.001, making it statistically insignificant from zero (SE= 0.0007). The estimate does not change qualitatively when also controlling for the fixed effects, which indicates the robustness of results to omitted variables. Measuring the gender gap using the proportion of females in the inventors' team or whether a female author is ranked first or second presents similar qualitative results of zero gender gap controlling the text embeddings. Appendix table A.2 displays the same regression results using a restricted sample and comparing only female vs. no female patents teams and single author patents.

OLS is one of many types of matching estimators. Under treatment effect heterogeneity, it need not coincide with other reweighting estimators. Table 4 assesses the sensitivity of my estimated mean gender gap to different reweighting schemes. Column 1 presents the unexplained component of an Oaxaca-Blinder (Oaxaca, 1973; Blinder, 1973) analysis by running a model of IA on embeddings only for the no-female applications and reports the difference between the mean men allowance rate and the fitted values for the men's regression on the mean embeddings vector of mixed gender patents. Kline (2011) shows that this imputation estimator equals the Average Treatment on the Treated (ATT) estimator if either the propensity score or the outcome equation is linear in embeddings. Columns 2 and 3 report the Inverse probability Weighting (IPW) estimate of the ATE and ATT. Column 4 reports the Doubly Robust Machine Learning (DML) Chernozhukov et al. (2018) partially linear regression model implemented using the Python package DOUBLEML, where initial allowance and female indicators are predicted with a neural network (NN) model.[23] DML estimator will yield more precise estimates without compromising consistency under an additional assumption of sparsity of the BERT embeddings in the outcome and gender equations. Table 4 suggests that the mean gender gap is robust to the reweighting scheme, where the estimates of the gender gap in all the models are neither statistically significant from zero nor significantly different from the gender gap estimated by OLS in Table 3.

Gender Bias Over Time

The patent application data includes patent applications filed from 2001, allowing an estimation of the evolution of the gender gap over time. I do so by estimating the following OLS regression:

$$IA_i = \sum_{t=2001}^{2014} \tau_t \cdot \mathbb{1}\{\text{year}_i = t\} + \sum_{t=2001}^{2014} \beta_t \cdot F_i \cdot \mathbb{1}\{\text{year}_i = t\} + C_i' \gamma + \epsilon_i$$

where β_t, plotted in Figure 11, is the main coefficient of interest measuring the gender gap in initial allowance for the patent applications filed in year t, year_i is the filing year of application i, and C_i are the tex embeddings. The blue estimates represent the uncontrolled gender gaps, documenting a substantial gender gap that fluctuated over time. The purple dots are my preferred estimates of the gaps after controlling for the text embeddings, and the green dots verify that the results are robust to the inclusion of art-unit-year and class fixed-effects as controls.

The controlled gaps show that while at the beginning of the 2000s, mixed-gender teams were significantly less likely to find their patent application initially allowed, the system-wide gender gap converged towards zero over time. Similar trends over time have also been found in the selection of Fellows of the Econometrics Society (Card et al., 2021). Conditional on achievements, the historical gender gap against women economists was substantial and

[23]The neural network's number of layers, nodes in each layer and the regularization parameter are chosen by cross-validation.

significant, but it shrunk to zero between 1980 to 2010 and has become positive in recent years.

Gender Bias by Examiner Years of experience

Figure 12 investigates the variation in gender gaps across examiners' years of experience using my preferred model controlling for word embeddings. Each point in the Figure presents the estimated gender gap and confidence interval by years of experience bins. Although the average gender gap in the sample is qualitatively zero, Figure 12 shows that there is substantial heterogeneity across examiners by years of experience. On the one hand, examiners with up to 5 years of experience are significantly more likely to be biased against men, having a 1 to 0.7 higher probability of allowing patent applications with at least one female in the inventors' team. On the other hand, examiners with more than 12 years of experience are significantly more likely to be biased against mixed-gender teams. Variation by years of experience could reflect either an age effect, where as examiners accumulate more years of experience they tend to be more biased against women, or a cohort effect, where different cohorts have different tastes or gender stereotypes. I formally test these two hypotheses in the next section.

7 Variation in Gender Gap Across Examiners

The previous section reveals a zero average gender gap together with heterogeneity across examiners' years of experience and some variation over time. In this section, I quantify this variation and measure the heterogeneity across examiners, examiners' start year, and art units.

7.1 Targeting the Variance

Under the identification assumptions outlined in section 5.2 and as long as examiners' parameters are not essentially heterogeneous with respect to the patent text, we can estimate β_j by running the following fixed effect regression:

$$IA_i = \alpha_{J(i)} + \beta_{J(i)} F_i + C_i' \gamma + \epsilon_i, \tag{6}$$

where F_i is a measure of the femaleness of the patent's inventors, usually an indicator for a mixed-gender team, α_j is the examiner base-level initial allowance rate, β_j is the examiner-level gender bias, and C_i are the 2,048 text embeddings representing the patent application text. I summarise the variability of examiner level leniency and gender bias with the following target variance parameters: $\sigma_\alpha, \sigma_\beta$, the sample standard deviations of α_j and β_j, respectively, across examiners, weighted by the examiner total number of patent applications, and $Corr(\alpha, \beta) = \frac{\sigma_{\alpha,\beta}}{\sigma_\alpha \sigma_\beta}$, the correlation between examiner leniency and bias. Likewise, to quantify the share of variation driven by the start year and art units of examiners, I estimate equation 6 and the α and β parameters across examiners start-year and art units, and their respective variance components.

7.2 Estimation

I briefly review the Kline et al. (2020) leave-out estimation procedure which enables consistent estimation of the variance of β_j and α_j in the presence of unrestricted heterogeneity that grows with the number of regressors. For further details, see Appendix C. Using a matrix representation, any variance component can be written in a quadratic form:

$$\sigma^2 = \delta' A \delta$$

where $\delta = (\alpha', \beta')'$ are the collected $\alpha = (\alpha_1, .., \alpha_J)'$, and $\beta = (\beta_1, .., \beta_J)'$ examiner level coefficients, and A is the relevant weighting matrix. Kline et al. (2020) suggest estimating:

$$\hat{\sigma^2} = \hat{\delta}' A \hat{\delta} - \sum_{i=1}^{n} B_{ii} \hat{\xi}_i^2$$

where B_{ii} measures the influence of the ith squared error ϵ_i^2 on σ^2, $\xi_i^2 = \mathbb{V}(\epsilon_i | X_i)$ is the variance of the ith error, X_i is the matrix of all the regressors in the model including the examiner fixed effects, examiners times gender fixed effects, and the embeddings, and $\hat{\xi}_i^2$ is an unbiased estimate of ξ_i^2 derived from cross-fitting. Kline et al. (2020) provide the conditions on the X_i matrix that ensure consistency of the bias-corrected estimator. To compute $\hat{\sigma^2}$, I exploit the random projection method of Lindenstrauss (1984) when approximating $\hat{\sigma^2}$, and use a variant of the preconditioner in Koutis et al. (2011) designed to accommodate the fact that X_i includes a dense high dimensional segment of text-embeddings.

8 Variation in Gender Bias

8.1 Overall Variation

Table 5 reports the estimates of the standard deviation, correlation, and mean of the examiner, start-year, and art-unit level β_j and α_j effects. The standard deviation is the squared root of the Kline et al. (2020) leave-out estimate, and the correlations are the ratio of the leave-out covariance estimate and the standard deviations. In column (1), F_i is an indicator for having at least one female in the inventors' team. In column (2), it is the proportion of females, and in column (3), it is an indicator for having a female ranked first or second. Below, I describe the results in column (1), where the conclusions regarding the variability of gender bias are qualitatively the same across different gender definitions.

In panel (i), I estimate equation 6 with α_j and β_j varying across examiners. The standard deviation estimates suggest substantial heterogeneity in examiners' leniency, consistent with the previous findings in the literature (Frakes and Wasserman, 2014; Sampat and Williams, 2019; Farre-Mensa et al., 2020). The standard deviation is 9 percentage points, which is 10% bigger than the mean initial allowance level in the sample of 8.3 percentage points reported in table 1. The leave-out estimate of the standard deviation of the gender bias is 2.4 percentage points, which is 29% of the mean initial allowance rate in the sample. These results suggest that although the average examiner exhibits no bias, there is substantial variation in examiners' tastes, with some being more likely to prefer a patent with a mixed-gender team and others not.

The third row in table 5 shows a strong and negative correlation between examiners' bias and leniency, suggesting that lenient examiners who are more likely to allow a patent in the first round of the examination process are also more likely to be biased against women. In Section 10, we fit a nonlinear model for initial allowance and find zero correlation. It suggests, as in Kline and Walters (2020), that the negative correlation we find reflects a mechanical boundary effect as examiners with low initial allowance rate probability have less opportunity to discriminate. The fourth row presents the weighted average of β_j, weighted by the number of applications per examiner. I find that the average gender bias is qualitatively identical to the estimated OLS mean gender gap estimated in table 3.

Panel (ii) presents the variance components of α_j and β_j across 38 unique start years from 1975 to 2013. The estimated standard deviation of α_j across start-years is 3.8 percentage points, and the gender bias β_j is 1.2. Since each examiner has a unique start year, following the law of total variance, we can conclude that 17% of the variation in examiners' leniency and 25% of the variation in examiners' gender bias is driven by variation between different cohorts of examiners. Interestingly, at a start-year level, I find that the negative correlation between the leniency level and gender bias is 50% stronger than the one across examiners. The strong negative correlation, together with the finding that more experienced examiners are more likely to be biased against mixed-gender patents (Figure 12) implies that experienced examiners are more lenient, as also found in Frakes and Wasserman (2014), and was attributed partially to differences in the deadlines by grade levels.

Following the same exercise, panel (iii) presents the estimated standard deviation of α_j and β_j across 587 art units. Examiners are not entirely nested within art units because some move between art units or serve in more than one at the same time. However, since "movers" account for only 10 percent of the examiners, I repeat the above exercise, finding that approximately 33% and 25% of the variation across examiners in leniency and gender bias is explained by variability across art units.

The variance component estimators are calculated from a linear model with multiple treatment margins and embedding controls. As mentioned in Section 6 the OLS estimator and other matching estimators don't have to agree. This issue is particularly pronounced when dealing with a model with multiple treatment margins (Goldsmith-Pinkham et al., 2022). To assess the sensitivity of our results to the estimand, In Appendix Table A.3, we present the variance components of examiners' leniency and gender bias using Inverse Probability Weighting (IPW). Specifically, restricting attention to examiners with at least 100 observations, we start by estimating a multinomial logit model of the propensity score of each examiner and gender $\Pr(J(i) = j, F_i = f | C_i)$. Then, we use the IPW weights to reweight the data and estimate the variance component with the reweighted microdata. For further details, see Appendix Section C.2. We find that the IPW weighted variance components are qualitatively similar, with an estimated standard deviation of examiners' leniency of 9.3 and a standard deviation of the gender bias of 1.7.

8.2 Start-Year vs. Experience Effect

The evidence from Figure 12 and Table 5 suggests substantial variability in gender bias across examiners with different years of experience. Such variation could be driven by either cohort effect, in which examiners of different cohorts have different gender tastes, or by age/experience effect, in which more years of experience causes examiners to change and become more biased against mixed-gender patents. To test these hypotheses, I run the following fixed-effect regression accounting for the variation of both examiner and years of experience:

$$IA_i = \alpha_{J(i)} + \alpha_{2,exp(i)} + \left(\beta_{J(i)} + \beta_{2,exp(i)}\right)F_i + x_i'\gamma + \epsilon_i \tag{7}$$

where $exp(i)$ is the number of years of experience of examiner $J(i)$ who examines application i. Therefore, $\alpha_{2,exp}$ measures the leniency levels of examiners with exp years of experience, and $\beta_{2,exp}$ measures the gender bias of examiners with exp years of experience beyond the examiner levels leniency and gender tastes, measured by α_j and β_j. As noted by Abowd et al. (2002) in the context of wage models with both firm and individual fixed effects (Abowd et al., 1999; Card et al., 2013), estimation of model 7 is feasible only among the set of examiners "connected" by the same years of experience.

Table 6 reports the dispersion of the estimated examiners' gender gap. In the first panel, I examine the stability of the standard deviation of examiner level gender gap when also accounting for years of experience gender gap across the 8,335 examiners satisfying the connectivity restrictions. Comparing the standard deviation of gender bias to the one reported in table 5, I find that controlling for examiner years of experience fixed effects has a negligible effect on the examiner level variation when F_i is an indicator for mixed gender teams (column 1) and when it is a continuous variable for the proportion female in the inventor team (column 2). Moreover, it does not impact the estimated standard deviation when using an indicator for female ranked first or second in the inventors' list. The estimated standard deviation of β_{exp} is modest, between 10 to 5 times smaller than the examiner level one for mixed gender and proportion female gender variables. Moreover, the estimated variance is negative for the indicator of female ranked first or second, suggesting this component is very small or zero.

In the second and third panels, I repeat the same exercise but when grouping patent applications based on examiner start-year and art units. Results are qualitatively similar to the examiner-level analysis. Accounting for the years of experience fixed effects has almost no impact on the variability of gender bias across start years and a moderate effect on the estimated variability across art units when measuring the femaleness by the proportion of females. Taken together, these results establish that the variability in gender bias across examiners with different years of experience is driven by a cohort effect rather than age/experience effect.

8.3 Characterixing Biased Examiners

The analysis so far establishes that gaps in initial allowance vary substantially across examiners. To further describe the type of examiners that exhibit bias, I report the coefficients from regressions of $\hat{\beta}_j$ from equation 6 when F_i is an indicator for mixed gender patents, on various examiner characteristics. While such relationships do not necessarily describe a causal effect of examiner attributes on gender bias, they offer a summary of which examiners are more likely to exhibit bias. I describe both the cross-sectional relationship across all examiners and the within-art-unit relationship of examiners' attributes and bias.

Table 7 reports the main results. Columns 1-5 present the estimated coefficient from a regression of estimated gender bias on one characteristic: whether the examiner holds a Ph.D. or higher degree, female examiner, Asian examiner, years of experience, and the share of mixed gender patent applications in the USPC class. Column 6 presents the coefficients from a regression where all the covariates are included simultaneously. Finally, column 7 presents the estimated coefficients in a regression that includes art-unit fixed effects.

Female examiners and examiners holding a Ph.D. are more likely to be biased against men, while there is no correlation between the gender gap and Examiners with Asian names. Furthermore, in line with the results from the previous section, examiners with more years of experience are more likely to be biased against mixed-gender patents. Lastly, I find that the higher the share of mixed-gender patents in the USPC class, the less likely the examiner is biased against mixed-gender patents. Column (6) reveals that most of these relationships persist when controlling for all the covariates simultaneously, besides the coefficient on examiner education and Asian examiners.

The relationship between gender bias and other examiner and application characteristics within are-units is different than that the cross-sectional relationship. First, I find that the negative relationship between years of experience and gender bias is robust to including other characteristics and art-unit fixed effects. On the other hand, there is no clear relationship between examiners' gender and bias within art units. Lastly, I find that while the cross-sectional relationship between the share of mixed-gender teams and bias is positive, it is negative and twice as large within art units. This pattern is consistent with the finding in Breda and Ly (2015) that male-dominated fields in the entrance exam for French higher education were biased in favor of women.

8.4 Within Group Variation

There is substantial variation in leniency and gender bias across examiners, and that bias is correlated with various examiner characteristics. Next, I turn to study whether also the second moment varies by examiners' attributes such as gender, start year, and time.

Examiner gender

Table 8 displays the estimated means and standard deviations of the gender gap separately by the gender of the examiner. The first panel presents the estimate of β, the average gender gap, estimated in equation 6 separately by examiners' gender.[24] The first row presents the raw gender gap without controlling for the text embeddings, and the second row presents my preferred estimate for the average gender gap, accounting for the text embeddings. The first panel reveals two facts. First, the raw gender gap varies by the examiners' gender, with male examiners having almost twice the gap. These differences could reflect either examiners' gender differences in attitudes or differences in the distribution of examiners' gender and applicants across fields. Second, while the controlled gender gap in the full sample is statistically insignificant from zero, on average male examiners are 0.2 (SE=0.001) percentage points less likely to initially allow a patent by mixed-gender teams. In contrast, the point estimate for the gender bias of female examiners is very small in magnitude, positive, and statistically insignificant. The estimates for the set of examiners with non-classified names are similar to the ones among the male examiners.

