JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS

doi:10.1017/S0022109025000213

# How Can Innovation Screening Be Improved? A Machine Learning Analysis with Economic Consequences for Firm Performance

Xiang Zheng
*University of Connecticut School of Business*
xiang.zheng@uconn.edu (corresponding author)

## Abstract

This study utilizes U.S. Patent Office data to explore potential improvements in the patent examination process through machine learning. It shows that integrating machine learning with human expertise can increase patent citations by up to 26%. Using machine learning predictions as benchmarks, I find that the early expiration rate of granted patents positively correlates with examiners' false acceptance rates. These errors negatively impact public companies' operational performance and reduce successful IPO or M&A exits for private firms. Overall, this study highlights significant social and economic benefits of incorporating machine learning as a robo-advisor in patent screening.

## I. Introduction

> The strength and vitality of the U.S. economy depends directly on effective mechanisms that protect new ideas and investments in innovation and creativity. (U.S. Patent and Trademark Office)

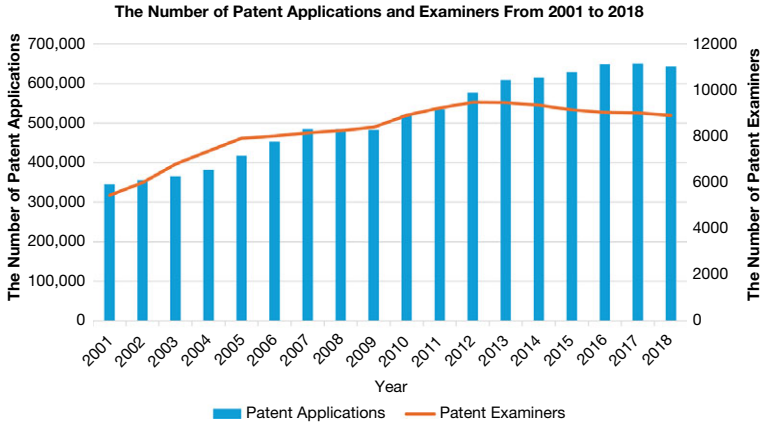
The patent system grants firms temporary monopoly rights over their inventions, providing crucial incentives for innovation and contributing to technological growth in the economy (see, e.g., Nordhaus (1969), Arrow (1972), and Mansfield (1986)). However, there has been considerable criticism of the U.S. patent system, alleging that it grants many low-quality patents through an inefficient screening process (see, e.g.,

I am grateful to Kai Li (the editor) and an anonymous referee for the constructive comments in the review process. I also thank Thomas Chemmanur, Francesco D'Acunto, Ran Duchin, Slava Fos, Edith Hotchkiss, Seung Jung Lee, Svetlana Petrova, Jonathan Reuter, Alberto Rossi, Yiming Qian, Ronnie Sadka, Philip Strahan, Wanli Zhao, and conference participants at the 2022 American Economic Association Annual Meeting, the 2020 Doctoral Consortium of the Financial Management Association Annual Meeting, the 2020 Southern Finance Association Annual Meeting, the 2021 Eastern Finance Association Annual Meeting, the 2021 Financial Management Association Annual Meeting, the 2022 Midwest Finance Association Annual Meeting, and seminar participants at Boston College, California State University at Fullerton, Chinese University of Hong Kong (Hong Kong & Shenzhen), City University of Hong Kong, Durham University, Fudan University, Georgetown University Global Virtual Seminar Series on Fintech, Luohan Academy, Oregon State University, Peking University, San Diego State University, Shanghai Jiaotong University, Shanghai University of Finance and Economics, University of Connecticut, and University of Florida for helpful comments. Any errors and omissions remain my own responsibility.

FIGURE 1

The Number of Patent Applications and Patent Examiners at the USPTO From 2001 to 2018

Figure 1 shows the number of patent applications and patent examiners at the USPTO from 2001 to 2018. Each blue bin represents the number of patent applications and the yellow line represents the number of patent examiners. Data source: Patent Statistics Chart and Patent Examination Data from the USPTO website.




Heller and Eisenberg (1998), Jaffe and Lerner (2011), Feng and Jaravel (2020), and Schankerman and Schuett (2022)). Critics argue that inefficient screening of patent applications reduces, instead of increasing, firms' incentives to innovate (see, e.g., Cornelli and Schankerman (1999), Lemley and Shapiro (2005), and Bessen and Maskin (2009)).

Several factors contribute to this problem. First, patent examiners at the U.S. Patent and Trademark Office (USPTO) face increasing time pressure. Patent applications surged from 345,732 in 2001 to 643,303 in 2018, as shown in Figure 1, but examiner numbers did not increase proportionally.[1] As a result, although examiners spend only about 19 hours per application, the average processing time is around 25 months (Frakes and Wasserman (2017)). Second, the USPTO struggles to recruit and retain top talent due to competition from the private sector (Jaffe and Lerner (2011)). Third, internal incentives at the USPTO tend to favor approvals over rejections; examiners are often evaluated based on the volume of applications processed, and approving an application typically requires less effort than rejecting it (Merges (1999), Frakes and Wasserman (2015)).

Motivated by these issues, this article explores whether combining human expertise with machine learning algorithms can improve the patent screening process. Machine learning's ability to quickly analyze large data sets could potentially alleviate time and resource constraints faced by examiners. Moreover, algorithms are not subject to human biases or agency problems like career or compensation concerns. Reflecting this potential, the USPTO has begun seeking help from artificial intelligence to enhance examination efficiency.[2]

[1]Data source: U.S. Patent Statistics Chart and Patent Examination Data from the USPTO website.

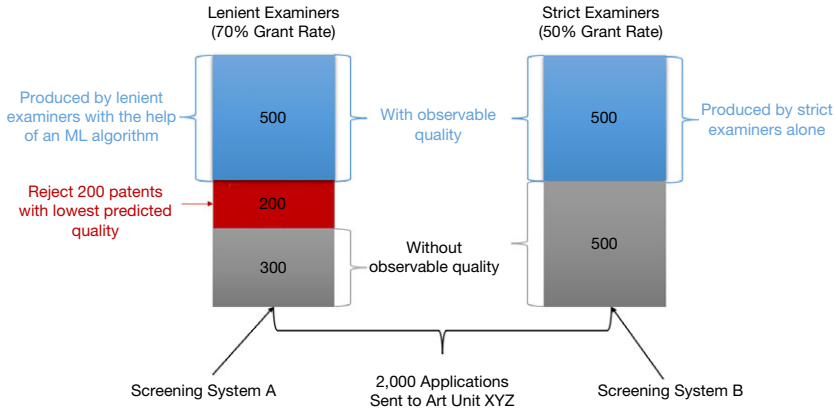
[2]USPTO director Andrei Iancu told the *Wall Street Journal*, "Our need is high and technology has advanced, so this is a good time to take advantage of these new tools to help our examiners." For the full news story, please see https://www.wsj.com/articles/patent-office-seeks-help-from-ai-11572297295.

https://doi.org/10.1017/S0022109025000213

FIGURE 2

An Illustrative Example of Using Examiner Leniency to Evaluate the Screening Performance of a Machine Learning Algorithm

Figure 2 provides an illustrative example of using examiner leniency to compare the performance of actual examiners and a machine learning algorithm.




Using detailed data on granted and rejected patent applications from the USPTO, I train a supervised machine learning algorithm (referred to as MLQuality) to predict patent quality based on application characteristics. These characteristics include textual features of patent claims, similarity measures to prior patents, backward citations, application originality, technological classes, and others. Patent quality is measured by forward citation counts, a standard measure in the literature (see, e.g., Trajtenberg, Henderson, and Jaffe (1997), Hall, Jaffe, and Trajtenberg (2005)). Out-of-sample predictions reveal that the current examination system grants many low-quality patents and rejects many high-quality ones within each art unit.[3]

Next, I test whether human examiners can do a better job with the help of Algorithm MLQuality. The main challenge here is the missing counterfactual: I do not observe actual quality information for applications rejected by humans but accepted by the algorithm. To address this selection issue, I exploit the quasi-random assignment of applications to examiners with different leniency levels within each art unit.[4] Following Kleinberg, Lakkaraju, Leskovec, Ludwig, and Mullainathan (2017), I divide examiners into two groups based on the median grant rate within each art unit, treating these two groups as two independent patent screening systems. Figure 2 illustrates this identification strategy using a hypothetical Art Unit XYZ, where lenient examiners approve 700 applications, and strict

[3]There are nine patent examining group centers, with each consisting of several art units examining patents in the relevant field.

[4]Because of this quasi-random assignment, I argue that the average quality of patent applications reviewed by examiners with different levels of leniency is similar. Several recent studies exploit this feature to make causal inferences (see, e.g., Maestas, Mullen, and Strand (2013), Sampat and Williams (2019), and Farre-Mensa, Hegde, and Ljungqvist (2020)).

examiners approve 500. In this example, I rank the 700 patents granted by lenient examiners based on predicted quality and use the algorithm to keep the 500 highest-quality patents. Then, I compare observable quality measures (e.g., forward citations) between these 500 patents granted by lenient examiners with the help of an algorithm and the 500 patents granted by strict examiners.

Intuitively, we would expect strict examiners to be the better examiners, given that they set a higher bar for approving patent applications. Nevertheless, lenient examiners, with the help of MLQuality, significantly outperform strict examiners, achieving a 26% increase in patent citations. This suggests that lenient (underperforming) examiners, when helped by the algorithm, can outperform strict (better-performing) examiners. Additionally, the analysis provides suggestive evidence on why human examiners may struggle with screening. Key predictors of patent citations include backward citations, text-based numerical vectors, similarity measures, and the originality of the application—factors that require considerable time and attention to assess properly.

Algorithm MLQuality assumes that the examiners' objective is to assess and grant higher-quality patent applications as measured by forward citation counts. To relax this assumption, I use examiners' past screening decisions as proxies for their objectives. Given the increasing time constraints on examiners, it is important to assess whether their screening decisions have deteriorated over time. To study this, I train another machine learning model, MLDecision, which maps patent application characteristics to examiners' past screening. I then apply MLDecision to predict screening decisions for more recent applications.

Evaluating MLDecision's performance faces the same missing counterfactual problem as MLQuality: the actual quality information for applications rejected by humans but accepted by the algorithm is not observable. To address this problem, I employ the same identification strategy as before. Starting with patents granted by lenient examiners in each art unit, I reject those with the highest predicted probability of rejection until the grant rate matches that of strict examiners. I then compare the quality of patents granted by lenient examiners assisted by MLDecision with those granted by strict examiners, both having observable citation counts and the same acceptance rate. This comparison reveals significant improvements in patent quality among recent applications in the out-of-sample test set: decisions made by lenient examiners assisted by MLDecision result in approximately a 17% increase in patent citations compared to those made by strict examiners.

Using the discrepancies between machine learning predictions and actual examiners' decisions, I further explore whether examiner-level characteristics, such as gender and workload, affect their screening performance. For each art unit and year, I calculate the number of applications accepted by examiners and rank all applications based on predicted citations (Algorithm MLQuality) or acceptance probabilities (Algorithm MLDecision). I then hypothetically accept the same number of applications as the examiners did, according to this ranking. Each application thus has two decisions: the examiner's actual decision and the algorithm's hypothetical decision. I define a "false accept" when the examiner accepts an application that the algorithm rejects and a "false reject" when the examiner rejects an application that the algorithm accepts. I find that busy examiners make more mistakes of both types, suggesting time constraints may reduce screening performance. More

experienced examiners make fewer false rejections but more false acceptances, possibly due to compensation incentives favoring acceptance. Lastly, I find no evidence that male examiners make more mistakes.

The findings so far suggest that human + machine could potentially outperform humans alone in granting higher-quality patents. Could human + machine also improve economic outcomes for firms applying for patents? To investigate this, I calculate the false acceptance rate of each patent examiner by making use of the disagreement in screening decisions between human examiners and Algorithm MLQuality (or MLDecision) among earlier applications. The empirical findings are summarized as follows: First, falsely accepted patents are more likely to expire early. Second, public firms holding patents granted by such examiners are more likely to have worse operating performance than other patent grantees. Third, affected private firms are less likely to exit successfully through initial public offerings (IPOs) or mergers and acquisitions (M&As). The above effects are economically significant. For example, if either Algorithm MLQuality or MLDecision were used to help examiners screen patent applications, public firms' annual return on assets (ROA) would increase by 35 to 48 basis points (bps), and private firms' probability of going public or being acquired in 3 years would increase by 0.9 to 1.6 percentage points. The above results are likely to be causal since patent applications are randomly assigned to patent examiners whose characteristics are unlikely to be correlated with firm characteristics.

