Are Patent Examiners Gender Neutral?*

Hadar Avivi†

December 27, 2024

Abstract

This paper studies the prevalence and evolution of gender bias in the United States Patent and Trademark Office (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, 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 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 approved patents by $12.6 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, Tim Christensen, Nick Flamang, Ingrid Haegele, Hilary Hoynes, 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, GSB Causal Panel Data Conference, MLESC24, and the Conference on Frontiers in Machine Learning and Economics. 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.

†Princeton University. Email: avivihadar@gmail.com

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, accounting for only 12% of the inventors in 2016 (Lissoni et al., 2018). While various explanations have been offered for this persistent gender gap, including differences in occupation choice (Hunt et al., 2013), rejection aversion (Aneja et al., 2024), and lack of exposure to innovation (Bell et al., 2019), there is a concern that gender bias causes patent examiners to over- or under-value contributions based on inventors' gender. 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, 2017; Sampat and Williams, 2019). Bias in patenting raises not only equity but also efficiency concerns, as misallocation of patent rights can have lasting economic effects, influencing growth (Jones, 2011; Bloom et al., 2013; Akcigit et al., 2017; Acemoglu et al., 2018), inventors' careers (Toivanen and Vaananen, 2012; Kline et al., 2019), and firm profits (Hall et al., 2005; Galasso and Schankerman, 2015; Kogan et al., 2017).

Previous attempts to estimate the importance of gender bias in the patent system suffer from several fundamental concerns. 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 Aneja et al. (2024) 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. 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 the gender composition of the inventors and the unobserved patent quality.

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 gender composition of the inventors, 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.

[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; USPTO, 1888).

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 examiner decisions, citations, and the stock market return of patent assignees, suggesting the text embeddings account for the patent quality. Moreover, I provide evidence that the assignment of applications to examiners is as good as random conditional on the text embeddings by showing that conditional on the text embeddings, the gender mix of the inventors and other non-text characteristics are *not* systematically correlated with examiner characteristics.

My target parameter – the gender gap in allowance conditional on the text – captures two distinct but related forms of disparities: disparate treatment and disparate impact. Disparate treatment arises when the decisions of the examiners are directly a function of the gender composition of the team of inventors, such as in models of taste-based discrimination (Becker, 1957), inaccurate stereotypes (Bordalo et al., 2016), or statistical discrimination (Aigner and Cain, 1977). Disparate impact results from decisions that are based on other non-text characteristics that are not the gender of the inventors but are correlated with it. This type of disparity was 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 the 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, after controlling for the text, the average allowance gap entirely disappears. This result is robust to different gender definitions, including thousands of art units,[2] year, and class fixed effects, and alternative matching estimates. Second, I document the evolution of the gender gaps in initial allowance from 2001 to 2013. 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 has decreased over time, and since 2005, it has converged to zero. This finding mirrors the recent evidence from academia (Card et al., 2021, 2022) and labor markets (Schaerer et al., 2023). Lastly, I show that, although the mean allowance gap is zero, there is substantial variation in gender gaps 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 in initial allowance is 2.3 percentage points, more than 25 percent of the mean initial allowance rate for no-female teams. Similar two-sided gender discrimination has also been found in audit experiments in the labor market (Arceo-Gomez and Campos-Vazquez, 2014; Kline and Walters, 2021; Kline

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

et al., 2022).

Examiner discretion plays a significant role in allowance decisions even beyond the gender of the inventors. After controlling for the patent text embeddings, I document substantial variation in examiner leniency, with a standard deviation of 110% of the average initial allowance rate. While previous literature has emphasized the importance of discretion in the examination process (Lemley and Sampat, 2012; Frakes and Wasserman, 2017; Sampat and Williams, 2019; Farre-Mensa et al., 2020), to my knowledge, this is the first estimate of examiner leniency that accounts for sampling error and conditions on the content of the patent.

Examiners' cohort explain 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. Time effects do not explain this result, ruling out the possibility that examiners changed their behavior over time. The art unit of examiners explains approximately 25% of the variation in gender bias, suggesting a wide across-field variation.

The importance of discretion and bias varies by examiner characteristics and over time. The standard deviation of the gender bias of male examiners is twice the size of the standard deviation of female examiners, suggesting that male examiners pose a higher risk in the system. Similarly, younger cohorts of examiners who joined the USPTO in later years have lower levels of variability in discretion and bias. Finally, combining these findings with the evidence that the mean preferences of younger examiners differ from those of senior examiners, I conclude that, over time, the risk of encountering an abnormally biased examiner has increased. This result suggests an increase in polarization within the community of examiners.

Exploring the legal ground behind initial rejections reveals that although most rejections are based on a lack of novelty or obviousness, the gender gaps in initial allowance are 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 a timely prior-art search (Frakes and Wasserman, 2017).

Much of the discrimination literature focuses primarily on average gaps, which only partially map to fairness and inefficiencies. First, even though there is no ex-ante bias on average, 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 returns of publicly traded firms. Since the stock market return is observed only for granted patents, I estimate a selection model (Heckman, 1979), exploiting the examiners as instruments, conditional on the patent content. My analysis estimates the annual cost of having a positive variance in gender bias among initially allowed patent applications assigned to publicly traded firms to be approximately \$1.67 million. Extrapolating this cost to granted applications, the results suggest that gender bias generates a loss of 12.6 million dollars per year. Finally, extending this cost to any form of examiner-level discretion reveals a cost of 76.6 million dollars per year, which is 31% of the value of the median public US firm in 2013.

This paper contributes to several strands of the literature. First, this paper is the first systematic 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); Hochberg et al. (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; Sarsons et al., 2021), and aligns with literature in economics studying discrimination and how it varies across gate-keepers (Abrams et al., 2012; Arnold et al., 2018; Dobbie et al., 2022; Feigenberg and Miller, 2022; Kline et al., 2022). Beyond documenting gender bias, this paper is among the studies that evaluate the economic effects of discrimination (Becker, 1957; Glover et al., 2017), and the consequences of inefficiencies in the innovation process (Bryan and Williams, 2021; Clemens and Rogers, 2023; Matcham and Schankerman, 2023).

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).[3] This paper provides a concrete and easy-to-implement method to account for the patent application content exploiting modern NLP methods. I demonstrate that the text embeddings can balance observables of patent applications across inventing teams with different gender mixes, suggesting that these comparisons are unconfounded. Therefore, the approach taken in this paper could serve as a useful framework to study other questions in both the patenting domain and in other settings where evaluations are mandated to depend on textual content.

The rest of the paper is organized as follows. Section 2 provides the institutional background, and Section 3 describes how I use the text data for analysis, Section 5 describes the patent data and the text embedding features, and Section 4 outlines the conceptual

[3]Examples for matching on text in social science include: Roberts et al. (2020) who model the text with latent Dirichlet allocation (LDA) model, Mozer et al. (2020) who match the text using distance metrics on the bag-of-words representation, and Zeng et al. (2022) which use bag-of-words methods to extract covariates from medical records to study the effect of different treatments on cancer. Keith et al. (2020) provides a review of using textual data to adjust for confounding from the computer science perspective.

framework and identification assumptions for the overall gender bias estimate. Section 6 reports the average gender bias overall and by years and examiners' years of experience. Section 7 discussed the identification assumptions for the examiner-level gender bias. Section 7.2 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

2.1 Patent Examination Process

A patent application is a form that describes 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 have been paid. The application claims are then classified and forwarded to the relevant USPTO technology center and art unit for examination.[4]

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, 2017; Sampat and Williams, 2019). However, other evidence also indicates within art unit specialization by class and subclass (Righi and Simcoe, 2019). Irrespective of whether the assignment is random within art units, all existing evidence suggests that it is the content of the application that plays a central role in how applications are assigned to art units and, in some cases, examiners.

The assigned examiner then evaluates the application content for compliance with law and regulation. She ensures that the patent claims include only a single invention,[5] that the claims clearly define the invention, and that the description adequately describes the invention. The examiner also conducts a prior art search by looking for related previous patents or non-patent literature to determine whether the claimed invention is novel and not obvious. 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

[4]Technology Centers are groups of examination units divided by a broad technology. Each examination unit is called an art unit.

[5]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.

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 of them is being objected to and/or rejected.

Upon receiving a Non-Final Rejection, the applicant is generally given three months to respond. The applicant's response is some combination of arguments and amendments regarding the claims, usually narrowing their scope. The applicant may also request a telephone or in person interview with the examiner.[6]

2.2 The Role of the Patent Application Text

The patent application text has a central role in the application process. Based on the text, patents are classified into classes and subject matter, which determine the specific technology centers, art units, and examiners to which they are assigned. Upon assignment, the Patent Act defines that examiners' assessments should be based primarily on the information in the patent applications.

An examiner can deny an invention on several grounds. One is lack of novelty and obviousness, which requires the examiner to compare the claimed invention with prior art. Other grounds for denial are missing statutory subject matter, non-usefulness of the proposed invention, inadequate writing, and failure of the application to satisfy disclosure requirements, all assessed based on the patent text and its content. All other non-patent-text information, such as existing prior art, citations from the non-patent literature, and personal knowledge, are allowed to be used only in conjunction with the patent text.

Although each patent application is required to disclose the inventors in the applications, the names of the inventors are not expected to be used to assess patentability. At the least, names cannot serve as reasonable grounds for rejection. Moreover, the role of uncertainty is also limited in the patent application process. Unlike academic papers, whose content could change dramatically during the review process, the invention is not allowed to be changed after a patent application has been submitted. 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. During an application process, amendments are allowed only to the claims by restricting their scope or adding to prior art citations. Examiners, of course, can still face uncertainty or have varying skills in identifying high-quality patents. However, any behavior that generates gender disparities conditional on the patent content would indicate a deviation from the stated mandate.

[6] 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.

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 match patent applications based on their 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, which inspired decades of research on numerical text representation. This paper adopts the state-of-the-art NLP approach of representing text as *text embeddings*—dense real-valued vectors of a finite dimension dwelling in some predefined vector space. Each dimension of the vector captures a specific feature of the text, and each text is represented by a point in that vector space such that words with the same meaning are closer to each other.

This approach to text representation builds on the *Distributional Hypothesis* from the field of distributional semantics in linguistics, an idea first popularized by Firth (1957). It stems from the semantic theory of language usage, which posits that lexical items (words, sentences, or paragraphs) used in similar contexts tend to convey similar meanings. This representation has achieved unprecedented success across various Artificial Intelligence (AI) tasks and is regarded as one of the key breakthroughs in recent NLP research.

In this section, I describe how I use the neural network language model BERT (Devlin et al., 2018) to convert the patent application text into embeddings, producing p continuous covariates $x_{i1}, ..., x_{ip}$, where p is of moderate size.

3.1 Contextualized Word Embeddings Using BERT Model

The Bidirectional Encoder Representations from Transformers (BERT) model (Devlin et al., 2018) is a deep neural network trained to perform two unsupervised tasks simultaneously. The first task involves predicting a randomly selected percentage of words in a paragraph, while the second task determines whether one sentence follows another in a given pair. This section briefly outlines the main structure of the BERT model and how it generates embeddings. For more detailed information, refer to the original paper (Devlin et al., 2018).

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 the model are the estimated parameters of each neural network layer. As such, there are several possible word embeddings representations. 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 5.2, 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 use 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 this language model was trained exclusively on patents, the estimated parameters represent the language structure, word distribution, and unique semantics of the specific domain of patents.

BERT embeddings, like those from other deep neural network NLP models, are renowned for their high quality text representation. In Section 5.2, I demonstrate that also in this context, the patent application text embeddings have high predictive power for the content of patents, allowance probability, and examiner decisions. Then I explain how we can be use them as covariates to answer causal questions.

4 Conceptual Framework - the Overall Gender Gap

Each patent application i filed in year t_i is characterized by its text T_i and by F_i, the gender composition of its inventor team. For ease of notation, I refer to F_i as an indicator equal to one when the team of inventors includes at least one female. In my analysis, I also explore other fractional versions of F_i, which do not affect the definitions described below. Since, after a non-final rejection, there are gender differences in persistence and the probability of resubmitting the patent application (Subramani et al., 2020), I restrict attention to the examination decision after the first round of the patent application process, denoted by IA_i.

Each patent application is assigned to one of $\mathcal{J}$ examiners, where the function $J : \{1, \ldots, n\} \rightarrow \{1, \ldots, \mathcal{J}\}$ indicates the examiner to which application i was assigned, and I define $Z_{ij} = \mathbb{1}\{J(i) = j\}$. 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.

The content of each patent application text is represented by a numerical vector of a finite dimension $C_i \equiv C(T_i) \in \mathbb{R}^p$. It describes the information encoded in the text that is relevant for examination decisions, where its vector representation allows it to vary on multiple, possibly large, number of dimensions.[7] As I describe in further detail below, my empirical analysis will assume that the patent content C_i is captured by the embedding representation generated by a BERT language model.

Following the instructions in the Patent Act, allowance decisions should be based on the patent content C_i. In practice, allowance decisions might also be based on other non-text characteristics, including the patent gender F_i. Thus, I model the decision rule of each examiner j as a function $IA_j(c, f, u)$ of the patent content $C_i = c$, the gender of the inventors $F_i = f$, and all the other non-content non-gender characteristics that affect the final decision, $U_{ij} = u$. This representation nests the model in which examiners extract the content of the patent, C_i, from the text, T_i, imperfectly and might vary by their skill to do so, where the errors are part of the U_{ij} term.[8]

With the initial allowance decision rule, the potential first-round examination outcome of patent application i from assignment to examiner j is $IA_{ij} \equiv IA_j(C_i, F_i, U_{ij})$. Consequently, the following function describes the system-wide initial allowance decision rule at the *overall*

[7] The existence of this representation is implied by the literature on distributional semantics described in Section 3.

[8] For example, rational examiners who infer the content of the patent from the text, a noisy signal, learn only a fraction $\lambda \in [0, 1]$ of the content: $\tilde{C}(T_i) = \lambda C(T_i)$ where λ is the signal to noise ratio. In that case the error, $C(T_i) - \tilde{C}(T_i)$ will be represented in the error term, U_{ij}.

USPTO level:

$$\tilde{IA}(C_i, F_i, \tilde{U}_i) \equiv \tilde{IA}(C_i, F_i, \underbrace{U_{iJ(i)}, Z_{iJ(i)}}_{\tilde{U}_i}) = \sum_i Z_{iJ(i)} IA_{J(i)}(C_i, F_i, U_{iJ(i)}),$$

where $\tilde{U}_i$ encompasses both examiner level non-content related tastes and shocks and the consequences of the assignment process on eventual allowance decision. Therefore, the observed initial allowance indicator of application i is $IA_i = \tilde{IA}(C_i, F_i, \tilde{U}_i)$ which is also the potential initial allowance of application i at the assigned examiner $IA_{iJ(i)}$.

The first parameter of interest is the content-adjusted overall gender gaps, which I define as:

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

where integrals are taken over the distribution of the patent application content $G(c)$ in the overall population. The function $\omega(c)$ gives the weights defining the estimand of interest. For example, if $\omega(c) = 1$, then β is the Average Treatment Effect (ATE). If $\mathbb{E}[IA_i|F_i = 0, C_i = c] - \mathbb{E}[IA_i|F_i = 1, C_i = c]$ does not vary with the patent content $C_i = c$, then β is the same regardless of the chosen weights. The parameter β describes the average USPTO-wide gender gap conditional on the patent application content.

The interest in the content-adjusted gender gap, β is motivated by the legal requirement that patentability decisions should be based primarily on the content and meaning of the patent invention. Any systematic variation in allowance among patents with identical content but different genders of inventors represents deviations from the stated mandate of the Patent Act. Such deviations could originate from various sources, which I discuss below.

4.1 Sources of Gender Gaps in Initial Allowance

This section outlines the potential sources for gender gap ($\beta \neq 0$) by modeling the reduced form initial allowance decision in terms of examiners' preferences and beliefs. To rationalize the behavior of each examiner, suppose that each patent application is characterized by a unique latent quality, denoted by $q_i \in \mathcal{Q} \subseteq \mathbb{R}$. The utility of each examiner, $v_j(d, q, f, \epsilon)$, depends on her allowance decision $d \in \{0, 1\}$, the patent quality q, and potentially the gender indicator $f \in \{0, 1\}$, and other features ϵ of either the patent, or the examiner, or both.[9] As elaborated next, these other factors could represent other potential biases of examiners,

[9]Examiner j's utility may vary over time, as patent content can hold different value depending on its novelty. I omit the subscript t from the model for simplicity but account for time-varying content effects in my analysis.

such as ones based on the race or ethnic makeup of the inventor team (Coluccia et al., 2023), examiner characteristics, limitations, such as time constraints (Frakes and Wasserman, 2017), and unexpected shocks.