The second panel of Table 8 investigates the within-group variability. I find that not only are female examiners less biased on average, but they also use less discretion and have much lower levels of variability in gender bias. The standard deviation of the leniency of female examiners is 7 percent, 72% of the standard deviation of the leniency levels of male and unknown examiners. The standard deviation of the gender gap among female examiners is

[24]This table uses F_i and an indicator for mixed-gender team. See Appendix tables A.4 and A.5 for the equivalent exercise with F_i being the share of female inventors and an indicator for having at least one female in the inventors' team.

only 1.3 percent, half of the standard deviation of the gender gap among unknown examiners. Taken together, I conclude that female examiners are substantially less likely to be biased both on average and mostly in terms of the risk they pose.

Within examiner start year variation

Next, I study how different cohorts of examiners differ by the risk they pose in the system by estimating the within start-year bins variation of α_j and β_j. Figure 13 displays the main results when the femaleness of the patent is measured by whether there is at least one female in the inventors' team and by the proportion of females in the inventors' team. The pink "X"s are the estimated standard deviation of α_j, and the blue dots are the estimated standard deviation of β_j. I find that the younger cohorts of examiners that joined the USPTO after 2003 are much less variable. Their discretion variability is almost 50% lower than the older cohorts, and the standard deviation of gender bias of the 2008+ cohort is around 40% of the standard deviation of the examiners that joined in 2000-2002. Using an indicator for mixed-gender patents reveals a slightly different story than when measuring the femaleness of the patent using the share of female inventors. The first presents stable behavior across the cohort that joined before 2000, and the latter suggests that the very old cohort was less variable. Taken together with the evidence from section 7, I conclude that younger cohorts pose less risk in the system and are biased in favor of women.

Within year variation

I find that the average gender gap decreased over time and converged to zero. Does the risk of encountering a biased examiner also change accordingly? To test, I estimate the standard deviation of β_j within bins of 3 to 4 years. To assess the trend in discretion over time, I also present the estimates of the standard deviation of α_j. The main results are presented in Figure 14. Although the average gender gap decreased over time, the variability in discretion and bias increased by almost 100%. Following the finding in table 6, we can conclude that the presented trend is driven by the change in the composition of examiners over time. The new young examiners are more likely to agree with each other but have different gender preferences than the seniors. As a result, we envision an increase in polarization within the examiners' community that results in increased variation in gender bias and increasing uncertainty regarding examiners' evaluation.

9 Robustness and Mechanisms

9.1 Rejection Reasons

To further investigate the roots behind initial rejections, I utilize the "USPTO Office Action Rejection data set" that includes the universe of mailed office actions and grounds for rejections from 2008 to 2017. A typical Non-Final Rejection Office action of examiners identifies the specific claims and the statutory or nonstatutory grounds on which those claims are objected to and/or rejected. Examiners can reject a patent on the following grounds: a 101-rejection, which reflects a violation of eligibility, double patenting, or lack of usefulness and credibility; a 102-rejection, which reflects a lack of novelty; a 103-rejection, which reflects a lack of obviousness; and a 112-rejection which describes not meeting the requirements regarding the adequacy of the disclosure of the invention.

I generate an indicator for whether the first round rejection had at least one office action on the ground of each of these categories. Notably, a rejection could have multiple grounds, and each ground could relate to more than one claim. As mentioned in Frakes and Wasserman (2014), rejections based on lack of novelty and obviousness are typically viewed as more time-consuming as they require a delicate prior art search and prior art comparison.

Table 9 presents the rejection reasons analysis. In panel (i), I run an OLS regression of an indicator for rejection reason on an indicator for having at least one female in the inventors' team on the sample of patent applications filed since 2008. The first row presents the uncontrolled gender gap, and the second row presents the gender gap after controlling for the patent text embeddings. Interestingly, the raw gap in obviousness and novelty rejection reasons is very small and indistinguishable from zero. These are the most prevalent rejection types that require effort and prior art search from the examiners. In contrast, on average, mixed-gender patents are more likely to be rejected because of lack of eligibility and writing - less common and simpler rejections. In line with previous findings, the second row of panel (i) shows that the gender gap disappears after controlling for the patent application text.

To study whether there are rejection reasons that are more likely to serve as grounds for rejection among biased examiners, I run the following stacked regressions together, clustering the standard errors by application id:

$$IA_i = \alpha_{J(i)} + \beta_{J(i)} F_i + C_i' \gamma + \epsilon_i$$
$$R_i = \alpha^R_{J(i)} + \beta^R_{J(i)} F_i + C_i' \gamma^R + \epsilon^R_i$$

The first equation is identical to the main regression I run in previous sections. In the second specification, R_i stands for an indicator for the reason of rejection, α^R_j measures the inclination of examiner j to initially reject on the ground of R, and β^R_j measures the extent to which examiner j is more likely to reject all male vs. mixed gender patents based on that ground. Using the estimates of gender bias β_j and β^R_j I estimate the variance component. For detailed information on the estimation, see Appendix Section C.[25]

Panel (ii.a) of Table 9 reports the corresponding estimated standard deviations and correlations. The standard deviation of the gender gap in initial allowance in 2008-2013 is 2.8 percentage points, slightly higher than the one in the full sample, reflecting again the increase in bias variability over time. The standard deviation of gender gap by rejection reason is around four percentage points, reflecting the higher base-level rejection rates. Also, as expected, I find a strong negative correlation between gaps in initial allowance and gaps in all the rejection reasons.

Estimates of the variance-covariance components of examiners' gender gap in initial allowance and rejection reasons can be used to estimate the infeasible OLS regression coefficients of initial allowance gender gap on rejections gender gaps:

$$\beta_j = \delta_0 + \delta_1 \beta^R_j + u_j$$

where the δ_1 coefficient is a function of the variance components: the ratio of the covariance between β_j and β^R_j and the variance of β^R_j. Then, given these estimates, I can calculate the implied R^2 from this regression measuring the share of gender bias variability in initial allowance that is explained by bias in each rejection reason. If examiners who perform a biased analysis choose the rejection reason proportional to the prevalence of each rejection reason, then the R^2 should be proportional to the likelihood of each rejection. In contrast, if they are more likely to use a particular rejection ground when performing a biased assessment, the proportions should not align with the distribution of rejections.

[25] Lachowska et al. (2022) run similar seemingly unrelated regressions with employer-employee data measuring the covariance between firm effects on wages and hours in Washington.

Panel (ii.b) in Table 9 presents the implied OLS δ_0 coefficient and R^2 from the infeasible regression mentioned above. I find that gaps in rejections based on Obviousness (103) and Writing (112) explain most of the initial allowance gender gap variability, accounting for 25 and 30 percent, respectively, of the variance of β_j. However, these rejections are not equally likely. Rejections based on obviousness are the most prevalent, accounting for almost 70 of the rejections in the first round of examination, being twice more likely than rejections based on writing. Evidently, biased evaluations are more likely to be based on writing grounds. This result aligns with the intuition that rejections based on writing are "simpler", requiring less effort compared to prior art rejections. A similar pattern was also found in Frakes and Wasserman (2014) in which examiners facing more time restrictions had a lower probability of performing obviousness rejections.

9.2 Other Non-text Characteristics

The objective of this paper is to estimate the gender bias parameter defined in equation 1. As explained in section 5, this parameter captures both disparate treatment and disparate impacts that are uncorrelated with text characteristics. To assess the extent to which other patent non-text characteristics, such as inventors' ethnicity and lawyer characteristics, translate into gender disparities, I rerun the main analysis, including other non-text characteristics. First, I re-estimate equation 5, adding one at a time each of these covariates: (i) an indicator for having at least one inventor with a foreign name, defined by a name that does not appear in the SSA name tables, an indicator for having at least one Asian in the inventors' team, (il) attorney gender, and (iii) Law company's years of experience.[26]

Pane (i) of Table 10 reports the main results. Column 1 replicates the results from table 3. Columns 2-5 present the coefficients on any female and the other characteristics estimated in separate regressions. In line with the finding in section 5.3.2, the inclusion of other non-text characteristics does not affect the measures of the gender gap in initial allowance. Moreover, the inventors' ethnicity and the attorney's gender do not affect allowance. In contrast, the law company's experience is positively correlated with initial allowance even after controlling for the patent application text. That could reflect either the causal effect of experienced law companies on the application process beyond their effect on writing or examiners' preferences

[26] See Appendix B for further detail on variable construction.

towards more established law companies.[27]

Second, in panel (ii), I estimate the following fixed-effect regression

$$IA_i = \alpha_{J(i)} + \beta_{J(i)} F_i + \eta_{J(i)} C_i + x_i' \gamma + \epsilon$$

where C_i is the additional non-text characteristics and η_j measures the variation in examiners tastes with respect to characteristic C_i. Then, using the estimated $\hat{\beta}_j$ and $\hat{\eta}_j$ parameters, panel (ii) displays their standard deviation and correlation. Echoing the results from panel (ii), the variance of gender bias is unaffected by the inclusion of other characteristics. I find that examiners exhibit variation with respect to these other characteristics comparable to the variability in gender bias, but the correlation between gender bias and these other biases is between 1 to 6 percent.

10 The Cost of Variance in Bias

The analysis conducted thus far has shed light on the prevalence of gender bias during the first round of the application examination process. Next, we turn to explore the broader implications of such bias and its potential effects on economic outcomes. Extensive research has demonstrated the significant contribution of innovation, including the acquisition of patents, to the success of firms and inventors (Kogan et al., 2017; Kline et al., 2019). In this section, I study the consequences of bias on economics outcomes by estimating the distribution of examiners' behavior and how it translates into differences in patents' stock market return as measured by Kogan et al. (2017)

Identification of the effect of bias on market return is challenging. Unlike the first round of the application process, Subramani et al. (2021) find that women are less likely to persist and resubmit their patent application if initially rejected. Hence gender gap in outcomes could also reflect differences in the females' behavior rather than the examiners'. In addition, market return is observed only for granted patents that were assigned to publicly traded firms, introducing a sample selection. To address this concern, this section adopts a sample selection correction approach inspired by the Heckman selection model (Heckman, 1979)

[27] That might also reflect that the text embeddings don't fully capture the patent text content. However, as can be seen in this table, its explanatory power of initial allowance is uncorrelated with the inventors' team gender and hence has no effect on the results.

using the examiners as instrumental variables, conditional on the patent text. I model examiners' decisions parametrically as a single index model and use that model to simulate different counterfactuals of examiners' behavior. Since examiners directly affect the outcomes of rejected patents by restricting the scope of their claims and adding prior art citations, I restrict attention to patents that were initially allowed, thereby not violating the exclusion restriction assumption.

10.1 Examiners' Decisions

Let q_i be the quality of patent i. I assume that q_i is an unobserved index, both by the examiners and by the analyst, and is agreed upon by all the examiners as the value that examiners seek to maximize through their decisions. Moreover, I assume that examiners observe the patent application text C_i together with a noisy signal, ω_{ij}, of the patent quality and the gender of the inventors' team F_i. Examiners are assumed to form an accurate posterior mean prediction of the patent quality given available information $E[q_i|\omega_{ij}, X_i, F_i = f]$. Finally, each examiner has a subjective cost, $\lambda_{jf}(C_i)$, associated with allowing a patent with gender $F_i = f$, yielding decision rule:

$$IA_{ij} = \mathbb{1}\{E[q_i|\omega_{ij}, C_i, F_i = f] \geq \lambda_{jf}(C_i)\} \tag{8}$$

$$= \mathbb{1}\{\mu_f(C_i) + u_{ij} \geq \lambda_{jf}(C_i)\} \tag{9}$$

Taste-based gender bias, as in Becker (1957), arises when examiners perceive differing social costs from allowing patents with different mixed-gender teams but the same expected posterior quality by applying different posterior quality thresholds by gender. Models of inaccurate stereotyping can result in observationally equivalent bias (Arnold et al., 2018; Bohren et al., 2022). Statistical discrimination as modeled in Aigner and Cain (1977) arises when gender affects examiners' posterior mean prediction of the patent quality due to differences in the prior distribution of patent quality by gender. As detailed in Section 5, all these models would produce an observed mean allowance gender gap for which $\Pr(IA = 1|C_i, F_i = 0) \neq \Pr(IA = 1|C_i, F_i = 1)$.

For every patent application, I model the potential log market return $R_{ij}(d)$ of patent application i with inventors' gender mix F_i and patent content C_i that was examined by examiner j as a function of whether the patent application was initially allowed $d \in \{0, 1\}$:

$$R_{ij}(IA_{ij}) = \psi_c(C_i, F_i)IA_{ij} + \delta_{cj}(C_i, F_i)(1 - IA_{ij}) + \epsilon_{ij} \tag{10}$$

Note that the expected market return among initially rejected patents is affected by the identity of the examiners because examiners can restrict the scope of the patent by requesting amendments to the patent claims and demanding the inventors cite additional prior art. However, among patents that were allowed in the first round of the patent application, there is no direct examiner causal effect. Therefore, the expected market return among initially allowed firms:

$$E[R_{ij}|F_i, C_i, J(i) = j, IA_{ji} = 1] = \psi_c(C_i, F_i) + E[\epsilon_{ij}|F_i, J(i) = j, IA_{ij} = 1] \tag{11}$$

where $E[\epsilon_{ij}|F_i, C_i, J(i) = j, IA_{ij} = 1]$, is the expected unobserved market return among initially allowed patents that were assigned to examiner j and whose inventor gender is F_i. If examiners' perceived quality is uncorrelated with observed market return ($Cov(\epsilon_{ij}, u_{ij}) = 0$), then we could estimate Equation 11 by running an OLS regression. Otherwise, $E[\epsilon_{ij}|F_i, C_i, J(i) = j, IA_i = 1]$ is *control function* that summarizes the selection bias. Note that finding that $\psi_c(C_i, 0) \neq \psi_c(C_i 1)$ need not signal bias unless we think that examiners seek to maximize market return. In any case in where $q \neq R$ and q and R are not very strongly correlated, it is possible that the average patents by mixed-gender authors have different market returns, while $\Pr(IA = 1|C_i, F_i = 0) \neq \Pr(IA = 1|C_i, F_i = 1)$ so applications by different gender-mix of authors receive the same treatment. Such disparities in market return conditional on patent content could reflect, for example, differences in the firms male and female authors work at or differences in how male and female inventors translate their inventions into valuable products.

The joint distribution of (u_{ij}, ϵ_{ij}) is modelled as jointly normal:

$$\begin{pmatrix} u_{ij} \\ \epsilon_{ij} \end{pmatrix} | F_i = f \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix} \right)$$

where ρ captures the extent to which examiners' decisions are correlated with patents' market return. When $\rho > 0$, market return is correlated with the measure of quality perceived by examiners. In that case, among examiners who are biased against females, the mean market return of patents by mixed-gender teams would be lower than their mean market return had these patents been judged as males' patents, the opposite will be true among examiners who are biased against all-male patents.

Identification - Selection models without excluded instruments are only identified by functional form restrictions (Heckman, 1990). Therefore, following the findings in section 5.3, I treat examiners as instruments that shift the probability of allowance but do not directly affect the market return of initially allowed patents. Formally, identification requires the following assumptions: Relevance, exclusion, and monotonicity. The validity of the first two assumptions has already been established in the previous sections. I find that examiners' discretion plays a significant role in the allowance decision and that, conditional on the patent application content, examiners' assignments are as good as random. Also, since I estimate the model only for initially allowed applications, examiners, by construction, have no direct causal effect on market return.

The third assumption, monotonicity, requires formally that for every pair of examiners j and j', either $\Pr(IA_{ij} = 1) \geq \Pr(IA_{ij'} = 1)\ \forall i$, or $\Pr(IA_{ij} = 1) \leq \Pr(IA_{ij'} = 1)\ \forall i$. In fact, this is a weaker assumption of the well-known strict monotonicity assumption by Imbens and Angrist (1994). Strict monotonicity does not allow for random violation of the common ordering that could arise when $u_{ij} \neq 0$. This relaxation of the strict monotonicity was introduced in Frandsen et al. (2023), allowing random violation in monotonicity as long as u_{ij} have the same variance for all examiners.

The monotonicity assumption requires that examiners agree on the ranking of applications on average, which imposes restrictions on the behavioral model of examiners' decisions and the skills of examiners. To relax this assumption, the model includes heterogeneity both by examiners, gender, and the content of the patent $\lambda_{jf}(x_i)$, thereby allowing examiners to rank patents with different content differently but still requiring the same applications ranking within gender and patent content.