The rest of the article is organized as follows: Section II discusses the relation of my article to the existing literature. Section III discusses the institutional background of the patent examination process. Section IV describes the patent application data and sample statistics. Section V discusses the empirical design and results of the machine learning analysis. Section VI describes the firm-level data and discusses the empirical analysis of firm performance. Section VII concludes.

## II. Relation to the Existing Literature

My article is related to several different strands of the literature. The first strand is the theoretical and legal literature that explores the question of how to improve the patent screening process by reforming the patent system itself. For example, Dreyfuss (2008) argues that the patent system systematically creates type II errors (i.e., erroneous grants) due to the resource constraints faced by patent examiners and the incentive structure at the USPTO. Dreyfuss (2008) proposes to increase the nonobviousness threshold in order to reduce the incidence of type II error (see also, e.g., Duffy (2008), Eisenberg (2008), and Mandel (2008)). In a similar vein, Schankerman and Schuett (2022) theoretically show that almost half of all innovations granted patents would be produced even without patent incentives and argue that the social value of the patent system would be larger if antitrust limits on licensing were implemented. On the other hand, Scherer (1972), and several other theoretical papers focus on reforming the optimal patent right (i.e., patent length and breadth) to improve innovation incentives and quality (see, e.g., Gilbert and Shapiro (1990), Matutes, Regibeau, and Rockett (1996)). Finally, a set of related papers studies the cost and benefit of the patent litigation system in affecting patent validity and scope (see, e.g., Meurer (1989), Choi (1998), Lanjouw and

Schankerman (2001), and Bessen and Meurer (2006)). This article presents another exercise in using machine learning to evaluate the performance of the current patent screening system and provides further discussions on improving the current system under resource constraints.

The second strand of related literature applies machine learning techniques to economics and finance research. For example, Athey and Imbens (2017) argue that although it has not yet been widely utilized in social science research, supervised machine learning has great potential for prediction problems. Several studies apply machine learning to issues in finance: for example, measuring asset risk premia (Gu, Kelly, and Xiu (2020)), predicting stock returns (Rossi (2018)), classifying fund types (Abis (2017)), and selecting boards of directors (Erel, Stern, Tan, and Weisbach (2021)). However, there are also challenges in applying machine learning in social science research. Kleinberg et al. (2017) use New York judges' decisions over bail cases as a setting to discuss unique potential endogeneity problems in applications of machine learning in social science and provide methodologies to address these problems using econometric identification schemes.[5] Several recent papers also discuss the potential benefits and challenges of using machine learning in patent examinations. Krishna, Feldman, Wolf, Gabel, Beliveau, and Beach (2016) developed an automated prior art search system and found that it reduces patent examiners' search effort. On the other hand, Choudhury, Starr, and Agarwal (2020) find that algorithms such as automated prior art searching systems may be biased from input incompleteness and argue that human expertise is complementary to machine learning in mitigating bias. Overall, my article complements the above literature by showing the potential efficiency gains from combining human expertise with machine predictions.[6]

Finally, my article also contributes to the empirical literature that studies the relationship between innovation, firm operating performance, and stock market performance. For example, Chemmanur, Gupta, and Simonyan (2022) show that private firms with a large number of patents and citations per patent have higher IPO valuations and future operating performance (see, e.g., Bowen III, Frésard, and Hoberg (2023)). In terms of stock market performance, Hall et al. (2005) empirically document that a larger number of citations per patent leads to higher market values for the firms holding the patents (see also, e.g., Zucker, Darby, and Armstrong (2002), Eberhart, Maxwell, and Siddique (2004)). Kogan, Papanikolaou, Seru, and Stoffman (2017) measure the economic value of a patent with the stock price announcement effect of the patent grant and study its relationship with aggregate economic growth and total factor productivity (TFP). Alternatively, Cohen, Diether, and Malloy (2013) show that the stock market does not take firms' past successes in innovation into consideration when valuing their future innovation. Fitzgerald, Balsmeier, Fleming, and Manso (2021) show that firms with exploitation innovation strategies are undervalued relative to firms with exploration

[5]See also Kleinberg, Ludwig, Mullainathan, and Obermeyer (2015), and Mullainathan and Spiess (2017) for detailed discussions on how to use machine learning as an applied econometrics tool.

[6]In addition, my paper discusses and addresses the forward-looking issue existing in both the training data and the training algorithm itself. For example, when the algorithm is trained with future information, it will result in unfair comparisons between humans and machines in the test set since humans in the test set are not able to access this future information.

innovation strategies. Some recent papers also study the effect of examiner characteristics and decisions on firm performance. For example, Kline, Petkova, Williams, and Zidar (2019) estimate the ex ante value of accepted and rejected patent applications and study the relationship between patent-induced shocks and labor productivity. Shu, Tian, and Zhan (2022) test whether the workload of each patent examiner can predict firms' future stock market returns and show that investors underreact to the negative effect of examiner workload on patent quality. My article complements the above literature by showing the potential economic gains in firm performance from combining human with machine intelligence in patent examinations.

## III. Patent Examination Process and Patentability

### A. Patent Examination Process

The patent examination process starts with the filing of a patent application with the USPTO; the USPTO then forwards the newly filed application to a relevant art unit for examination. Next, that patent application is assigned to a patent examiner, a specialized technology employee with training and experience pertinent to the invention, for examination. Though there are no explicit policies regarding how patent applications are assigned, many recent studies show that patent applications are randomly assigned to examiners within each art unit: applications filed first are assigned to the first available examiners (see, e.g., Maestas et al. (2013), Farre-Mensa et al. (2020), and Sampat and Williams (2019)).

After receiving a patent application, examiners first evaluate the claimed invention against the existing state of knowledge in the "prior art," consisting of patent documents and scientific and commercial literature, to determine whether the invention satisfies legal requirements for patentability. If an invention fails the patentability requirement, the examiner issues an office action rejecting that application for unpatentability and explains the reasons for the rejection. Following such a rejection, the inventor may revise the application and submit it again or withdraw it. My article focuses only on the first round of all regular nonprovisional utility patent applications to mitigate the concern that subsequent applications may not be randomly assigned (Righi and Simcoe (2019)).

### B. Legal Requirements for Patentability

The Patent and Copyright Clause of the Constitution (Article I, Section 8, Clause 8) grants Congress the power "to promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries." To fulfill this mandate, the U.S. Patent Act (35 U.S. Code §101) sets the requirements for patent protection as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvements thereof, may obtain a patent, subject to the conditions and requirements of this title.

Under the U.S. Patent Act, an invention is patentable if it satisfies the following three criteria: novelty, usefulness, and nonobviousness. Specifically, the novelty

requirement (35 U.S. Code §102) states that an invention cannot be patented if the invention has been publicly disclosed before the applicant filed for patent protection, and the usefulness requirement states that the subject matter must be useful. Usually, a patent application can easily pass both the novelty and usefulness requirements. However, the nonobviousness requirement (35 U.S. Code §103), according to which the invention must represent a nonobvious improvement over the prior art, is an ambiguous threshold that has attracted much criticism from the law literature alleging that it leads to the approval of many low-quality patents (see, e.g., Duffy (2008), Dreyfuss (2008), Eisenberg (2008), and Mandel (2008)).

Since the goal of the U.S. Patent Act is to reward patent applicants for providing the public with new discoveries with a limited exclusive right to their invention, I argue that, at the very minimum, the system should not discard high-quality applications in favor of low-quality applications. For example, the USPTO itself states in its 2018–2022 strategic plan that the most important goal for the office is to continue optimizing patent quality.[7] To capture this goal, I first trained a machine learning algorithm by assuming the objective of examiners is to screen against patent quality. Nevertheless, the above assumption may fail to capture other aspects of the examiners' objective. To relax this assumption, I also trained a machine learning algorithm based on the screening decisions made by examiners for earlier patent applications.

### C. Measuring Patent Quality

The existing literature measures patent quality using either the scientific or the economic value of a patent. A patent's scientific value is usually constructed based on the number of citations that a patent receives after it is granted (see, e.g., Trajtenberg (1990), Trajtenberg et al. (1997), and Hall et al. (2005)), while its economic value is constructed based on the announcement return from the patent grant news to the patent owner (see, e.g., Kogan et al. (2017)). Since the economic value of a patent can be measured only if the patent is owned by a public firm, I use citation-based measures as my measures of patent quality. More specifically, I use patent forward citation as my primary quality measure: the number of citations of a patent in the 4 years after the patent is granted.

## IV. Patent Application Data and Sample Selection

### A. Patent Application Data

I collect data on patent applications from the USPTO website, which provides various research data sets.[8] In particular, I collect patent application examination data from the Patent Examination Research Dataset (Graham, Marco, and Miller (2018), Marco, Toole, Miller, and Frumkin (2017)), the patent application claims data from the Patent Claims Research Dataset (Marco, Sarnoff, and Charles (2019))

[7]For the full USPTO 2018–2022 strategic plan, please see https://www.uspto.gov/sites/default/files/documents/USPTO_2018-2022_Strategic_Plan.pdf.

[8]For a complete list of research data sets provided by the USPTO, please see https://www.uspto.gov/ip-policy/economic-research/research-datasets.

and PatentsView, patent application citation data from the Office Action Research Dataset for Patents (Lu, Myers, and Beliveau (2017)) and PatentsView, and patent assignment data from the Patent Assignment Dataset (Marco, Myers, Graham, D'Agostino, and Apple (2015b)).[9]

### 1. Turning Patent Claims Text into Numerical Variables

The claim section in each patent application defines the extent of the protection sought. A typical patent contains several claims, each representing an original contribution, which can be considered a good measure of the actual invention (Tong and Frame (1994)). If the claims in a patent application are very similar or are close to the claims in earlier patents, I would expect this patent application's quality (innovativeness) to be low. To capture the similarity of each patent application filed in a given technological class to all prior patents in that technological class, I convert claim texts into a vector of 50 dimensions through word embedding and create similarity-based novelty measures based on application claims in a similar spirit of Arts, Cassiman, and Gomez (2018).

First, I compiled the claims texts from each patent application in the training sample, as well as from all patents granted before 2001 within the same technological class, to create one corpus for each technological class. Using the given corpus in each technological class, I produced a 50-dimensional vector from the claims text with the *Word2vec* algorithm.[10] Regarding technological classes, the USPTO has developed its own U.S. Patent Classification (USPC) system, which includes over 450 unique classes and 150,000 subclasses. However, the USPC classes do not directly correspond to established product and industry classifications (Marco, Carley, Jackson, and Myers (2015a)). Hall, Jaffe, and Trajtenberg (2001) developed a hierarchical classification (the NBER classification) by aggregating USPC classes into 37 (2-digit) subcategories.[11]

Next, I calculated the pairwise cosine similarity between patent application $i$ and patent $j$, where the patent $j$ was granted before the filing date of application $i$ but within the same technological class: $Sim\left(P_i,P_j\right)=\frac{P_iP_j}{|P_i||P_j|}$. $P_i$ and $P_j$ are the vectors of 50 dimensions of patent application $i$ and patent $j$. For each new application $i$ in the training data, I computed the pairwise cosine similarity between it and each prior patent granted within the same technological class.

Finally, I constructed three sets of similarity measures based on pairwise cosine similarities between each focal application and prior patents within the same

[9]Public PAIR (Patent Application Information Retrieval) data were available from the USPTO website until recently. Though not all patent applications received by the USPTO were included in Public PAIR, the database included more than 83% of all patent applications granted after the implementation of the American Inventors Protection Act (AIPA) in late 2000. For the regular utility patent applications on which this paper focuses, the coverage in Public PAIR from 2001 increases to 95% as a consequence of the AIPA, according to Graham et al. (2018).

[10]The *Word2vec* algorithm learns vector representations of words from the input text corpus and places words that share a similar context in the corpus in close proximity to one another in the vector space, where the vector space is set to 50 dimensions (see, e.g., Mikolov, Chen, Corrado, and Dean (2013a), Mikolov, Le, and Sutskever (2013b), and Mikolov, Yih, and Zweig (2013c) for details).

[11]The NBER classification comes from the NBER Patent Data Project: https://sites.google.com/site/patentdataproject.

technological class. Specifically, for each focal application, I calculated the average, maximum, and minimum cosine similarities with patents in four groups: the 10 most recent patents, the 100 most recent patents, the 1,000 most recent patents, and all available prior patents. This resulted in a total of twelve similarity measures —four average similarities, four maximum similarities, and four minimum similarities—each corresponding to one of the four groups of prior patents.