Since patent quality is unobserved, the examiner forms beliefs about its distribution based on the observed signals. These signals include the patent application text, from which examiners extract its content, $C_i = c$, and additional non-text characteristics, including the gender of the inventors $F_i = f$, and other factors, $U_{ij} = u$, that influence the allowance decision.[10]

Conditional on these signals, examiners hold beliefs about the conditional distribution of quality given $F_i = f$ and $U_{ij} = u$, represented by $\tilde{\mathcal{F}}_{f,u}(q)$. Assuming rational behavior, each examiner forms decisions based on the respective posterior distribution $\tilde{\mathcal{F}}_{f,u}(q|c)$ of patent quality after observing the application content c, alongside the other characteristics f and u. The initial allowance decision is then chosen to maximize the examiner's expected utility, integrated over this posterior distribution:

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

Note that this representation allows the beliefs of the examiner regarding the distribution of quality to diverge from the actual distribution of quality $\mathcal{F}_{f,u}(q)$.

Disparate treatment: A decision rule that directly depends on the gender of an inventor, be it due to preferences or beliefs, constitutes a form of disparate treatment. Such a decision rule implies that there exists $c \in \mathbb{R}^p$, and $u \in \mathbb{R}$, for which $IA_j(c, 1, u) \neq IA_j(c, 0, u)$. The canonical disparate treatment model is Becker (1957)'s taste-based discrimination. In this model, the utility function of examiners $v_j(d, q, f, u)$ varies directly with the gender of the inventors. Another form of bias is statistical discrimination, emerging when the beliefs of the examiner are shaped by the actual distribution of quality $\mathcal{F}_{f,u}(q)$, and when this distribution varies with the gender of the inventors (Aigner and Cain, 1977). Finally, bias may also arise from inaccurate beliefs regarding the prior distribution of quality that vary systematically with the gender of the inventor team $\tilde{\mathcal{F}}_{f,u}(q) \neq \mathcal{F}_{f,u}(q)$ (Bordalo et al., 2016).

[10] Both U and ϵ are observed by the examiner during the evaluation of the patens. ϵ represents all the non-patent-quality characteristics that affect examiners' decisions beyond the gender mix of the inventors, and U represents all the non-patent-text characteristics that affect examiners' decisions beyond the gender of the inventors. Both include, for example, race discrimination, institutional constraints, or temporal shocks. However, tastes related to the patent text, topics, and styles would not be considered in U, although they are part of what ϵ captures.

Disparate impact: Gender gaps could arise additionally due to a relationship between the other factors (U_{ij}) that affect the allowance decision of examiners that are correlated with the gender of the inventors. For example, if examiners directly discriminate based on ethnicity, and ethnicity is associated with the gender mix of the inventors, this would result in a non-zero value of β.[11] Gender gaps at the system-wide USPTO level could also result from a systematic assignment of patents of a certain gender to less lenient examiners (correlation with Z_{ij}).

In Section 7, I provide supporting evidence that the assignment of applications to examiners process does not drive gender gaps conditional on the patent application content. Although Section 9.2 shows that the paper's findings are robust to controlling a few other non-textual characteristics like ethnicity and country of origin, it is important to note that, at the examiner level, it is not feasible to distinctly separate the gaps stemming from disparate impact and disparate treatment since the gender of inventors is not randomly assigned. Nevertheless, non-zero estimates of content-adjusted gender gaps are consistent with behavior that violates the stated examination rules since they are based on nontext characteristics.

Another example of disparate impact in my setting is disparities arising from examiners' preferences for different writing styles, as explored in Levitskaya et al. (2022). Since the mapping from text to content is not one-to-one—identical inventions can be described with different word combinations—disparities may emerge from such stylistic preferences. Assessing the extent to which BERT text embeddings capture these factors is an interesting direction for future research.

4.2 Identification

Since we cannot directly control for the patent text, I measure differences in allowance probability conditional on the BERT text embedding vector, denoted by $X_i \equiv X_i(T) \in \mathbb{R}^k$. This section discusses the assumption regarding the BERT embeddings vector that is required for the identification of β.[12]

Assumption A1. (BERT embeddings). The BERT embedding vector, X_i, is sufficient for

[11] Disparate impacts have been recently studied in the context of the US judicial system by Arnold et al. (2022) and conceptualized in Bohren et al. (2022).

[12] Throughout the analysis, I assume that the BERT embeddings are fixed and measured with no error. The extent to which measurement error in the representation of unstructured data biases the results is an area of active research (Sellam et al., 2021; Battaglia et al., 2024).

the patent content C_i.

$$E[Y_i|C_i, X_i] = E[Y_i|X_i]$$

for any $Y_i \in \{IA_i, F_i\}$

Assumption A1 requires the BERT embeddings X_i to sufficiently cover all the information in C_i. While C_i can be thought of as the minimal text representation that encompasses the patent content and that is essential for the decisions examiners make, this assumption requires that the least, the BERT embeddings representation is a finer representation of the text than the content C_i. This is analogous to controlling for a fully saturated set of controls in a stratified experiment rather than the coarsest balancing score, the propensity score. Notably, this assumption does not require that the BERT embeddings capture all the possible information in the text. It requires it only to represent all the essential elements of the examination process, such as what it claims, how it functions, and what is new about it. Moreover, since examiners are expected to determine whether the description and the claims are written clearly, the embedding vector should also capture the language and writing clarity. While I cannot test this assumption directly, I provide evidence that the patent embeddings are strongly predictive of examiners' decisions, patent quality proxies, and the gender mix of the parent inventors.

Under Assumption A1, the overall gender bias (1) is non-parametrically identified. To estimate this parameter from finite data, 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 6 that the results are robust to other matching techniques.[13]

5 Data

The primary 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) (Miller, 2020).

[13]The OLS coefficent on a treatment indicator from a model with controls is a weighted average of conditional average treatment effects. Angrist (1995) demonstrated it first in a model with a single binary control, and Angrist and Krueger (1999) and Goldsmith-Pinkham et al. (2022) extend this result to a general set of control variables.

I restrict the sample to utility[14] patent applications filed after November 29th, 2000[15] 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. To avoid detecting differential behavior to non-US inventors, I include only patent applications submitted by US inventors.

I merge this dataset with several other datasets: 1) The USPTO Patents View data, which includes detailed information on both granted patents and patent applications. 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 pull the abstract and description text of each patent application; 4) Examiners' roster, pay scale, and education levels from Frakes and Wasserman (2017) 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 see Appendix Section B.

Applicants and Examiner Characteristics: I assume examiners infer inventors' gender from their names. Therefore, I classify the gender of each inventor based on distribution of the first name by gender from U.S. Social Security Administration (SSA) data, under the assumption that U.S.-based examiners recognize common U.S. names. This method classifies 75% of names in my sample, with unlisted SSA names treated as foreign and of unknown gender.

To infer examiners' true gender, I take an improved approach that maximizes name coverage, including foreign-sounding names, as my ultimate goal is to infer their true, rather than implied, gender. In addition to U.S. SSA data, I incorporate gender information from the UK Intellectual Property Office, the World Intellectual Property Organization (WIPO), the gender-guesser Python package, and Genderize.io. This comprehensive approach identifies the gender of 85% of examiners' names.

I measure each examiner's years of experience and education level based on the FOIA roster tables provided by Frakes and Wasserman (2017), which date back to 1992 and

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

[15] 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.

end in 2012. For examiners who joined before 1992 or after 2012, I supplement this data with information from the first office action and validate the approach using administrative records. See Appendix Section B for more details.

5.1 Descriptive Statistics

Table 1 provides summary statistics of the patent applications satisfying the sample restrictions described above. The sample includes over 1.2 million patent applications; only 16% of them include at least one female inventor, and only 11% have a female inventor as the first or second author in the list inventors.[16] The average inventor team has 2.5 inventors, whereas teams with at least one female inventor are larger. 35% of the patent applications are of sole inventors, most of whom are male or unknown.

Table 1: Descriptive statistics of patent applications

	Full sample (1)	All male or unk (2)	At least 1 female (3)	Female ranked 1st or 2nd (4)
# of observations	1220831	1034932	185899	128609
Team size	2.478	2.277	3.595	2.952
Proportion female	0.068	0.000	0.444	0.532
Sole inventor	0.351	0.391	0.129	0.187
Sole female inventor	0.020	0.000	0.129	0.187
Initial allowance (IA)	0.083	0.086	0.066	0.063
Ever granted	0.648	0.658	0.593	0.567

Note: This table presents the descriptive statistics of the 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.

An initial allowance of a patent application is a rare event: only 8.3% of the patent applications are allowed in the first round of examination, while 64.6% of the patent applications are eventually granted. Additionally, the raw gender gap in initial allowance and granting is substantial, as seen in the last two rows of Table 1. Patent applications with at least one female inventor are almost two percentage points less likely to find their patent initially allowed (23% less compared to all male or unknown patent applications) and 6.1 percentage points less likely to eventually have their patents granted (10% less compared to all male of unknown grant rate).

[16] The rank of each inventor in the list of inventors doesn't necessarily indicate the contribution to the invention. I use this measure to proxy the visibility of females in the team of inventors.

Table 2: Examiners' descriptive statistics

		By examiner gender		
	All (1)	Female (2)	Male (3)	Unknwon (4)
# of examiners	8519	2045	5107	1367
By start-year				
< 1995	0.203	0.201	0.211	0.176
1996-2001	0.314	0.330	0.304	0.326
> 2001	0.483	0.469	0.485	0.498
Initial allowance rates				
Mean	0.067	0.053	0.072	0.068
10th percentile	0.000	0.000	0.000	0.000
90th percentile	0.176	0.142	0.190	0.175

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

Table 2 reports the descriptive statistics of patent examiners. There are 8,519 examiners in the sample: 2,045 were classified as female, 5,107 as male, and 1,367 could not be classified by gender based on their name. Twenty percent of the examiners joined the USPTO before 1995, 31% joined between 1996 and 2001, and the rest joined after 2001, with no differential trend by examiner gender. Lastly, examiners vary substantially in their propensity to initially allow patents, a phenomenon I measure formally in Section 8.1. The mean examiner-level initial allowance rate is 6.7 percentage points, with male examiners having a 1.9 percentage-point higher allowance probability. Additionally, there is substantial heterogeneity even within gender, with examiners at the 90th percentile being three times more likely to initially allow patents than the mean.

5.2 Selection and Visualization of the Patent Text Embeddings

The BERT model has several layers, each potentially serving as text embeddings. To choose the list of embeddings, I adhere to the following protocol: For every part of the patent application text, i.e., description and claims, I generate a unique embedding vector using the [CLS] token. Since BERT inputs are limited to 512 tokens, I split the text into paragraphs and compute the [CLS] embeddings for each, averaging them to represent the entire text. 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 yields (3×2) six possible embedding vectors, each with a total of 1,023 features.[17]

The preferred set of embeddings is selected to achieve the largest improvement in adjusted R^2 when predicting Initial Allowance (IA) and any female (F) in linear regression, subject to my machine memory constraint.[18] These adjusted R^2's are presented in subfigure 1a,[19] where darker circles indicate that more embedding vectors were included. To compare with previous literature, the "X" point displays the adjusted R^2 values obtained by regressing IA and any female on the set of art-unit-year and class fixed-effects, which are typically used to control for confounding (e.g., Jensen et al., 2018; Choi et al., 2019).

Subfigure 1a shows that all the possible combinations of embedding vectors dominate the art-unit-year and class fixed effects in their predictive power of mixed-gender teams. In addition, the set of embedding vectors that include at least two vectors is more likely to dominate the fixed effects in predicting IA. This result is striking since the fixed-effects control for more than 8,000 covariates, while the sets of two embeddings vectors comprised only 2,046 covariates. Additionally, it seems that the improvement in the predictive power is limited 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, Subfigures 1b and 1c presents the partial adjusted R^2 within art-unit-year and class fixed effects and examiner fixed effect.[20] These figures emphasize that the embedding vector introduces new information beyond the fine text classification to classes and sub-classes.

Describing embeddings: I conduct two exercises to describe the embedding vector. First, I apply the Uniform Manifold Approximation and Projection (UMAP) algorithm (McInnes et al., 2018) to reduce the vector's dimensionality and present the results in Figure 2. This figure presents the 2-dimensional UMAP reduction on a random sample of 50% of the patent applications, with the differing colors representing different technology centers.

This figure communicates two pieces of information. First, although the embeddings were not directly trained to predict theology centers and fields, they are nonetheless able to predict

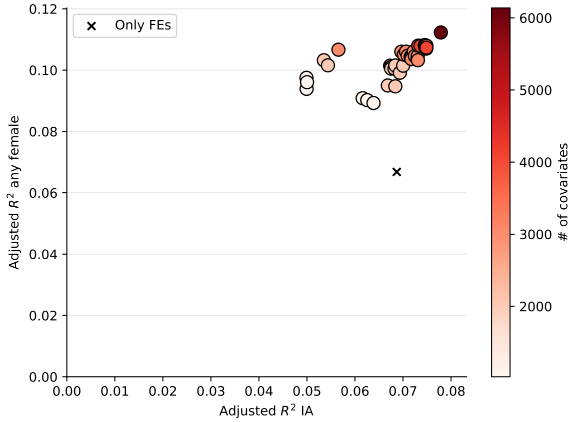
[17] Each original embedding vector has 1024 features mapped into a probability space and, therefore, sum to one. Hence, one of the features is excluded to avoid multicollinearity.

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

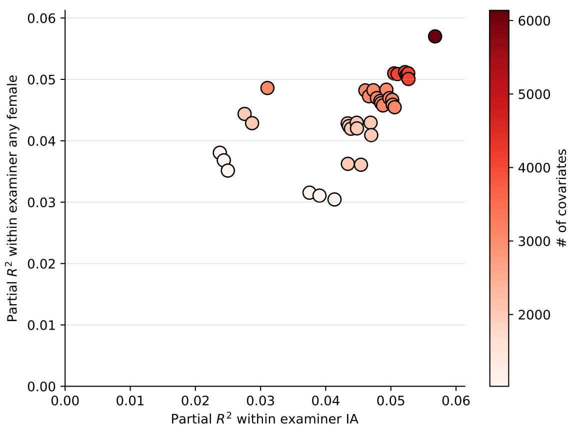
[19] For a detailed description of the adjusted R^2 values of each combination of embeddings vector see appendix Table A.1

[20] 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.

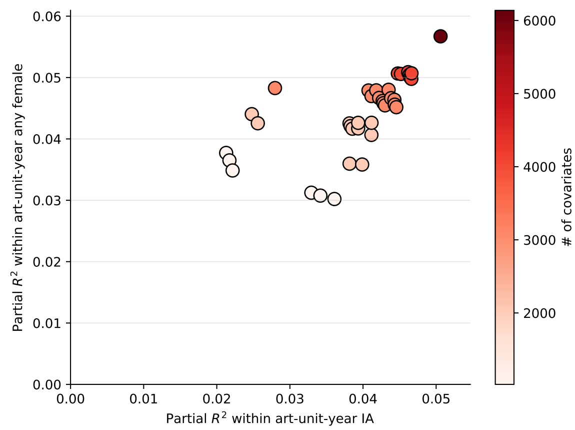
Figure 1: The predictive power of the BERT embeddings for initial allowance and gender



(a) Adj. R^2



(b) Within examiner

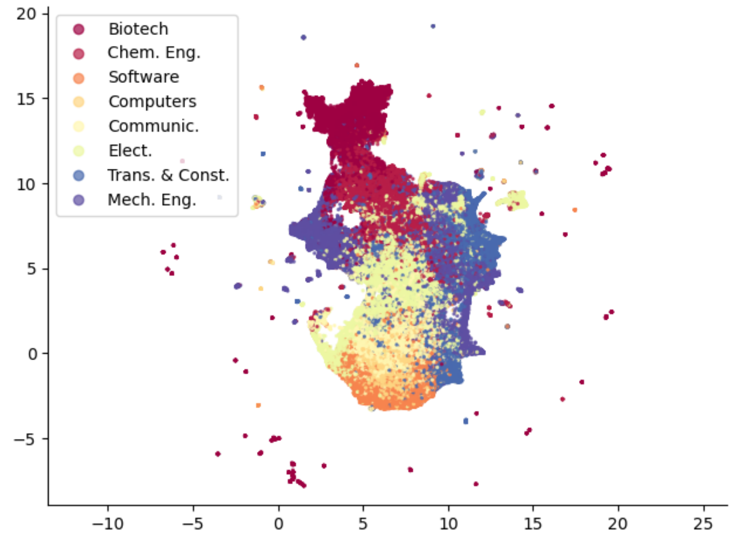


(c) Within art-unit-year & class

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. Subfigure (a) plots the adjusted R^2, and subfigures (b) and (c) plot the partial adjusted R^2 within examiners and art-unit-year and class fixed effects, accordingly. Different dots represent different combinations of the embedding layers of the patent claims and description. The exact point estimates behind this Figure are available in Appendix Table A.1. Darker dots represent a model with more embeddings, and the "X" symbol in subfigure (a) represents the adjusted R^2 for a model with art-unit-year and class fixed effects, which includes more than 8,000 fixed effects.

these dimensions. Second, as a continuous index, the embeddings provide richer information than discrete field indicators. For example, patents from mechanical engineering technology centers that are similar to those from the computer and communication technology centers are represented as points that are closer to each other on the graph. In contrast, mechanical engineering patents that are more similar to patents from the chemical engineering technology center would appear as points on the other side of the mechanical engineering clutter. Such

Figure 2: 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.

similarities provide a nuanced measure of content that mimicks the way we perceive ideas as a continuous domain.