Rather than controlling for over 2,000 text embeddings, I control for a lower dimensional representation of the patent content to relax the monotonicity assumption, and in the next section, I explain how I leverage the propensity score from Section 8.1 to reweight the maximum likelihood so it accounts for embeddings. I use the first two components of the UMAP dimensionality reduction presented in Figure 4, which I denote by $\tilde{x}$. Therefore, the random thresholds and the parameters in the market return and initial allowance equation are modeled as:

$$\mu_F(\tilde{x}_i) = \mu_0 + \tilde{x}_i'\mu_{0x} + (\mu_1 + \tilde{x}_i'\mu_{1x})F$$
$$\lambda_{jF}(\tilde{x}) = \lambda_{0j} + \tilde{xj}_i'\lambda_{0xj} + (\lambda_1 + \tilde{x}_i'\lambda_{1xj})F$$
$$\psi_c(\tilde{x}_i, F_i) = \psi_0 + \psi_1 F + \tilde{x}_i'\psi_x,$$

where I normalize $\tilde{x}$ to be mean zero with a standard deviation of one. I describe next

Likelihood - Define $\theta_j = (\lambda_{0j}, \lambda_{1j}, \lambda_{0xj}', \lambda_{1xj}')$. For every patent application i assigned to examiner $J(i)$, I define the following probabilities:

$$p_i(\theta_{J(i)}) = Pr(IA_{iJ(i)} = 1 | F_i, \tilde{x}, \theta_{J(i)})$$
$$f_i(\theta_{J(i)}) = Pr(R_i = r | IA_i = 1, F_i, \tilde{x}, \theta_{J(i)})$$

where $p_i(\theta_j)$ is the probability of initial allowance, and $f_i(\theta_j)$ captures the probability density of observing $R_i = r$ for initially allowed patents. Let $I_j \equiv \{i : J(i) = j\}$ be the set of applications that were assigned to examiner j. The likelihood of observing $\{IA_i, R_i, F_i, \tilde{x}_i\}_{i \in I_j}$ for the patent applications assigned to examiner j is

$$l_j(\{IA_i, R_i, F_i, \tilde{x}_i\}_{i \in I_j} | \theta_j) = \prod_{i \in I_j} (1 - p_i(\theta_j))^{1 - IA_i} \cdot f_i(\theta_j)^{IA_i}.$$

I account for differences in patent content beyond the first two UMAP components by reweighting my likelihood to match the distribution of examiner-gender characteristics. I begin with the propensity scores $ps_{ij} = \Pr(J(i) = j, F_i = f | C_i)$ from section 8.1. Then, I normalize the weight of each application in two steps. In the first step I normalize $w_{ij} = (\frac{1}{ps_{ij}})/(\sum_{i \in I_j} \frac{1}{ps_{ij}})$ to sum to one within each examiner. Then, to hold the number of applications per examiner constant, I form the adjusted weights:

$$\tilde{w}_{ij} = w_{ij} \times N_j$$

where N_j is the total number of applications assigned to examiner j. Therefore, the likelihood of observing $\{IA_i, R_i, F_i, \tilde{x}_i\}_{i \in I_j}$ for the applications assigned to examiner $J(i) = j$ is:

$$l_j(\{IA_i, R_i, F_i, \tilde{x}_i\}_{i \in I_j} | \theta_j) = \prod_{i \in I_j} \left[(1 - p_i(\theta_j))^{1 - IA_i} \cdot f_i(\theta_j)^{IA_i} \right]^{\tilde{w}_{ij}}.$$

This content-adjusted likelihood ensures that examiners with more applications have more influence on the estimation, which is in line with my estimation strategy of the variance components in section 7.

I treat the examiner-level thresholds θ_j as draws from a prior normal distribution with mean zero and variance-covariance matrix Σ:

$$\theta_j \sim \mathcal{N}(0, \Sigma).$$

Finally, the log-likelihood of the data given the parameter vector $\Theta = (\mu_0, \mu_{0x}, \mu_1, \mu_{1x}, \psi_0, \psi_x, \psi_1, \Sigma)$ is:

10.2 Results

Table 11 presents the main parameter estimates of Θ. Panel (i) displays the estimates for the initial allowance regression. μ_0 represents the mean initial allowance rate for all-male patent applications and μ_1 the differences between all-male and mixed-gender. In line with our findings using a linear model, the estimate of μ_1 is very small and insignificant from zero, suggesting that the is no average difference in allowance probability between mixed-gender and all-male patent applications.

Panel (iii) presents the variance of examiners' allowance thresholds, where σ_0 represents the base-level variability in allowance decisions of examiners, and $\sigma_{\lambda 1}$ describes the extent to which examiners' allowance threshold is different between all-male vs. mixed-gender patent applications, i.e., the variability in gender bias. Translating the $\sigma_{\lambda 1}$ to allowance probability Table 11 implies that the variance of gender bias for the average patent text is 0.022, which accords with the findings in Table 5. Moreover, we find that the correlation between λ_1 and λ_0 is zero, suggesting that the negative correlation we found in Table 5 reflected a mechanical boundary relationship.

Panel (ii) presents the estimates of the log market return equation conditional on initial allowance and after correcting for selection bias, where ψ_0 describes the mean market return on all-male patents and ψ_1 describes the difference between all-male and mixed-gender inventor teams. I find that, on average, mixed-gender patents generate 247 thousand dollars higher market return than no-female patents with equivalent patent content and text. This disparity could reflect the differences in productivity of the firms female inventors work at or the productivity of the teams female inventors participate in.

Nevertheless, even if mixed-gender patents generate more valuable inventions for firms, mean zero gender bias together with positive variance doesn't necessarily imply that the behavior examiners will generate inefficiencies. Variance in bias generates lower total market return value if the objectives of examiners, i.e., the object they maximize, need to be correlated with market return. This aspect is evident from the estimate of ρ in panel (iv), which shows that the correlation between initial allowance decision and market returns is 0.1 (SE=0.004).

10.3 Counterfactuals

I evaluate the welfare implications of positive variance in bias together with mean zero gender bias via a simulation exercise. Using the estimated parameters in Tables 11, I simulate a set of 6,000 examiners, each observing 1,500 applications with average patent text embeddings and 15% mixed gender teams. My goal is to compare the average market return of two cases. First, the status quo with mean zero gender bias but positive variance. I denote the potential initial allowance of each application i under this scenario by $IA_i(\sigma_{\lambda_1} > 0)$. The second scenario I consider is a new initial allowance decision rule that keeps the same examiner-level initial allowance rate, therefore allowing for examiner heterogeneity in leniency, but enforces uniform zero bias for all the examiners. Formally, for every examiner j, I find the scalar $\tilde{t}_j$ such that

$$\Pr(u_{ij} \geq \lambda_{j0} - \mu_0 + (\lambda_{j1} - \mu_1)F_i) = \Pr(u_{ij} \geq \tilde{t}_j).$$

I label the potential allowance that would have arisen under this decision rule as $IA_i(\sigma_{\lambda_1} = 0)$. Then, for every scenario, I report the stock market return of initially allowed patens, where the main focus is on the compliers, i.e., those applications that were not allowed in the status quo but are allowed under uniform zero bias $(IA(\sigma_{\lambda_1} > 0) = 0 \to IA(\sigma_{\lambda_1} = 0) = 1)$, and those there were allowed in the status quo but are not allowed under uniform zero bias $(IA(\sigma_{\lambda_1} > 0) = 1 \to IA(\sigma_{\lambda_1} > 0) = 0)$. If the total stock market return of the first is greater than the second, we conclude that mean zero bias with positive variance generates economic loss.

Table 12 presents the mean market return of the two groups of compilers by the gender of the inventors' team. Mixed-gender patents have a higher market return on average, which is also reflected in the table. 1.1% of the female applications would have been affected by the zero variance policy, resulting in an increase of 179 thousand dollars of stock market return per average patent application. Among all male applications, 0.22 percent of the applications would be affected by the policy, which would result in an increase of 4,300 dollars per application on average.

With an average of 40,000 patent applications assigned to publicly traded firms a year and an average of 9% initial allowance rate, the total loss in stock market return amounts to 1.13 million dollars for female patents and 289 thousand dollars for all male patents a year. This total cost of 1.419 million dollars a year reflects only a lower bound of the social cost of positive variance of bias because it evaluates only the loss of applications in the first round of patent applications, which are less than 20 percent of the eventually allowed applications.

11 Conclusion

Women are underrepresented in the patent system. Yet, the extent to which observed disparities arise from discriminatory practices is unclear. In order to shed light on this, this study adopts a novel identification strategy and analyzes the gender gap in the first round of the examination process conditional on the patent application text. This paper finds that while the average gender bias is zero, it masks substantial heterogeneity across examiners, where some examiners are biased against women and some others are biased in favor of women. This paper finds that the start-year of the examiner explains 25 percent of the variance of gender bias, where male or senior examiners are more likely to be biased against mixed-gender patents, and young or female examiners are more likely to be biased against no-female patents. Lastly, studying the dynamics over time, I find that, on average, the system turned from being biased against women in the years 2001-2003 to unbiased on average. However, due to the changes in examiners' cohort composition, that variance in gender bias has increased over time, reflecting an increase in the risk of encountering an abnormally biased examiner.

Most of the discrimination literature dating back at least to Oaxaca (1973) primarily fixates on average gaps, which map only partially to fairness and inefficiencies. Firstly, even though there is no ex-ante bias, heterogeneity in bias undermines ex-post horizontal equity where patents of equivalent quality but different inventor team compositions experience disparate odds of approval. Secondly, an exclusive focus on mean bias overlooks the detrimental ramifications of misallocation, which may manifest even if the mean bias is zero. Utilizing Kogan et al. (2017))'s stock market return model for patents, our analysis estimates the lower bound of the annual cost of having a positive gender bias variance among patent applications assigned to publicly traded firms to be approximately $1.5 million.

The findings of this paper thus necessitate a critical reevaluation of the patent application process, particularly the aspect that allows examiner access to inventors' names. To counter the escalating influence of discretion and bias in patent examinations, the implementation of blind reviews and/or other computer-based methods could provide potential remedies.

References

Abowd, J. M., Creecy, R. H., Kramarz, F., et al. (2002). Computing person and firm effects using linked longitudinal employer-employee data. Technical report, Center for Economic Studies, US Census Bureau.

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Abrams, D. S., Bertrand, M., and Mullainathan, S. (2012). Do judges vary in their treatment of race? *The Journal of Legal Studies*, 41(2):347–383.

Acemoglu, D., Akcigit, U., Alp, H., Bloom, N., and Kerr, W. (2018). Innovation, reallocation, and growth. *American Economic Review*, 108(11):3450–91.

Acikalin, U. U., Caskurlu, T., Hoberg, G., and Phillips, G. M. (2022). Intellectual property protection lost and competition: An examination using machine learning. Technical report, National Bureau of Economic Research.

Aigner, D. J. and Cain, G. G. (1977). Statistical theories of discrimination in labor markets. *Ilr Review*, 30(2):175–187.

Akcigit, U., Grigsby, J., and Nicholas, T. (2017). The rise of american ingenuity: Innovation and inventors of the golden age. Technical report, National Bureau of Economic Research.

Arceo-Gomez, E. O. and Campos-Vazquez, R. M. (2014). Race and marriage in the labor market: A discrimination correspondence study in a developing country. *American Economic Review*, 104(5):376–80.

Arnold, D., Dobbie, W., and Hull, P. (2022). Measuring racial discrimination in bail decisions. *American Economic Review*, 112(9):2992–3038.

Arnold, D., Dobbie, W., and Yang, C. S. (2018). Racial bias in bail decisions. *The Quarterly Journal of Economics*, 133(4):1885–1932.

Arnold, D., Dobbie, W. S., and Hull, P. (2020). Measuring racial discrimination in bail decisions. Technical report, National Bureau of Economic Research.

Becker, G. (1957). 1971. the economics of discrimination.

Bell, A., Chetty, R., Jaravel, X., Petkova, N., and Van Reenen, J. (2019). Who becomes an inventor in america? the importance of exposure to innovation. *The Quarterly Journal of Economics*, 134(2):647–713.

Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, pages 436–455.

Bohren, J. A., Hull, P., and Imas, A. (2022). Systemic discrimination: Theory and measurement. Technical report, National Bureau of Economic Research.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2016). Stereotypes. *The Quarterly Journal of Economics*, 131(4):1753–1794.

Breda, T. and Ly, S. T. (2015). Professors in core science fields are not always biased against women: Evidence from france. *American Economic Journal: Applied Economics*, 7(4):53–75.

Card, D., DellaVigna, S., Funk, P., and Iriberri, N. (2021). Gender differences in peer recognition by economists. Technical report, National Bureau of Economic Research.

Card, D., DellaVigna, S., Funk, P., and Iriberri, N. (2022). Gender gaps at the academies. Technical report, National Bureau of Economic Research.

Card, D., Heining, J., and Kline, P. (2013). Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly journal of economics*, 128(3):967–1015.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters.

Choi, J., Stein, C., and Williams, H. (2019). Are patent examiners gender neutral?

Coluccia, D. M., Dossi, G., and Ottinger, S. (2023). Racial discrimination and lost innovation.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Dobbie, W., Goldin, J., and Yang, C. S. (2018). The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review*, 108(2):201–40.

Ellinas, E. H., Rebello, E., Chandrabose, R. K., Shillcutt, S. K., Hernandez, M., and Silver, J. K. (2019). Distinguished service awards in anesthesiology specialty societies: analysis of gender differences. *Anesthesia & Analgesia*, 129(4):e130–e134.

Farre-Mensa, J., Hegde, D., and Ljungqvist, A. (2020). What is a patent worth? evidence from the us patent "lottery". *The Journal of Finance*, 75(2):639–682.

Feigenberg, B. and Miller, C. (2022). Would eliminating racial disparities in motor vehicle searches have efficiency costs? *The Quarterly Journal of Economics*, 137(1):49–113.

Firth, J. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in linguistic analysis*, pages 10–32.

Frakes, M. D. and Wasserman, M. F. (2014). Is the time allocated to review patent applications inducing examiners to grant invalid patents?: Evidence from micro-level application data. Technical report, National Bureau of Economic Research.

Frandsen, B., Lefgren, L., and Leslie, E. (2023). Judging judge fixed effects. *American Economic Review*, 113(1):253–277.

Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3):535–74.

Goldsmith-Pinkham, P., Hull, P., and Kolesar, M. (2022). Contamination bias in linear regressions. Technical report, National Bureau of Economic Research.

Government, U. S. (1888). *Women Inventors to Whom Patents Have Been Granted by the United States Government*. Government Prinitng Office.

Graham, S. J., Marco, A. C., and Miller, R. (2015). The uspto patent examination research dataset: A window on the process of patent examination. *Georgia Tech Scheller College of Business Research Paper No. WP*, 43.

Hall, B. H., Jaffe, A., and Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of economics*, pages 16–38.

Heckman, J. (1990). Varieties of selection bias. *The American Economic Review*, 80(2):313–318.

Hunt, J., Garant, J.-P., Herman, H., and Munroe, D. J. (2013). Why are women underrepresented amongst patentees? *Research Policy*, 42(4):831–843.

Imbens, G. and Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.

Jensen, K., Kovacs, B., and Sorenson, O. (2018). Gender differences in obtaining and maintaining patent rights. *Nature biotechnology*, 36(4):307–309.

Khan, B. Z. (1996). Married women's property laws and female commercial activity: Evidence from united states patent records, 1790–1895. *The Journal of Economic History*, 56(2):356–388.

Kline, P. (2011). Oaxaca-blinder as a reweighting estimator. *American Economic Review*, 101(3):532–37.

Kline, P., Petkova, N., Williams, H., and Zidar, O. (2019). Who profits from patents? rent-sharing at innovative firms. *The Quarterly Journal of Economics*, 134(3):1343–1404.

Kline, P., Rose, E. K., and Walters, C. R. (2022). Systemic discrimination among large us employers. *The Quarterly Journal of Economics*, 137(4):1963–2036.

Kline, P., Saggio, R., and Sølvsten, M. (2020). Leave-out estimation of variance components. *Econometrica*, 88(5):1859–1898.

Kline, P. M. and Walters, C. R. (2020). Reasonable doubt: Experimental detection of job-level employment discrimination. Technical report, National Bureau of Economic Research.

Koffi, M. (2021). Innovative ideas and gender inequality. Technical report, Working Paper Series.

Kogan, L., Papanikolaou, D., Seru, A., and Stoffman, N. (2017). Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics*, 132(2):665–712.

Lachowska, M., Mas, A., Saggio, R., and Woodbury, S. A. (2022). Worker and employer heterogeneity in hours determination. 40(S1):S469–S493.

Lanjouw, J. O. and Schankerman, M. (2001). Characteristics of patent litigation: a window on competition. *RAND journal of economics*, pages 129–151.