I used the 50-dimensional vector, twelve similarity measures, and numerical statistics of claims as input variables, along with other patent application characteristics discussed in a later section, to train the Algorithm MLQuality or MLDecision.

To construct the 50-dimensional vector and its derived similarity measures for applications in the test sample, I used all claims text in each patent application in the test sample and all patents granted prior to 2010 as a new corpus and repeated the above steps. This ensures that the trained algorithm never uses any information from the test sample, avoiding potential temporal leakage.

### B. Summary Statistics

Table 1 reports summary statistics for all patent applications of the numerical variables used in my machine-learning prediction. Out of 638,159 applications, 434,960 (68.2%) are approved: the average number of 4-year forward citations is 1.886. In terms of numerical statistics of claims, each patent application, on average, has 2.791 independent claims and 15.527 dependent claims, where the average length of an independent claim (around 138 words) tends to be longer than that of a dependent claim (around 42 words). I also compute the originality index for each patent application: $Originality_i = 1 - \sum_j^{n_i} s_{ij}^2$, where $s_{ij}$ denotes the fraction of backward citations made by application $i$ in patent class $j$ from the total number of patent classes $n_i$ and $\sum_j^{n_i} s_{ij}^2$ is the Herfindahl–Hirschman index (Hirschman (1980)). By definition, the originality index captures the technological dispersion of prior patents utilized in patent application $i$. The average number of backward patent citations and originality index are 8.505 and 0.165, respectively. In addition to citing prior patents, a patent application may cite previous applications, scientific literature, and foreign patents. The average numbers of backward citations of patent applications, citations of scientific literature, and citations of foreign patents are 2.752, 3.835, and 2.903, respectively.[12] In addition to patent application characteristics, 26.9% of patent applications are submitted by small entities, and 43.9% of primary inventors are from the U.S.

## V. Machine Learning Prediction Design and Results

The empirical design to analyze the efficiency of the patent screening process follows 3 steps (Kleinberg et al. (2017)). First, I partition my sample into training and test sets, as described in Section V.A. Second, I separately train the two algorithms using the training set by mapping a patent application's characteristics to its quality (Algorithm MLQuality) and its screening outcome (Algorithm MLDecision).

[12]I exclude citations made by examiners when I count backward citations for each patent application.

https://doi.org/10.1017/S0022109025000213

TABLE 1

Summary Statistics (Patent Applications)

Table 1 shows descriptive statistics for the sample of patent applications from 2001 to 2013 used in my machine learning analysis. *ForwardCitations* counts the number of future citations that each patent has received over a 4-year period after it is granted. *NumberIndepClaims* and *NumberDepClaims* count the number of independent claims and dependent claims for each patent application. *NumberWordsIndepClaims* and *NumberWordsDepClaims* count the total number of words in independent claims and dependent claims for each patent application. *MinNumberWordsIndepClaims* and *MinNumberWordsDepClaims* count the minimum number of words in independent claims and dependent claims for each patent application. *AvgNumberWordsIndepClaims* and *AvgNumberWordsDepClaims* count the average number of words per independent claim and per dependent claim for each patent application. *NumberCitedForeignPatents* counts the number of foreign patents that each patent application has cited. *NumberCitedLiterature* counts the number of scientific literature that each patent application has cited. *NumberCitedApplications* counts the number of patent applications that each patent application has cited. *OriginalityApplication* captures the industry dispersion of backward cited patent applications that each patent application has made, which equals to 1 minus the Herfindahl–Hirschman index of industries that cited patent applications belong to. *NumberCitedPatents* counts the number of patents that each patent application has cited. *OriginalityPatent* captures the industry dispersion of backward cited patents that each patent application has made, which equals to 1 minus the Herfindahl–Hirschman index of industries that cited patents belong to. *USInventorDummy* is a dummy variable indicating whether an investor is from the U.S. or not. *SmallEntityDummy* is a dummy variable indicating whether a patent application is from a small entity or not.

| | N | Mean | Median | p10 | p90 | Std. Dev. |
|---|---|---|---|---|---|---|
| *Panel A. Patent Application Quality Variables* | | | | | | |
| ForwardCitations | 434,960 | 1.886 | 1 | 0 | 4 | 5.242 |
| *Panel B. Patent Application Characteristics* | | | | | | |
| SmallEntityDummy | 638,159 | 0.269 | 0 | 0 | 1 | 0.444 |
| NumberClaims | 638,159 | 2.791 | 2 | 1 | 5 | 2.545 |
| NumberWordsInClaims | 638,159 | 361.526 | 258 | 85 | 695 | 499.204 |
| MinNumberWordsInClaims | 638,159 | 115.744 | 92 | 32 | 210 | 130.599 |
| NumberDepClaims | 638,159 | 15.527 | 14 | 4 | 27 | 13.436 |
| NumberWordsInDepClaims | 638,159 | 601.319 | 475 | 135 | 1135 | 879.589 |
| MinNumberWordsInDepClaims | 638,159 | 21.841 | 17 | 11 | 30 | 64.494 |
| AvgNumberWordsInClaims | 638,159 | 138.196 | 114 | 51.500 | 235.333 | 136.412 |
| AvgNumberWordsInDepClaims | 638,159 | 42.356 | 34.125 | 20.875 | 64.500 | 69.755 |
| NumberCitedForeignPatents | 638,159 | 2.903 | 0 | 0 | 7 | 10.674 |
| NumberCitedLiterature | 638,159 | 3.835 | 0 | 0 | 6 | 22.242 |
| NumberCitedApplications | 638,159 | 2.752 | 0 | 0 | 5 | 15.437 |
| OriginalityApplication | 638,159 | 0.155 | 0 | 0 | 0.769 | 0.309 |
| NumberCitedPatents | 638,159 | 8.505 | 0 | 0 | 18 | 35.568 |
| OriginalityPatent | 638,159 | 0.165 | 0 | 0 | 0.618 | 0.259 |
| USInventorDummy | 638,159 | 0.439 | 0 | 0 | 1 | 0.496 |
| AvgSimilarity (All prior patents) | 638,159 | 0.629 | 0.647 | 0.478 | 0.756 | 0.113 |
| MaxSimilarity (All prior patents) | 638,159 | 0.959 | 0.964 | 0.934 | 0.981 | 0.025 |
| MinSimilarity (All prior patents) | 638,159 | 0.048 | 0.054 | −0.080 | 0.167 | 0.099 |
| AvgSimilarity (Prior 10 patents) | 638,159 | 0.613 | 0.632 | 0.435 | 0.767 | 0.132 |
| MaxSimilarity (Prior 10 patents) | 638,159 | 0.797 | 0.825 | 0.640 | 0.914 | 0.118 |
| MinSimilarity (Prior 10 patents) | 638,159 | 0.403 | 0.409 | 0.179 | 0.619 | 0.167 |
| AvgSimilarity (Prior 100 patents) | 638,159 | 0.626 | 0.644 | 0.475 | 0.751 | 0.112 |
| MaxSimilarity (Prior 100 patents) | 638,159 | 0.895 | 0.907 | 0.826 | 0.951 | 0.058 |
| MinSimilarity (Prior 100 patents) | 638,159 | 0.269 | 0.273 | 0.088 | 0.444 | 0.139 |
| AvgSimilarity (Prior 1,000 patents) | 638,159 | 0.629 | 0.647 | 0.478 | 0.753 | 0.110 |
| MaxSimilarity (Prior 1,000 patents) | 638,159 | 0.932 | 0.939 | 0.891 | 0.967 | 0.035 |
| MinSimilarity (Prior 1,000 patents) | 638,159 | 0.174 | 0.178 | 0.017 | 0.328 | 0.122 |

I present the results in Section V.B. Third, I evaluate the prediction accuracy of these two algorithms using patent applications in the out-of-sample test set and present the results in Section V.C. Last, I test whether these two algorithms can improve the screening decisions of actual patent examiners by comparing the predicted decisions to those of patent examiners and presenting relevant results in Section V.D.

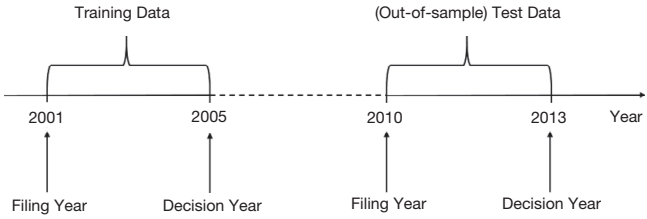
## A. Sample Partition

I use the unique application number to merge across different data sets and obtain an initial sample of 3,473,251 patent applications that have screening

FIGURE 3

Training and Testing Data Used for My Machine Learning Prediction

Figure 3 shows the partition for the training and test data used for my machine-learning prediction. I select applications filed from 2001 to 2005 with screening status available before the beginning of 2006 into the training set, and applications filed from 2010 to 2013 with screening status available before the beginning of 2014 into the test set. The training set is used to form the algorithm for my prediction and the test set is used to evaluate all of my results. The final sample used in my machine learning prediction consists of 280,243 patent applications in the training set and 357,101 patent applications in the test set.




outcomes available (i.e., the patent was either granted or rejected) and that were filed at the USPTO from 2001 to 2014. When I train a machine learning algorithm to compare its predictions with human decisions, I have to make sure the data used to train the algorithm is ex ante available for actual examiners in the test set in order to make fair comparisons. In my setting, I use patent application characteristics and patent outcomes of earlier applications in the training set. Since my outcome variable for training the algorithm, the 4-year forward citations, and available only 4 years after each application is granted, I set a 4-year gap between the training and test sample periods. Specifically, I use applications filed from 2001 to 2005 whose screening status is available before 2006 for the training sample to train the machine learning algorithm and applications filed from 2010 to 2013 whose status is available before 2014 for the test sample to evaluate the algorithm.[13]

When I partition my sample in this way, both my trained machine learning algorithm and the patent application quality measure in the training sample are available for the period from the beginning of 2010. In other words, the information needed to train the algorithm is also available for patent examiners in the test sample. This sample partition approach allows me to make a fair comparison of the decisions of the algorithm and the actual examiners in terms of screening any patent application in the test set. Figure 3 presents the sample partition along the timeline. The final sample in my machine learning prediction consists of 280,690 patent applications in the training set and 357,471 patent applications in the test set.

## B. Training Machine Learning Algorithms

To train Algorithm MLQuality, I need both input variables of patent application characteristics and an output variable of patent application quality from applications in the training data: the output variable $y$ is the 4-year forward citations of

[13]I also partition my sample in alternative ways: by partitioning the whole sample randomly into a training sample and a test sample and by partitioning the whole sample over time but without the 4-year gap. Though these alternatively partitioned samples are subject to the concerns raised in this section, the results based on these alternative partitioning methods are similar to the main findings in this article.

each patent described in Section III.C, and the input variables $X$ include the numerical statistics of the claims texts described in Section IV.B; the text-based numerical vector of claims; backward citations of prior patents, patent applications, foreign patents, and scientific literature; filing year dummies; an inventor nationality dummy; a small entity dummy; 37 NBER classes dummies; and 597 art unit dummies. As I mentioned earlier in Section V.A, my training set consists of 280,690 patent applications, including 81,428 rejected applications and 199,262 accepted applications. Since the rejected applications do not have information on forward citations, the 199,262 accepted applications with an available forward citation are used for training Algorithm MLQuality.

I train the prediction function called extreme gradient boosting, an ensemble method of decision trees based on tree boosting.[14] A decision tree is a tree-like prediction function that can be trained by splitting the data set into subsets based on particular values of input variables, where the process is repeated until splitting no longer adds value to the predictions (see, e.g., Rokach and Maimon (2008)). Since a single decision tree may produce a weak learning function subject to noise, gradient boosting algorithms optimize a cost function by iteratively choosing a weak learning function that follows the negative gradient direction to produce a strong learning function (see, e.g., Friedman (2001), Chen and Guestrin (2016)). The strength of an extreme gradient boosting algorithm is finding the best feature across different subsamples. In addition, I implement 5-fold cross-validation when training the algorithm to alleviate the in-sample over-fitting problem.