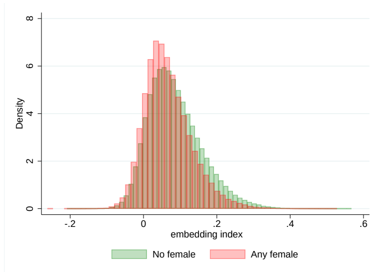
In the second exercise, I depict the distribution of the explanatory power of text embeddings in predicting initial allowance across gender groups by running the following regression:

$$IA_i = \alpha + \beta F_i + X_{it}'\gamma + \epsilon_i,$$

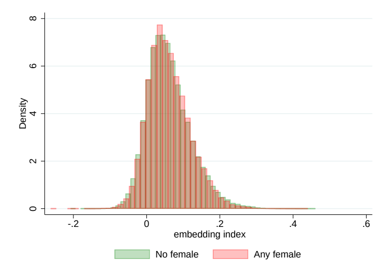
where IA_i is an indicator that equals one if patent application i was allowed at the first round of the examination process, F_i is an indicator for the presence of at least one woman in the team of inventors, and X_{it} is the embedding vector. Since the same content could have diffrent value in diffrent years X_{it} additioonally includes the first 100 UMAP componenets interacted with a linear time trend. One can think of the embeddings index, $X_{it}'\hat{\gamma}$, as measuring the patent application quality.

The distribution of the embedding index for mixed-gender applications, as shown in subfigure 3a, is skewed left, possibly suggesting that the proportion of low-quality mixed-gender patents exceeds that of all-male patents. However, subfigures 3b-3i, which display the distribution of the embedding index across technology centers, reveal substantial heterogeneity. For instance, in the biotechnology center, which has the highest proportion of mixed-gender

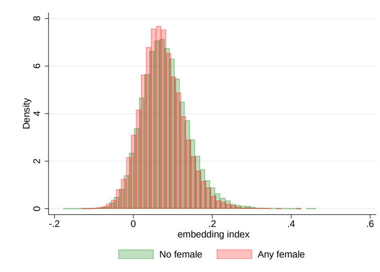
Figure 3: Histogram of embeddings index by gender and technology center



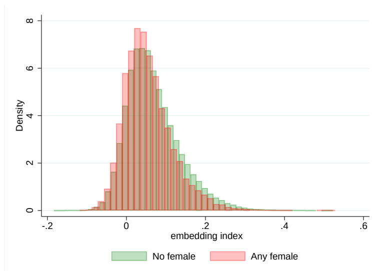
(a) All



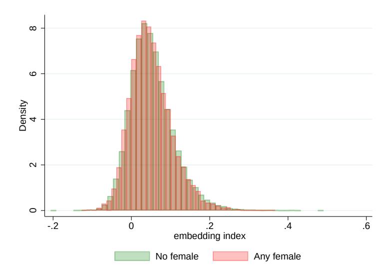
(b) Biotechnology (16)



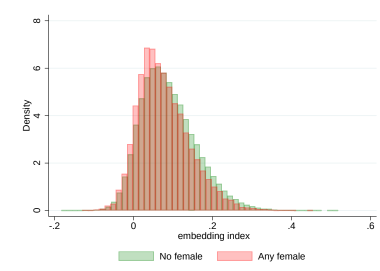
(c) Engineering (17)



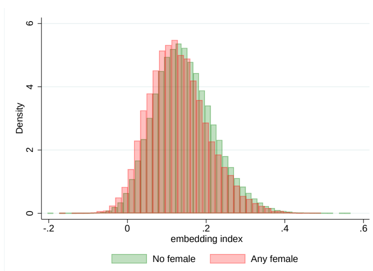
(d) Software (21)



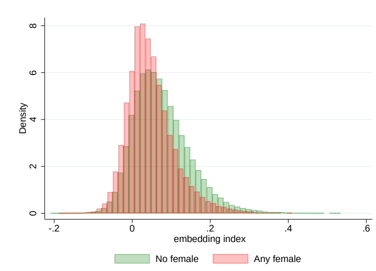
(e) Computers (24)



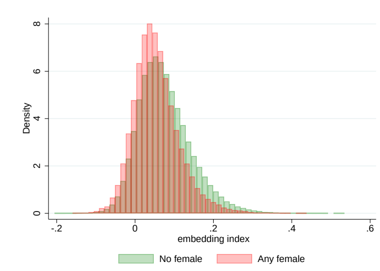
(f) Communications (26)



(g) Optical (28)



(h) Transportation (36)



(i) 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 $IA_i = \alpha + \beta F_i + X_{it}'\gamma + \epsilon_i$ where IA_i is an indicator for initial allowance, F_i is a mixed gender team indicator, and X_{it} are my preferred 2,046 text embeddings and a linear interaction of the first 100 UMAP componenets and filing year. The embedding index is the estimated linear combination $X_{it}'\hat{\gamma}$.

applications, the distributions for mixed-gender and all-male embedding indices are identical. This pattern could result from a Roy-type model where women choose technologies based on comparative advantage or from a model with spillovers where women's productivity improves with an increased share of women.

This pattern could alternatively be explained by bias related to the patent's content, which arises when examiners undervalue feminine topics (Bohren et al., 2018). As I elaborate in Section 4, In this paper, I focus on analyzing bias conditional on the patent application content. Studying biases in writing styles and topics could make an interesting avenue for future research beyond the scope of this paper.

5.3 Justifying Identification Assumptions

5.3.1 BERT Embeddings are Predictive of Patent Quality

To assess whether the embedding vector effectively represents the content and quality of the text, Appendix Figure A.2 displays a split sample binned scatter plot of the OLS predictions of allowance decisions and the patent quality proxies. Subfigures (a)-(d) depict forecasts for rejection reasons defined by patent law: obviousness, novelty, writing, and eligibility violations. The first two require understanding the invention's uniqueness and contribution, while the latter two are more technical and require following several guidelines the law defines. The plots reveal that even the simplest linear model strongly predicts examination decisions.

BERT embeddings also predict proxies for patent quality. Subfigures (e) and (f) apply the same method to two widely recognized quality metrics: log number of citations and log Kogan et al. (2017)'s stock market return. Similarly, the BERT embeddings are found to be strongly predictive of these variables, where the figure shows that the linear model provides a good approximation of the relationship between embeddings and the outcomes.

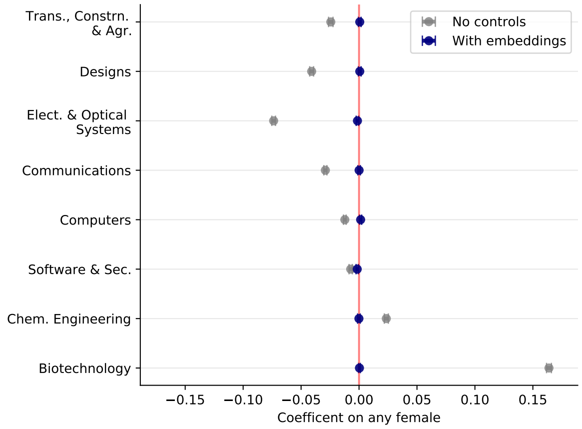
5.3.2 Balance Tests

Assumption A1 requires that the BERT embeddings adequately encode the content of the patent application. This assumption implies that, conditional on embeddings, gender gaps in allowance should not be driven by any correlation between the gender of the inventors and the content of the patent, including its field and class. To test this assumption, I run an OLS regression of proxies of the patent content, such as technology center indicators and patent claims statistics, on various variables that describe the gender of the inventors.

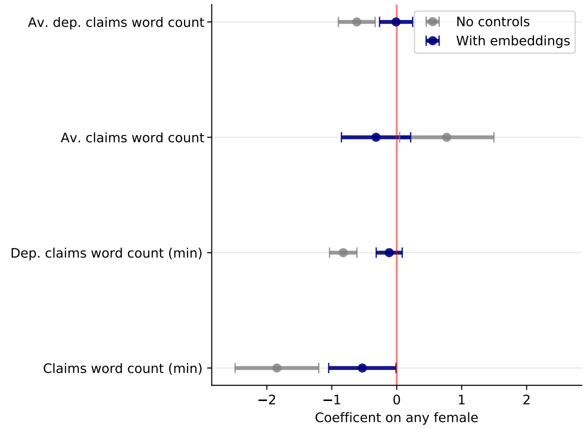
Figure 4a displays the relationship between technology centers and the presence of at least one female inventor in the team, both with and without controlling for text embeddings. The gray dots present the uncontrolled relationship between mixed-gender patents and each technology center, while the blue dots show the relationship after accounting for embeddings. Although mixed-gender applications are not uniformly distributed across technology centers, text embeddings effectively account for this variation. Appendix Figure A.7 displays the same analysis using other measures for the femaleness of the patent application.

Similarly, Subfigure 4b shows that BERT embeddings also balance the relationship between examiner gender claim statistics, such as the average and minimum number of words per

Figure 4: Balance test of technology centers and claim statistics



(a) Technology centers



(b) Claim statistics

Note: This Figure plots the relationship between technology center indicators (subfigure (a)), claim statistics (subfigure (b)), and the gender composition of the team of inventors with and without controlling for the patent application text embeddings. In subfigure (b) the claim statistics include the mean and min number of words in all the patents claims, and among the dependent claims. The gray dots display the estimates from an OLS regression with no controls, and the blue dots display the estimates from an OLS regression controlling linearly for the patent text embeddings and the interaction of the first 100 UMAP components and linear time-trend. Bars indicate 95% confidence intervals based on robust standard errors.

claim, across all claims and among independent claims.[21]. Previous research has shown that these statistics strongly indicate the patent scope (Lanjouw and Schankerman, 2001; Marco et al., 2019). Lastly, Appendix Figure A.8 tests whether the results are time-sensitive. This figure replicates the same balance tests separately by five-year time intervals. In line with previous findings, conditional on the BERT embeddings, the relationship between the gender composition of the team of inventors and text characteristics is balanced within years.

6 Overall Gender Gap in Initial Allowance

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

$$IA_i = \alpha + \beta F_i + x_i'\gamma + \epsilon_i \tag{2}$$

[21] 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.

where F_i is a variable characterizing the femaleness of the team of inventors, 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 5.2 and an interaction of the first 100 UMAP components with linear time tred,[22] 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 inventor, 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 patent list of inventors.

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

	IA (1)	IA (2)	IA (3)	IA (4)
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
AUC	0.5170	0.7619	0.7500	0.8065
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
AUC	0.5175	0.7619	0.7500	0.8065
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
AUC	0.5137	0.7619	0.7500	0.8065
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 the OLS coefficients, adjusted R^2, and Area Under the Curve (AUC) from regressions of an indicator for initial allowance on the gender of the inventors' team. The gender of the inventors is 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 application list of inventors (panel (iii)). Robust standard errors are reported in parentheses.

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 which

[22]To reduce the dimensionality of our model, we include time-trend interactions only with the first 100 UMAP components. Appendix Section XX shows that this restriction is robust to controlling for the interaction of the filing year and all the 2048 embeddings.

are around 25 percent of the mean no-female initial allowance rates. 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 gap to 0.001, making it statistically insignificant from zero (SE = 0.0007). Lastly, in a model that includes both fixed effects and embbedings, has no qualitatively diffrent gedenr gaps, suggesting for robustness to the inclusion of these controls.

These findings are robust to using alternative measures for the femalness of the patent. Panels (ii) and (iii) show that using proportion of females or whether a female author is ranked first or second as F_i provides similar qualitative results of zero bias when controlling the text embeddings. Appendix Table A.3 presents similar findings for the restricted sample which compares only female vs. no female patents teams and single author patents, and, similarly, Appendix Figure A.3 studies the nonlinear relationship between the number of female inventors and the share of inventors in the team and initial allowance probability.

OLS is one of many types of matching estimators. With treatment effect heterogeneity, it need not coincide with other reweighting schemes. Table 4 assesses the sensitivity of the OLS gender gap to other reweighting estimators. 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 estimator equals the Average Treatment on the Treated (ATT) if either the propensity score or the outcome equation is linear in embeddings. Column 1 confirms that the mean gender gap remains zero.[23]

The mean gender gap remins zero when estimating it by inverse probability weighting. Columns 2 and 3 report the Augmented Inverse probability Weighting (AIPW) estimate of the ATE and ATT, estimating the propensity score for F_i with linear regression, and 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. DML estimator will yield more precise estimates without compromising consistency under an additional assumption of sparsity of the BERT embeddings in the IA and gender equations. Table 4 suggests that the mean overall gender bias 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.

[23]In Appendix Figure A.4, we plot the coefficients of embeddings from that separate regressions of IA on embeddings for the sample of any-female patents against the coefficients from the sample of all-male or unknown patents. Accordingly, we find that after accounting for excess noise in the estimated coefficients, the two sets are strongly correlated ($\rho = 0.925$).

Table 4: Non-linear models for the mean gender gap in initial allowance

	Oaxaca-Blinder	AIPW		DML
	ATT	ATE	ATT	NN
	(1)	(2)	(3)	(4)
Any female	-0.0019	-0.0012	-0.0009	0.0006
	(0.0013)	(0.0009)	(0.0006)	(0.0007)

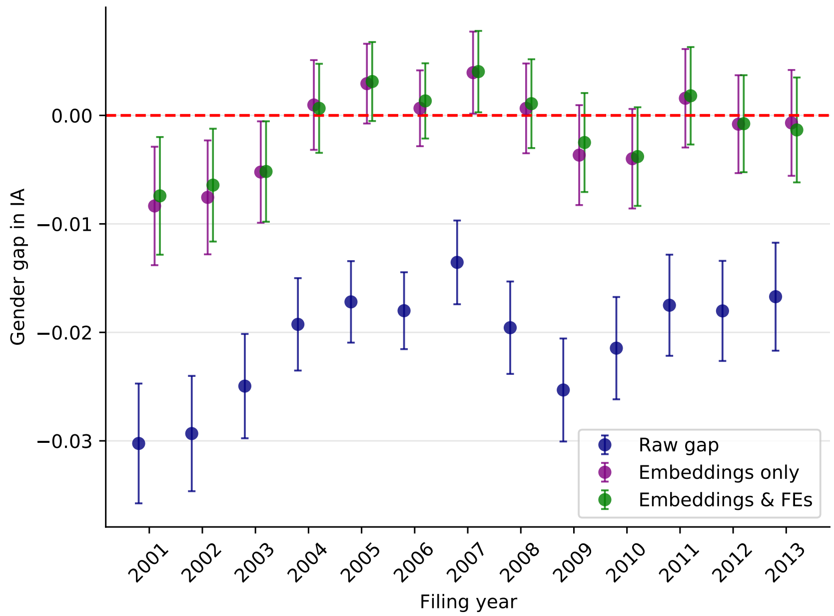
Note: This table reports the estimates of the overall gender gap in initial allowance between mixed-gender teams and no female teams conditional on BERT embeddings. Column 1 reports the fully interacted Oaxaca (1973); Blinder (1973) ATT estimate by running an OLS regression of initial allowance on embeddings and the interaction of the first 100 UMAP components and linear time-trend 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 Augmented Inverse Probability Weighting (AIPW) estimates of the ATE and the ATT, estimating the propensity score with a logit regression. 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.