Levitskaya, E., Kedrick, K., and Funk, R. J. (2022). Investigating writing style as a contributor to gender gaps in science and technology. *arXiv preprint arXiv:2204.13805*.

Li, Z. and Liu, S. (2023). Gender difference in innovation recognition: A textual analysis approach. *Available at SSRN 4527066*.

Lindenstrauss, W. J. J. (1984). Extensions of lipschitz maps into a hilbert space. *Contemp. Math*, 26(189-206):2.

Lissoni, F., Miguelez, E., Toole, A., Myers, A., Breschi, S., Ferruci, E., Sterzi, V., Tarasconi, G., et al. (2018). Progress and potential-a profile of women inventors on us patents, in office of the chief economist ip data highlights (n° 2). Technical report.

Marco, A. C., Sarnoff, J. D., and Charles, A. (2019). Patent claims and patent scope. *Research Policy*, 48(9):103790.

Marco, A. C., Toole, A., Miller, R., and Frumkin, J. (2017). Uspto patent prosecution and examiner performance appraisal.

McInnes, L., Healy, J., and Melville, J. (2018). Uniform manifold approximation and projection for dimension reduction. arxiv. *arXiv preprint arXiv:1802.03426*.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, pages 693–709.

Parasurama, P. (2021). racebert–a transformer-based model for predicting race and ethnicity from names. *arXiv e-prints*, pages arXiv–2112.

Righi, C. and Simcoe, T. (2019). Patent examiner specialization. *Research Policy*, 48(1):137–148.

Rossiter, M. W. (1993). The matthew matilda effect in science. *Social studies of science*, 23(2):325–341.

Sampat, B. and Williams, H. L. (2019). How do patents affect follow-on innovation? evidence from the human genome. *American Economic Review*, 109(1):203–36.

Silver, J. K., Bank, A. M., Slocum, C. S., Blauwet, C. A., Bhatnagar, S., Poorman, J. A., Goldstein, R., Reilly, J. M., and Zafonte, R. D. (2018). Women physicians underrepresented in american academy of neurology recognition awards. *Neurology*, 91(7):e603–e614.

Srebrovic, R. and Yonamine, J. (2020). Leveraging the bert algorithm forpatents with tensorflow and bigquery. Technical report, https://github.com/google-research/bert.

Subramani, G., Aneja, A., and Reshef, O. (2021). Try, try, try again? persistence and the gender innovation gap.

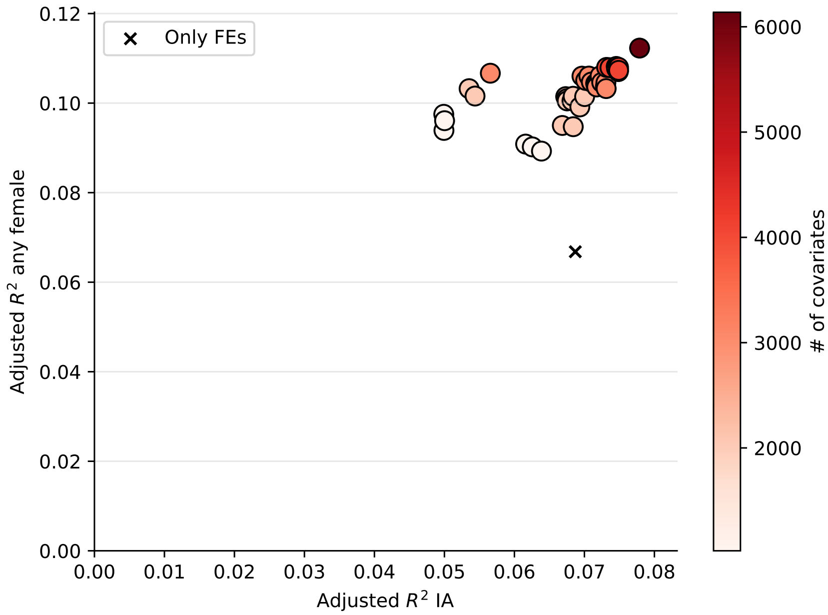
Suzgun, M., Melas-Kyriazi, L., Sarkar, S. K., Kominers, S. D., and Shieber, S. M. (2022). The harvard uspto patent dataset: A large-scale, well-structured, and multi-purpose corpus of patent applications. *arXiv preprint arXiv:2207.04043*.

Toivanen, O. and Vaananen, L. (2012). Returns to inventors. *Review of Economics and Statistics*, 94(4):1173–1190.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.

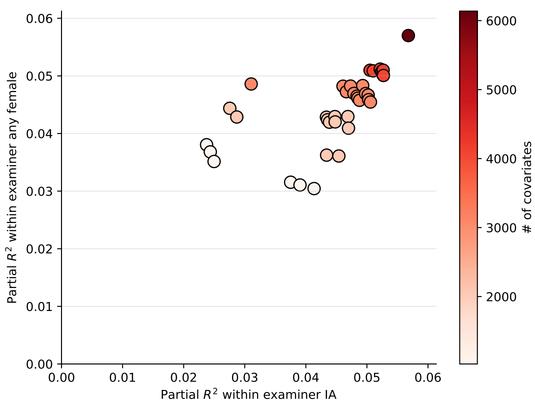
Figures

Figure 1: Adjusted R^2 of initial allowance and any female indicator on patent text embeddings

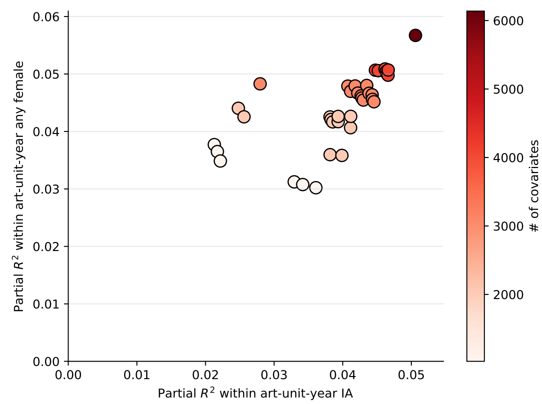


Note: This Figure plots the adjusted R^2 from regressing initial allowance (the horizontal axis) and mixed gender team indicator (the vertical axis) on different combinations of text embeddings. Different dots in the Figure represent different combinations of the embedding layers and different embedding representations of the patent's claims and the patent's description section. Section 4.2 provides information on the embeddings layer, and the exact point estimates generating this Figure are available in Appendix table A.1. Darker dots represent a model with more covariates, and the "X" symbol represents the adjusted R^2 for a model with art-unit-year and class fixed effects, which includes more than 8,000 fixed effects.

Figure 2: Partial adjusted R^2 of initial allowance and any female indicator on patent text embeddings



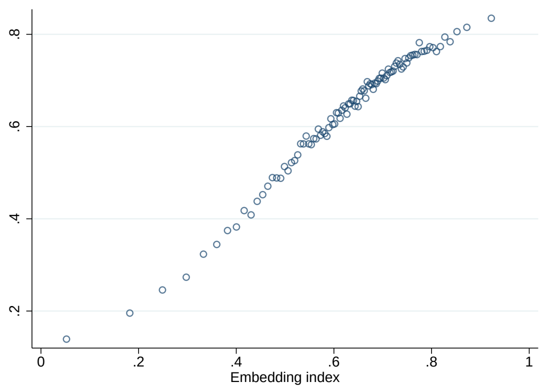
(a) Within examiner



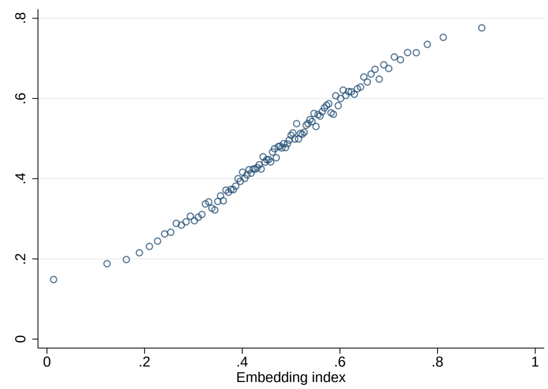
(b) Within art-unit-year class

Note: This Figure plots the partial adjusted R^2 from regressing initial allowance (the horizontal axis) and mixed gender team indicator (the vertical axis) on different combinations of text embeddings. Sub-Figure (a) measures the predictive power of embeddings within examiners, and Sub-Figure (b) measures the predictive power of embeddings after controlling for art-unit-year and class fixed effects. Different dots in the Figure represent different combinations of the embedding layers and of the embedding representations of the patent's claims and the patent's description section. Section 4.2 provides information on the embeddings layer and the exact point estimates generating this Figure are available in Appendix table A.1. Darker dots represent a model with more covariates

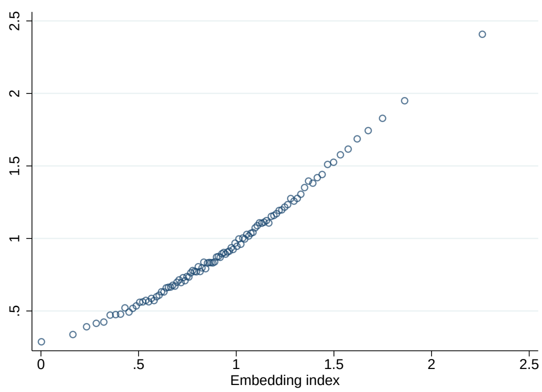
Figure 3: Split sample binned scatter plot of examiners' decisions and patent quality proxies on embeddings



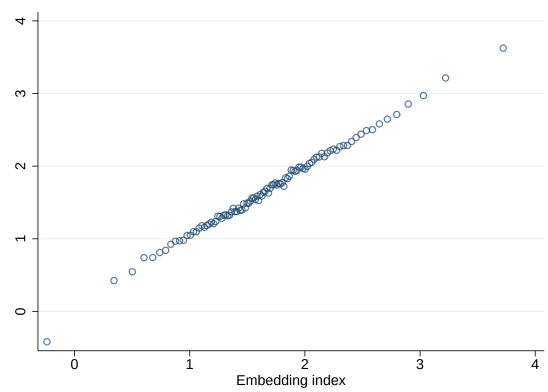
(a) Rejections based on non-obviousness



(b) Rejections based lack of novelty



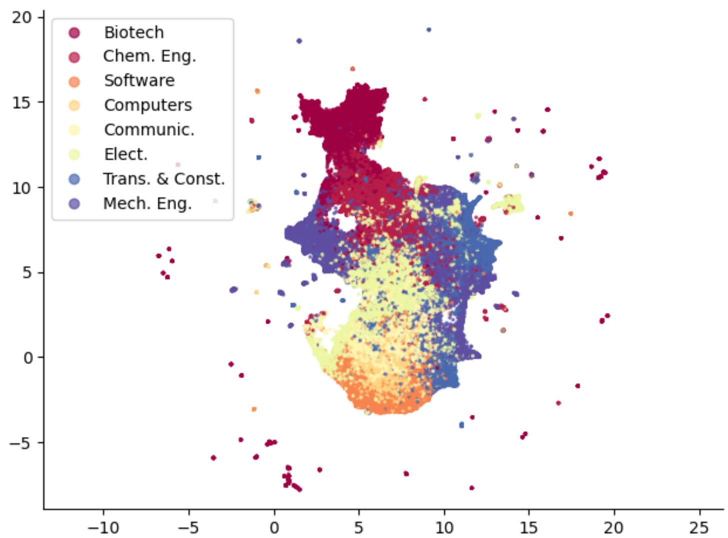
(c) Log citations



(d) Log market return

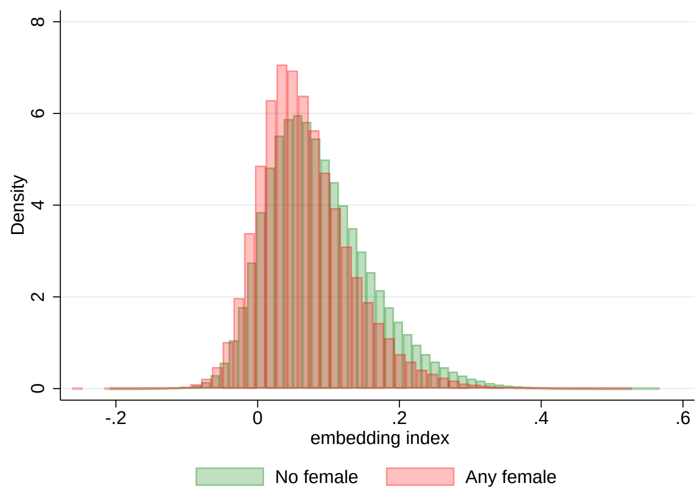
Note: This Figure plots a binned scatter with 100 bins of a split sample prediction of examiners' decisions, log citations, and log market return with the patent text embeddings. Sub-Figure (a) plots the results of an indicator for examiner rejection on the ground of non-obviousness, sub-Figure (b) plots the results of an indicator for examiner rejection on the ground of lack of novelty, panel (c) plots the results on log number of citations after granting among granted patents, and sub-Figure (d) plots the results of Kogan et al. (2017)'s measure for patent market return.

Figure 4: UMAP visualization of the patent text embeddings by technology centers



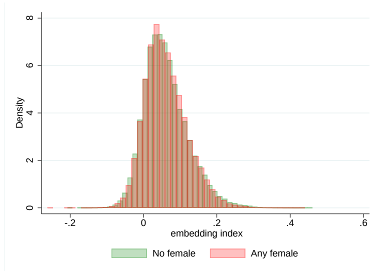
Note: This Figure plots the Uniform Manifold Approximation and Projection (UMAP) McInnes et al. (2018) visualization of patent text embeddings and its relationship with patent technology centers. The Figure was generated using a random sample of 30% of the patent application. Different colors represent different USPTO technology centers.

Figure 5: Histogram of embeddings index by the gender mix of inventors

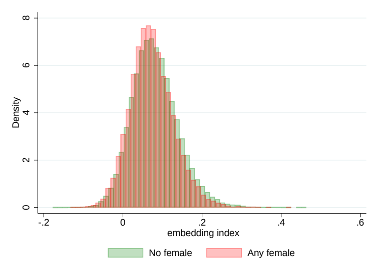


Note: This Figure plots the distribution of the embedding index separately for patent applications written by mixed-gender patents and by teams with no female inventors. To generate this Figure, I estimate an OLS regression of the form $IA_i = \alpha + \beta F_i + C_i'\gamma + \epsilon_i$ where IA_i is an indicator for initial allowance, F_i is a mixed gender team indicator, and C_i are my preferred 2,046 text embeddings. The embedding index is the estimated linear combination $C_i'\hat{\gamma}$.