Figure 4 presents the important features identified by Algorithm MLQuality in predicting patent citation. Feature importance is determined by calculating the relative contribution of each feature to the model's predictive power: whether that feature was selected to split on during the tree-building process, and how much the overall squared error decreased. The most important feature, backward citations made by patent applications, accounts for 28.6% of the total predictive power of the algorithm. Additionally, a set of 50-dimensional vectors collectively explain 22.2% of the model's predictive power. Several art units also play a significant role in citation forecasting. For example, Art Unit 3672, which focuses on mining and earth engineering, is responsible for 3.1% of the algorithm's predictive power. This is followed by Art Unit 2829, which focuses on semiconductors and memory, contributing 1.0%. Furthermore, various technological classes significantly contribute to prediction accuracy: Information Storage (NBER Classification 24) accounts for 1.9% of the total predictive power. This is complemented by contributions from Organic Compounds (NBER Classification 14), Drugs (NBER Classification 31), and Semiconductor Devices (NBER Classification 46), each contributing over 1%. Two similarity measures derived from the 50-dimensional vectors—Maximum similarity (with respect to the prior 1,000 patents) and Minimum similarity (with respect to the prior 100 patents)—each contribute 0.8% to the predictive power. Other important features include patent application originality and the number of

[14]Section IA.1 of the Supplementary Material provides a detailed discussion of the supervised machine learning problem and the extreme gradient boosting algorithm.

cited scientific literature, cited foreign patents, cited patent applications, claims, and words in claims. Interestingly, whether an inventor is based in the U.S. also explains 1.8% of the predictive power, possibly reflecting a language advantage for U.S. inventors.[15]

To train the Algorithm MLDecision, I followed the same procedure as above and replaced the output variable $y$ with the screening decision made by the actual examiner. Since both accepted and rejected applications have information on their screening outcomes, all 280,690 patent applications in the training sample are used for the training Algorithm MLDecision.

## C. Evaluating the Out-of-Sample Prediction Performance of Machine Learning Algorithms

In this section, I present the out-of-sample prediction performance of the machine learning model. The out-of-sample root mean square error (RMSE) of Algorithm MLQuality is 3.3. Since RMSE correlates with the magnitude of the outcome variable, it is challenging to assess whether an RMSE of 3.3 is small or large. To provide context, predicting the mean citations for all patents in the out-of-sample test set yields an RMSE of 6.4, while predicting the mean citations for all patents within each art unit results in an RMSE of 6.2. Additionally, to visualize the predictive accuracy, Figure IA.1 in the Supplementary Material plots the predicted citations from Algorithm MLQuality against the actual citations of granted patents. The majority of data points cluster around the 45-degree line, indicating high accuracy in the out-of-sample predictions.

Evaluating the out-of-sample prediction performance for Algorithm MLDecision is less straightforward. If Algorithm MLDecision predicts examiner decisions perfectly in the out-of-sample test set, it has no use in terms of improving examiner decisions since it agrees perfectly with examiners. To that extent, I report the differences between the predicted decisions and the actual decisions for applications in the out-of-sample test set. The out-of-sample RMSE of Algorithm MLDecision is 0.29. For comparison, predicting the mean grant rates for all applications in the out-of-sample test set yields an RMSE of 0.47, while predicting the mean grant rates for all applications within each art unit results in an RMSE of 0.45. Two possibilities may explain the discrepancies between predicted and actual decisions: i) Algorithm MLDecision performs poorly in out-of-sample prediction. If this is the case, we would expect the quality of the predicted accepted applications to be worse than that of the actual accepted applications. ii) Examiners are performing worse over time, possibly due to the time and resource constraints discussed earlier. If this is the case, we would expect the quality of the predicted accepted applications to be better than that of the actual accepted applications. Section V.D.2 provides tests for these two possibilities.

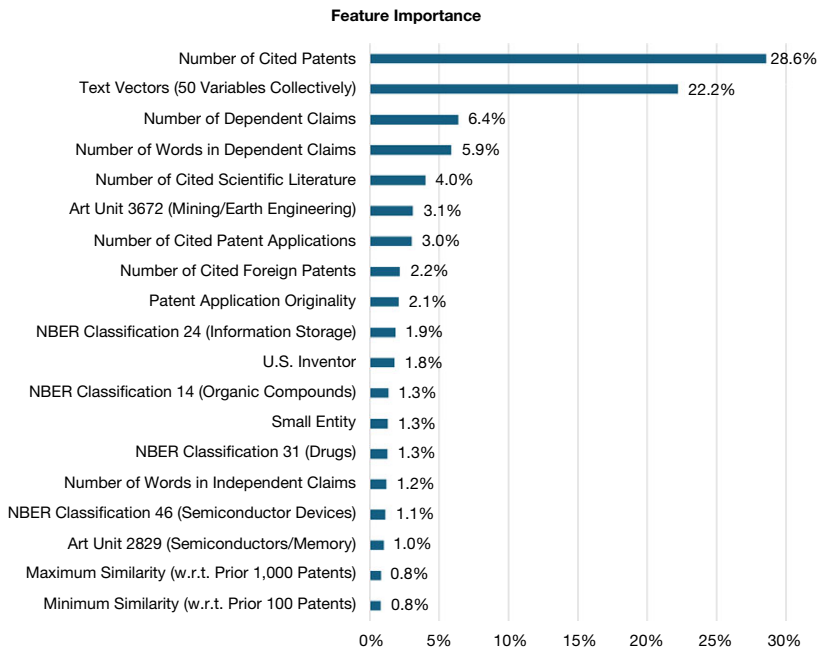
[15]Figures IA.2 and IA.3 in the Supplementary Material plot the top 10 important features based on the mean absolute Shapley (SHAP) values and the distribution of SHAP values for these 10 features (Shapley (1953)). The top important features based on the mean absolute SHAP value highly overlap with those identified in Figure 4.

https://doi.org/10.1017/S0022109025000213

FIGURE 4

Important Features Identified by Algorithm MLQuality

Figure 4 lists the 20 most important features identified by Algorithm MLQuality. The predictive power of each feature measured as the percentage of total predictive power is on the *x*-axis. The name of each of the features is on the *y*-axis.




## D. Improving Screening Decisions with the Help of a Machine Learning Algorithm

### 1. Do Examiners Accept Low-Quality Patents?

To answer the question of how often low-quality patents pass examinations, I examine the grant rates of actual examiners across patent applications of different predicted citations. To visualize the results, I divide the patent applications in the test set equally into 1,000 bins based on the patents' predicted citation and compute the grant rates of patent applications for actual examiners in each of these 1,000 bins. Figure 5 plots the correlation between the grant rates of actual examiners and the average predicted citation of patent applications in each bin. I find that while examiners do indeed reject more applications in bins with lower quality, they still accept around 50% of the patents in the bins with the lowest quality. These patents, on average, receive 0 citations going forward and have very limited social values in terms of promoting science and useful arts.

### 2. Using Variation in Examiner Leniency to Quantify the Improvement to Screening Decisions from Machine Learning Algorithms

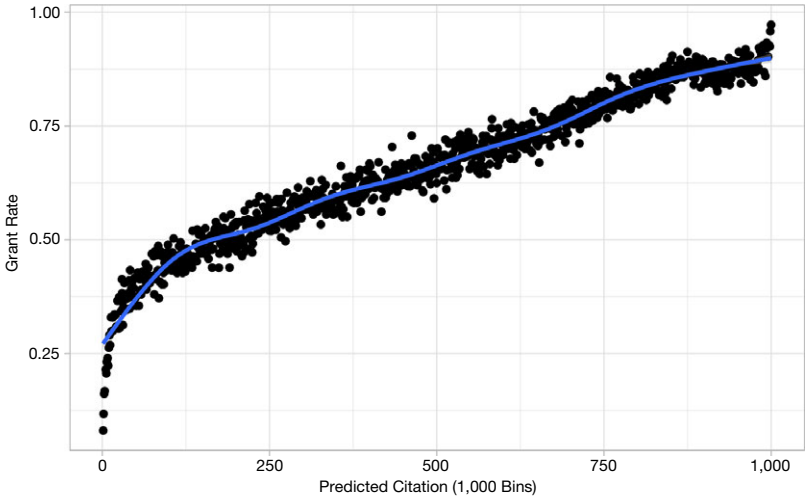
One way to quantify the potential quality gain achieved by algorithms is to rank all patent applications within the same art unit based on my predicted citation/

https://doi.org/10.1017/S0022109025000213

FIGURE 5

The Relation Between Predicted Citation by Algorithm MLQuality and Actual Examiner Grant Decisions

Figure 5 shows the relation between predicted citations by Algorithm MLQuality and actual examiner grant decisions. The rank of the average predicted citation of all patent applications in each pin is on the *x*-axis. The grant rate is on the *y*-axis.




grant probability and then set the grant rate of an algorithm to be the same as that of examiners. I can then compare the average citation of all patent applications granted by an algorithm to the average citation of the granted patents. However, measuring the improvement in this way may be misleading since I do not have information on the actual citations of the patent applications rejected by examiners but approved by an algorithm. To address this issue, I make use of the fact that patent applications are randomly assigned to examiners with different grant rates within the same art unit: lenient examiners (i.e., those with an above-median grant rate) accept approximately 77.6% of patent applications and strict examiners accept 49.5% of patent applications. Thus, given all patents granted by lenient examiners in a given art unit, I can reject additional applications based on predicted citation/grant probability to match the grant rate of strict examiners within that art unit (i.e., those with a below-median grant rate). Then, I can compare the average actual citation of patents granted by the algorithm to those granted by strict examiners.

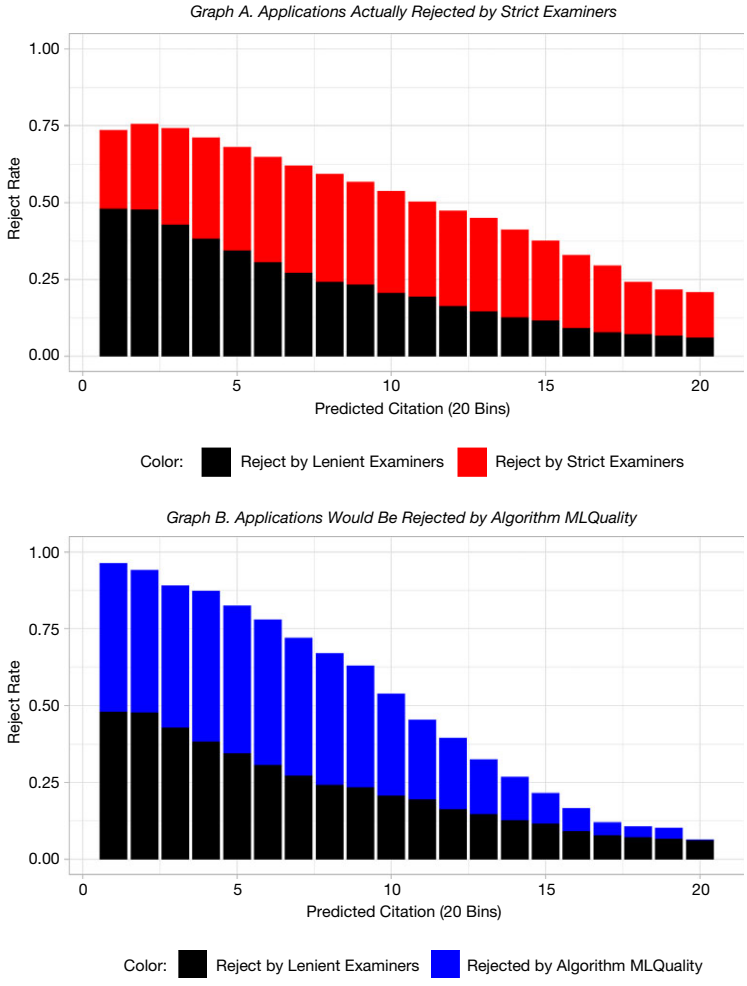
More importantly, comparing the decisions of examiners with different levels of leniency allows me to track the quality (citation) of marginal applications that are rejected. Figure 6 shows the results of these comparisons based on Algorithm MLQuality. I sort patent applications by predicted citation and divide them into 20 bins. At the bottom of a given bin, the black bar shows the fraction of patent applications rejected by lenient examiners. The red bar on top of the black bar in a given bin shows the fraction of additional applications rejected by strict examiners, while the blue bar on top of the black bar in a given bin shows the share of additional applications that were rejected by Algorithm MLQuality. Graph A of Figure 6 shows that strict examiners would reject additional applications from patent

https://doi.org/10.1017/S0022109025000213

FIGURE 6

Comparison Between Applications Rejected by Strict Examiners and Applications Rejected by Algorithm MLQuality

Figure 6 compares applications rejected by strict examiners and those rejected by Algorithm MLQuality within each art unit. Strict (lenient) examiners are defined as those with an above (below) median reject rate in each art unit. I divide patent applications in the test set into 20 bins by the predicted number of forward citations (*x*-axis). In both graphs, the black bar at the bottom of each bin shows the fraction of patent applications rejected by lenient examiners. The red bar in Graph A shows which applications strict examiners actually reject. The blue bar in Graph B shows which applications Algorithm MLQuality would reject to match the grant rate of strict examiners.




applications in both the low- and high-quality bins. However, Graph B of Figure 6 shows that Algorithm MLQuality would reject additional applications starting from the lowest quality of predicted citation.