Gender Gaps Over Time

Figure 5 plots the evolution of gender gap over time suggesting that the gender gap has changed over time. The blue estimates represent the uncontrolled gender gaps, documenting substantial raw gaps 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 content-adusted 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. Appendix Figure A.5 confirms that the above pattern is robust to estimating the gender gap separately by year.

These results align with recent findings. Similar time trends have 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 significant, but it shrunk to zero between 1980 and 2010 and has become positive in recent years. It also mirrors the trends in gender discrimination in hiring decisions estimated in audit experiments (see Schaerer et al., 2023, for meta analysis). A dynamic reversal of bias over time could reflect changes in the composition of examiners or shifts in their behavior (e.g., Bohren et al., 2019), a key question explored in the analysis that follows.

Figure 5: Gender gap in initial allowance by filing year



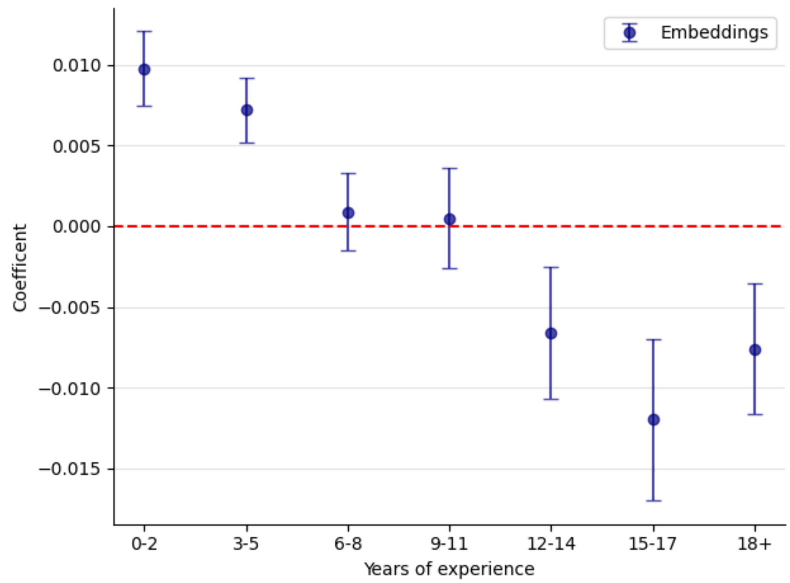
Note: This figure plots the estimated gender gap and 95% confidence intervals in initial allowance of patent applications filed between the years 2001 and 2013 by filing year. Blue dots plot the uncontrolled gender gap, purple dots plot the gender gap conditional on the BERT embeddings and the interaction of the first 100 UMAP components and linear time-trend. The 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.

Gender Gaps by Examiner Years of experience

Figure 6 investigates the variation in gender gaps across examiners' years of experience using my preferred model that controls for the text embeddings. Each point in the figure presents the estimated gender gap and confidence interval within bins of approximately three years experience. Although the average gender gap in the sample is statistically indistinguishable from zero, Figure 6 shows 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 0.7 to 1 percentage point higher probability of allowing patent applications with at least one female in the team of inventors. 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/experience effect, where examiners become more biased against women as they gain experience, or a cohort effect, where different cohorts have different tastes or gender stereotypes. It could also reflect a nonrandom assignment of patents to examiners. In the next sections, I formally test these

Figure 6: Gender gap by examiners' years of experience



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

hypotheses.

7 Variation Across Examiners

The previous section suggests substantial heterogeneity in gender gaps across examiners. Building on this, I turn to formally study this heterogeneity and its economic consequences. To do so, in this section, I begin by describing the new target parameters, identification assumptions, and statistical framework.

7.1 Parameter of Intrest and Identification

To study heterogeneity across examiners, my second parameter of interest is the examiner-level content-adjusted initial allowance gap between applications submitted by male and female teams:

$$\beta_j = \int \omega(c)(\mathbb{E}[IA_{ij}|F_i = 0, C_i = c] - \mathbb{E}[IA_{ij}|F_i = 1, C_i = c]) dG(c), \tag{3}$$

where, similar to Equation (1), integrals are taken over the distribution of the patent application content $G(c)$ in the overall population. β_j represents the tendency of examiner j's decision to vary with the gender of the inventors.

The assignment process of applications to examiners depicted in section 2 suggests that the patent content is the main instrument through which applications are assigned to examiners. Therefore, the second identification assumption exploits this phenomenon.

Assumption A2. (Conditional Independence)

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

This assumption requires that conditional on the patent application content and filing year, there is no systematic relationship between the assignment of applications to examiners and non-text features that affect decisions, including the gender of the inventors. It does not require random assignment of examiners conditional on the text. This assumption requires that after controlling for the content of the patent application, all the other aspects that influence the assignment process, have no meaningful effect on the decision process. This assumption relies on the nature of the assignment process depicted in Section 2, and in the next Section, I present several tests for its validity.

Combining assumptions A1 and A2, each β_j is non parametrically identified by:

$$\beta_j = \int \omega(x)(\mathbb{E}[IA_i|F_i = 0, X_i = x, Z_{ij} = 1] - \mathbb{E}[IA_i|F_i = 1, X_i = x, Z_{ij} = 1])dG(x),$$

where $G(x)$ is the empirical distribution of patent content X. Similar to the way I estimate the overall gender bias, my main analysis imposes further parametric assumptions and estimates each β_j using linear regression, controlling for the text embeddings in an additive separable way. To assess the sensitivity of my estimates to this functional form, I verify that my findings are robust to using alternative non-linear models.

7.2 Variance Componenet

Under Assumptions A1 and A2, and as long as the within examiner heterogeneity with respect to the patent text is restricted, one can estimate β_j by running the following fixed effect regression:

$$IA_i = \alpha_{J(i)} + \beta_{J(i)} F_i + X_{it}' \gamma + \epsilon_i, \tag{4}$$

where F_i measures the femaleness of patent application i's—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 gap, and X_{it} are the 2,046 text embeddings representing the patent application content together. To account for the changing value of content over time—what was once considered novel may no longer be viewed as original—X_{it} also includes the first 100 UMAP components interacted with a linear time trend.

I summarise the heterogeneity of examiner level leniency and gender gaps 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 explained by the start year of examiners and art units, I estimate Equation (4) and the α and β parameters across the start-year of examiners and art units, and their respective variance components.

I estimate the variance components by applying a variant of the Kline et al. (2020)'s leave-one-out bias correction that accommodates the fact that my design matrix includes a dense high dimensional segment of text-embeddings. For further details, see Appendix C.

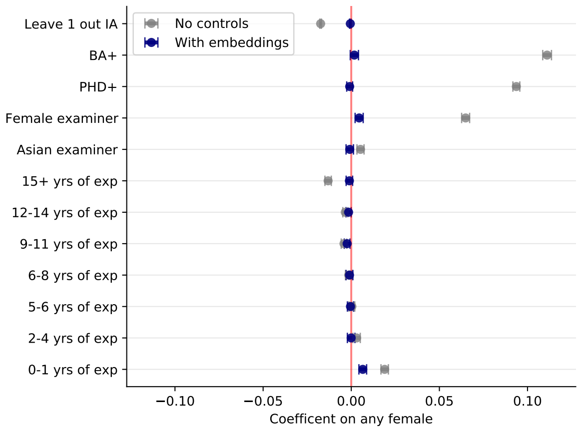
7.3 Evidence for Conditional Independence Assumption

7.3.1 Balance Test

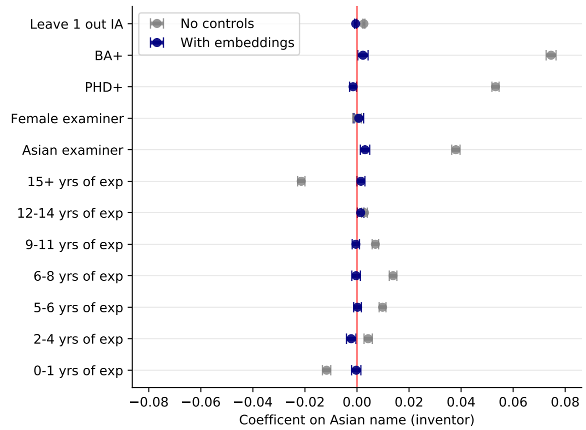
Assumption A2 implies that the assignment of examiners to applications (Z_{ij}) should not be systematically correlated with non-text characteristics (U_{ij}, F_i) conditional on the patent content. Figure 7 tests that assumption by plotting the relationship between various examiner characteristics and non-text characteristics after controlling for the BERT embeddings. The non-text characteristics I explore include the presence of at least one female inventor (Figure 7a), at least one inventor with an Asian sounding name (Figure 7b), and the percentage of females in the attorney team (Figure 7c).

The first examiner characteristic in each figure is the leave-one-out mean of the initial allowance rate, which is often used in the literature to test for random assignment of judges to cases (Arnold et al., 2018; Dobbie et al., 2018, 2022). The analysis also includes other examiner characteristics such as 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.6 verifies that the results are robust to other definitions of the femaleness of the patent, and Appendix Figure A.8 shows that they are not time-sensitive.

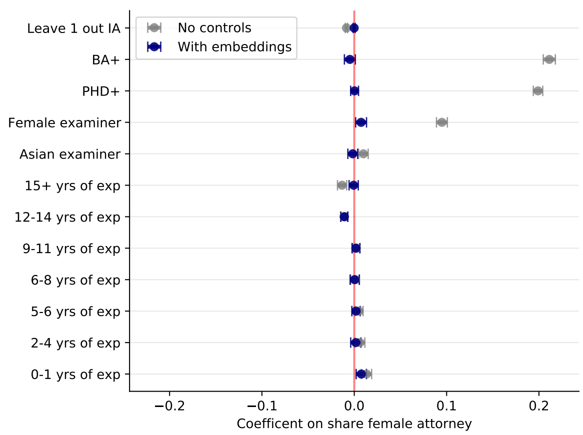
Figure 7: Balance test of examiner characteristics



(a) Mixed gender team



(b) Any Asian inventor



(c) Share female attorney

Note: This Figure plots the relationship between examiner characteristics and patent application non-text characteristics with and without controlling for the patent application BERT embeddings. The gray dots present the estimates from an OLS regression with no controls, and the blue dots present the estimates from an OLS regression that includes BERT embeddings controls and an interaction of the first 100 UMAP components and linear time-trend. Subfigure (a) displays the relationship between examiner characteristics and an indicator for at least one female inventor, subfigure (b) presents that relationship with an indicator for at least one Asian inventor, and subfigure (c) displays that relationship with the share of female attorneys. Bars indicate 95% confidence intervals based on robust standard errors.

7.3.2 Omitted Variable Bias - Long and Short Regressions

Under assumptions A1 and A2, and as long as the treatment effect heterogeneity with respect to the patent text embedding is limited, the estimates of examiner-level gender bias should be invariant to the inclusion of other non-text characteristics that are potentially correlated with examination outcome.[24] To assess this hypothesis, I run the following "short" regression:

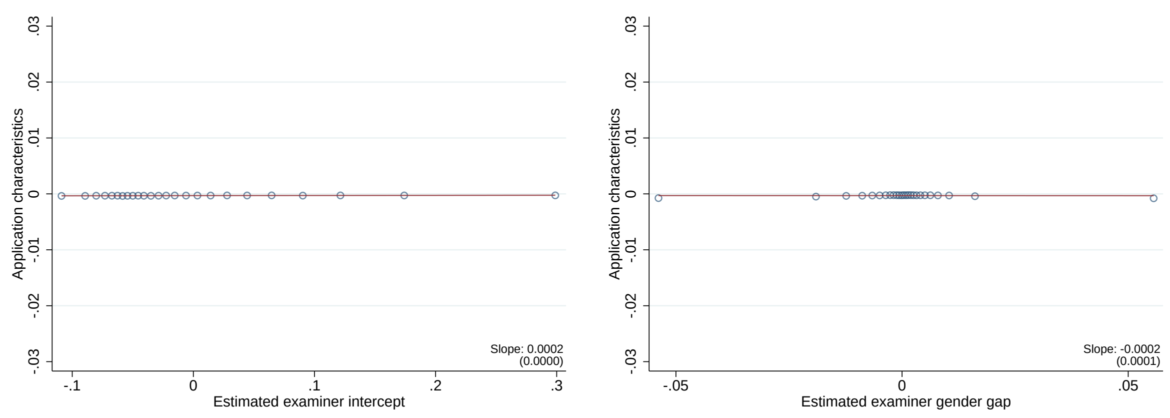
$$IA_i = \alpha_{J(i)} + \beta_{J(i)} F_i + X_i' \eta + \epsilon_i, \tag{5}$$

and "long" regression

$$IA_i = \tilde{\alpha}_{J(i)} + \tilde{\beta}_{J(i)} F_i + X_i' \tilde{\eta} + w_i \gamma + \epsilon_i, \tag{6}$$

where F_i is an indicator for a mixed-gender patent, and w_i includes patent attorney gender, ethnicity, and years of experience, patent inventor team size, and an indicator for foreign priority. α_j and $\tilde{\alpha}_j$ are examiner fixed effects measuring the leniency of examiner j, and β_j and $\tilde{\beta}_j$ are examiner level tendency to prefer mixed-gender patents. My objective is to assess whether the estimates of α_j and β_j from the short regression in Equation 5 are sensitive to the inclusion of the additional controls w_i.

Figure 8: 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 4) 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 slopes indicate that examiner effects are not affected by the inclusion of these covariates.

Figure 8 presents the main findings suggesting for little evidence for ommited variable

[24] 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 gender gap.

bias. 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 (6), 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 5, and Appendix Table A.2 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.[25]

8 Gender Gap Heterogeneity

8.1 Overall Variance Component Estimates

Table 5 reports the standard deviation, correlation, and mean estimates 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 team of inventors; in column (2), it is the proportion of females; and in column (3), F_i is an indicator for having a female ranked first or second in the list of patent inventors. Below, I describe the results in column (1), where the conclusions regarding the heterogeneity of the gender gaps are qualitatively the same across the different measures of F_i.

The standard deviation of α_j, presented in Panel (i), suggests substantial heterogeneity in examiners' leniency. While previous literature has emphasized the importance of discretion in the examination process (Lemley and Sampat, 2012; Frakes and Wasserman, 2017; Sampat and Williams, 2019; Farre-Mensa et al., 2020), to my knowledge, this is the first estimate of that leniency that acounts for the patent content and corrects for measurement error. The standard deviation of α_j is 9 percentage points, which is 110% of the mean initial allowance level reported in Table 1. It implies that an application assigned to an examiner with one standard deviation higher leniency compared to the mean would more than double its chances of allowance in the first round of the examination process.

The standard deviation of β_j, presented in the second row of Panel (i), implies that the allowance probabilities of examiners vary substantially by the gender of the inventors. The leave-out estimate of the standard deviation of the gender bias is 2.4 percentage points,

[25] 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 $0.023 \times 0.0002 = 0.0000046$.

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 bias-corrected standard deviation and correlations of examiners, examiner start-year, and art units leniency and gender bias, estimated in a linear regression controlling linearly for patent text embeddings. Different columns use 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 list of inventors. All variance components are weighted by the number of patent applications.

which is 29% of the mean initial allowance rate in the sample. This implies that a patent application with mixed-gender or all-female inventors assigned to an examiner with one standard deviation higher bias against mixed-gender patents would have 29% lower probability of being allowed compared to an application with identical content but male-only inventors and an average-leniency examiner.

Taken together, the evidence in Table 5 suggests examiner discretion and gender bias play a crucial role in allowance decisions violating both horizontal and vertical equity. First, although the probability of initial allowance is low on average, there are examiners who are likely to allow a substantial share of the applications they examine. Second, although the average examiner exhibits no bias, there is substantial variation in examiners' tastes, with some favoring patents with female inventors and others not.

The third row in Table 5 shows a strong negative correlation between examiner-level bias and leniency. In Section 10, I fit a nonlinear model for initial allowance and find zero correlation between the two. It suggests, as noted in Kline and Walters (2021), that the negative correlation I find here 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. In line with the OLS gender gap estimated in Table 3, the mean β_j is zero.