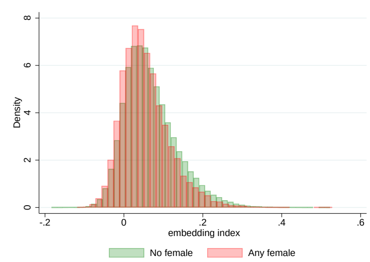
Figure 6: Histogram of embeddings index by gender and technology center



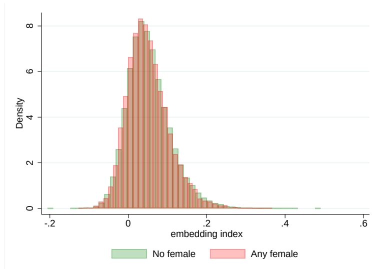
(a) Biotechnology (16)



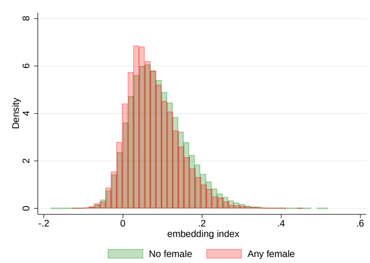
(b) Engineering (17)



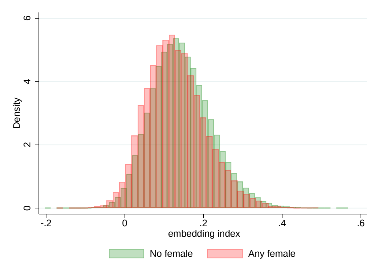
(c) Software (21)



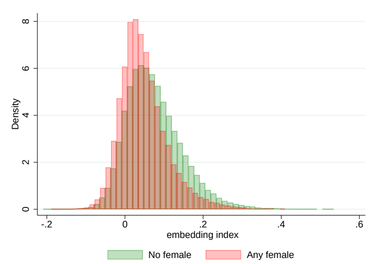
(d) Computers (24)



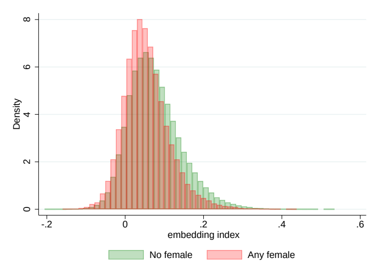
(e) Communications (26)



(f) Optical (28)



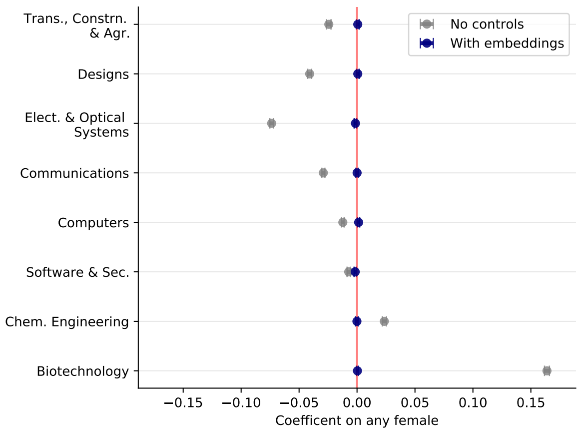
(g) Transportation (36)



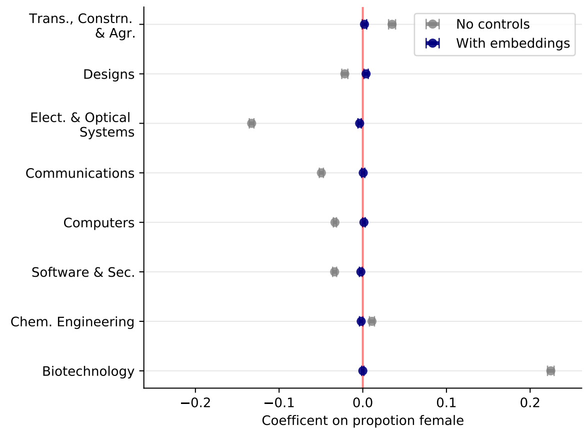
(h) Mech. Engineering (37)

Note: This Figure plots the distribution of the embedding index separately for patent applications written by mixed-gender patents by teams with no female inventors and by technology centers. To generate this Figure, I estimate an OLS regression, separately for every technology center, of the form $IA_i = \alpha + \beta F_i + C_i' \gamma + \epsilon_i$ where IA_i is an indicator for initial allowance, F_i is a mixed gender team indicator, and C_i are my preferred 2,046 text embeddings. The embedding index is the estimated linear combination $C_i' \hat{\gamma}$.

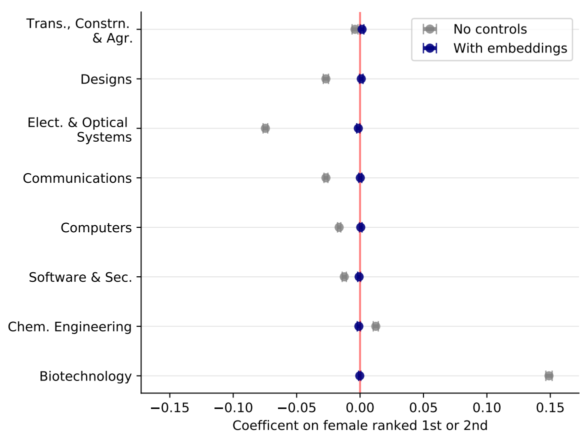
Figure 7: Balance test of technology centers



(a) Any female



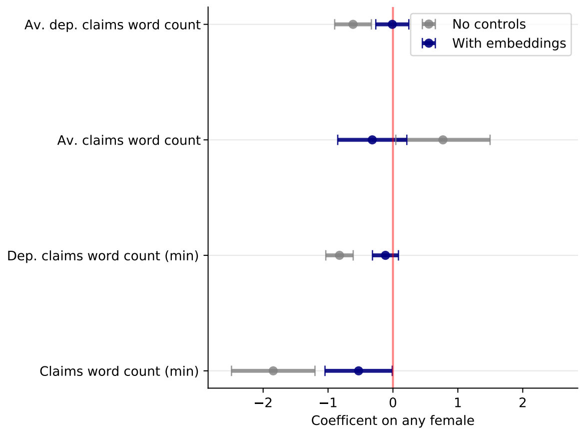
(b) Proportion female



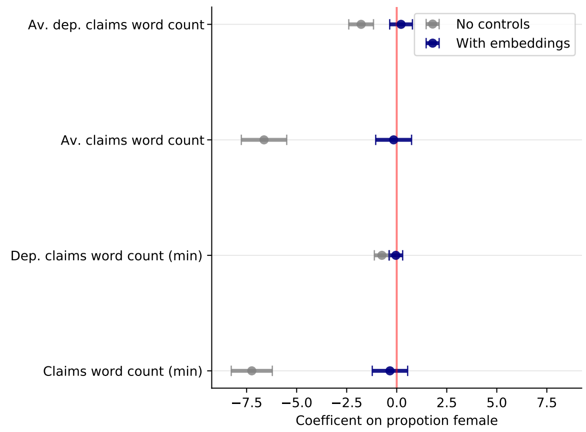
(c) Female 1st or 2nd

Note: This Figure plots the relationship between technology center indicators and the gender of the inventors' team with and without controlling for the patent application text embeddings. The gray dots are the estimates from an uncontrolled OLS regression, and the blue dots are the estimates from an OLS regression controlling linearly for the patent text embeddings. Bars indicate 95% confidence intervals based on robust standard errors.

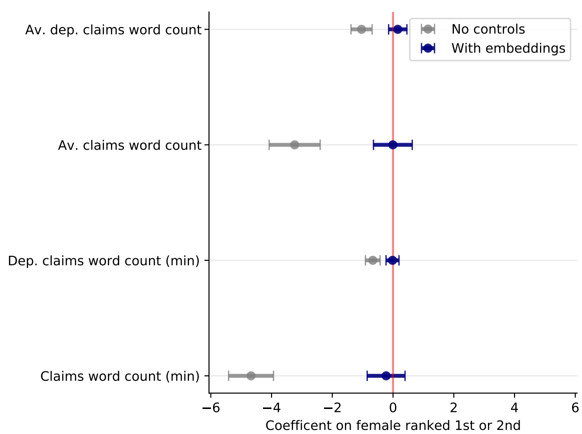
Figure 8: Balance test for claim variables



(a) Any female



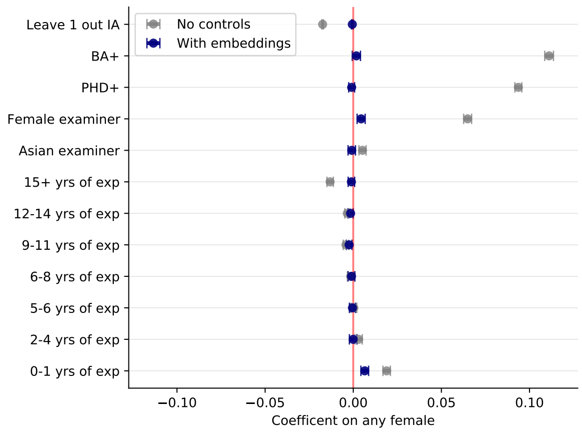
(b) Proportion female



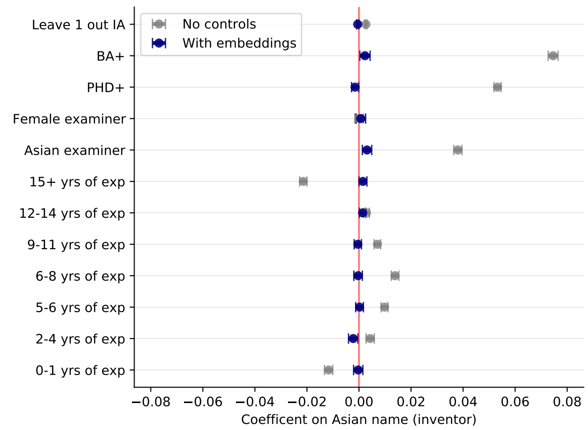
(c) Female 1st or 2nd

Note: This Figure plots the relationship between patent application claim counts and the gender of the inventors' team with and without controlling for the patent application text embeddings. The gray dots are the estimates from an uncontrolled OLS regression, and the blue dots are the estimates from an OLS regression controlling linearly for the patent text embeddings. Bars indicate 95% confidence intervals based on robust standard errors.

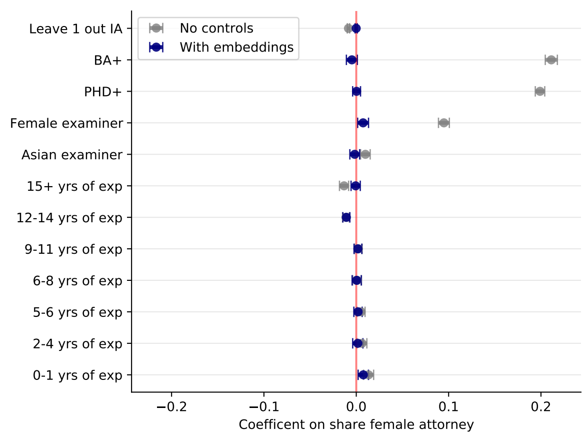
Figure 9: Balance test of examiner characteristics



(a) Mixed gender team



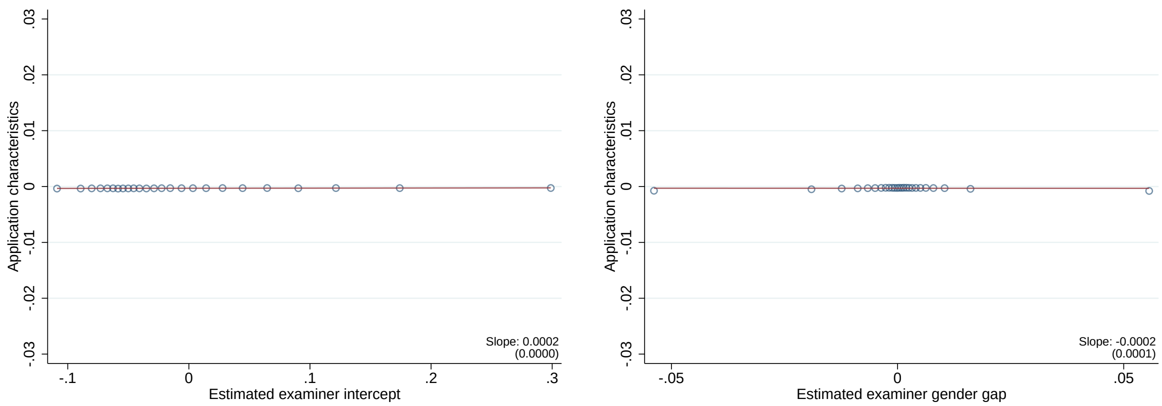
(b) Any Asian inventor



(c) Share female attorney

Note: This Figure plots the relationship between examiner characteristics and non-text characteristics of inventors and attorneys with and without controlling for the patent application text embeddings. The gray dots are the estimates from an uncontrolled OLS regression, and the blue dots are the estimates from an OLS regression controlling linearly for the patent text embeddings. Bars indicate 95% confidence intervals based on robust standard errors.

Figure 10: Omitted variable bias test in examiner leniency and gender bias

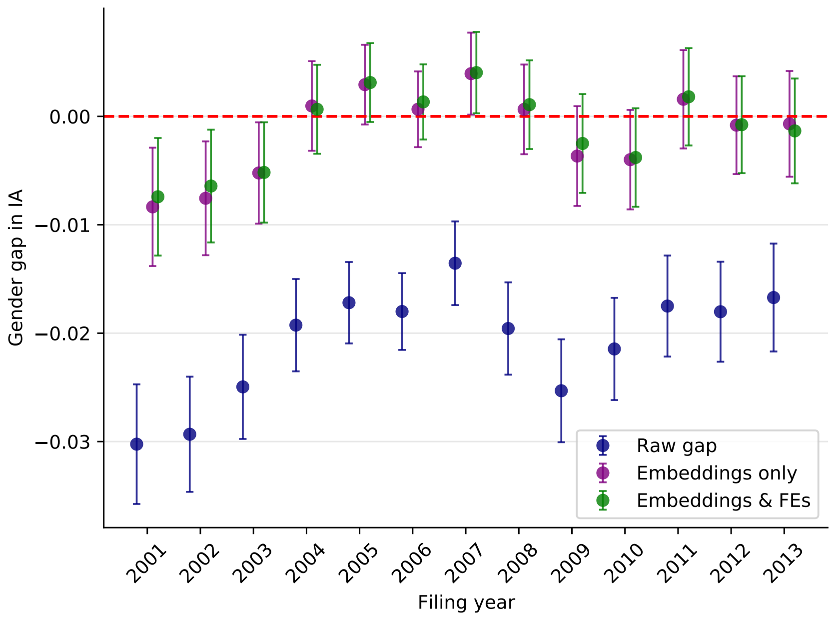


(a) Relationship w/ α_j

(b) Relationship w/ β_j

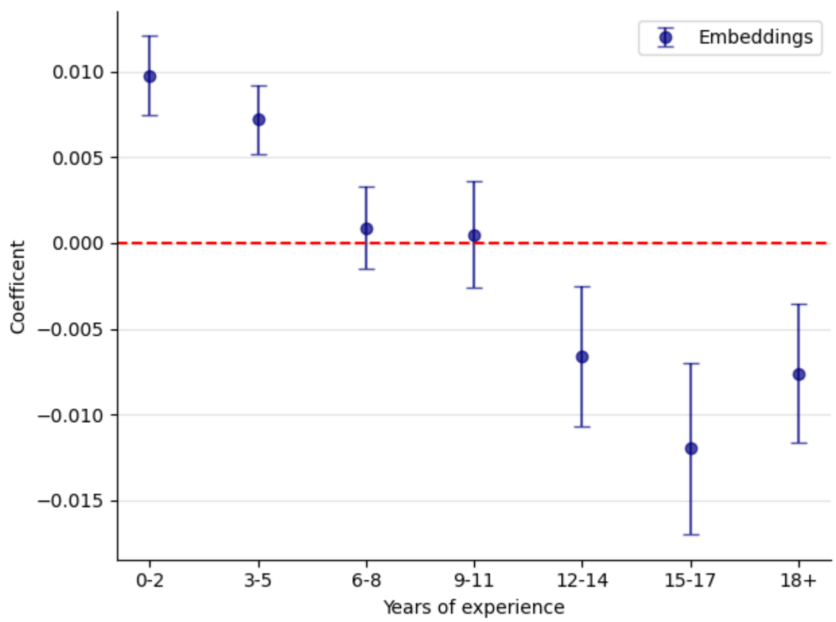
Note: This Figure plots the test for whether the estimated examiner leniency and gender bias effects ($\hat{\alpha}_j$ and $\hat{\beta}_j$ from equation 6) are correlated with other non-text characteristics that predict initial allowance: law firm experience, team size, an indicator for foreign priority, indicator for at least one Asian inventor, and proportion female attorney. The flat slops indicate that examiner effects are not affected by the inclusion of these covariates.

Figure 11: Gender Gap in initial allowance by filing Year



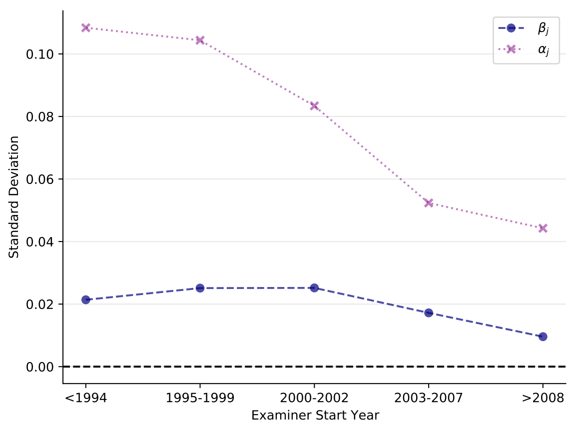
Note: This Figure plots the estimated gender gap and 95% confidence intervals in the initial allowance of patent applications filed between the years 2001 and 2013 by filing year. Blue dots plot the uncontrolled gender gap, purple dots plot the gender gap conditional on the text embeddings, and green dots plot the gender gap conditional on both text embeddings and art-unit-year and class fixed effects. Confidence intervals based on robust standard errors.