Next, I quantify the quality gain under the above exercise by comparing the actual outcome of strict examiner decisions to the decisions made by lenient examiners + Algorithm MLQuality. On average, patents approved by strict examiners receive 1.79 citations over 4 years, whereas those approved by lenient

examiners + Algorithm MLQuality average 2.55 citations. Excluding patents that received no citations, the average for patents reviewed by strict examiners increases to 3.53 but still trails behind the 4.47 average for patents granted by lenient examiners with Algorithm MLQuality. These comparisons indicate a citation increase ranging from 26% to 42%.

I also employ the same identification strategy to quantify the potential quality gain achieved by Algorithm MLDecision. Figure 7 compares the decisions of strict examiners to those made by lenient examines and Algorithm MLDecision, which shows similar results as above. Quantitatively, the combination of lenient examiner decisions and Algorithm MLDecision results in an average of 2.24 citations per patent – or 4.15 for patents excluding those with 0 citations. This represents an increase of 25% in overall citations and 17% among patents receiving citations, compared to those evaluated by strict examiners alone.

### 3. Do Examiners' Characteristics Affect Their Performance?

To explore whether examiner characteristics are related to their screening performance, I measure examiners' screening performance based on whether there is disagreement between machine learning predictions and actual patent examiners' decisions. Specifically, I first calculate the number of applications accepted by examiners within each art unit for a given year. Next, I rank all patent applications (both accepted and rejected applications) within each art unit filed in that same year based on the predicted citation by Algorithm MLQuality (or predicted acceptance probability by Algorithm MLDecision). I then hypothetically accepted the same number of applications as the examiners did in that year. At this point, each patent application has two decisions: the actual decision made by the examiner and a hypothetical decision made by the machine learning algorithm. I define a "false accept" (FalseAccept) as a case where the patent is accepted by the examiner but rejected by the algorithm. Conversely, a "false reject" (FalseReject) occurs when the examiner rejects the patent application but the algorithm would have accepted it.

Using these definitions, I construct four metrics to evaluate examiner screening performance for each year: i) the number of false reject cases, ii) the false rejection rate, iii) the number of false accept cases, and iv) the false acceptance rate. I also construct the following three measures of examiner characteristics: work experience in a given year, workload in a given year, and gender. To identify the gender of each examiner, I make use of a Social Security Administration data set, namely, national data on the relative frequency of given names in the population of U.S. births where the individual has a Social Security number.[16] This data set contains all given names and their associated genders with a population >5. I match examiners' first names with the names in this data set to obtain examiners' gender.[17]

[16]To access this data set, please see https://www.ssa.gov/oact/babynames/limits.html.

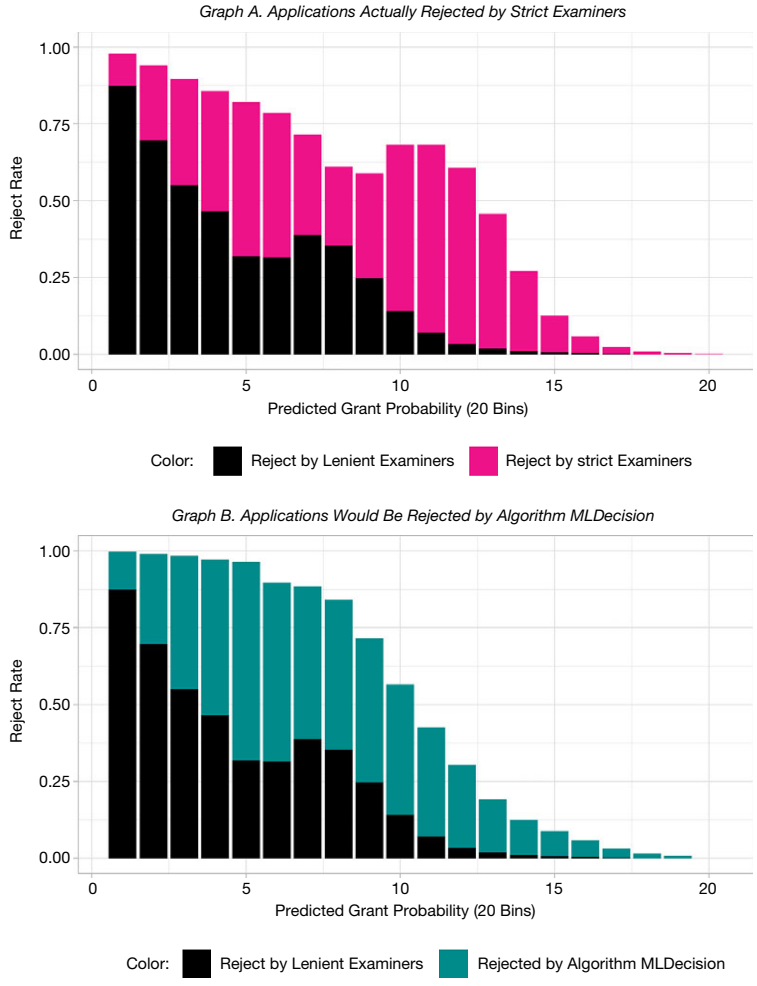
[17]When a given name is associated with both genders, I first calculate its probability of being specific to a certain gender based on the gender-specific population and assign the name to the gender with the probability >90%.

https://doi.org/10.1017/S0022109025000213

FIGURE 7

**Comparison Between Applications Rejected by Strict Examiners and Applications Rejected by Algorithm MLDecision**

Figure 7 compares applications rejected by strict examiners and those rejected by Algorithm MLDecision within each art unit. Strict (lenient) examiners are defined as those with an above (below) median reject rate in each art unit. I divide patent applications in the test set into 20 bins by predicted acceptance probability (*x*-axis). In both graphs, the black bar at the bottom of a given bin shows the fraction of patent applications being rejected by lenient examiners. The pink bar in Graph A shows which applications strict examiners actually reject. The green bar in Graph B shows which applications Algorithm MLDecision would reject to match the grant rate of strict examiners.




I test the relationship between examiner characteristics and screening performance with the following regression:

$$(1)\qquad y_{i,t} = \alpha + \beta_1 WorkLoad_{i,t} + \beta_2 WorkExperience_{i,t} + \beta_3 MaleExaminer_i + ArtUnit_a + StatusYear_t + \epsilon_{i,t},$$

where $i$ indexes the patent examiner, $a$ the art unit, and $t$ the issue year of a patent. $y$ includes # False Rejection, False Rejection Rate, # False Acceptance, and False

Acceptance Rate. $WorkLoad_{i,t}$ measures the workload of examiner $i$ in year $t$ and is calculated as the natural logarithm of the number of patent applications reviewed by that examiner. $WorkExperience_{i,t}$ measures the work experience of a given examiner in year $t$ and is calculated as the natural logarithm of the number of years worked in the patent office by examiner $i$. $MaleExaminer_i$ is a dummy variable that equals 1 if the gender of examiner $i$ is male, and 0 otherwise. $Art\,Unit_a$ and $Status\,Year_t$ indicate the art unit fixed effect and status year fixed effect.

Panel A of Table 2 presents the results of the above regression using Algorithm MLQuality as the benchmark. First, busy examiners tend to make more mistakes of both types, as suggested by the positive coefficient of *WorkLoad* in columns 1, 3, and 4 of Table 2. These findings are consistent with increasing time constraints faced by patent examiners, reducing their screening performance. Second, the negative coefficients of *WorkExperience* in columns 1 and 2 of Table 2 suggest that patent examiners with more experience tend to make fewer false rejection mistakes. However, they tend to make more false acceptance mistakes, as suggested by the positive coefficient of *WorkExperience* in columns 3 and 4 of Table 2. These

TABLE 2

Relationship Between Patent Examiner Characteristics and Screening Performance

The sample shown in Table 2 consists of patent examiners in the out-of-sample test set. *# False Rejections* (*# False Acceptances*) counts the number of false rejections (false acceptances) made by a given examiner in each year, where *False Rejection* (*False Acceptance*) equals to 1 if a patent application is rejected (accepted) by that examiner but accepted (rejected) by the algorithm, and 0 otherwise. *False Rejection Rate* (*False Acceptance Rate*) measures the percentage of false rejections (acceptances) over reviewed applications by a given examiner in each year. *WorkExperience* measures the work experience of a given examiner in a given year and is calculated as the natural logarithm of the number of years worked in the patent office for that examiner. *WorkLoad* measures the workload of a given patent examiner in a given year and is calculated as the natural logarithm of the number of patent applications reviewed by that examiner. *MaleExaminer* is a dummy variable that equals 1 if the gender of a given examiner is male, and 0 otherwise. Art Unit fixed effects and issue year fixed effects are included in all regressions. *t*-statistics are in parentheses. *** indicates significance at the 1% level.

| | Dependent Variable | | | |
|---|---|---|---|---|
| | # False Rejections | False Rejection Rate | # False Acceptances | False Acceptance Rate |
| | 1 | 2 | 3 | 4 |
| *Panel A. False Acceptance/Rejection Based on Algorithm MLQuality* | | | | |
| WorkLoad | 1.014*** | −0.034*** | 1.719*** | 0.016*** |
| | (32.52) | (−12.50) | (41.31) | (7.01) |
| WorkExperience | −0.122*** | −0.008*** | 0.136*** | 0.003*** |
| | (−19.90) | (−15.67) | (16.72) | (7.13) |
| MaleExaminer | 0.038 | −0.000 | 0.224*** | 0.001 |
| | (0.91) | (−0.02) | (4.02) | (0.26) |
| Art Unit FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| $R^2$ | 0.432 | 0.104 | 0.395 | 0.073 |
| No. of obs. | 18,013 | 18,013 | 18,013 | 18,013 |
| *Panel B. False Acceptance/Rejection Based on Algorithm MLDecision* | | | | |
| WorkLoad | 0.644*** | −0.024*** | 1.122*** | 0.009*** |
| | (28.57) | (−10.38) | (40.10) | (4.93) |
| WorkExperience | −0.080*** | −0.005*** | 0.091*** | 0.002*** |
| | (−18.19) | (−11.91) | (16.58) | (5.80) |
| MaleExaminer | 0.025 | −0.001 | 0.124*** | 0.000 |
| | (0.84) | (−0.30) | (3.32) | (0.20) |
| Art Unit FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| $R^2$ | 0.379 | 0.079 | 0.387 | 0.060 |
| No. of obs. | 18,013 | 18,013 | 18,013 | 18,013 |

results confirm the findings in the existing literature that more experienced examiners have a greater grant rate (see, e.g., Lemley (2009), deGrazia, Pairolero, and Teodorescu (2021)), suggesting potential agency problems induced by the compensation structure in the patent office: examiners obtain higher compensation by accepting more patent applications. Lastly, I find no evidence of more mistakes made by male examiners. Panel B of Table 2 reports similar results with Algorithm MLDecision as the benchmark.

## 4. Do Disagreements Between Humans and Machine Algorithms Predict Early Patent Expiration?

In this section, I test whether granted patents from the out-of-sample test set that the algorithm would have rejected would expire early. To measure disagreements between the machine learning predictions and actual screening decisions of patent examiners, I label a patent falsely accepted if it is accepted by an examiner but rejected by the algorithm.

Section 154 of the U.S. Patent Law (35 U.S. Code §154 (a)) sets the term of a utility patent filed on or after June 8, 1995, in the U.S. to 20 years from the earliest filing date of the application for the granted patent. Section 41 of the U.S. Patent Law (35 U.S. Code §41 (b) & (c)) states that maintenance fees must be paid at regular intervals to keep utility patents active.[18] If these falsely accepted patents should not have been granted in the first place, we would expect them to be more likely to expire early as a result of delays or defaults in the payment of maintenance fees. Specifically, I test whether falsely accepted patents are properly maintained with the following regression:

$$y_i = \alpha + \beta FalseAccept_i + ArtUnit_a + IssueYear_t + Small\&MicroEntity_s + USPC_j + \epsilon_i, \tag{2}$$

where $i$ indexes the patent, $a$ the art unit, $t$ the issue year of a patent, $s$ the size of the patentee, and $j$ the USPC class. $y$ represents patent maintenance-related dummies indicating the following four aspects: payment of the 4th-year maintenance fee, payment of the 8th-year maintenance fee, mailing of a maintenance fee reminder, and patent expiry for unpaid maintenance fees. $FalseAccept_i$ is a dummy variable equal to 1 if a patent is accepted by actual examiners but would be rejected by the algorithm. $ArtUnit_a$, $IssueYear_t$, $SmallEntity_s$, and $USPC_j$ represent art unit fixed effects, issue year fixed effects, the small entity dummy, and USPC class fixed effects.[19]

Panel A (B) of Table 3 presents the results of regressing equation (2) based on the disagreement between examiners and Algorithm MLQuality (MLDecision). The negative coefficients of *FalseAccept* in columns 1 and 2 suggest that falsely accepted patents are less likely to be maintained by their holders 4 years and 8 years after being granted. The positive coefficient of *FalseAccept* in column 3 suggests

[18]The patentee must pay maintenance fees before the 4th, 8th, and 12th years to keep the patent active.