Panel (ii) presents the variance components of α_j and β_j across the 38 unique start years of examiners, spanning from 1975 to 2013. The estimated standard deviation of α_j across start-years is 3.8 percentage points, and that of the gender bias β_j is 1.2. Since examiners are nested within start years, following the law of total variance, I conclude that 17% of the variation in the leniency of examiners and 25% of the variation in examiner-level gender bias is explained by the variation across cohorts of examiners. 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 6) implies that experienced examiners are more lenient, as was also documented in Frakes and Wasserman (2017) and was attributed partially to differences in time constraints examiners of different grade levels face.

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. Nevertheless, since "movers" account for only 10 percent of the examiners, we can approximately state that 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 discussed in Section 6, the OLS estimator and other matching estimators do not necessarily 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 my results to the estimand, In Appendix Table D.1, I present the variance components of leniency and gender bias using Inverse Probability Weighting (IPW). Specifically, restricting attention to examiners with at least 100 observations, I start by estimating the propensity score of each examiner $\Pr(J(i) = j | X_i, F_i = f)$ with an OLS regression. Then, I use the IPW weights to reweight the data and estimate the variance component with the reweighted microdata. See Appendix Section D for further details.

8.2 Start-Year vs. Experience Effect

The evidence from Figure 6 and Table 5 suggests substantial heterogeneity in gender gaps across examiners with different years of experience. Such variation could be driven by either cohort effects, 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:

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

where $exp(i)$ is the number of years of experience of examiner $J(i)$ at the time of assignment to patent 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 variance component estimates from Equation 7, suggesting it is a cohort, rather than age/experience effects, that drives the results in Figure 6. In the first panel, I examine the stability of the standard deviation of examiner level gender gap after accounting for years of experience gender gaps 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 at all when F_i is an indicator for a female ranked first or second in the list of inventors. The change in 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. Furthermore, the estimated variance for β_{exp} is negative when F_i is an indicator of a female ranked first or second, suggesting that this variance component is very small or zero.

The second and third panels of Table 6 provide further evidence that experience effects do not drive the variation in gender gaps across cohorts of examiners and art units. Years of experience fixed effects have almost no impact on the standard deviation of gender gaps across start years and a moderate effect on the standard deviation of gaps across art units.

Table 6: Heterogeneity across examiners and years of 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 bias-corrected standard deviation and correlations of examiners, examiners' start-year, and art units gender gaps, 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 use 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 list of 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.

Taken together, these results establish that the heterogeneity in gender gaps across examiners with different years of experience is driven mainly by cohort effects, where different cohorts of examiners have different preferences toward the gender of the team of inventors.

8.3 Characterizing 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 4 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. Since bias varies across technologies and art units, 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 coefficients from a nuivariate regression on: 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.

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 examiner-level gender bias in initial allowance $\hat{\beta}_j$ and examiner characteristics estimated with an OLS regression. Columns 1-6 include only the examiner characteristics listed in the table, and column 7 reports the coefficients from a regression that additionally controls for art unit fixed effects. Robust standard errors are in parentheses.

Female examiners and examiners holding a Ph.D. are more likely to be biased against men, while there is no correlation between the bias 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.

Within are-units, the relationship between gender bias and examiner characteristics differs from the cross-sectional relationship. On the one hand, 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 the gender of examiners 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) documenting bias in favor of women in male-dominated fields in the entrance exam for French higher education.

8.4 Within-Group Heterogeneity

Examiner gender

Table 8 reports 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 4 separately by examiner gender.[26] The first row reports the raw gender gap without controlling for the text embeddings, and the second row reports 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 gender of examiners, with a gap among male examiners twice the size of that among female examiners. These differences could reflect either gender differences in bias or differences in the distribution of examiners and applicants across fields and genders. 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 suggests substantial heterogeneity by examiner gender. 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 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, thereby posing less risk for bias in the system.

[26] 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 team of inventors.

Table 8: Heterogeneity in gender gap by examiner gender

	Examiner gender		
	Female (1)	Male (2)	Unknown (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

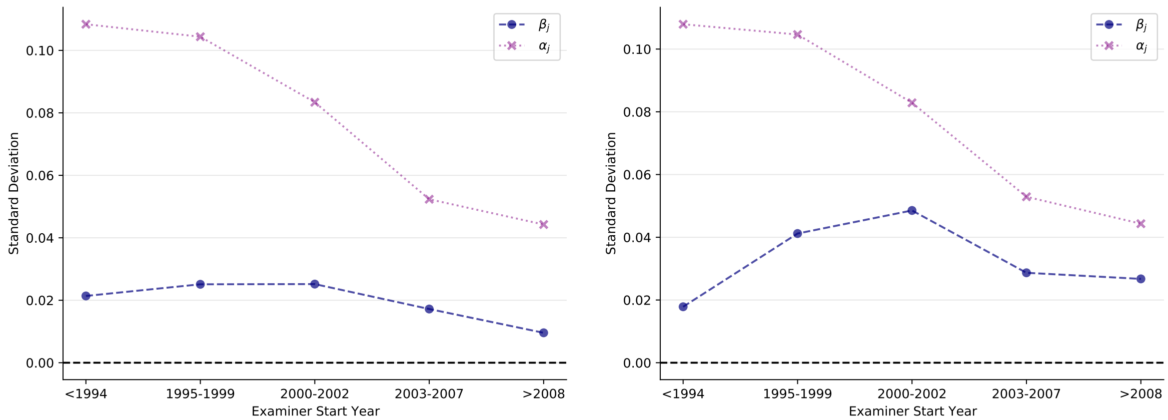
Note: This table presents the distribution of gender gaps by gender of the examiner. Panel (i) reports the mean gender bias separately by examiner gender with and without controlling for the text embeddings. Panel (ii) reports the bias-corrected standard deviation of examiners' gender bias and leniency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

Examiner start year

Different cohorts of examiners differ not only in their average gender preferences but also in the probability of performing extreme levels of gender bias. Figure 9 studies that phenomenon by estimating the within start-year bins variance components of α_j and β_j when the femaleness of the patent is measured by whether there is at least one female in the team of inventors and by the proportion of females in the team of inventors. The pink "X"s are the estimated standard deviation of α_j, and the blue dots are the estimated standard deviation of β_j.

The figure shows that younger cohorts of examiners who joined the USPTO after 2003 exhibit lower levels of variability in discretion, with a standard deviation of α_j nearly 50% lower than that of older cohorts. While these younger cohorts demonstrate reduced discretion, the variance in gender gaps is more stable across diffrent cohorts.

Figure 9: Standard deviation of examiner gender gap and leniency by examiner start year



(a) Any female inventor

(b) Proportion female

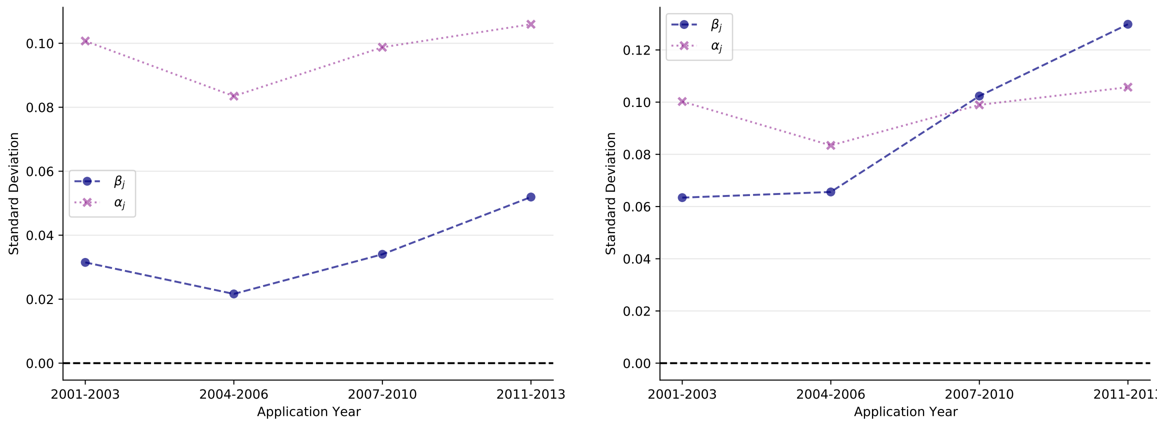
Note: This Figure plots the bias-corrected standard deviation of examiner gender gap and leniency. The blue dots are the estimated standard deviation of examiners' gender gap, and the pink "X"s are the estimated standard deviations of examiners' leniency. All variance estimates are weighted by the number of applications per examiner.

Filing year

The average gender bias decreased over time and converged to zero. This leads to the follow-up question: does the risk of encountering a biased examiner also change accordingly? To test that, Figure 10 plots the standard deviations of β_j within bins of 3 to 4 years. To assess the trend in discretion over time, Figure 10 additionally plots the estimates of the standard deviation of α_j.

Although the average gender gap converged to zero over time, the variance in discretion and bias nearly doubled. Combined with the findings in Table 6, this suggests that changes in examiner composition drive these trends. Younger examiners are more aligned in their decisions but differ in gender preferences from senior examiners, leading to increased polarization in preferences for gender and greater uncertainty regarding the outcome of the examination process.

Figure 10: Standard deviation of gender gap and leniency over time



(a) Any female inventor

(b) Proportion female

Note: This Figure plots the bias-corrected standard deviation of examiner gender gap and leniency. 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.

9 Robustness and Mechanisms

9.1 Rejection Reasons

To investigate the underlying roots behind initial rejections, I utilize the "USPTO Office Action Rejection data set" which covers the universe of mailed office actions and grounds for rejections from 2008 to 2017. A typical Non-Final Rejection Office action by an examiner 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 failure to meet the requirements regarding the adequacy of the disclosure of the invention.

For every rejection ground category, I generate four indicators, one for each rejection ground, which equals one if the patent application was rejected at the first round of examination on the basis of that category. Notably, a rejection could have multiple grounds, and nd each ground may apply to more than one claim. As discussed in Frakes and Wasserman (2017), 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. Panel (i) presents the coefficient from an OLS regression of an indicator for rejection reason on an indicator for at least one female inventor on the sample of patent applications filed after 2008. The first row reports the uncontrolled gender gap, while the second row shows the gap after controlling for 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. 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 average gender gap disappears after controlling for the patent application text.

To study which rejection reasons 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:

$$\begin{aligned} IA_i =& \alpha_{J(i)} + \beta_{J(i)} F_i + X_{it}' \gamma + \epsilon_i \\ IR_i =& \alpha^R_{J(i)} + \beta^R_{J(i)} F_i + X_{it}' \gamma^R + \epsilon^R_i . \end{aligned}$$

The first equation is identical to the main regression I present in previous sections. In the second equation, IR_i is an indicator for the reason of rejection, α_j^R measures the inclination of examiner j to initially reject on the ground of reason IR, and β_j^R measures the extent to which examiner j is more likely to reject all male vs. mixed gender patents based on that ground. With the estimates of gender bias β_j and β_j^R, I estimate the variance-covariance matrix of (β_j, β_j^R). For detailed information on the estimation, see Appendix Section C.[27]

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 the higher variance level of gender gaps in recent years documented in Figure 10. As expected, there is a strong negative correlation between examiner-level gender gaps in initial allowance and examiner-level gender gaps in all the rejection reasons.

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

$$\beta_j = \delta_0 + \delta_1 \beta_j^R + u_j ,$$

[27] 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.

where the δ_1 coefficient is a function of the variance components, i.e., the ratio of the covariance between β_j and β_j^R and the variance of β_j^R. Given these estimates, I calculate the implied R^2 from this regression that measures the share of gender bias variability in initial allowance explained by bias in each rejection reason. If a biased examiner chooses the rejection reason proportional to the prevalence of each rejection reason, then the R^2 should be proportional to the likelihood of each rejection reason. 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.

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

	Rejection type			
	Obviousness (103) (1)	Novelty (102) (2)	Eligibility (101) (3)	Writing (112) (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) Variance componenets				
$Std(\beta_j)$	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 and examiner-level heterogeneity in gender gaps for initial allowances and initial rejections by rejection reasons. Panel (i) reports the mean gender gap in rejection reasons, estimated using OLS regressions with and without controlling for text embeddings. Panel (ii) shows the bias-corrected standard deviation of examiners' gender gaps, estimated in a stacked regression with initial allowances and rejection reasons as outcomes, as described in Appendix Section C, controlling for patent text embeddings. Variance components are weighted by the number of patent applications.

Panel (ii.b) in Table 9 presents the implied OLS δ_0 coefficient and R^2 from the infeasible regression mentioned above, revealing that biased examiners are disproportionally more likely to perform writing-based rejections. The gender 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 writing-based rejections.

Taken together, these indicate that biased evaluations are more likely to be based on writing. This result aligns with the reasoning that such rejections are "simpler" because they require less effort than prior-art-based rejections. A similar pattern was identified in Frakes and Wasserman (2017), in which examiners facing more time restrictions had a lower probability of performing obviousness rejections.

9.2 Other Non-text Characteristics

To assess the extent to which other non-text patent characteristics, such as inventor ethnicity and lawyer characteristics, translate into gender disparities due to disparate impacts, I rerun the main analysis, including additional non-text characteristics. Specifically, I re-estimate Equation 2, sequentially adding the following covariates: (i) an indicator for the presence of at least one inventor with a foreign name, defined as not appearing in the SSA name tables; (ii) an indicator for the presence of at least one Asian inventor on the team; and (iii) the years of experience of the attorney's office.[28]

Panel (i) of Appendix Table A.6 suggests that the mean overall gender bias estimates are robust to the inclusion of other non-text characteristics. Column 1 replicates the results from Table 3, and columns 2-5 report the coefficients from a regression that includes an additional non-text characteristic, suggesting that the including other non-text characteristics does not affect the measures of the gender gap in initial allowance. Moreover, I find no average allowance gap by inventor ethnicity or attorney's gender. Contrarily, however, the lawyer's years of experience are positively related to the 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 differences in examiners' preferences towards more established law companies.

Second, Panel (ii) reports the variance components from the following fixed-effect regression

$$IA_i = \alpha_{J(i)} + \beta_{J(i)} F_i + \eta_{J(i)} w_i + X'_{it}\gamma + \epsilon_i,$$

where w_i is the additional non-text characteristics and η_j measures the variation in examiners tastes with respect to characteristic w_i. Then, Panel (ii) reports the standard deviation and

[28] See Appendix B for details on variable construction.

correlation of (β_l, η_j). Echoing the results from Panel (i), the variance of gender bias is unaffected by the inclusion of other characteristics. Results suggest that examiners exhibit variation with respect to these other characteristics at a comparable level to the variability in gender bias, but the correlation between gender bias and these other biases is very low, amounting between 1 to 6 percent.

9.3 Robustness

One of the grounds for rejecting patents is lack of novelty, which relates to the extent to which a particular innovation differs from the existing stock of knowledge at that time. Mismeasurement of this value by the BERT embeddings, in a way correlated with the gender composition of the patent, could lead to biased estimates. To address this, we measure the novelty of a patent using an approach similar to Kelly et al. (2021). For each patent application, we calculate the mean cosine similarity between it and all other patent applications filed in the same technology center during the three years preceding its filing. The closer the patent is to prior knowledge, the less novel it is, whereas greater distance suggests a higher likelihood of being a breakthrough invention. We then use this measure as an additional control when estimating the gender gap in initial allowance rates between mixed-gender and all-male patent applications.

The similarity index we use strongly predicts allowance decisions, as shown in Appendix Figure A.7. However, its inclusion in the regression does not affect the estimated gender gaps, regardless of whether embedding controls are included. This suggests that there is no significal correlation between breakthrough patents and the gender composition of the inventors. Additionally, we find little heterogeneity across applications with different values of similarity indices. Among the most innovative patents (those with the lowest cosine similarity values), the raw gender gap is 25% lower than the overall raw gender gap. However, this heterogeneity disappears entirely after controlling for text embeddings, resulting in zero effective gender gaps, regardless of patent novelty.

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 examination process. Next, I turn to explore the broader implications of such bias and its potential effects on economic outcomes by estimating how bias and discretion at the examiner level translate into differences in granted patents' stock market return as measured by Kogan et al. (2017).