Figure 12: Gender gap by examiners' years of experience

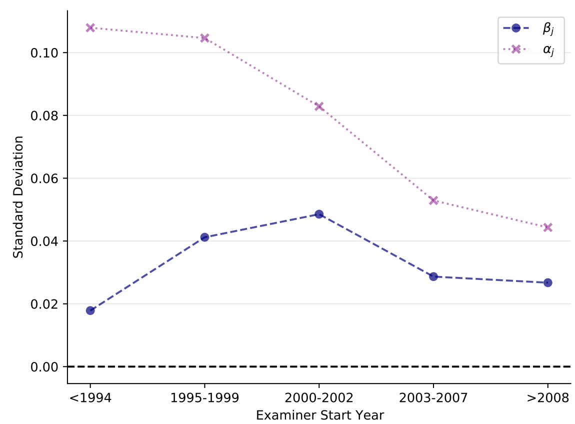


Note: This Figure plots the estimated gender gap and 95% confidence intervals in the initial allowance of patent applications filed between the years 2001 and 2013 by examiners' years of experience. The dots plot the gender gap conditional on the patent text embeddings. Confidence intervals based on robust standard errors.

Figure 13: Standard deviation of examiner gender gap and leniency by examiner start year



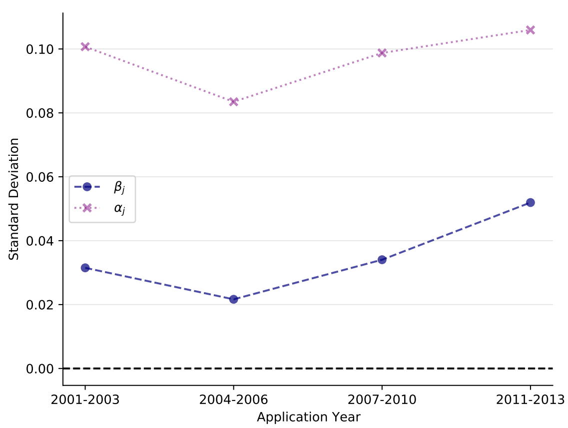
(a) Any female inventor



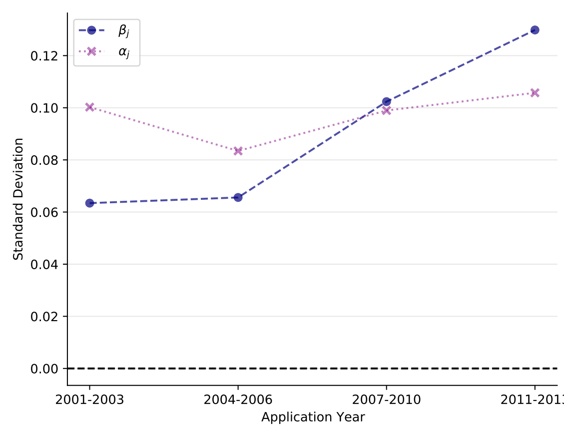
(b) Proportion female

Note: This Figure plots the biased corrected standard deviation of examiner bias and leniency. The variance component is estimated in a linear regression with examiner fixed effects and examiner times application gender fixed effects separately by bins of examiner start year. The blue dots are the estimated standard deviation of examiners' gender gap, and the pink "X"s are the estimated standard deviations of examiners' leniency. All variance estimates are weighted by the number of applications per examiner.

Figure 14: Standard deviation of gender gap and leniency over time



(a) Any female inventor



(b) Proportion female

Note: This Figure plots the biased corrected standard deviation of examiner bias and leniency. The variance component is estimated in a linear regression with examiner fixed effects and examiner times application gender fixed effects separately by bins of patent application filing year. The blue dots are the estimated standard deviation of examiners' gender gap, and the pink "X"s are the estimated standard deviations of examiners' leniency. All variance estimates are weighted by the number of applications per examiner.

Tables

Table 1: Patent applications' descriptive statistics

	Full sample (1)	All male or unk (2)	At least 1 female (3)	Female ranked 1st or 2nd (4)
# of observations	1224315	1019495	204820	143118
Team size	2.477	2.257	3.573	2.948
Proportion female	0.075	0.000	0.450	0.536
Sole inventor	0.351	0.395	0.131	0.188
Sole female inventor	0.022	0.000	0.131	0.188
Initial allowance (IA)	0.083	0.086	0.067	0.065
Ever granted	0.646	0.656	0.595	0.571

Note: This table presents the descriptive statistics of the patent applications for the main sample of US patent applications filed between the years 2001-2013. Column 1 presents the counts and means of the full sample, column 2 of the no female applications, column 3 of the set of patent applications with at least one female author, and column 4 for the set of patent applications with at least one female ranked first or second in the application list of inventors.

Table 2: Examiners' descriptive statistics

		By examiner gender		
	All	Female	Male	Unknown
	(1)	(2)	(3)	(4)
# of examiners	8550	2055	5128	1367
Initial allowance rates	0.067	0.053	0.072	0.068
By start-year				
<1995	0.203	0.200	0.211	0.176
1996-2001	0.314	0.330	0.304	0.326
>2001	0.483	0.470	0.485	0.498

Note: This table presents the descriptive statistics of the 8,550 examiners in my sample by examiners' gender and years of experience.

Table 3: OLS estimates of the mean gender gap in initial allowances

	(1)	(2)	(3)	(4)
	IA	IA	IA	IA
(i) Any female	-0.0200	-0.0035	-0.0010	-0.0007
	(0.0006)	(0.0006)	(0.0007)	(0.0006)
Adj. R^2	0.0007	0.0687	0.0682	0.1038
(ii) Proportion female	-0.0407	-0.0088	-0.0013	-0.0006
	(0.0011)	(0.0011)	(0.0012)	(0.0011)
Adj. R^2	0.0008	0.0687	0.0682	0.1038
(iii) Female 1st or 2nd	-0.0222	-0.0042	-0.0008	-0.0004
	(0.0007)	(0.0007)	(0.0007)	(0.0007)
Adj. R^2	0.0006	0.0687	0.0682	0.1038
Art-unit-year class FE	No	Yes	No	Yes
Embeddings	No	No	Yes	Yes
# of applications	1,220,512	1,220,393	1,220,512	1,220,393
# of examiners	8,519	8,519	8,519	8,519

Note: This table reports OLS estimates of regressions of an indicator for initial allowance on the gender of the inventors' team, represented by an indicator for having at least one female inventor (panel (i)), proportion female (panel (ii)), and an indicator for having female ranked first or second in the inventors' team (panel (iii)). Robust standard errors are reported in parentheses.

Table 4: Oaxaca-Blinder and Inverse Probability reweighting estimates for the mean gender gap in initial allowances

	Oaxaca-Blinder	IPW		DML
	ATT	ATE	ATT	NN
	(1)	(2)	(3)	(4)
Any female	-0.0019	-0.0015	-0.0006	0.0006
	(0.0013)	(0.0009)	(0.0006)	(0.0007)

Note: This table reports matching estimates of the overall gender gap in initial allowance between mixed-gender teams and no female teams conditional on BERT embedings. Column 1 reports the fully interacted (Oaxaca, 1973; Blinder, 1973) ATT estimate by running an OLS regression of initial allowance on embeddings only in the no-female sample and reporting the difference between the mean initial allowance rates of men and the fitted values on the no female regression on the mean embeddings among mixed-gender patents. Columns 2-3 report the Inverse Probability Weighting (IPW) estimates of the ATE and the ATT, estimating the propensity score with a logit model. Column (4) reports the Doubly Robust Machine Learning (DML) (Chernozhukov et al., 2018) partially linear regression implemented using the Python package DOUBLEML. The model predicts initial allowance and gender with a neural network (NN) model, where the network's number of layers, nodes in each layer, and regularization parameters were chosen by cross-validation. Robust standard errors are reported in parentheses.

Table 5: Heterogeneity in gender bias in initial allowances

	(1) Any Female	(2) Proportion Female	(3) Female 1st or 2nd
(i) By examiner			
$Std(\alpha_j)$	0.090	0.090	0.090
$Std(\beta_j)$	0.024	0.037	0.022
$Corr(\alpha_j, \beta_j)$	-0.649	-0.840	-0.714
$\bar{\beta}_j$	-0.0009	-0.0014	-0.0009
# of examiners	8335	8335	8147
# of obs	1216346	1216346	1207069
(ii) By start-year			
$Std(\alpha_j)$	0.038	0.038	0.038
$Std(\beta_j)$	0.012	0.020	0.012
$Corr(\alpha_j, \beta_j)$	-0.802	-0.774	-0.764
$\bar{\beta}_j$	-0.0011	-0.0008	-0.0009
# of cohorts	38	38	38
# of obs	1216346	1216346	1216346
(iii) By art-unit			
$Std(\alpha_j)$	0.052	0.052	0.052
$Std(\beta_j)$	0.012	0.020	0.010
$Corr(\alpha_j, \beta_j)$	-0.333	-0.407	-0.405
$\bar{\beta}_j$	-0.0013	-0.0025	-0.0016
# of art-units	587	587	585
# of obs	1216329	1216329	1216278

Note: This table presents the biased corrected standard deviation and correlations of examiners, examiners' start-year, and art units estimated in a linear regression controlling for patent text embeddings. Different columns present estimates using different measures of the femaleness of the application. Column (1) uses an indicator for mixed-gender patents, column (2) uses the proportion female, and column (3) uses an indicator for having a female ranked first or second in the inventors' list. All variance components are weighted by the number of patent applications.

Table 6: Heterogeneity in examiner, start-year, and art-unit gender bias explained by examiner experience

	(1) Any Female	(2) Proportion Female	(3) Female 1st or 2nd
(i) Examiner gap			
$Std(\beta_j)$	0.023	0.036	0.022
$Std(\beta_{exp})$	0.002	0.006	.
$Corr(\beta_j, \beta_{exp})$	0.125	0.025	.
# of examiners	8335	8335	8147
# of obs	1216315	1216315	1207038
(ii) Start-year gap			
$Std(\beta_j)$	0.010	0.016	0.011
$Std(\beta_{exp})$	0.003	0.008	.
$Corr(\beta_j, \beta_{exp})$	0.550	0.298	.
# of start-years	38	38	38
# of obs	1216315	1216315	1216315
(iii) Art-unit gap			
$Std(\beta_j)$	0.012	0.020	0.010
$Std(\beta_{exp})$	0.009	0.015	0.008
$Corr(\beta_j, \beta_{exp})$	-0.046	-0.050	-0.043
# of art-units	587	587	585
# of obs	1216298	1216298	1216247

Note: This table presents the biased corrected standard deviation and correlations of examiners, examiners' start-year, and art-units gender gap estimated in a linear regression controlling for patent text embeddings using both units fixed effect and examiners years of experience fixed effect as described in equation 7. Different columns present estimates using different measures of the femaleness of the application. Column (1) uses an indicator for mixed-gender teams, column (2) uses the proportion female, and column (3) uses an indicator for having a female ranked first or second in the list of patent inventors. All variance components are weighted by the number of patent applications. Dots indicate an estimated negative variance and, therefore, undefined standard deviation and correlation.

Table 7: Relationship between gender gap and examiner characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PhD+	0.0011***					0.0002	0.0000
	(0.0002)					(0.0002)	(0.0002)
Female		0.0010***				0.0004**	-0.0001
		(0.0001)				(0.0001)	(0.0001)
Asian name			-0.0000			-0.0004*	-0.0007***
			(0.0002)			(0.0002)	(0.0002)
Yrs of experience				-0.0004***		-0.0003***	-0.0003***
				(0.0000)		(0.0000)	(0.0000)
Sh. mixed gender apps					0.0068***	0.0044***	-0.0087***
in USPC class					(0.0006)	(0.0007)	(0.0014)
Art-unit FE	No	No	No	No	No	No	Yes
Applications	1218137	1218137	1218137	1218137	1218137	1218137	1218137
Examiners	8335	8335	8335	8335	8335	8335	8335

Note: This table presents the relationship between the estimated gender gap in initial allowance $\hat{\beta}_j$ and examiner characteristics estimated using an OLS regression. Columns 1-6 include only the examiner characteristics listed in the table, and column 7 also controls for art-unit fixed effects. Robust standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 8: Within examiner gender variability in gender gap

	Examiner gender		
	Female	Male	Unknown
	(1)	(2)	(3)
(i) OLS gap			
w/o embeddings	-0.0129	-0.0221	-0.0229
	(0.0011)	(0.0009)	(0.0018)
w/ embeddings	0.0007	-0.0020	-0.0021
	(0.0011)	(0.0010)	(0.0018)
(ii) Examiners fixed-effects			
$Std(\alpha_j)$	0.070	0.094	0.097
$Std(\beta_j)$	0.013	0.026	0.025
$\bar{\beta}_j$	-0.000	-0.001	-0.001
# of examiners	1,995	5,001	1,339
# of apps	289,703	732,605	194,038
Mean IA	0.062	0.090	0.091
Share mixed teams	0.187	0.138	0.153

Note: This table presents the distribution of gender bias by examiners' gender. Panel (i) presents the mean gender bias separately by examiner gender with and without controlling for the text embeddings. Panel (ii) presents the biased corrected standard deviation of examiners' gender gap and leniency estimated in a linear regression controlling for patent text embeddings. All variance components are weighted by the number of patent applications.

Table 9: Gender bias in initial allowance and the rejection reason

	Rejection type			
	Obviousness (103)	Novelty (102)	Eligibility (101)	Writing (112)
	(1)	(2)	(3)	(4)
(i) OLS gap				
w/o embeddings	0.001	-0.014	0.022	0.041
	(0.002)	(0.002)	(0.001)	(0.002)
w/ embeddings	0.002	0.000	-0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
(ii) Examiner fixed-effects				
(a) Variability				
$Std(\beta_j)$	0.028	0.028	0.028	0.026
$Std(\beta_j^R)$	0.046	0.043	0.039	0.041
$Corr(\beta_j, \beta_j^R)$	-0.493	-0.326	-0.255	-0.530
$\bar{\beta_j}$	0.000	0.000	0.000	0.000
$\bar{\beta_j^R}$	0.003	0.001	-0.000	-0.003
(b) Implied OLS				
coefficient	-0.298	-0.211	-0.184	-0.345
Implied R^2	0.254	0.107	0.065	0.297
# of examiners	6,996	6,996	6,996	6,996
# of apps	504,884	504,884	504,884	504,884
Mean IA	0.098	0.098	0.098	0.098
Mean rejection outcome	0.694	0.519	0.145	0.354

Note: This table presents the mean gender gap and its variability in initial allowance and the grounds for rejections. Panel (i) presents the mean gender gap of rejection reasons estimated in an OLS regression with and without text embeddings. Panel (ii) presents the biased corrected standard deviation of examiners' gender gap estimated in a stacked regression having both initial allowance and other rejection reasons as outcomes, as described in Appendix section C. All the regressions include controls for the patent text embeddings. All variance components are weighted by the number of patent applications.

Table 10: Estimated gender gap and the relationship with other non-text characteristics

	(1)	(2)	(3)	(4)
(i) OLS gap				
Any female	-0.0010	-0.0010	-0.0010	-0.0011
	(0.0007)	(0.0007)	(0.0007)	(0.0007)
Other effect		-0.0002	-0.0008	0.0056
		(0.0006)	(0.0006)	(0.0009)
# of applications	1,220,512	1,220,512	1,220,512	1,220,512
# of examiners	8,519	8,519	8,519	8,519
(ii) Fixed-effect regression				
Std. any female	0.023	0.022	0.022	0.023
Std. other effect		0.022	0.023	0.039
Correlation		0.024	0.008	0.031
# of applications		1,215,230	1,214,970	1,216,346
# of examiners		8,300	8,300	8,335
Characteristic	-	Any foreign	Asian name	Lawyer experience

Note: This table presents the mean gender gap, its variability, and its sensitivity to the inclusion of other non-text characteristics. Panel (i) presents the coefficients on mixed-gender teams and the other characteristics estimated in an OLS regression controlling for text embeddings. Panel (ii) presents the biased corrected standard deviation of examiner gender gap and examiner level sensitivity to other characteristics estimated in a regression on initial allowance on examiner fixed effects, examiner time gender fixed effects, and examiner times other characteristic fixed effects. All the regressions include controls for the patent text embeddings. All variance components are weighted by the number of patent applications. The other characteristic's name is displayed in the last row.