[19]A patentee that is a small or micro entity needs to pay only 1/2 or 1/4 of the maintenance fees paid by a large entity.

TABLE 3

Relationship Between Weak Patent Screening and Subsequent Patent Maintenance

The sample shown in Table 3 consists of granted patents in the out-of-sample test set. *FalseAccept* equals 1 if a patent is accepted by an actual examiner but rejected by the algorithm, and 0 otherwise, as described in Section V.D.4. Small & Micro Entity Dummies, Art Unit fixed effects, issue year fixed effects, and patent USPC class fixed effects are included in all regressions. *t*-statistics are in parentheses. *** indicates significance at the 1% level.

| | Dependent Variable | | | |
|---|---|---|---|---|
| | Payment of Maintenance Fee in the 4th Year | Payment of Maintenance Fee in the 8th Year | Maintenance Fee Reminder Mailed | Patent Expired for Failure to Pay Maintenance Fees |
| | 1 | 2 | 3 | 4 |
| *Panel A. False Acceptance Based on Algorithm MLQuality* | | | | |
| FalseAccept | −0.048*** | −0.018*** | 0.046*** | 0.050*** |
| | (−26.87) | (−12.04) | (21.88) | (26.60) |
| Small & Micro Entity Dummies | Yes | Yes | Yes | Yes |
| Art Unit, Patent USPC Class, & Year FE | Yes | Yes | Yes | Yes |
| $R^2$ | 0.094 | 0.458 | 0.083 | 0.058 |
| No. of obs. | 235,597 | 235,597 | 235,597 | 235,597 |
| *Panel B. False Acceptance Based on Algorithm MLDecision* | | | | |
| FalseAccept | −0.052*** | −0.015*** | 0.062*** | 0.053*** |
| | (−24.44) | (−8.44) | (25.36) | (23.97) |
| Small & Micro Entity Dummies | Yes | Yes | Yes | Yes |
| Art Unit, Patent USPC Class, & Year FE | Yes | Yes | Yes | Yes |
| $R^2$ | 0.093 | 0.458 | 0.083 | 0.058 |
| No. of obs. | 235,597 | 235,597 | 235,597 | 235,597 |

that patentees holding falsely accepted patents are more likely to receive maintenance fee reminders. Further, the positive coefficient of *FalseAccept* in column 4 indicates that falsely accepted patents are more likely to expire due to unpaid maintenance fees. These results collectively show that falsely accepted patents turn out to be not very useful to their holders.

## VI. Innovation Screening and Firm Performance

This section extends my empirical analysis to study the (potential) economic consequences of the current patent screening procedure for firm performance. First, I describe firm data and an ex ante measure of the innovation screening performance of patent examiners in Section VI.A. Second, I discuss empirical results on the effect of innovation screening on the subsequent operating performance of public firms in Section VI.B. Lastly, I examine the effect of innovation screening on subsequent exits of private firms in Section VI.C.

### A. Firm Data and Sample Selection

I use all patent applications filed since 2010 with screening results available by 2018 in my analysis. In addition to the data on patent applications and patent examiners used in the previous section, I collect data on patent assignees from the USPTO website, accounting, and financial data for public firms from Compustat, firm characteristics, and VC financing data for private firms from VentureXpert. I match each data set with firm names standardized by the NBER patent data

name standardization routine.[20] By construction, both public and private firms analyzed in this section should have at least one patent application filed since 2010.

### 1. Measure of Innovation Screening Performance

I calculate the false acceptance rate of each patent examiner based on the cases of disagreement between the machine learning predictions and screening decisions of a given examiner. Specifically, I calculate the false acceptance rate of examiner $e$ in art unit $a$ who reviews patent applications $p$ at date $t$ as follows:

$$ExaminerFalseAcceptRate_{p,e,t,a} = \frac{\#FalseAccept_{e,t,a}}{\#Reviewed_{e,t,a}}, \tag{3}$$

where $\#Reviewed_{e,t,a}$ is the number of patent applications reviewed by examiner $j$ prior to the date $t$; $\#FalseAccept_{e,t,a}$ is the number of patent applications and falsely accepted by examiner $j$ prior to date $t$ as defined in Section V.D.3.[21]

To match the time horizon of financial and accounting data on firm performance, I further measure the patent screening performance of the examiners associated with each firm in each quarter by averaging the false acceptance rates over the past 3 years of examiners who have examined that firm's patent applications using a 3-year rolling window.[22] For example, the false acceptance rate of firm $i$ in quarter $q$ is calculated as follows:

$$AvgExaminerFalseAcceptRate_{i,q} = \frac{1}{N}\sum_{a=1}^{N}\left(\sum_{t=q-13}^{q-1} ExaminerFalseAcceptRate_{p,e,t,a}\right), \tag{4}$$

where $ExaminerFalseAcceptRate_{p,e,t,a}$ is the false acceptance rate over the past 3 years of examiner $e$ reviewing firm $i$'s patent application $p$ and $N$ is the total number of patent applications filed by firm $i$ with screening results that became available in the past 3 years.

By construction, the false acceptance rate of an individual examiner is available ex ante for any newly filed patent application in my sample. More importantly, these measures are unlikely to be correlated with firm characteristics due to the quasi-random assignment of patent applications to patent examiners within each art unit.

### 2. Summary Statistics

Table 4 reports summary statistics for my measures of innovation screening performance and firm characteristics. Panel A presents the summary statistics of

[20]The name standardization routine comes from the NBER Patent Data Project: https://sites.google.com/site/patentdataproject.

[21]I exclude the patent application $p$ from both the numerator and the denominator. I also exclude firms whose patent applications are assigned to patent examiners who reviewed fewer than 10 patent applications prior to patent application $p$. All results in this section are robust to removing the above exclusions.

[22]I use different time windows to measure firm-level innovation screening performance (i.e., 1-quarter, 1-year, and 2-year windows), and all empirical results in this section remain qualitatively similar.

TABLE 4

Summary Statistics (Firms)

Table 4 shows descriptive statistics for the sample of both public and private firms that have at least one patent application filed since 2010 and with status available before (and including) 2018. Panel A reports summary statistics for the sample of public firms; Panel B reports summary statistics for the sample of private firms. *AvgExaminerFalseAcceptRate* is defined as the average false acceptance rates of examiners that are related to all granted and rejected applications for each firm in a given past 3-year rolling window as described in Section VI.A.1, where the false acceptance rate of an examiner associated with each patent application is defined as the ratio of falsely accepted applications over all applications he/she has made decisions prior to that patent application. A patent application is falsely accepted if it is accepted by the actual examiner but rejected by the machine learning algorithm. *#ApplicationsReviewed* counts the number of patent applications being reviewed for each firm in a given past 3-year rolling window. *ROA* is the ratio of quarterly net income over book assets. *Cash Flow* is the quarterly cash flow over book assets. *R&D Expenditures* are the quarterly R&D expenditures over book assets. *FirmSize* is the natural logarithm of book assets. *Leverage* is the total debt (both current liability and long-term debt) over book assets. *Ln (M/B)* is the natural logarithm of the market-to-book ratio. *SuccessfulExit* is a dummy, which equals 1 if a given private firm has exited through an IPO or M&A by the end of my sample period, and 0 otherwise. *LnVCFinancingAmount* is the natural logarithm of the quarterly investment amount for each firm. *LnNumberFundInvested* is the natural logarithm of the quarterly number of invested funds for each firm. *LnFirmAge* is the natural logarithm of firm age, which equals the current year minus the firm founding year plus 1. All accounting variables (i.e., *ROA*, *Cash Flow*, *R&D Expenditures*, *Leverage*, *Ln*(*M*/*B*)) are winsorized at 1% and 99%.

| | *N* | Mean | Median | p10 | p90 | Std. Dev. |
|---|---|---|---|---|---|---|
| *Panel A. Public Firm Sample (firm-quarter level)* | | | | | | |
| AvgExaminerFalseAcceptRate (MLQuality) | 19,573 | 0.132 | 0.129 | 0 | 0.222 | 0.084 |
| AvgExaminerFalseAcceptRate (MLDecision) | 19,573 | 0.083 | 0.080 | 0 | 0.148 | 0.064 |
| #ApplicationsReviewed | 19,573 | 2.577 | 2.197 | 0.693 | 4.796 | 1.606 |
| ROA | 19,437 | −0.023 | 0.006 | −0.126 | 0.033 | 0.084 |
| FirmSize | 19,464 | 6.931 | 6.656 | 3.969 | 10.335 | 2.449 |
| Leverage | 18,704 | 0.195 | 0.159 | 0 | 0.461 | 0.205 |
| Ln(M/B) | 18,570 | 1.092 | 1.022 | 0.055 | 2.200 | 0.858 |
| R&D Expenditures | 19,464 | 0.033 | 0.018 | 0 | 0.088 | 0.044 |
| *Panel B. Private Firm Sample (firm-quarter level)* | | | | | | |
| AvgExaminerFalseAcceptRate (MLQuality) | 5,896 | 0.132 | 0.124 | 0 | 0.231 | 0.096 |
| AvgExaminerFalseAcceptRate (MLDecision) | 5,896 | 0.080 | 0.074 | 0 | 0.154 | 0.068 |
| #ApplicationsReviewed | 5,896 | 1.840 | 1.792 | 0.693 | 3.091 | 0.919 |
| SuccessExit | 5,896 | 0.259 | 0 | 0 | 1 | 0.438 |
| FirmAge | 5,896 | 10.731 | 10 | 6 | 16 | 4.610 |
| LnVCFinancingAmount | 5,896 | 0.217 | 0 | 0 | 0 | 0.767 |
| NumberFundInvested | 5,896 | 0.371 | 0 | 0 | 1 | 1.295 |

firm performance and firm characteristics for public firms at the firm-quarter level. The average false acceptance rate for public firms based on algorithms MLQuality and MLDecision are 14.1% and 8.3%. The median number of patent applications filed by public firms in a given 3-year window is 14; the median quarterly *ROA*, defined as net income divided by total assets, is 0.6%. Public firms, on average, have log book assets of 6.9, a leverage ratio of 0.2, a log market-to-book ratio of 1.1, and R&D expenditures of 3.3%.[23]

Panel B of Table 4 presents the summary statistics of firm performance and firm characteristics for private firms at the firm-quarter level. The average false acceptance rate for private firms based on algorithms MLQuality and MLDecision are 13.6% and 8.0%. The median number of patent applications filed by private firms in a given 3-year window is 5, much lower than the figures for public firms. In terms of firm characteristics, the average age of private firms is 10.7; private firms have a log quarterly VC financing amount of 0.2 and a quarter

[23] All accounting variables (i.e., *ROA*, *R&D Expenditures*, *Leverage*, and *Ln(M/B)*) are winsorized at 1% and 99%. All regression results are robust to winsorizing with different thresholds.

number of VC funds of 0.4. Finally, the average rate of successful exits through IPOs or M&As is 25.9%.

## B. Innovation Screening and the Subsequent Operating Performance of Public Firms

In this section, I empirically examine whether the false acceptance rates of patent examiners have any effect on the subsequent operating performance of public firms using the following regression:

$$ROA_{i,q+n} = \alpha + \beta AvgExaminerFalseAcceptRate_{i,q} + \gamma X_{i,q} + Firm_i + YearQuarter_q + \epsilon_{i,q}, \quad (5)$$

where $i$ indexes the firm, $j$ the industry, and $q$ the quarter and $n$ equals 1, 4, 8, or 12. $y$ is the operating performance of each public firm, measured using either *ROA* or *Cash Flow*. For example, $ROA_{i,q+4}$ measures the subsequent 4-quarter (or 1-year) operating performance of each public firm. $AvgExaminerFalseAcceptRate_{i,q}$ is the average false acceptance rate in the past 3 years (or 12 quarters) of the examiners who have examined firm $i$'s patent applications as described in Section VI.A.1. $X$ is a vector of control variables, including the number of patents reviewed in the past 3 years, firm size in quarter $t$, leverage in quarter $t$, market-to-book ratio in quarter $t$, and R&D expenditures in quarter $t$ as described in Section VI.A.2. $Firm_i$ and $YearQuarter_q$ represent firm fixed effects and year-quarter fixed effects. All standard errors are double-clustered at the firm and quarter levels.