Identification of the effect of bias on the market return of granted patents is challenging. Unlike the first round of the application process, Aneja et al. (2024) find that women are less likely to persist and resubmit their patent application if initially rejected. Hence, a gender gap in outcomes of granted patents could also reflect differences in the female behavior rather than the examiners. In addition, market return is observed only for granted patents 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) exploiting examiners as instrumental variables, conditional on the patent text. I model the allowance decision of examiners 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 initially allowed patents, thereby not violating the exclusion restriction assumption.

10.1 Examiner Decision

I start by modeling the initial allowance decision of examiners from Section 4.1 structurally. I assume examiners form an accurate posterior mean prediction, $E[q_i|\epsilon_{ij}, C_i, F_i = f]$, of the patent quality q_i given the available information on patent content C_i, the gender mix of the inventors F_i, and a noisy signal ϵ_{ij}. To make an allowance decision, each examiner compares the posterior mean quality to a subjective cost, $\tau_j(C_i, f)$, for allowing a patent with gender $F_i = f$ and content C_i. Thus, this model yields the following decision rule:

$$\begin{aligned} IA_{ij} &= \mathbb{1}\{E[q_i|\epsilon_{ij}, C_i, F_i = f] \geq \tau_j(C_i, f)\} \\ &= \mathbb{1}\{\mu(C_i, f) + u_{ij} \geq \tau_j(C_i, f)\}. \end{aligned} \tag{8}$$

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 the posterior mean prediction of patent quality due to differences in the prior distribution of patent quality by gender. As detailed in Section 4, all of these models would result in a gender gap in the probability of initial allowance where $\Pr(IA = 1|C_i, F_i = 0) \neq \Pr(IA = 1|C_i, F_i = 1)$.

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

$$R_{ij}(IA_{ij}) = \psi_c(C_i, F_i)IA_{ij} + \delta_{cj}(C_i, F_i)(1 - IA_{ij}) + \omega_{ij}, \tag{9}$$

where $\psi_c(C_i, F_i)$ is the mean log market return among initially allowed patents, $\delta_{cj}(C_i, F_i)$ is the mean log market return among initially rejected patents, and ω_{ij} is a mean zero unobserved component. 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 examination, there is no direct examiner causal effect. Therefore, the expected market return of firms assigned to initially allowed patents is:

$$E[R_{ij}|F_i, C_i, J(i) = j, IA_{ji} = 1] = \psi_c(C_i, F_i) + E[\omega_{ij}|F_i, C_i J(i) = j, IA_{ij} = 1], \tag{10}$$

where $E[\omega_{ij}|F_i, C_i, J(i) = j, IA_{ij} = 1]$, is the expected unobserved component of market return among initially allowed patents with gender mix F_i and patent content C_i that were assigned to examiner j. If the perceived quality of examiners is uncorrelated with market return ($Cov(\omega_{ij}, u_{ij}) = 0$), then one could estimate Equation (10) by running an OLS regression. Otherwise, $E[\omega_{ij}|F_i, C_i, J(i) = j, IA_i = 1]$ is the *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 returns (Canay et al., 2024). In any case 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 $\Pr(IA = 1|C_i = c, F_i = 0) = \Pr(IA = 1|C_i = c, F_i = 1)$ so applications with the same contents and 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 inventors work at or differences in how male and female inventors translate their inventions into valuable products.

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

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

where ρ captures the extent to which the decisions of examiners are correlated with patents' stock 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 total market return of allowed patents by mixed-gender teams would be lower than the total market return had these patents been judged by the same examiner as an all-male team of inventors, and vice versa.

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 without directly affecting 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 discretion plays a significant role in the allowance decision of examiners and that, conditional on the patent application content, the assignment of applications to examiners is as good as random. Also, since I estimate the model only among 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 ℓ, either $\Pr(IA_{ij} = 1) \geq \Pr(IA_{i\ell} = 1)\ \forall i$, or $\Pr(IA_{ij} = 1) \leq \Pr(IA_{i\ell} = 1)\ \forall i$. 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.

Conceptually, the monotonicity assumption requires that examiners agree on the ranking of applications on average, therefore, this assumption imposes restrictions on the underlying behavioral model and skills of examiners. To relax this assumption, I allow the preferences of examiners to vary across patent applications by inventor gender and by the content of the patent $\tau_j(C_i, F_i)$. By doing so, I allow examiners to rank patents with different content differently, requiring monotonicity only among applications with the same 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 2, which I denote by $\tilde{x}$. Thus, the random thresholds and the parameters in the market return and initial allowance equation are modeled as follows:

$$
\begin{aligned}
\mu_(\tilde{x}_i, F_i) &= m_0 + \tilde{x}_i' m_{0x} + (m_1 + \tilde{x}_i' m_{1x}) F_i \\
\tau_j(\tilde{x}_i, F_i) &= \tau_{0j} + \tilde{x}_i' \tau_{0xj} + (\tau_1 + \tilde{x}_i' \tau_{1xj}) F_i \\
\psi(\tilde{x}_i, F_i) &= \psi_0 + \psi_1 F_i + \tilde{x}_i' \psi_x,
\end{aligned}
$$

where $\tilde{x}$ is normalized to mean zero with a standard deviation of one.

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

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

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 $ps_{ij} = \Pr(J(i) = j | C_i, F_i = f)$ 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 aligns with my estimation strategy of the variance components in Section 7.2. A similar reweighting approach was taken in Chan et al. (2022).

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:

$$\mathcal{L}(\Theta | IA_i, R_i, F_i, C_i) = \sum_j \log \int_\theta l_j(\{IA_i, R_i, F_i, \tilde{x}_i\}_{i \in I_j} | \theta_j) \phi(\theta_j | \Theta) d\theta_j,$$

where $\phi(\theta_j|\Theta)$ represents the prior normal distribution described above, and the integral is computed by simulated maximum likelihood.

10.2 Results

Table 10 presents the main parameter estimates of Θ. Panel (i) displays the estimates for the initial allowance regression. μ_0 governs the mean initial allowance rate for all-male patent applications and μ_1 the differences between all-male and mixed-gender. In line with my findings using a linear model, the estimate of μ_1 is very small and insignificant from zero, suggesting that there is no average difference in allowance probability between mixed-gender and all-male patent applications.

Panel (iii) presents the variance of the allowance threshold of examiners, where σ_{λ_0} represents the standard deviation of the leniency of examiners, and σ_{λ_1} describes the extent to which the allowance threshold of examiners differ between all-male vs. mixed-gender patent applications, i.e., the standard deviation in gender bias. Translating the σ_{λ_1} to allowance probability space, the estimates in Table 10 imply that the variance of gender bias for the average patent text is 0.022, which accords with the findings in Table 5. Additionally, I find that the correlation between λ_1 and λ_0 is zero, suggesting that the negative correlation I 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 log market return among all-male patents and ψ_1 describes the difference between all-male and mixed-gender teams of inventors. On average, mixed-gender patents generate 0.247 higher log market return than no-female patents with equivalent patent content and text. This disparity could reflect differences in the productivity of the firms female inventors work at or differences

Table 10: Sample selection parameter estimates

	log(R)
	(1)
(i) IA	
μ_0	-1.749
	(0.010)
Any female (μ_1)	-0.010
	(0.017)
(ii) Outcome	
ψ_0	1.275
	(0.067)
Any female (ψ_1)	0.247
	(0.025)
(iii) Random effects	
σ_{τ_0}	0.650
	(0.003)
σ_{τ_1}	0.338
	(0.007)
$Corr(\tau_1, \tau_0)$	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 reports the estimated parameter from a selection model (Heckman, 1979) of the joint distribution of initial allowance and stock market return. μ_0 and $\sigma_{\lambda 0}$ determine the distribution of examiners' initial allowance decision for all male applications, and μ_1 and $\sigma_{\lambda 1}$ determine the distribution of examiners' 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 error terms.

in 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 don't necessarily imply that the behavior examiners generate inefficiencies. Variance in bias generates lower total returns if the objectives of examiners, i.e., the object they maximize, are 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). The finding that the decision of examiners is moderately positively correlated with economic outcomes aligns with the findings of Matcham and Schankerman (2023) who model the full screening and renegotiation decisions of examiners and applicants.

10.3 Counterfactuals

I evaluate the economic implications of positive variance in bias together with mean zero gender bias via a simulation exercise. Using the parameters in Tables 10, I simulate a set of 6,000 examiners, each examining 1,500 applications with an average patent text embedding value, enforcing 15% of the teams to be mixed-gender teams. I evaluate the impact of bias and discretion on stock market return using two counterfactual exercises. In both, I compare the status quo distribution of examiner behavior in which examiners exhibit substantial variation in discretion, mean zero gender bias, and positive variation in bias. I compare the status quo to two scenarios. The first considers a new initial allowance decision rule that maintains the same examiner-level initial allowance rate but enforces uniform zero gender bias for all the examiners. Therefore, this simulation allows for examiner heterogeneity in leniency but shuts down only the gender bias component while maintaining the same number of allowed applications as in the status quo. Formally, for every examiner j, I find the scalar $\tilde{t}_j$ such that

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

The second scenario considers the initial allowance decision rule in which there is no discretion across examiners. To maintain the same total number of applications accepted across counterfactual simulations, I set the examiner threshold to be the one that attains the same baseline rate of initial allowance. Then, for every scenario, I report the stock market returns of initially allowed patents, where the main focus is on the compliers, i.e., those applications that were not allowed in the status quo but are allowed under the counterfactual exercise and the opposite. If the total stock market return of the first is greater than the second, I conclude that the observed levels of discretion generate economic loss.

Variation in bias and discretion generates a substantial economic loss, as reported in Table 11. Panel (A) reports the evaluation of uniform zero bias. It shows that 1.1% of the female applications would have been affected by the zero variance in bias policy, increasing the average stock market return of allowed patents by 231 thousand dollars. Among all male applications, 0.22 percent would be affected by the policy, resulting in an average increase of 47,000 dollars per application. With an average of 40,000 patent applications per year assigned to publicly traded firms and an average of 8.5% initial allowance rate, column (4) shows the total loss in stock market return amounts to 1.4 million dollars for mixed gender patents and 299 thousand dollars for all male patents a year. This total cost of 1.7 million dollars per year reflects only a lower bound of the social cost of positive variance of gender bias because it accounts only for the loss in the first round of the patent examination process.

Under the assumption that examiners follow the same behavior in later rounds, column (5) shows that this total cost reaches 12.6 million dollars per year for eventually granted patents.

Table 11: Counterfactuals under uniformly zero bias

	(1) Allowance given	(2) Allowance taken	(3) Difference	(4) Total market loss	(5)
				IA	Ever granted
(A) Uniform zero bias					
(i) Mixed-gender apps					
Av. market return	5.867	5.688	0.231	1.372	10.333
Share	0.0116	0.0117			
(ii) All-male apps					
Av. market return	4.572	4.529	0.047	0.299	2.250
Share	0.0022	0.0022			
(B) Zero discretion					
(i) Mixed-gender patents					
Av. market return	6.270	5.522	0.231	4.712	35.570
Share	0.0401	0.0400			
(ii) All male patents					
Av. market return	4.972	4.476	0.047	5.447	41.011
Share	0.0310	0.0303			

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 10 and under two additional counterfactual exercises. Panel (A) reports the counterfactual outcomes under uniform zero bias while allowing for variation in discretion across examiners. Panel (B) reports the results under no variation in discretion across examiners, setting all examiners' threshold to the one that results with the mean initial allowance rate. Column (1) reports the average market return among the patent applications that were not allowed under the status quo but are allowed under the uniform zero bias/zero discretion simulation. Column (2) reports the average market return among the applications allowed under the status quo but not under the uniform zero bias/zero discretion simulation. Column (3) reports the difference between column (1) and column (2). Columns (4)-(5) report the total yearly difference in market return for 40,000 applications assigned to publicly traded firms, with 15% mixed-gender teams. Column (4) reports the total market return loss among initially allowed patent applications, and column (5) reports the total loss for ever-granted patent applications, assuming the first round of examination behavior and later rounds of examination distribute similarly.

The economic cost from any variation in discretion across examiners is almost six times larger. In Panel B, I report the mean market return of compliers, comparing the status quo to the no-discretion scenario. The share of compliers is higher under this exercise. However, since I enforce a symmetric distribution of discretion, the average market return loss per patent application is identical to the one found in Panel A. The larger share of compliers eventually results in a larger total loss of 10.15 million dollars per year among initially allowed patent applications and a total of 76.6 million dollar loss per year for eventually granted patent applications. To have a sense of this magnitude, it amounts to 31% of the value of the median public US firm in 2013.

11 Conclusion

Women are underrepresented in the patent system. Yet, the extent to which observed disparities arise from discriminatory practices is unclear. To shed light on this, this study 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. The start-year of the examiner explains 25 percent of the variance of gender bias, where senior examiners are more likely to be biased against mixed-gender patents, and young examiners are more likely to be biased against no-female patents. Lastly, studying the dynamics over time, I find that, on average, the system evolved from being discriminatory against women in the years 2001-2003 to being unbiased on average. However, due to the changes in the cohort composition of examiners, that variance in gender bias has increased over time, reflecting an increase in the risk of encountering an abnormally biased examiner.

Much of the discrimination literature primarily fixates 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 where patents of equivalent quality but different inventor team compositions experience disparate odds of approval. Second, an exclusive focus on mean bias overlooks the detrimental effects of misallocation, which may manifest even if the mean bias is zero. Utilizing Kogan et al. (2017)'s stock market return model for patents, my analysis estimates the annual cost of having a positive variance in gender bias among initially allowed patent applications assigned to publicly traded firms to be approximately $1.7 million. Extrapolating the implied cost for granted application, I report that bias generates a loss of 12.6 million dollars per year.

The findings in this paper call for reassessing the patent examination process, particularly regarding examiners' access to inventors' names. Implementing blind reviews, potentially supported by computer-based methods, could provide effective remedies. Additionally, since only a small number of examiners exhibit significant bias, the USPTO could identify these evaluators and investigate their work (Avivi et al., 2021; Kline et al., 2022).

An interesting topic for future research is to assess the extent to which examiner discretion and bias influence inventors' behavior. If women's motivation is more negatively affected by bias, even a symmetric, two-sided bias could lead to larger disparities in the patent system (Aneja et al., 2024). Understanding the disparate impact of discretion is also critical for evaluating the welfare costs of bias, especially if differential behavior changes the distribution of patent quality of granted patents. Moreover, if women are risk-averse and aware of the high levels of discretion and bias in the system, this awareness could discourage them from pursuing careers as inventors or submitting patents. Determining the causes of selection into science and the role of information remains an important and active area of research (Bell et al., 2019).

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.

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.

Aneja, A., Reshef, O., and Subramani, G. (2024). Attrition and the gender patenting gap. *Review of Economics and Statistics*, pages 1–31.

Angrist, J. (1995). Estimating the labor market impact of voluntary military service using social security data on military applicants.

Angrist, J. D. and Krueger, A. B. (1999). Empirical strategies in labor economics. In *Handbook of labor economics*, volume 3, pages 1277–1366. Elsevier.

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.

Avivi, H., Kline, P., Rose, E., and Walters, C. (2021). Adaptive correspondence experiments. In *AEA Papers and Proceedings*, volume 111, pages 43–48. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Battaglia, L., Christensen, T., Hansen, S., and Sacher, S. (2024). Inference for regression with variables generated from unstructured data. *arXiv preprint arXiv:2402.15585*.

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.

Bloom, N., Schankerman, M., and Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, 81(4):1347–1393.

Bohren, A., Imas, A., and Rosenberg, M. (2018). The language of discrimination: Using experimental versus observational data. In *AEA Papers and Proceedings*, volume 108, pages 169–174. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

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

Bohren, J. A., Imas, A., and Rosenberg, M. (2019). The dynamics of discrimination: Theory and evidence. *American economic review*, 109(10):3395–3436.

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.

Bryan, K. A. and Williams, H. L. (2021). Innovation: market failures and public policies. In *Handbook of industrial organization*, volume 5, pages 281–388. Elsevier.

Canay, I. A., Mogstad, M., and Mountjoy, J. (2024). On the use of outcome tests for detecting bias in decision making. *Review of Economic Studies*, 91(4):2135–2167.

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.

Chan, D. C., Gentzkow, M., and Yu, C. (2022). Selection with variation in diagnostic skill: Evidence from radiologists. *The Quarterly Journal of Economics*, 137(2):729–783.

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?