Table 11: Sample selection parameter estimates

	log(R)
	(1)
(i) IA	
μ_0	-1.749
	(0.010)
μ_1	-0.010
	(0.017)
(ii) Outcome	
ψ_0	1.275
	(0.067)
ψ_1	0.247
	(0.025)
(iii) Random effects	
σ_{λ_0}	0.650
	(0.003)
σ_{λ_1}	0.338
	(0.007)
$Corr(\lambda_1, \lambda_0)$	0.039
	(0.220)
(iv) Outcome-IA dist.	
ρ	0.105
	(0.041)
σ	1.484
	(0.007)
Likelihood	-122581.4
# of parameters	19
# of apps	472408
# of examiners	7550

Note: This table presents the parameter estimates of the distribution of gender bias in the first round of the patent examination process and stock market return by estimating a selection model. μ_0 and $\sigma_{\lambda 0}$ determine the distribution of initial allowance decision for all male applications, and μ_1 and $\sigma_{\lambda 1}$ determine the distribution of initial allowance decision of mixed gender applications. ψ_0 describes the mean Kogan et al. (2017) stock market return of all male applications, and ψ_1 describes the difference between all male and mixed-gender patents. $Corr(\lambda_1, \lambda_0)$ describes the correlation between examiners allowance threshold of all male vs. mixed gender patents. Panel (iv) reports the parameters of the joint distribution of initial allowance and market return. The model includes the first two demeaned UMAP components of the text embeddings and their interaction with gender as controls, allowing the random thresholds to vary by these dimensions. We assume that the random thresholds that vary with the UMAP embeddings of all male and mixed-gender applications are uncorrelated. Observations include only patents assigned to publicly traded firms, and the market return model is estimated for initially allowed patents. Robust standard errors are presented in parentheses.

Table 12: Counterfactuals under uniformly zero bias

	(1) Allowance given $IA_i(\sigma_{\lambda_1} > 0) = 0 \rightarrow IA_i(\sigma_{\lambda_1} = 0) = 1$	(2) Allowance taken $IA_i(\sigma_{\lambda_1} > 0) = 1 \rightarrow IA_i(\sigma_{\lambda_1} = 0) = 0$	(3) Difference	(4) Total market loss
(i) Mixed-gender apps				
Av. market return	5.867	5.688	0.179	1.130
Share	0.0116	0.0117		
(ii) All-male apps				
Av. market return	4.572	4.529	0.043	0.289
Share	0.0022	0.0022		

Note: This table reports the average market return (measured in a million US dollars) of all-male and mixed-gender patent applications under the status-quo distribution of examiners' gender bias reported in Table 11 and a counterfactual exercise uniform zero bias. Column (1) reports the average market return among the patent applications that were not allowed under the status quo ($IA(\sigma_{\lambda_1} > 0) = 0$) but are allowed under the uniform zero bias simulation ($IA(\sigma_{\lambda_1} = 0) = 1$). Column (2) reports the average market return among the applications that are allowed under the status quo ($IA(\sigma_{\lambda_1} > 0) = 1$) but not allowed under the uniform zero bias ($IA(\sigma_{\lambda_1} = 0) = 0$). Column (3) reports the difference between column (1) and column (2). Column (4) reports the total yearly difference in market return for 40,000 applications assigned to publicly traded firms, with 15% mixed-gender teams and the rest all-male teams.

A Appendix Figures and Tables

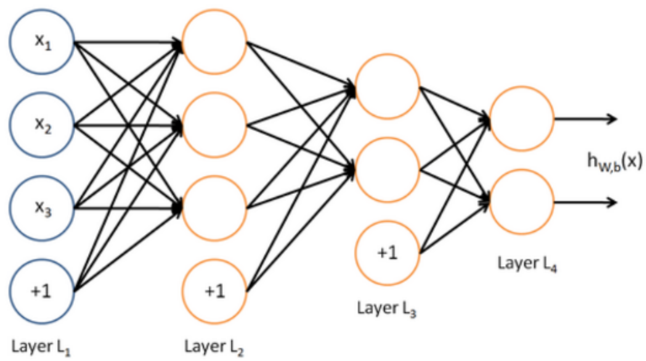
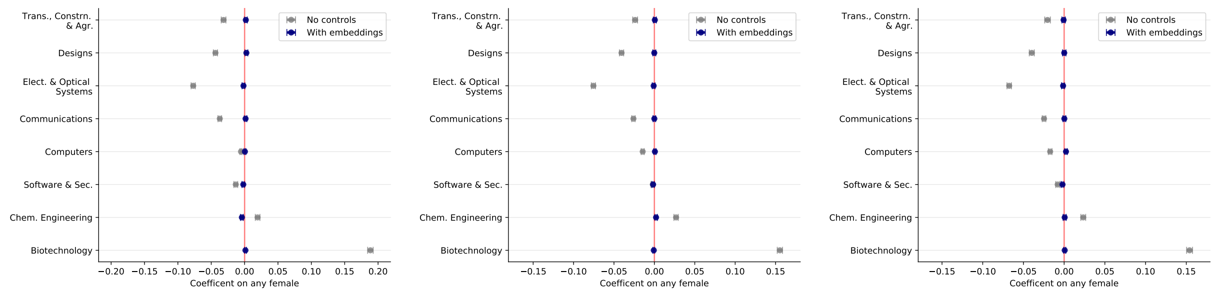


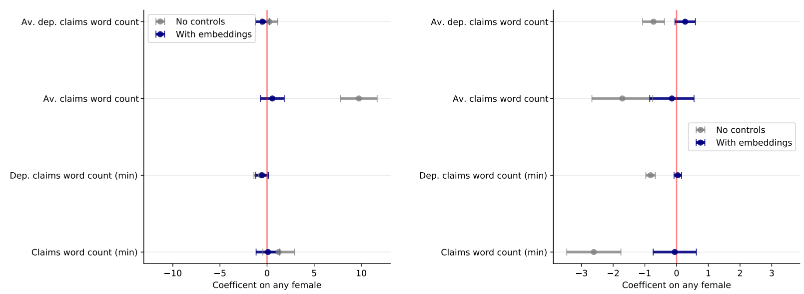
Figure A.1: Illustration of Generic Neural Network Model

Note: This Figure illustrates an example of a generic neural network model with 3 hidden layers.

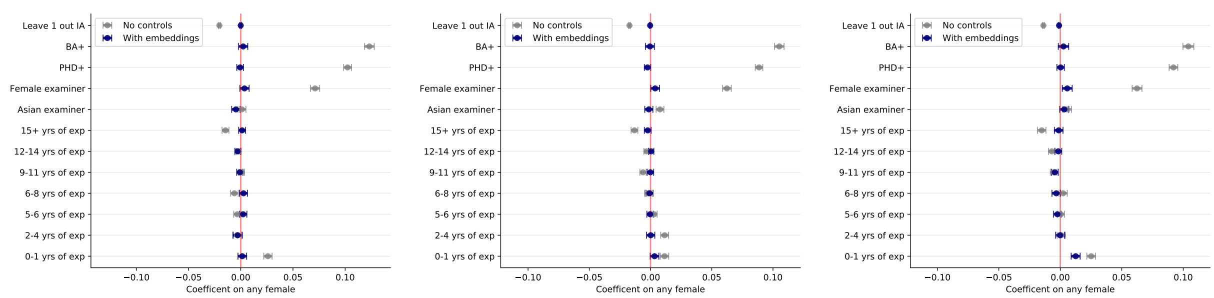
Figure A.2: Balance test by year



(a) 2001-2004 (b) 2005-2009 (c) 2010-2013



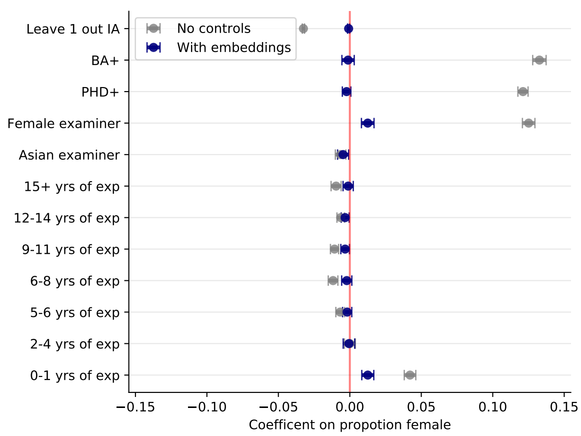
(d) 2001-2004 (e) 2005-2009 (f) 2010-2013



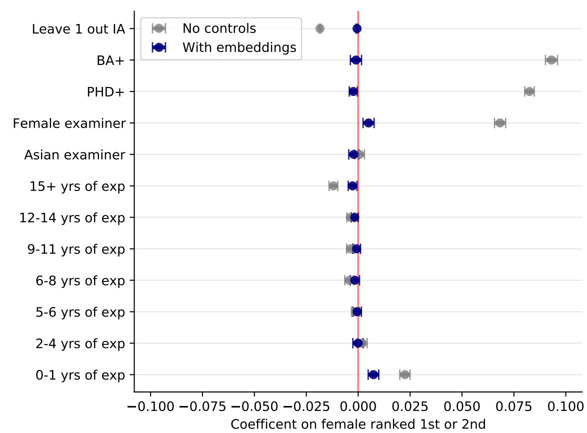
(g) 2001-2004 (h) 2005-2009 (i) 2010-2013

Note: This Figure plots the relationship between patent application and examiner characteristics and the gender of the inventors' team with and without controlling for the patent application text embeddings. The gray dots are the estimates from an uncontrolled OLS regression, and the blue dots are the estimates from an OLS regression controlling linearly for the patent text embeddings. Bars indicate 95% confidence intervals based on robust standard errors.

Figure A.3: Balance test for examiner characteristics using alternative female definitions



(a) Proportion female



(b) Female 1st or 2nd

Note: This Figure plots the relationship between examiner characteristics and technology centers with different measures of patent "femaleness" with and without controlling for the patent application text embeddings. The gray dots are the estimates from an uncontrolled OLS regression and the blue dots are the estimates from an OLS regression controlling linearly for the patent text embeddings. Bars indicate 95% confidence intervals based on robust standard errors.

Table A.1: Adjusted R^2 on initial allowance and any female by embedding layer

			Adj. R^2		Partial R^2, art-unit-yr class		Partial R^2, examiner	
Claim layer (1)	Description layer (2)	# of embeddings (3)	IA (4)	Any female (5)	IA (6)	Any female (7)	IA (8)	Any female (9)
(0, 0, 0)	(1, 0, 0)	1023	0.0499	0.0975	0.0213	0.0377	0.0238	0.0381
(0, 0, 0)	(0, 0, 1)	1023	0.0499	0.0939	0.0222	0.0349	0.0250	0.0351
(0, 0, 0)	(0, 1, 0)	1023	0.0500	0.0961	0.0217	0.0365	0.0244	0.0368
(0, 0, 0)	(1, 1, 0)	2046	0.0535	0.1032	0.0248	0.0440	0.0276	0.0444
(0, 0, 0)	(0, 1, 1)	2046	0.0544	0.1016	0.0256	0.0426	0.0287	0.0429
(0, 0, 0)	(1, 1, 1)	3069	0.0566	0.1067	0.0280	0.0483	0.0311	0.0486
(1, 0, 0)	(0, 0, 0)	1023	0.0616	0.0908	0.0329	0.0312	0.0375	0.0315
(0, 1, 0)	(0, 0, 0)	1023	0.0626	0.0902	0.0342	0.0308	0.0390	0.0311
(0, 0, 1)	(0, 0, 0)	1023	0.0639	0.0893	0.0361	0.0302	0.0414	0.0304
(1, 1, 0)	(0, 0, 0)	2046	0.0668	0.0950	0.0382	0.0360	0.0434	0.0362
(1, 0, 0)	(1, 0, 0)	2046	0.0673	0.1014	0.0382	0.0425	0.0434	0.0428
(1, 0, 0)	(0, 1, 0)	2046	0.0675	0.1010	0.0383	0.0421	0.0436	0.0424
(1, 0, 0)	(0, 0, 1)	2046	0.0675	0.1005	0.0385	0.0417	0.0439	0.0420
(0, 1, 0)	(0, 1, 0)	2046	0.0682	0.1005	0.0393	0.0417	0.0448	0.0420
(0, 1, 0)	(1, 0, 0)	2046	0.0684	0.1015	0.0393	0.0426	0.0448	0.0429
(0, 1, 1)	(0, 0, 0)	2046	0.0684	0.0947	0.0399	0.0358	0.0454	0.0361
(0, 0, 1)	(0, 0, 1)	2046	0.0693	0.0991	0.0411	0.0407	0.0470	0.0409
(1, 0, 0)	(1, 1, 0)	3069	0.0696	0.1060	0.0407	0.0479	0.0461	0.0482
(1, 0, 0)	(0, 1, 1)	3069	0.0702	0.1051	0.0412	0.0470	0.0466	0.0473
(0, 1, 0)	(1, 1, 0)	3069	0.0706	0.1060	0.0418	0.0479	0.0473	0.0482
(0, 1, 0)	(0, 1, 1)	3069	0.0710	0.1047	0.0422	0.0467	0.0479	0.0470
(1, 1, 0)	(1, 0, 0)	3069	0.0716	0.1044	0.0427	0.0462	0.0484	0.0465
(1, 1, 0)	(0, 1, 0)	3069	0.0717	0.1040	0.0428	0.0458	0.0485	0.0461
(1, 1, 0)	(0, 0, 1)	3069	0.0718	0.1036	0.0430	0.0455	0.0488	0.0458
(0, 0, 1)	(1, 1, 0)	3069	0.0722	0.1061	0.0435	0.0480	0.0493	0.0483
(0, 0, 1)	(0, 1, 1)	3069	0.0725	0.1046	0.0438	0.0466	0.0498	0.0469
(0, 1, 1)	(0, 1, 0)	3069	0.0730	0.1037	0.0443	0.0456	0.0503	0.0459
(0, 1, 1)	(1, 0, 0)	3069	0.0731	0.1046	0.0443	0.0464	0.0502	0.0467
(0, 1, 1)	(0, 0, 1)	3069	0.0731	0.1032	0.0446	0.0452	0.0506	0.0455
(1, 1, 0)	(1, 1, 0)	4092	0.0732	0.1080	0.0448	0.0507	0.0505	0.0510
(1, 1, 0)	(1, 0, 1)	4092	0.0736	0.1079	0.0452	0.0506	0.0510	0.0509
(1, 0, 1)	(1, 1, 0)	4092	0.0745	0.1082	0.0462	0.0509	0.0522	0.0512
(1, 0, 1)	(1, 0, 1)	4092	0.0746	0.1077	0.0463	0.0504	0.0524	0.0507
(0, 1, 1)	(0, 1, 1)	4092	0.0748	0.1070	0.0466	0.0498	0.0527	0.0501
(0, 1, 1)	(1, 0, 1)	4092	0.0749	0.1080	0.0466	0.0507	0.0527	0.0510
(1, 0, 1)	(0, 1, 1)	4092	0.0749	0.1073	0.0465	0.0500	0.0526	0.0503
(1, 1, 1)	(1, 1, 1)	6138	0.0779	0.1123	0.0506	0.0567	0.0568	0.0570

Note: This table presents the adjusted R^2 and the partial adjusted R^2 from running regressions of initial allowance and a mixed gender indicator on different layers of text embeddings. Column 1 displays the claims text embeddings layer, column 2 presents that description text embeddings layer and column 3 displays the total number of embeddings in the estimated regression where each embedding layer includes 1023 features.

Table A.2: Gender gap in initial-allowance for single gender teams

	IA (1)	IA (2)	IA (3)	IA (4)
(i) All female vs no female	-0.0364	-0.0112	-0.0010	-0.0002
	(0.0013)	(0.0014)	(0.0014)	(0.0014)
Adj. R^2	0.0004	0.0688	0.0693	0.1050
# of applications	1,062,510	1,062,369	1,062,510	1,062,369
# of examiners	8,519	8,519	8,519	8,519
(ii) Sole inventor	-0.0345	-0.0109	-0.0013	-0.0006
	(0.0015)	(0.0015)	(0.0016)	(0.0016)
Adj. R^2	0.0008	0.0691	0.0743	0.1085
# of applications	428,330	428,040	428,330	428,040
# of examiners	8,516	8,516	8,516	8,516
Art-unit-year class FE	No	Yes	No	Yes
Embeddings	No	No	Yes	Yes

Note: This table reports OLS estimates of regressions of an indicator for initial allowance on the gender of the inventors' team, using only the sample of patent applications with all female inventors teams vs. no female inventors' teams (panel (i)), and the sample of sole inventors (panel (ii)). Robust standard errors are reported in parentheses.