The regression results using *ROA* as the dependent variable are reported in Table 5. Panel A uses the false acceptance rate based on Algorithm MLQuality and shows that the coefficient of *AvgExaminerFalseAcceptRate* is negative in all 4 columns and becomes statistically significant in column 4. These results suggest that public firms whose patent applications are reviewed by examiners with higher past false acceptance rates perform worse starting 2 years or more. Panel B uses the false acceptance rate based on Algorithm MLDecision and shows similar results. These results are also economically significant. For example, a 1-standard-deviation increase in *AvgExaminerFalseAcceptRate* decreases annual *ROA* by 22 bps to 37 bps. In other words, if all patent applications were screened by Algorithm MLQuality (MLDecision) (i.e., if *AvgExaminerFalseAcceptRate* reduces to 0), annual ROA would increase by 35 bps (48 bps). Moreover, the effect of the current patent screening system on firm performance is likely to be causal due to the quasi-random assignment of patent applications to patent examiners.[24] Regarding potential channels, these results are consistent with the finding in Section V.D.4 that firms are more likely to allow falsely

[24]I focus on the impact of false acceptances on firm performance for several reasons. First, false acceptance is the main problem facing the USPTO, with an average acceptance rate of more than 68%. Second, analyzing the impact of false rejections on firm performance poses empirical challenges: i) we do not observe actual quality information for applications rejected by humans but accepted by the algorithm, and ii) it is difficult to isolate the negative impact of correct rejections on firm performance.

TABLE 5

Relationship Between Screening Performance of Patent Examiners and Subsequent Operating Performance of Public Firms

The sample in Table 5 consists of firms that have at least one patent application filed since 2010 and with application outcome available by 2018. *ROA* is the ratio of quarterly net income over book assets. *AvgExaminerFalseAcceptRate* is defined as the average false acceptance rates of examiners that are related to all granted and rejected applications for each firm in a given past 3-year rolling window as described in Section VI.A.1, where the false acceptance rate of an examiner associated with each patent application is defined as the ratio of falsely accepted applications over all applications he/she has made decisions prior to that patent application. A patent application is falsely accepted if it is accepted by the actual examiner but rejected by the machine learning algorithm. *#ApplicationsReviewed* counts the number of patent applications being reviewed for each firm in a given past 3-year rolling window. All regressions include *FirmSize*, *Leverage*, *Ln(M/B)*, and *R&D Expenditures* as control variables. All accounting variables (i.e., *ROA*, *Cash Flow*, *R&D Expenditures*, *Leverage*, *Ln*(*M*/*B*)) are winsorized at 1% and 99%. Firm and year-quarter fixed effects are included in all regressions. *t*-statistics are in parentheses. *** and ** indicate significance at the 1% and 5% levels, respectively.

| | Dependent Variable: Subsequent ROA | | | |
|---|---|---|---|---|
| | 1 Quarter | 1 Year | 2 Years | 3 Years |
| | 1 | 2 | 3 | 4 |
| *Panel A. False Acceptance Rate of Patent Examiners Based on Algorithm MLQuality* | | | | |
| AvgExaminerFalseAcceptRate | −0.004 | −0.014 | −0.034 | −0.080** |
| | (−0.53) | (−0.71) | (−1.18) | (−2.37) |
| #ApplicationsReviewed | 0.000 | 0.003 | 0.012*** | 0.014*** |
| | (0.16) | (1.41) | (3.85) | (3.76) |
| Controls | Yes | Yes | Yes | Yes |
| Firm & YearQuarter FE | Yes | Yes | Yes | Yes |
| $R^2$ | 0.719 | 0.880 | 0.940 | 0.970 |
| No. of obs. | 17,417 | 14,958 | 11,714 | 8,711 |
| *Panel B. False Acceptance Rate of Patent Examiners Based on Algorithm MLDecision* | | | | |
| AvgExaminerFalseAcceptRate | 0.000 | −0.023 | −0.103** | −0.175*** |
| | (0.01) | (−0.83) | (−2.44) | (−3.50) |
| #ApplicationsReviewed | 0.000 | 0.003 | 0.012*** | 0.014*** |
| | (0.18) | (1.42) | (3.84) | (3.79) |
| Controls | Yes | Yes | Yes | Yes |
| Firm & YearQuarter FE | Yes | Yes | Yes | Yes |
| $R^2$ | 0.719 | 0.880 | 0.940 | 0.970 |
| No. of obs. | 17,417 | 14,958 | 11,714 | 8,711 |

granted patents to expire early, possibly due to the failed commercialization of these patents.[25]

## C. Innovation Screening and the Subsequent Performance of Private Firms

In this section, I study the relationship between innovation screening and the subsequent performance of private firms with the following specifications:

$$y_{i,q+n} = \alpha + \beta AvgExaminerFalseAcceptRate_{i,q} + \gamma Z_{i,q} + State_s + Industry_j + YearQuarter_q + \epsilon_{i,q}, \quad (6)$$

[25]I also find that firms are more likely to be sued in patent litigation cases if their patent applications are screened by examiners with higher false acceptance rates based on Algorithm MLQuality as shown in Table IA.1 in the Supplementary Material. However, the results on patent litigation are not significant based on Algorithm MLDecision.

https://doi.org/10.1017/S0022109025000213

where $y$ is the performance of each private firm, measured using either subsequent *LnVCFinancingAmount* or *SuccessfulExit*. For example, $LnVCFinancingAmount_{i,q+4}$ is the natural logarithm of the VC investment amount received by each firm in the following 1-year (4-quarter) period; $SuccessfulExit_{i,q+4}$ is a dummy variable that equals 1 if a firm successfully exits by either an IPO or M&A in the following 1-year (4-quarter) period, and 0 otherwise. $AvgExaminerFalseAcceptRate_{i,q}$ is my screening performance measure for the examiners who have examined firm $i$'s patent applications in the past 3 years (12 quarters) as described in Section VI.A.1. $Z$ is a vector of control variables, including the number of patents reviewed and granted in the past 3 years, firm age in quarter $t$, VC funding received in quarter $t$, and the number of funds invested in quarter $t$ as described in Section VI.A.2. $State_s$, $Industry_j$, and $YearQuarter_q$ represent state-of-incorporation, 2-digit SIC industry, and year-quarter fixed effects. Standard errors are clustered at the state level.

Table 6 reports the regression results using *SuccessfulExit* as the dependent variable. Panel A of Table 6 uses the false acceptance rate based on Algorithm MLQuality and shows that the coefficient of *AvgExaminerFalseAcceptRate* is negative and statistically significant in all 4 columns, suggesting that private firms whose patent applications are reviewed by examiners with higher past false

TABLE 6

Relationship Between Screening Performance of Patent Examiners and Subsequent Exits of Private Firms

The sample in Table 6 consists of firms that have at least one patent application filed since 2010 and with application outcome available by 2018. *SuccessfulExit* is a dummy, which equals 1 if a given private firm has exited through an IPO or M&A by the end of my sample period, and 0 otherwise. *AvgExaminerFalseAcceptRate* is defined as the average false acceptance rates of examiners that are related to all granted and rejected applications for each firm in a given past 3-year rolling window as described in Section VI.A.1. *#ApplicationsReviewed* counts the number of patent applications being reviewed for each firm in a given past 3-year rolling window. The following control variables are included in all regressions: *LnVCFinancingAmount*, *TotalFundingToDate*, *LnNumberFundInvested*, and *LnFirmAge*. Year-quarter fixed effects, industry (2-digit SIC code) fixed effects, and state fixed effects are included in all regressions. *t*-statistics are in parentheses. All regressions are OLS regressions with standard errors clustered at the state level. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

| | Dependent Variable: Subsequent SuccessfulExit | | | |
|---|---|---|---|---|
| | 1 Quarter | 1 Year | 2 Years | 3 Years |
| | 1 | 2 | 3 | 4 |
| *Panel A. False Acceptance Rate of Patent Examiners Based on Algorithm MLQuality* | | | | |
| AvgExaminerFalseAcceptRate | −0.005 | −0.056** | −0.126*** | −0.173*** |
| | (−0.45) | (−2.51) | (−3.27) | (−3.37) |
| #ApplicationsReviewed | 0.008*** | 0.017*** | 0.024*** | 0.024* |
| | (3.89) | (3.84) | (2.78) | (1.93) |
| Controls | Yes | Yes | Yes | Yes |
| Industry, YearQuarter, & State FE | Yes | Yes | Yes | Yes |
| $R^2$ | 0.016 | 0.038 | 0.059 | 0.070 |
| No. of obs. | 5,888 | 5,888 | 5,888 | 5,242 |
| *Panel B. False Acceptance Rate of Patent Examiners Based on Algorithm MLDecision* | | | | |
| AvgExaminerFalseAcceptRate | 0.003 | −0.052* | −0.107** | −0.129 |
| | (0.22) | (−1.95) | (−2.17) | (−1.68) |
| #ApplicationsReviewed | 0.008*** | 0.017*** | 0.025*** | 0.025** |
| | (3.87) | (3.92) | (2.84) | (2.03) |
| Controls | Yes | Yes | Yes | Yes |
| Industry, YearQuarter, & State FE | Yes | Yes | Yes | Yes |
| $R^2$ | 0.016 | 0.037 | 0.058 | 0.069 |
| No. of obs. | 5,888 | 5,888 | 5,888 | 5,242 |

acceptance rates are less likely to exit successfully by either an IPO or M&A in subsequent quarters. Panel B of Table 6 uses the false acceptance rate based on Algorithm MLDecision and shows similar but slightly weaker results. Both results are also economically significant: a 1-standard-deviation increase in *AvgExaminerFalseAcceptRate* decreases the following 2-year probabilities of exiting successfully by an IPO or M&A by 1.2% (0.7%). In other words, if all patent applications were screened by Algorithm MLQuality (MLDecision) (i.e., if *AvgExaminerFalseAcceptRate* reduces to 0), the probability of exiting successfully by an IPO or M&A over the following 2-year period would increase by 1.6% (0.9%). Overall, these results suggest that weak innovation screenings could negatively affect the probability of subsequent exits through IPOs or M&As for private firms.

## VII. Conclusion

In this article, I examine whether the patent screening process can be improved under the current patent system in terms of granting higher-quality patents. I argue that examiners may not process relevant information efficiently enough to screen out low-quality applications, possibly due to the examiners' increasing time constraints and conflicting incentives. However, machine learning algorithms have much larger capacities than humans to process information efficiently and can potentially reduce human-specific agency issues. Using utility patent applications filed at the USPTO from 2001 to 2018, I trained two separate machine learning algorithms to learn application quality and examiner decision using earlier patent applications and predict the quality and examiner decision of more recent patent applications out of sample. I show that the current patent system accepts many low-quality patents. To compare the performance of a human-only system and a human + machine system, I make use of the quasi-random assignment of patent applications to examiners with different levels of leniency within each art unit. I find that a human + machine system could significantly improve the quality of the granted patents.

The analysis reveals that busy and more experienced examiners are more prone to improperly grant patents, pointing to the influence of both resource constraints and agency problems. Regression analysis suggests that such inappropriately granted patents often expire prematurely, indicating their limited benefit to the owners. To examine the potential economic gains that could be achieved under the current patent screening system, I construct an innovation screening performance measure by calculating the false acceptance rate of each patent examiner. I find that the worse innovation screening performance of examiners negatively impacts firm performance. For example, public firms whose patent applications are reviewed by examiners with higher false acceptance rates are likely to have lower operating performance (measured by ROA). Similarly, private firms under such examiners have a reduced likelihood of successful exits via IPOs or M&As. Economically, implementing a hypothetical human-plus-machine screening process could increase public firms' annual ROA by 35 to 48 bps and enhance private firms' 3-year probability of successful exits by 0.9 to 1.6 percentage points.

It is important to acknowledge the potential limitations of my analyses. First, the missing counterfactual issue pervades machine learning applications in screening decisions: I do not observe actual outcomes for applications rejected by human examiners but accepted by the algorithm. Recognizing this, I show that this issue

can be addressed using quasi-random assignment of patent applications to patent examiners. Second, using forward citations as the sole measure of patent quality may capture only parts of what examiners value. To mitigate this, I use examiners' past screening decisions as proxies for their screening objectives to train a second algorithm. Nevertheless, a key benefit of machine learning algorithms is that they are not subject to the agency conflicts and resource constraints currently faced by human examiners at USPTO.

To summarize, despite the great promise of augmenting human expertise with machine learning in patent screenings, replacing human examiners entirely might lead to unforeseen consequences, such as strategic patent filings (Choudhury et al. (2020)). Instead, machine learning could serve as a complementary tool, aiding examiners in refining their decision-making process—perhaps as an audit mechanism to reevaluate their assessments.[26] Based on the findings in this article, combining the expertise of human examiners and the strength of machine learning may potentially yield better screening outcomes.

## Supplementary Material

To view supplementary material for this article, please visit http://doi.org/10.1017/S0022109025000213.

## Funding Statement

This work is not supported by any funding.

## References

Abis, S. "Man vs. Machine: Quantitative and Discretionary Equity Management." Working Paper, Columbia Business School (2017).

Arrow, K. J. "Economic Welfare and the Allocation of Resources for Invention." In *Readings in Industrial Economics*, Vol 1, C. K. Rowley, ed. Berlin, Germany: Springer (1972), 219–236.

Arts, S.; B. Cassiman; and J. C. Gomez. "Text Matching to Measure Patent Similarity." *Strategic Management Journal*, 39 (2018), 62–84.

Athey, S., and G. W. Imbens. "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives*, 31 (2017), 3–32.

Bessen, J., and E. Maskin. "Sequential Innovation, Patents, and Imitation." *RAND Journal of Economics*, 40 (2009), 611–635.

Bessen, J. E., and M. J. Meurer. "Patent Litigation with Endogenous Disputes." *American Economic Review*, 96 (2006), 77–81.

Bowen III, D. E.; L. Frésard; and G. Hoberg. "Rapidly Evolving Technologies and Startup Exits." *Management Science*, 69 (2023), 940–967.

Chemmanur, T. J.; M. Gupta; and K. Simonyan. "Top Management Team Quality and Innovation in Venture-Backed Private Firms and IPO Market Rewards to Innovative Activity." *Entrepreneurship Theory and Practice*, 46 (2022), 920–951.

Chen, T., and C. Guestrin. "Xgboost: A Scalable Tree Boosting System." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM (2016) 785–794.

Choi, J. P. "Patent Litigation as an Information-Transmission Mechanism." *American Economic Review*, 88 (1998), 1249–1263.

[26]While human examiners may or may not change their decisions after reexaminations of patent applications, such a reexamination process may potentially reduce human bias from behavioral issues or the increasing time constraints that examiners face.

Choudhury, P.; E. Starr; and R. Agarwal. "Machine Learning and Human Capital Complementarities: Experimental Evidence on Bias Mitigation." *Strategic Management Journal*, 41 (2020), 1381–1411.
Cohen, L.; K. Diether; and C. Malloy. "Misvaluing Innovation." *Review of Financial Studies*, 26 (2013), 635–666.
Cornelli, F., and M. Schankerman. "Patent Renewals and R&D Incentives." *RAND Journal of Economics*, 30 (1999), 197–213.
deGrazia, C.; N. A. Pairolero; and M. Teodorescu. "Shorter Patent Pendency Without Sacrificing Quality: The Use of Examiner's Amendments at the USPTO." *Research Policy*, 50 (2021), 2019-03.
Dreyfuss, R. C. "Nonobviousness: A Comment on Three Learned Papers." *Lewis & Clark Law Review*, 12 (2008), 431.
Duffy, J. F. "A Timing Approach to Patentability." *Lewis & Clark Law Review*, 12 (2008), 343.
Eberhart, A. C.; W. F. Maxwell; and A. R. Siddique. "An Examination of Long-Term Abnormal Stock Returns and Operating Performance Following R&D Increases." *Journal of Finance*, 59 (2004), 623–650.
Eisenberg, R. S. "Pharma's Nonobvious Problem." *Lewis & Clark Law Review*, 12 (2008), 375.
Erel, I.; L. H. Stern; C. Tan; and M. S. Weisbach. "Selecting Directors Using Machine Learning." *Review of Financial Studies*, 34 (2021), 3226–3264.
Farre-Mensa, J.; D. Hegde; and A. Ljungqvist. "What Is a Patent Worth? Evidence from the US Patent "Lottery"." *Journal of Finance*, 75 (2020), 639–682.
Feng, J., and X. Jaravel. "Crafting Intellectual Property Rights: Implications for Patent Assertion Entities, Litigation, and Innovation." *American Economic Journal: Applied Economics*, 12 (2020), 140–181.
Fitzgerald, T.; B. Balsmeier; L. Fleming; and G. Manso. "Innovation Search Strategy and Predictable Returns." *Management Science*, 67 (2021), 1109–1137.
Frakes, M. D., and M. F. Wasserman. "Does the U.S. Patent and Trademark Office Grant Too Many Bad Patents? Evidence from a Quasi-Experiment." *Stanford Law Review*, 67 (2015), 613.
Frakes, M. D., and M. F. Wasserman. "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 (2017), 550–563.
Friedman, J. H. "Greedy Function Approximation: A Gradient Boosting Machine." *Annals of Statistics*, 29 (2001), 1189–1232.
Gilbert, R., and C. Shapiro. "Optimal Patent Length and Breadth." *RAND Journal of Economics*, 21 (1990), 106–112.
Graham, S. J.; A. C. Marco; and R. Miller. "The USPTO Patent Examination Research Dataset: A Window on Patent Processing." *Journal of Economics & Management Strategy*, 27 (2018), 554–578.
Gu, S.; B. Kelly; and D. Xiu. "Empirical Asset Pricing via Machine Learning." *Review of Financial Studies*, 33 (2020), 2223–2273.
Hall, B. H.; A. Jaffe; and M. Trajtenberg. "Market Value and Patent Citations." *RAND Journal of Economics*, 36 (2005), 16–38.
Hall, B. H.; A. B. Jaffe; and M. Trajtenberg. "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." NBER Working Paper No. 8498 (2001).
Heller, M. A., and R. S. Eisenberg. "Can Patents Deter Innovation? The Anticommons in Biomedical Research." *Science*, 280 (1998), 698–701.
Hirschman, A. O. *National Power and the Structure of Foreign Trade*, Vol. 105. Berkeley, CA: University of California Press (1980).
Jaffe, A. B., and J. Lerner. *Innovation and Its Discontents: How Our Broken Patent System Is Endangering Innovation and Progress, and What to Do About It*. Princeton, NJ: Princeton University Press (2011).
Kleinberg, J.; H. Lakkaraju; J. Leskovec; J. Ludwig; and S. Mullainathan. "Human Decisions and Machine Predictions." *Quarterly Journal of Economics*, 133 (2017), 237–293.
Kleinberg, J.; J. Ludwig; S. Mullainathan; and Z. Obermeyer. "Prediction Policy Problems." *American Economic Review*, 105 (2015), 491–495.
Kline, P.; N. Petkova; H. Williams; and O. Zidar. "Who Profits from Patents? Rent-Sharing at Innovative Firms." *Quarterly Journal of Economics*, 134 (2019), 1343–1404.
Kogan, L.; D. Papanikolaou; A. Seru; and N. Stoffman. "Technological Innovation, Resource Allocation, and Growth." *Quarterly Journal of Economics*, 132 (2017), 665–712.
Krishna, A. M.; B. Feldman; J. Wolf; G. Gabel; S. Beliveau; and T. Beach. "User Interface for Customizing Patents Search: An Exploratory Study." *International Conference on Human-Computer Interaction*. Berlin, Germany: Springer (2016), 264–269.
Lanjouw, J. O., and M. Schankerman. "Characteristics of Patent Litigation: A Window on Competition." *RAND Journal of Economics*, 32 (2001), 129–151.
Lemley, M. A. "Examiner Characteristics and the Patent Grant Rate." (2009).
Lemley, M. A., and C. Shapiro. "Probabilistic Patents." *Journal of Economic Perspectives*, 19 (2005), 75–98.

Lu, Q.; A. Myers; and S. Beliveau. "USPTO Patent Prosecution Research Data: Unlocking Office Action Traits." USPTO Economic Working Paper No. 10 (2017).
Maestas, N.; K. J. Mullen; and A. Strand. "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt." *American Economic Review*, 103 (2013), 1797–1829.
Mandel, G. N. "Another Missed Opportunity: The Supreme Court's Failure to Define Nonobviousness or Combat Hindsight Bias in KSR v. Teleflex." *Lewis & Clark Law Review*, 12 (2008), 323.
Mansfield, E. "Patents and Innovation: An Empirical Study." *Management Science*, 32 (1986), 173–181.
Marco, A.; M. Carley; S. Jackson; and A. Myers. "The USPTO Historical Patent Data Files: Two Centuries of Invention." USPTO Economic Working Paper No. 2015-1 (2015a).
Marco, A. C.; A. Myers; S. J. Graham; P. D'Agostino; and K. Apple. "The USPTO Patent Assignment Dataset: Descriptions and Analysis." USPTO Economic Working Paper No. 2015-2 (2015b).
Marco, A. C.; J. D. Sarnoff; and A. Charles. "Patent Claims and Patent Scope." *Research Policy*, 48 (2019), 103,790.
Marco, A. C.; A. A. Toole; R. Miller; and J. Frumkin. "USPTO Patent Prosecution and Examiner Performance Appraisal." USPTO Economic Working Paper No. 2017-08 (2017).
Matutes, C.; P. Regibeau; and K. Rockett. "Optimal Patent Design and the Diffusion of Innovations." *The RAND Journal of Economics*, 27 (1996), 60–83.
Merges, R. P. "As Many as Six Impossible Patents Before Breakfast: Property Rights for Business Concepts and Patent System Reform." *Berkeley Technology Law Journal*, 14 (1999), 577.
Meurer, M. J. "The Settlement of Patent Litigation." *RAND Journal of Economics*, 20 (1989), 77–91.
Mikolov, T.; K. Chen; G. Corrado; and J. Dean. "Efficient Estimation of Word Representations in Vector Space." arXiv preprint arXiv:1301.3781 (2013a).
Mikolov, T.; Q. V. Le; and I. Sutskever. "Exploiting Similarities Among Languages for Machine Translation." arXiv preprint arXiv:1309.4168 (2013b).
Mikolov, T.; W.-T. Yih; and G. Zweig. "Linguistic Regularities in Continuous Space Word Representations." In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (2013c) 746–751.
Mullainathan, S., and J. Spiess. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives*, 31 (2017), 87–106.
Nordhaus, W. D. "An Economic Theory of Technological Change." *American Economic Review*, 59 (1969), 18–28.
Righi, C., and T. Simcoe. "Patent Examiner Specialization." *Research Policy*, 48 (2019), 137–148.
Rokach, L., and O. Z. Maimon. *Data Mining with Decision Trees: Theory and Applications*, 2nd edition. Singapore: World Scientific (2008).
Rossi, A. "Predicting Stock Market Returns with Machine Learning." Working Paper, Georgetown University (2018).
Sampat, B., and H. L. Williams. "How Do Patents Affect Follow-On Innovation? Evidence from the Human Genome." *American Economic Review*, 109 (2019), 203–236.
Schankerman, M., and F. Schuett. "Patent Screening, Innovation, and Welfare." *Review of Economic Studies*, 89 (2022), 2101–2148.
Scherer, F. M. "Nordhaus' Theory of Optimal Patent Life: A Geometric Reinterpretation." *American Economic Review*, 62 (1972), 422–427.
Shapley, L. S. "A Value for n-Person Games." *Contribution to the Theory of Games*, 2 (1953), 307–317.
Shu, T.; X. Tian; and X. Zhan. "Patent Quality, Firm Value, and Investor Underreaction: Evidence from Patent Examiner Busyness." *Journal of Financial Economics*, 143 (2022), 1043–1069.
Tong, X., and J. D. Frame. "Measuring National Technological Performance with Patent Claims Data." *Research Policy*, 23 (1994), 133–141.
Trajtenberg, M. "A Penny for Your Quotes: Patent Citations and the Value of Innovations." *RAND Journal of Economics*, 21 (1990), 172–187.
Trajtenberg, M.; R. Henderson; and A. Jaffe. "University Versus Corporate Patents: A Window on the Basicness of Invention." *Economics of Innovation and New Technology*, 5 (1997), 19–50.
Zucker, L. G.; M. R. Darby; and J. S. Armstrong. "Commercializing Knowledge: University Science, Knowledge Capture, and Firm Performance in Biotechnology." *Management Science*, 48 (2002), 138–153.