Clemens, J. and Rogers, P. (2023). Demand shocks, procurement policies, and the nature of medical innovation: Evidence from wartime prosthetic device patents. *Review of Economics and Statistics*, pages 1–45.

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.

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

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.

Frakes, M. D. and Wasserman, M. F. (2017). Is the time allocated to review patent applications inducing examiners to grant invalid patents? evidence from microlevel application data. *Review of Economics and Statistics*, 99(3):550–563.

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

Galasso, A. and Schankerman, M. (2015). Patents and cumulative innovation: Causal evidence from the courts. *The Quarterly Journal of Economics*, 130(1):317–369.

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

Glover, D., Pallais, A., and Pariente, W. (2017). Discrimination as a self-fulfilling prophecy: Evidence from french grocery stores. *The Quarterly Journal of Economics*, 132(3):1219–1260.

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

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.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, pages 153–161.

Hochberg, Y., Kakhbod, A., Li, P., and Sachdeva, K. (2023). Inventor gender and patent undercitation: Evidence from causal text estimation. Technical report, National Bureau of Economic Research.

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.

Jones, C. I. (2011). Misallocation, economic growth, and input-output economics. Technical report, National bureau of economic research.

Keith, K. A., Jensen, D., and O'Connor, B. (2020). Text and causal inference: A review of using text to remove confounding from causal estimates. *arXiv preprint arXiv:2005.00649*.

Kelly, B., Papanikolaou, D., Seru, A., and Taddy, M. (2021). Measuring technological innovation over the long run. *American Economic Review: Insights*, 3(3):303–320.

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. and Walters, C. (2021). Reasonable doubt: Experimental detection of job-level employment discrimination. *Econometrica*, 89(2):765–792.

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.

Lemley, M. A. and Sampat, B. (2012). Examiner characteristics and patent office outcomes. *Review of economics and statistics*, 94(3):817–827.

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.

Matcham, W. and Schankerman, M. (2023). Screening property rights for innovation.

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

Miller, R. D. (2020). Technical documentation for the 2019 patent examination research dataset (patex) release. Technical Report 2020-4, USPTO Economic Working Paper.

Mozer, R., Miratrix, L., Kaufman, A. R., and Anastasopoulos, L. J. (2020). Matching with text data: An experimental evaluation of methods for matching documents and of measuring match quality. *Political Analysis*, 28(4):445–468.

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.

Roberts, M. E., Stewart, B. M., and Nielsen, R. A. (2020). Adjusting for confounding with text matching. *American Journal of Political Science*, 64(4):887–903.

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.

Sarsons, H., Gërxhani, K., Reuben, E., and Schram, A. (2021). Gender differences in recognition for group work. *Journal of Political economy*, 129(1):101–147.

Schaerer, M., du Plessis, C., Nguyen, M. H. B., van Aert, R. C., Tiokhin, L., Lakens, D., Clemente, E. G., Pfeiffer, T., Dreber, A., Johannesson, M., et al. (2023). On the trajectory of discrimination: A meta-analysis and forecasting survey capturing 44 years of field experiments on gender and hiring decisions. *Organizational Behavior and Human Decision Processes*, 179:104280.

Sellam, T., Yadlowsky, S., Wei, J., Saphra, N., D'Amour, A., Linzen, T., Bastings, J., Turc, I., Eisenstein, J., Das, D., et al. (2021). The multiberts: Bert reproductions for robustness analysis. *arXiv preprint arXiv:2106.16163*.

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., Reshef, O., and Louis, W. S. (2020). Persistence and the gender innovation gap.

Sur, P. and Candès, E. J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proceedings of the National Academy of Sciences*, 116(29):14516–14525.

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

USPTO (1888). *Women Inventors: To Whom Patents Have Been Granted by the United States Government, 1790 to July 1, 1888: Also Included Appendix 1, July 1, 1888-Oct. 1, 1892; Appendix 2, Oct. 1, 1892-Mar. 1, 1895.* US Government Printing Office.

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.

Yadlowsky, S., Yun, T., McLean, C. Y., and D'Amour, A. (2021). Sloe: A faster method for statistical inference in high-dimensional logistic regression. *Advances in neural information processing systems*, 34:29517–29528.

Zeng, J., Gensheimer, M. F., Rubin, D. L., Athey, S., and Shachter, R. D. (2022). Uncovering interpretable potential confounders in electronic medical records. *Nature communications*, 13(1):1–14.

A Appendix Figures and Tables

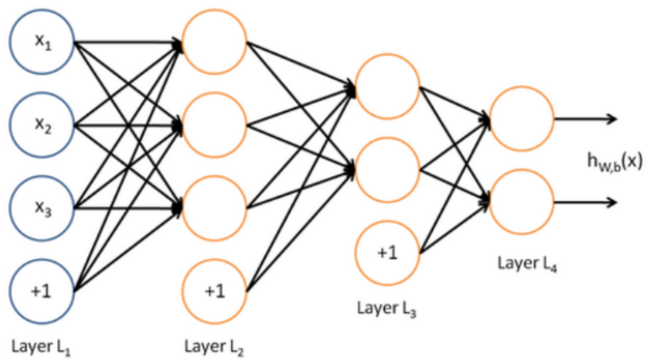
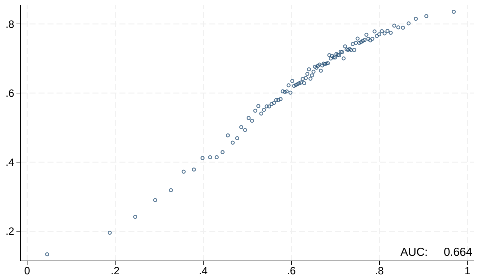


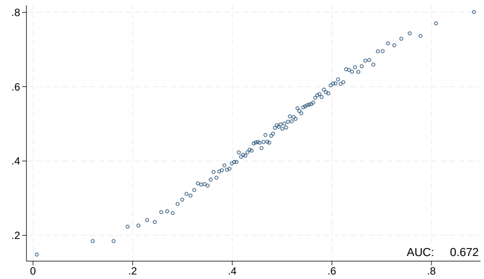
Figure A.1: Visual Illustration of a Neural Network Model

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

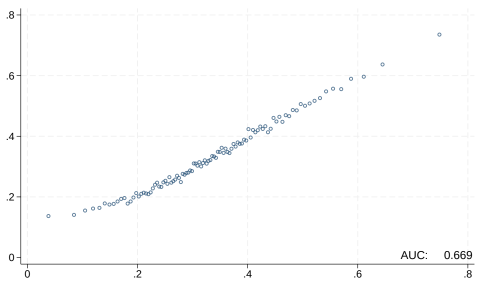
Figure A.2: Split sample binned scatter of embeddings prediction of patent quality proxies



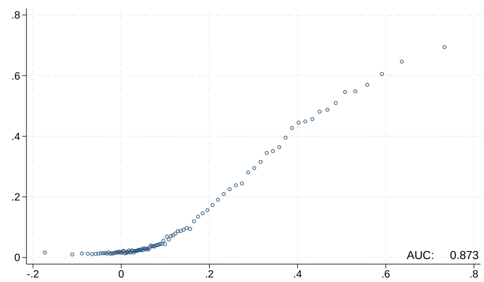
(a) Rejections - non-obviousness



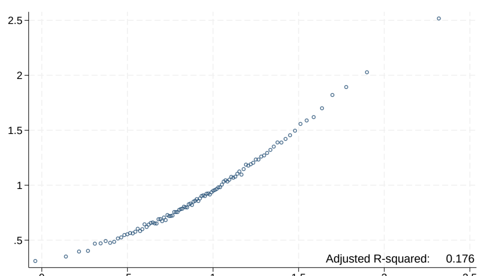
(b) Rejections - lack of novelty



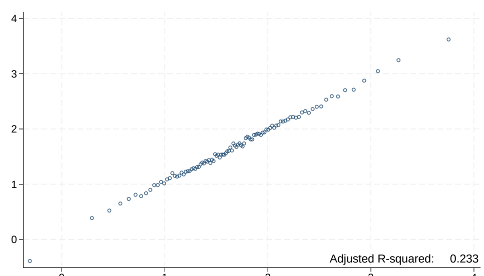
(c) Rejections - writing



(d) Rejections - eligibility



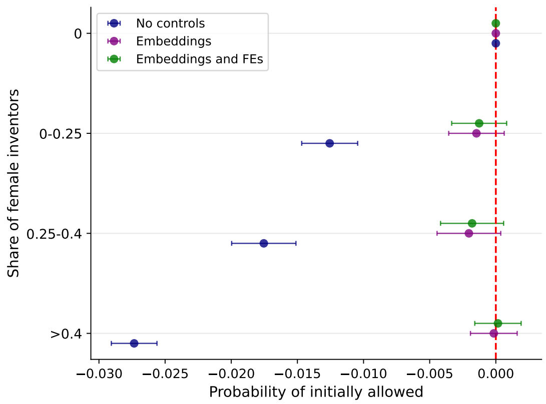
(e) Log citations



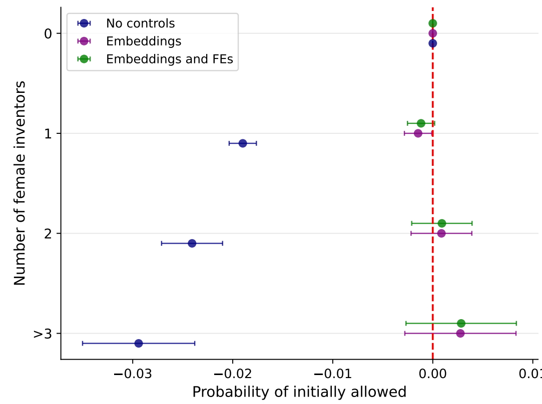
(f) Log market return

Note: This Figure plots a binned scatter with 100 bins of a split sample OLS prediction of examiners' decisions, log citations, and log market return with the patent text embeddings against the real values. Subfigure (a) plots the prediction for rejection based on the ground of non-obviousness, subfigure (b) plots the prediction for rejection based on the ground of lack of novelty, subfigure (c) plots the prediction for rejection based on the ground of inadequate writing, subfigure (d) plots the prediction for rejection based on the ground of not following the eligibility criteria, panel (e) plots the prediction of the log number of citations after granting among granted patents, and subfigure (f) plots the prediction of Kogan et al. (2017)'s measure for patent market return. AUC stands for Area Under the Curve, which measures the overall diagnostic accuracy of the predictor.

Figure A.3: The gender gap in initial allowance, by the number and share of female inventors



(a) By share female inventors



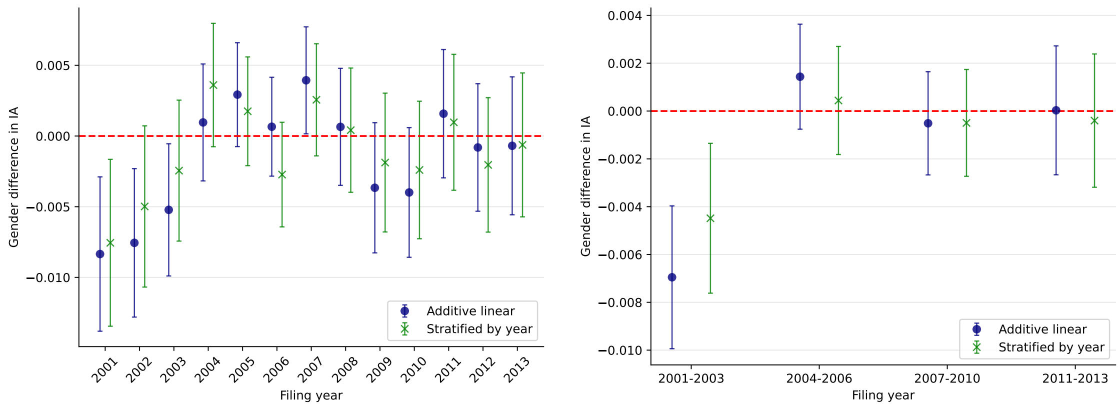
(b) By number of female inventors

Note: This Figure plots the difference in initial allowance probability between mixed gender to all male or unknown teams. Subfigure (a) plots this relationship by share female inventors, and subfigure (b) plots this relationship by the number of female inventors. Bars indicate 95% confidence intervals.

Figure A.4: Fully Interacted Regression of Ebbedings with Gender

Note: This figure plots the estimated coefficients of the 2048 BERT embeddings from a regression of an indicator for initial allowance on embeddings, separately for patents with at least one female inventor (horizontal axis) and patents with all male inventors or unknown gender (vertical axis). The green line shows the naïve regression line, while the black line shows the bias-corrected slope, accounting for excess noise in the coefficient estimates. "Rho" in the bottom-right corner refers to the correlation, and "rho UB" refers to the bias-corrected correlation.

Figure A.5: The gender gap in initial allowance over time

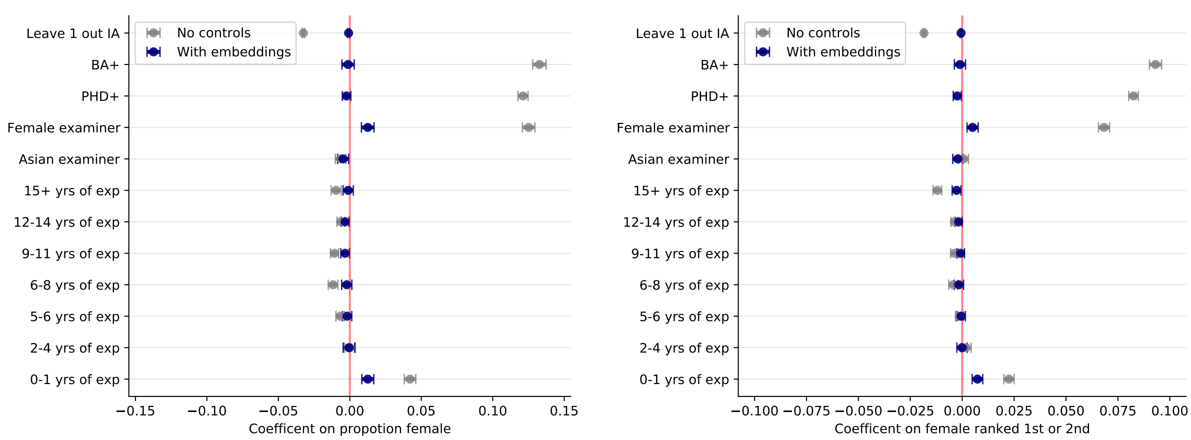


(a) By year

(b) By 3-4 years interval

Note: This Figure plots the difference in initial allowance probability between mixed gender to all male or unknown teams over time. The blue dots present the estimated coefficient from the linear regression described in Equation (**??**), and the green dots present the coefficient on any female inventor indicator for a separate regression on the sample of applications filled in that year, controlling for the patent BERT embeddings. Bars indicate 95% confidence intervals.

Figure A.6: Balance test for examiner characteristics using alternative female definitions

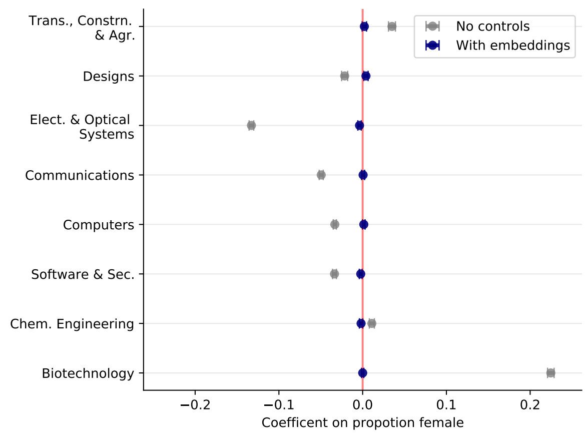


(a) Proportion female

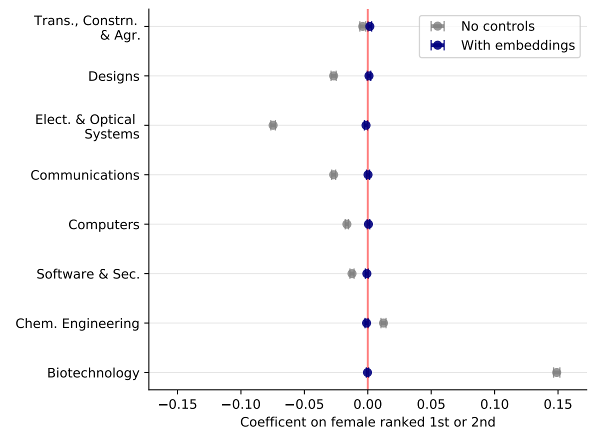
(b) Female 1st or 2nd

Note: This Figure plots the relationship between examiner characteristics 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.

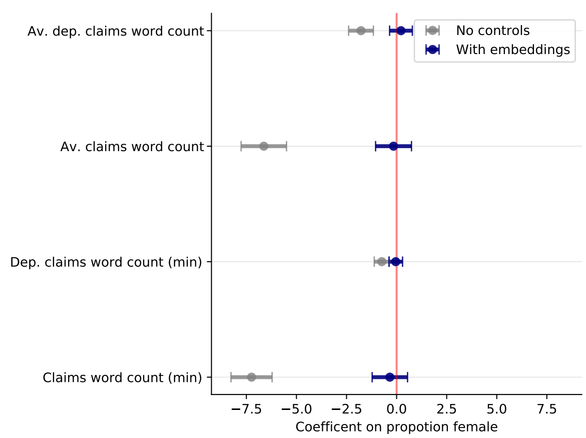
Figure A.7: Balance test, alternative female variables



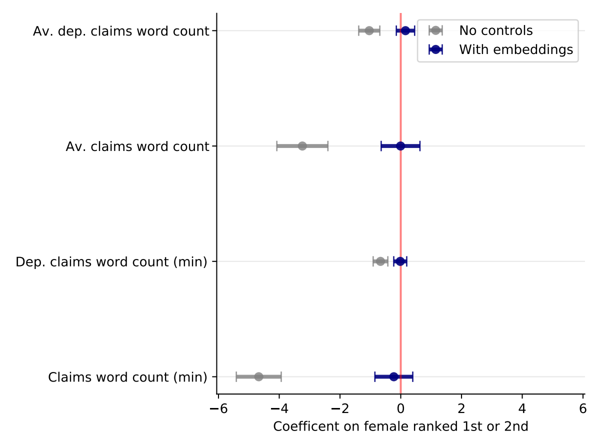
(a) Proportion female



(b) Female 1st or 2nd



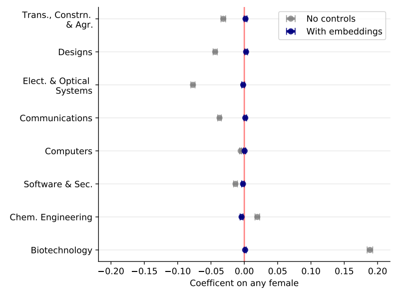
(c) Proportion female



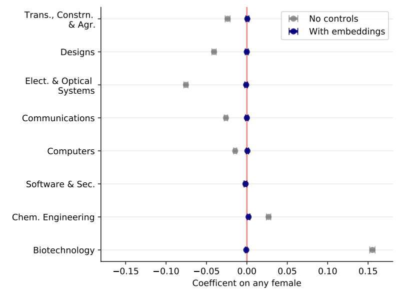
(d) Female 1st or 2nd

Note: This Figure plots the relationship between technology center indicators (subfigures (a) and (b)), claim statistics (subfigures (c) and (d)), and the gender composition of the team of inventors with and without controlling for the patent application text embeddings. The gray dots display the estimates from an uncontrolled OLS regression, and the blue dots display the estimates from an OLS regression controlling linearly for the patent text embeddings. Bars indicate 95% confidence intervals based on robust standard errors.

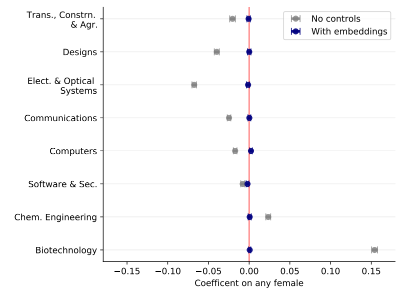
Figure A.8: Balance test by year



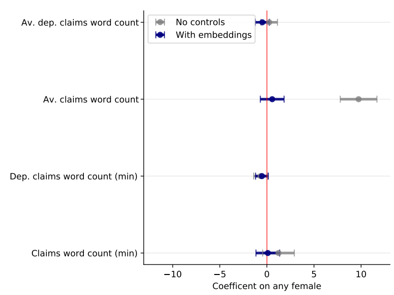
(a) 2001-2004



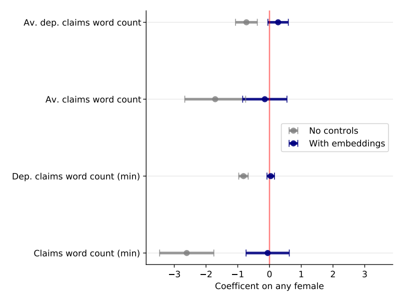
(b) 2005-2009



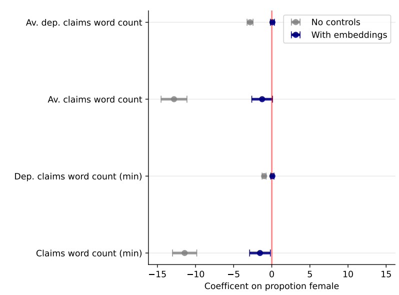
(c) 2010-2013



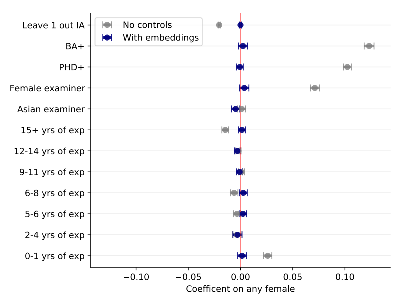
(d) 2001-2004



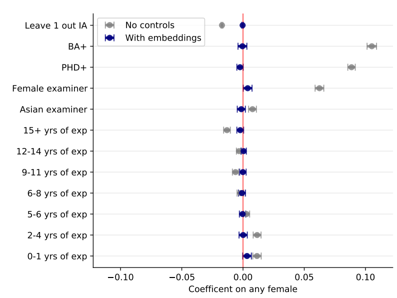
(e) 2005-2009



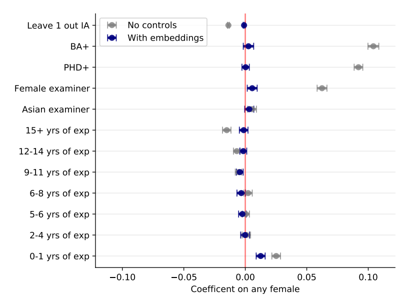
(f) 2010-2013



(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.

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	Description layer	# of embeddings	IA	Any female	IA	Any female	IA	Any female
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(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 the 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: Relationship between omitted covariates and initial allowance

	(1) IA
Female attorney	-0.0018
	(0.0008)
Asian attorney	-0.0009
	(0.0005)
Attorney experience	0.0028
	(0.0006)
Team size	-0.0008
	(0.0002)
foreign_p	0.0015
	(0.0012)
# of applications	1220508

Note: This table reports the relationship between omitted covariation and initial allowance when estimating the examiner-level gender bias conditional of text embeddings. The table presents the coefficients on female patent attorney gender, Asian attorney, number of previously granted patents, inventors' team size, and indicator for foreign priority. The model includes a fixed effect for every examiner, and a fixed effect for every examiner times an indicator for having at least one female in the inventors' team. Robust standard errors are reported in parentheses.

Table A.3: Mean gender gap in initial allowance for single gender teams

	IA (1)	IA (2)	IA (3)	IA (4)
All female vs no female	-0.0364	-0.0112	-0.0010	-0.0002
	(0.0013)	(0.0014)	(0.0014)	(0.0014)
AUC	0.5170	0.7612	0.7490	0.8071
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
Sole inventors	-0.0345	-0.0109	-0.0013	-0.0006
	(0.0015)	(0.0015)	(0.0016)	(0.0016)
AUC	0.5170	0.7753	0.7420	0.8239
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 the OLS coefficients adjusted R^2, and Area Under the Curve (AUC) from regressions of an indicator for initial allowance on the gender of the team of inventors, using subsamples with single-gender teams. Panel (i) reports the estimates from the subsample of patent applications with all female inventors teams vs. no female inventors, and panel (ii) reports the subsample of sole inventors. Robust standard errors are reported in parentheses.

Table A.4: Within examiner variance in gender bias, proportion female

	Female (1)	Male (2)	Unknown (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 bias by the gender of the examiner using the proportion of females in the team of inventors as a measure for the "femaleness" of the patent application. Panel (i) presents the mean gender gap separately by examiner gender with and without controlling for the text embeddings. Panel (ii) presents the Kline et al. (2020) bias-corrected variance components of the gender bias 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 variance in gender bias, female ranked first or second

	Female (1)	Male (2)	Unknown (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 bias by the gender of the examiner 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) bias-corrected variance components of examiners' gender bias and leniency estimated in a linear regression controlling for patent text embeddings. All variance components are weighted by the number of patent applications. Robust standard errors are reported in parentheses.

Table A.6: Robustness to other controling for 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) Examiner heterogeneity				
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 reports the robustness of the mean gender bias and its variability across examiners to the inclusion of other non-text characteristics. Panel (i) reports the coefficients on mixed-gender teams and the other characteristics estimated in an OLS regression controlling for text embeddings. Panel (ii) reports the bias-corrected standard deviation of the 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 names of the other characteristics are displayed in the last row.

Table A.7: Robustness to controling for text similairy

	IA (1)	IA (2)	IA (3)	IA (4)
(i) ALL (N=1,220,512)				
Any female	-0.02004	-0.02089	-0.00105	-0.00102
	(0.00064)	(0.00064)	(0.00065)	(0.00065)
Similarity index		-0.43677		0.04813
		(0.00816)		(0.01431)
(ii) High similarity (N= 406,745)				
Any female	-0.02453	-0.02471	-0.00045	-0.00047
	(0.00116)	(0.00116)	(0.00118)	(0.00118)
Similarity index		-0.19638		0.09698
		(0.01618)		(0.02775)
(iii) Mid similarity (N= 406,741)				
Any female	-0.02237	-0.02240	-0.00202	-0.00202
	(0.00113)	(0.00113)	(0.00116)	(0.00116)
Similarity index		-0.59124		-0.08136
		(0.06712)		(0.06979)
(iv) Low similarity (N= 406,742)				
Any female	-0.01489	-0.01518	-0.00029	-0.00028
	(0.00101)	(0.00101)	(0.00104)	(0.00104)
Similarity index		-1.21495		-0.17161
		(0.03148)		(0.04601)
Embeddings	No	No	Yes	Yes

Note: This table illustrates the robustness of the gender gap estimates to controlling for the cosine similarity between the patent application and all the patent applications filed in the same technology center during the three years preceding its filing. Column (1) report the mean uncontrol allowance gap between mixed gender and all patent applications; column (2) reports the estimate for a model that additionally controls for the patent cosine similarity; column (3) reports the gender gap conditional on the on the BERT embeddings, and column (4) reports the gender gap conditional on both the BERT embeddings and the cosign similarity index.

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).[29] 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.[30] 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).[31] 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 value using a series of event study designs of the stock market return of patents among publicly traded patents.

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

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

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

6. Patent Maintenance Fee Events.[32] 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.[33] '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[34] patent applications filed after November 29th, 2000[35] 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).[36] In cases where the first name is missing or includes only the first letter, I assign gender to the middle name.[37] Applications with at least one inventor with missing first and middle names are excluded from the analysis.

I classify a name 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

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

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

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

[35] 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.

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

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

80%,[38] 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 75% 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 in the following order:

I assign gender to examiners using the following data sources by order in which they are used:

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.[39] 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 the detection of non-East-Asian names because they are known to not be accurate for Asian names.[40]

As a result, I could identify 80% 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).

[38] 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.

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

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

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 I use two sources of information. First, I use the roster data from Frakes and Wasserman (2014), which includes the years 1992-2014. My sample starts before 1992, so I fill in the missing information by identifying the first office action by examiners 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 the 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, I modify the start year of examiners that have suspiciously long "gap years", meaning they have no assigned patents after the start years. I do so by defining the start year only if the examiner doesn't have a gap size of a certain size. I 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

I estimate the following OLS regression:

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

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} from 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) relies on MATLAB's preconditioned conjugate gradient routine *pcg* which solves systems of linear equations in large sparse problems. However, in my settings, X_i includes a dense embedding component preventing the algorithm from converging. 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:

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

where IA_i is an indicator for initial allowance and IR_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:

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

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 $\hat{\xi}^2_{i12} = \frac{IR_i(IA_i - X_i\hat{\eta})}{1-P_{ii}}$ is the leave-i-out estimator of the covariance of the i's error in the two regression models.

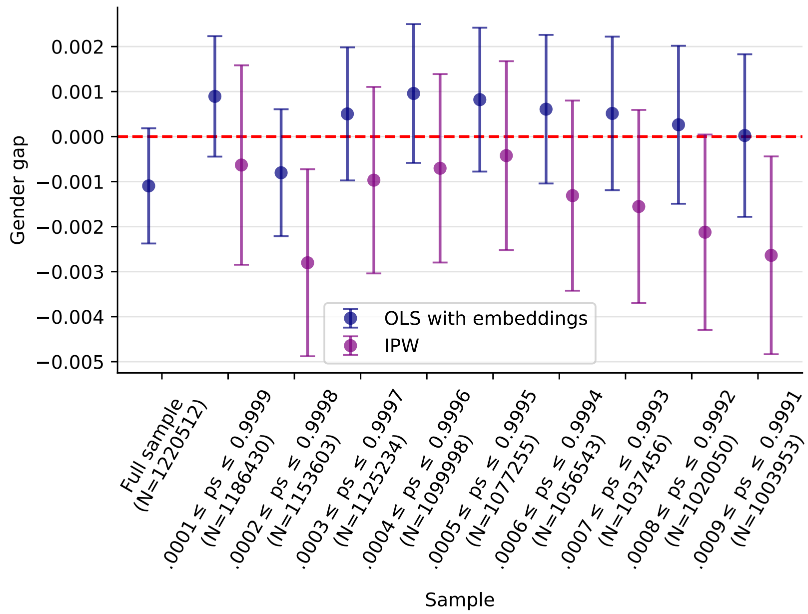
D Propensity Score Estimation

I estimate the propensity score probability of assignment to each examiner j using an OLS regression. The advantage of using OLS instead of nonlinear models like logit or probit is that OLS is less susceptible to bias caused by high dimensionality (Sur and Candès, 2019).[41] To avoid overfitting, I estimate the model using cross-fitting: the sample is split into five random groups, the model is trained on $\frac{4}{5}$ of the sample, and the propensity scores are calculated for the remaining fifth. Throughout the analysis, I trim propensity scores with

[41] Future drafts will compare the OLS propensity score with a logit propensity score that incorporates the correction proposed in Yadlowsky et al. (2021).

values lower than 0.0003% or higher than 99.9997%.[42] Figure D.1 shows that the results are robust to the timing cutoff. Figure XX provides evidence for overlap across examiners by selected art units. In the next subsection, I explain how I estimate the variance component with the inverse probability weights.

Figure D.1: Gender gap in initial allowance by trimming cuttof



Note: This figure plots the estimated gender gap and 95% confidence intervals in initial allowance. The blue dots present the gender gap from an OLS regression that controls for embeddings linearly. The purple dots present the results from an IPW model, where the propensity score of each examiner was estimated by an OLS regression.

D.1 Variance Component With IPW

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:

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

[42] A small cutoff is used to avoid dropping too many observations and altering the analysis sample.

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_{if} for females and w_{im} for males, the variance of the gender gap is:

$$\begin{aligned}
\tilde{n}_{jg} &= \sum w_{ig} \\
\widehat{\theta_{jg}^2} &= \frac{1}{\tilde{n}_{jf}\tilde{n}_{jm}}\sum_{l=2}^{n_{jf}}\sum_{k=1}^{n_{jm}} w_{lf}w_{km}y_{ljg}y_{kjg} \\
\widehat{\theta_{jf}\theta_{jm}} &= \frac{1}{\tilde{n}_{jf}\tilde{n}_{jm}}\sum_{l=1}^{n_{jf}}\sum_{k=1}^{n_{jm}} w_{lf}w_{km}y_{ljf}y_{kjm}
\end{aligned}$$

Table D.1 shows the main results.

Table D.1: 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 times gender fixed effect controlling for embeddings linearly. 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, reweighting the observation using the propensity score.