Table A.3: Variance components using inverse probability weighting

	(1) KSS	(2) IPW
$Std(\alpha_j)$	0.090	0.0930
$Std(\beta_j)$	0.024	0.0174
$\bar{\beta}_j$	-0.001	-0.0051
# of examiners	8335	7550
# of applications	1,216,346	1,220,195

Note: This table reports the variance components of examiners leniency (α_j) and examiners gender bias (β_j). Column (1) reports the variance components from Table 5, by estimating a linear model with examiners and examiners time gender fixed effect controlling for embeddings linierly. Column (2) reports the variance component by first estimating the propensity score of each examiner and application gender only among examiners with at least 100 applications. And then estimating the variance component using the leave-one-out formula (for further detail, see Appendix Section C.2), reweighting the observation using the propensity score.

Table A.4: Within examiner gender variability in gender gap, proportion female

	Female	Male	Unknown
	(1)	(2)	(3)
(i) OLS gap			
w/o emmbeddings	-0.0288	-0.0439	-0.0459
	(0.0018)	(0.0017)	(0.0033)
w/ emmbeddings	0.0001	-0.0021	-0.0008
	(0.0019)	(0.0017)	(0.0033)
(ii) Fixed-effect regression			
$Std(\alpha_j)$	0.071	0.094	0.097
$Std(\beta_j)$	.	0.032	0.073
$\bar{\beta}_j$	-0.001	-0.002	-0.000
# of examiners	1,995	5,001	1,339
# of apps	289,703	732,605	194,038
Mean IA	0.062	0.090	0.091
Share mixed teams	0.087	0.061	0.065

Note: This table presents the distribution of gender gap by examiners' gender using the proportion of female in the inventors' team as a measure for the "femaleness" of the patent application. Panel (i) presents the mean gender gap separately by examiners gender with and without controlling for the text embeddings. Panel (ii) presents the Kline et al. (2020) biased corrected variance components of examiners gender gap and leniency estimated in a linear regression controlling for patent text embeddings. All variance components are weighted by the number of patent applications.

Table A.5: Within examiner gender variability in gender gap, female ranked first or second

	Female	Male	Unknown
	(1)	(2)	(3)
OLS gap			
w/o emmbeddings	-0.0163	-0.0238	-0.0253
	(0.0012)	(0.0011)	(0.0020)
w/ emmbeddings	-0.0003	-0.0012	-0.0016
	(0.0013)	(0.0011)	(0.0020)
Fixed-effect regression			
$Std(\alpha_j)$	0.070	0.094	0.097
$Std(\beta_j)$	0.013	0.025	0.017
$\bar{\beta}_j$	-0.002	-0.001	0.000
# of examiners	1,953	4,875	1,319
# of apps	288,120	726,023	192,926
Mean IA	0.062	0.090	0.091
Share mixed teams	0.133	0.096	0.104

Note: This table presents the distribution of gender gap by examiners' gender using an indicator for having at least one female ranked first or second in the inventors' team as a measure for the "femaleness" of the patent application. Panel (i) presents the mean gender gap separately by examiners gender with and without controlling for the text embeddings. Panel (ii) presents the Kline et al. (2020) biased corrected variance components of examiners gender gap and leniency estimated in a linear regression controlling for patent text embeddings. All variance components are weighted by the number of patent applications.

B Data Appendix

B.1 Patent Data

The main data source is the USPTO Patent Examination Research Dataset (Graham et al., 2015) which includes the universe of all public patent applications available online in the Public Patent Application Information Retrieval system (Public PAIR).[28] For every patent application, the Public PAIR data includes information on inventors' first and last name together with additional variables such as country, application number, publication number, the grant date if granted, and examiners, art-unit, and technological classes and sub-classes identifiers.

This dataset is merged to several other datasets:

1. The USPTO Patents View data.[29] It includes detailed information on both granted patents and patent applications. Specifically, this source includes the list of all the patents it cites, including granted and non-granted patents, and an implied identifier for inventors.

2. The Patent Claims Research Dataset (Marco et al., 2019).[30] This data-set includes detailed information on the number of claims and change in claims of patent applications and granted patent.

3. "Google Patents Research Data" from which I merged the abstract and description of each patent application.

4. Examiners roster, pay scale, and education data. Frakes and Wasserman (2014) generously provided me with detailed roster data and pay scale dated from 1994 they received through FOIA requests. These data are used to determine the start year of examiners (see detail below) and their years of experience.

5. Kogan et al. (2017) patent market value data. They provide estimates of the market

[28] The data can be found here: https://www.uspto.gov/learning-and-resources/electronic-data-products/patent-examination-research-dataset-public-pair

[29] https://www.patentsview.org/download/

[30] https://www.uspto.gov/learning-and-resources/electronic-data-products/patent-claims-research-dataset

value using a series of event study designs of the stock market return of patents among publicly traded patents.

6. Patent Maintenance Fee Events.[31] This dataset records all the maintenance fee events for granted patents granted from September 1, 1981. These fees are due 4, 8, and 12 years after the patent grant and are increasing over time.

7. USPTO Office Action Rejection.[32] ‘Office action” is a written notification to the applicant of the examiner’s decision on patentability. It generally discloses the reasons for any rejections, objections, or requirements and includes relevant information or references that the applicant may find useful for responding to the examiner and deciding whether to continue prosecuting the application. This data set includes all the mailed office actions from 2008 to 2017. It includes information on the grounds for rejections raised.

Sample restriction: I restrict the sample to utility[33] patent applications filed after November 29th, 2000[34] and before January 1st, 2014. To avoid detecting differential behavior to non-US inventors, and since the gender identification relies on the gender distributions of first names provided by the US Social Security Administration, I include only patent applications written by US inventors in my analysis.

B.2 Variable Construction

B.2.1 Inventors’ name coding

The application data includes for every patent application the first name, middle name, and last name of each inventor. To assign gender and an indicator for a foreign name I follow the following procedure. For every first name, I assign a probability of being a woman based on the gender name distribution provided by the US Security Administration (SSA).[35] In

[31] https://www.uspto.gov/learning-and-resources/electronic-data-products/additional-patent-data-products

[32] https://developer.uspto.gov/product/patent-application-office-actions-data-stata-dta-and-ms-excel-c

[33] Utility patents are granted for the “invention of a new and useful process, machine, manufacture, or composition of matter” (USPTO 2010).

[34] Since the American Inventors Protection Act of 1999 almost all the USPTO patent applications filed after 29 November 2000 were published online, regardless of whether they are granted or not.

[35] https://www.ssa.gov/oact/babynames/limits.html

cases where the first name is missing or includes only the first letter, I assign gender to the middle name.[36] Applications with at least one inventor with missing first and middle names are excluded from the analysis.

A name is classified as male if the threshold probability for male names in the SSA data is higher than 90%. Since women make up roughly 11% of the inventors and men account for 80%,[37] I set the women threshold to be higher, at 98.5%, roughly equating the type one error across gender, assuming the distribution of names in the general population is the same in the inventor population. Using this protocol I could assign gender to XX% of the names in my final sample.

B.2.2 Examiner Gender Coding

Unlike with inventors' gender coding, my goal was to assign gender to all possible examiners. Therefore I use a collection of data sources by assigning gender to each name by the following order:

1. The US SSA administrative baby names by gender.
2. Name gender published by the United Kingdom Intellectual Property Organization. This dataset is binary, a name is classified as either male or female.
3. WIPO Dataset.[38] Name gender dataset published by World Intellectual Property Organization. Its main advantage is that it includes names of different languages across countries.
4. gender-guesser Python package.
5. After exhausting all the datasets mentioned before, I use https://genderize.io/Genderize.io for detection of non-East-Asian names because they are known to not be accurate for Asian names.[39]

[36] Although the middle name could potentially be of a gender different from the first-name, I assume that for the examiners, this is the name that embodies the gender signal.

[37] Lissoni et al. (2018) find that in 2016 women account for 12% of the inventors. Following their analysis and using their data that identified the gender of inventors of granted patents using by country name distribution, I find that in the time period of this study, women account for 11% of the inventors.

[38] https://www.wipo.int/publications/en/details.jsp?id=4125

[39] For more details see https://jmla.pitt.edu/ojs/jmla/article/view/1289

As a result, I could identify X% of the examiners in my data set.

B.2.3 Ethnicity

To identify the ethnicity of examiners and inventors I apply the raceBERT algorithm (Parasurama, 2021) that was trained on the U.S. Florida voter registration data set using a BERT architecture. The model predicts the likelihood of a name belonging to 5 U.S. census race categories (White, Black, Hispanic, Asian Pacific Islander, American Indian Alaskan Native). Using this algorithm I classify an Asian name as either east Asian or Japanese. Using this algorithm I find that 46% of the unique inventor names in my data are Asians.

B.2.4 Examiners' Start Year and Years of Experience

To calculate the years of experience of each examiner and start year we use two sources of information. First, the roster data from Frakes and Wasserman (2014) includes years 1992-2014. Because the data is censored from below and we fill in the missing examiners and years by identifying the first office action using the transaction data set provided in the public PAIR data-set. Specifically, any examiner application transactions with "DOCK" record indicates an assignment or a change in assignment of an application's examiner, therefore for every application, I use the date of the most recent "DOCK" record to indicate the date on which the application is docketed to the current examiner.

There is a concern that the start year classification using the transaction data is downward biased as the examiner records in Public PAIR data set assign the application's examiner as the one who was assigned to process or archive the application at the time of disposal. In an effort to account for that, we modify the start year of examiners that have suspiciously long "gap years", meaning they have no assigned patents after the start years. We do so by defining the start year only if the examiner doesn't have a gap size of a certain size. We find the optimal gap size by minimizing the distance between the start year from the administrative FOIA records and the implied start year for examiners whose start year is greater than 1994 and apply this rule to the examiners with missing information.

B.2.5 Attorneys

The Public PAIR data set has a table named "attorney_agent" which records the first name, last name, and practice category of patent attorney(s) in each application. Interestingly, the filing of patent applications seems to have a large number of attorneys/agents involved where the average number of attorneys/agents is 28.20.

C Leave out Estimation of the Variance Component

The model in the reduced form analysis is an OLS regression of the form:

$$y_i = \alpha_{j(i)} + \beta_{j(i)} F_i + x_i' \gamma + \epsilon_i$$

where y_i is the outcome of interest, usually an indicator for initial allowance, F_i indicates the femaleness of the patent application, usually an indicator for a mixed gender patent, α_j are examiner fixed effects, β_j is the examiner level tendency to overvalue patent written by female inventors, and x_i is a vector of over 2,000 continuous embeddings. This specification can be written as:

$$y_i = X_i' \eta + \epsilon_i$$

where X_i collects the vectors of examiner indicators, examiner times gender indicators, and the embedding features. Using that matrix representation, any variance component can be written in a quadratic form:

$$\sigma^2 = \delta' A \delta$$

where $\delta = (\alpha', \beta')'$ are the collected $\alpha = (\alpha_1, .., \alpha_J)'$, and $\beta = (\beta_1, .., \beta_J)'$ examiner level coefficients, and A is the relevant weighting matrix. Kline et al. (2020) suggest estimating:

$$\hat{\sigma^2} = \hat{\delta}' A \hat{\delta} - \sum_{i=1}^{n} B_{ii} \hat{\xi}_i^2$$

where $B_{ii} = X_i' S_{ii}^{-1} A S_{ii}^{-1} X_i$ measures the influence of the i's squared error ϵ_i^2, $S_{ii} = \sum_i X_i X_i'$, $\xi_i^2 = \mathbb{V}(\epsilon_i | X_i)$ is the variance of the i's error, $\hat{\xi}_i^2 = \frac{y_i(y_i - X_i'\eta)}{1 - P_{ii}}$ is a the leave-i-out estimator described in Kline et al. (2020), and $P_{ii} = X_i' S_{ii}^{-1} X_i$ is the leverage of the i's observation on the estimate of $\hat{\eta}$. Kline et al. (2020) provide the conditions on the X_i matrix that ensure consistency of the bias corrected estimator.

Computation of $\hat{\sigma^2}$ is intensive as it requires computing the B_{ii} and P_{ii} with a model with over 18 thousand parameters. Therefore, as suggested in Kline et al. (2020) I exploit the random projection method by Lindenstrauss (1984) when approximating $\hat{\sigma^2}$. The Matlab code provided by Kline et al. (2020) relays on MATLAB's preconditioned conjugate gradient routine *pcg* which solves big linear equations in large sparse problems. However, in my settings X_i includes a dense embedding component preventing the algorithm to converge. To accommodate this problem I orthogonalize the embeddings component matrix $E_i = QR$ using QR decomposition so $Q_i'Q_i$ equals to identity matrix that can be represented as a sparse matrix and enables convergence of the *pcg* function.

C.1 Covariance Across Different Regressions

In Section 9.1 I estimate the following model:

$$IA_i = \alpha_{j(i)} + \beta_{j(i)} F_i + x_i'\gamma + \epsilon_i$$
$$R_i = \alpha^R_{j(i)} + \beta^R_{j(i)} F_i + x_i'\gamma^R + \epsilon^R_i$$

where IA_i is an indicator for initial allowance and R_i is an indicator for the rejection reason in the first round of examination. The variance component of interest is $cov(\beta_{j(i)}, \beta^R_{j(i)})$, the covariance between the gender gap in initial allowance and the gender gap in that particular rejection reason. Writing this model in a matrix representation:

$$IA_i = X_i'\eta + \epsilon_i$$
$$R_i = X_i'\eta^R + \epsilon^R_i$$

we can easily see that both of the regressions share the same design matrix X_i described above. As described in Lachowska et al. (2022) the estimator of the covariance using the leave-out procedure is therefore:

$$\widehat{cov(\beta_{j(i)}, \beta^R_{j(i)})} = \hat{\beta}' A \hat{\beta}^R - \sum_{i=1}^{n} B_{ii} \hat{\xi}^2_{i12}$$

where $\beta = (\beta_1, .., \beta_J)'$ are the collected examiner level gender bias in initial allowance and $\beta^R = (\beta^R_1, .., \beta^R_J)'$ are the collected examiner level gender bias R_i, A_i is the relevant weighting matrix, B_{ii} is identical to the one described in the previous section and $\hat{\xi}^2_{i12} = \frac{R_i(IA_i - X_i\hat{\eta})}{1-P_{ii}}$ is the leave-i-out estimator of the covariance of the i's error in the two regression models.

C.2 Variance component of reweighted observations

There are J examiners, each examiner examines n_j patents, n_{jf} of female inventors, and n_{jm} of male inventors. denote y_{ij} the (residualized) outcome of each patent examined by examiner j. Assume $(y_{ij})_{i=1}^{n_j}$ are independent. Denote y_{ijg} the outcome of gender $g \in \{0,1\}$

Let $\theta_{jf} = \mathbb{E}[y_{ji}|F=1]$, $\theta_{jm} = \mathbb{E}[y_{ji}|F=0]$, and define for every examiner the gender gap:

$$\beta_j = \theta_{jf} - \theta_{jm}$$

The variance of examiners gender gaps:

$$\begin{aligned}\mathbb{V}(\beta_j) &= \frac{1}{J}\sum_{j=1}^{J}\beta_j^2 - \left(\frac{1}{J}\sum_{j=1}^{J}\beta_j\right) \\ &= \frac{J-1}{J^2}\sum_{j=1}^{J}\beta_j^2 - \frac{2}{J^2}\sum_{j=2}^{J}\sum_{k=1}^{j-1}\beta_j\beta_k\end{aligned}$$

The unbiased estimate analog is:

$$\widehat{\mathbb{V}(\beta_j)} = \frac{J-1}{J^2}\sum_{j=1}^{J}\hat{\beta}_j^2 - \frac{2}{J^2}\sum_{j=2}^{J}\sum_{k=1}^{j-1}\hat{\beta}_j\hat{\beta}_k$$

where

$$\widehat{\beta_j} = \frac{1}{n_{jf}}\sum_{i=1}^{n_{jf}} y_{ijf} - \frac{1}{n_{jm}}\sum_{i=1}^{n_{jm}} y_{ijm}$$

and

$$\widehat{\beta_j^2} = \widehat{\theta_{jf}^2} + \widehat{\theta_{jm}^2} - 2\widehat{\theta_{jf}\theta_{jm}}$$

where the unbiased estimator for each element is:

$$\widehat{\theta_{jg}^2} = \binom{n_{jg}}{2}\sum_{l=2}^{n_{jg}}\sum_{k=1}^{l-1} y_{ljg}y_{kjg}$$

for every group $g \in \{0,1\}$, and

$$\widehat{\theta_{jf}\theta_{jm}} = \frac{1}{n_{jf}n_{jm}}\sum_{l=1}^{n_{jf}}\sum_{k=1}^{n_{jm}} y_{ljf}y_{kjm}$$

With a set of weights w_i, the variance component of the bias is: