Playing the Patent Lottery

Cameron M. Ellis*, Johannes G. Jaspersen† & Valentin Luz‡

November 18, 2025

Abstract

We investigate whether the U.S. patent examiner lottery, which is intended to ensure random assignment and equal treatment of applicants, can be strategically manipulated. Using data on six million applications from 2001–2020, we derive the null distribution of examiner leniency under true randomization and show that many law firms and in-house counsel systematically obtain far more lenient examiners than chance would permit. We document multiple mechanisms consistent with strategic manipulation, including leveraging institutional knowledge and the exploitation of publicly observable examiner workloads. These deviations from randomness have economically meaningful consequences: sorting firms on examiner leniency at publication produces sizable and statistically significant return spreads, implying that markets only partially incorporate assignment-based variation in approval likelihood. We develop a simple model of manipulation investments that explains which firms are most likely to game the patent lottery and show empirical support for its predictions. Our findings challenge the fairness of the patent lottery and call into question empirical designs that rely on examiner randomization for causal identification.

JEL Codes: **032** · **034** · **L51** · **G14**

Keywords: Patents · Innovation · Leniency Designs · Law Firms

*Tippie College of Business, University of Iowa (cameron-ellis@uiowa.edu)
†LMU Munich School of Management, LMU Munich (jaspersen@lmu.de)
‡LMU Munich School of Management, LMU Munich (luz@lmu.de)

1 Introduction

The patent system exists because competitive markets do not properly incentivize investment into private research (Budish et al., 2016). To reward firms for their innovation, and to share the idea of the innovation itself with the public, firms are granted exclusive monopoly power over their inventions for a limited time. This monopoly power is extremely valuable, with even "marginal" innovations generating substantial economic rents (Farre-Mensa et al., 2020). In order to ensure that this value is allocated to true innovation, patents are examined for novelty and non-obviousness by patent examiners. Because they are human, these examiners vary in their level of expertise and their willingness to grant patents. To ensure a level playing field, examiners should be assigned to applications randomly in a process known as the patent "lottery".

In this paper we ask, and answer, four questions about the patent lottery. (1) Can it be gamed? We derive the theoretical distribution of examiner leniency under random assignment and show that certain law firms and in-house counsel receive lenient examiners at a rate far beyond what could be attributed to chance. (2) How do you game the patent lottery? We identify multiple mechanisms through which sophisticated applicants can influence examiner assignment, from strategic timing to leveraging institutional knowledge. (3) When can we expect companies to game the patent lottery? We adapt the models of Nordhaus (1969) and Budish et al. (2016) and show that large firms are most likely to manipulate the system. Because the lottery by its nature is zero-sum, gaming by large firms makes it less likely that small firms have their patents approved. (4) Is there value in getting assigned a more lenient patent examiner? By implementing a trading strategy that exploits variation in examiner leniency at the publication date, we show that investors can earn sizable and statistically significant abnormal returns.

The USPTO receives over 600,000 patent applications annually and employs approximately 8,000 examiners to evaluate them (Farre-Mensa et al., 2020). Each examiner specializes in a specific technology area—called an "art unit"—and within each art unit, applications are distributed to examiners through a computerized assignment system. This system, introduced to eliminate favoritism and ensure fairness, operates as a lottery: applications enter a queue and are assigned to the next available examiner with the relevant technical expertise. The assumption underlying this system is that random assignment produces equal treatment across applicants. Yet examiners differ substantially in their grant rates, with some approving patents at rates 40 percentage points higher than their peers within the same art unit (Farre-Mensa et al., 2020). This variation in examiner leniency, combined with the high value of patent rights, creates powerful incentives to try to subvert the randomization process at the core of the patent lottery.

Can the patent lottery be gamed? To answer this question, we construct the null distribution of examiner leniency under true randomization. Using data on 6 million patent applications from 2001-2020, we calculate each examiner's grant rate relative to their art unit average. Under random assignment, the distribution of examiner leniency received by any entity providing legal counsel should converge to the population distribution as the number of applications increases. We analyze 1,392 entities with at least 1,000 applications, of which 974 are law firms. We find that the empirical distribution of entity fixed effects differs significantly from what would be expected under true randomization (Kolmogorov-Smirnov test: $D = 0.16$, $p < 0.001$). Many entities show patterns of examiner assignment that cannot be reconciled with random allocation, translating to substantial advantages in patent approval rates.

When can we expect companies to game the patent lottery? We demonstrate that strategic timing of patent submissions can predict application success. Using a simple predictor based on the average leniency of available examiners in the ten days prior to filing, we show that for about 97% of art units, the area-under-curve (AUC) exceeds the random-assignment benchmark of 0.5, indicating that success can be predicted with relatively little uncertainty. This suggests that sophisticated applicants with access to public USPTO data on examiner grant rates and workloads can time their submissions strategically to increase their chances of drawing lenient examiners. Our empirical analysis also reveals that small entities receive examiners who are 1.13 percentage points less lenient on average, while applications with more claims—a proxy for patent value—systematically receive more lenient examiners. These patterns are inconsistent with truly random assignment and suggest multiple pathways through which the lottery can be gamed.

To complement our evidence on the economic relevance of the patent examination process, we also examine whether an investor could earn abnormal returns by trading on information revealed during the patent application process. For this reason, we implement a trading strategy as it provides a powerful way to benchmark the economic magnitude of our findings. If the cross-sectional patterns we document are sufficiently strong, they should translate into predictable variation in stock returns around the disclosure and eventual outcomes of patent applications. Consistent with this prediction, we find that sorting firms into portfolios on the day after patent-application publication based on examiner leniency yields a pronounced return spread across portfolios, and a long–short strategy delivers positive and statistically significant alphas. The magnitude of this spread varies across industries, consistent with the notion that art units map into different technological and industry classifications, and some art units' average success probabilities are more predictable than others. These patterns imply that an investor who conditions on this information could earn economically meaningful abnormal returns by following such a strategy.

What companies should we expect to game the lottery? Our theoretical model addresses this question by adapting Nordhaus (1969) to introduce heterogeneous examiner quality and a costly manipulation technology. Firms compete over lenient examiners using manipulation investments. Firms' investments are determined by their idiosyncratic circumstances. Based on the unique Nash equilibrium solution to the model, we derive two empirically testable predictions. First, firms with lower marginal cost of capital will invest more to manipulate the patent lottery. Because larger firms often have lower cost of capital, this implies that smaller firms will spend less money on manipulation and thus receive examiners with a lower leniency. Second, firms with higher value patent applications have more to gain from a positive grant decision and thus invest more into manipulation. These predictions bear out in the data: we find that applications from smaller entities and those with fewer claims get assigned to examiners with significantly smaller leniency.

Our findings also have important implications for existing empirical research on patents. A widely-used identification strategy in the literature relies on examiner leniency as an instrumental variable for patent grants, assuming that examiner assignment is random (Lemley and Sampat, 2012; Sampat and Williams, 2019). This approach has been employed to study the causal effects of patents on follow-on innovation (Sampat and Williams, 2019), firm performance and entrepreneurial outcomes (Farre-Mensa et al., 2020), inventor mobility (Melero et al., 2020), and various other economic outcomes. We revisit this influential approach and demonstrate that it may produce spurious results. Using the reported coefficients from Sampat and Williams (2019), we construct a data generating process where patents have zero causal effect on follow-on innovation, yet their instrumental variable strategy still yields statistically significant negative effects. The problem arises because unobserved patent quality influences both examiner assignment and innovation outcomes. Even small correlations between quality and examiner leniency, precisely what we document when the lottery is gamed, can generate false positives in two-stage least squares estimation. This calls into question not only the fairness of the patent system but also the validity of a research design that has been employed across dozens of studies examining innovation spillovers, patent value, and firm performance.

This paper makes three contributions to the literature on patent systems and institutional design. First, we provide the first systematic evidence that the patent examiner lottery—long assumed to ensure equal treatment—can be gamed by sophisticated players. Previous work has documented examiner heterogeneity (Sampat and Williams, 2019) and its effects on innovation (Farre-Mensa et al., 2020), but has maintained the assumption of random assignment. Moreover, this randomness assumption has been widely exploited by econometricians as an identification strategy to examine important economic topics ranging from the value of patents to innovation spillovers.[1] We show this assumption fails for a substantial subset of applicants. Our statistical tests reveal patterns impossible to reconcile with true randomization, calling into question both a fundamental premise of patent system fairness and the validity of this common research design.

Second, we develop a theory of when gaming randomized government programs becomes profitable. While the industrial organization literature has extensively studied regulatory capture and lobbying, less attention has focused on manipulation of supposedly random allocation mechanisms. Our model shows that even small advantages in examiner assignment can generate large returns when patent values are high and competition is concentrated. The framework applies beyond patents to any setting where government randomly allocates valuable rights or assigns cases

[1] The patent lottery is a canonical example in guides for using leniency designs (e.g., Goldsmith-Pinkham et al., 2025).

to heterogeneous decision-makers, including immigration courts, disability determinations, and regulatory approvals. While most theoretical work on the patent system has focused on constructing institutions to maximize societal welfare under an assumption of no manipulation, a separate literature examines opportunities for exploitation, commonly through so-called "patent trolls". We add to this literature by identifying a new avenue for exploiting inefficiencies in the system: strategic manipulation of examiner assignment within the patent lottery.

Third, our findings have immediate policy implications for patent system integrity and innovation incentives. Gaming the examiner lottery redistributes innovative rewards from small inventors to large corporations, potentially reducing the incentive for breakthrough innovation by startups and independent inventors. Our evidence shows that small entities systematically receive less lenient examiners, while high-value patents (those with more claims) receive more lenient treatment—patterns that advantage sophisticated repeat players over individual inventors. Simple fixes, such as automated assignment systems with better randomization, regular audits of assignment patterns, and restrictions on examiner-applicant communications, could restore fairness at minimal cost. More broadly, our results suggest that any high-stakes government lottery requires active monitoring to prevent sophisticated actors from subverting randomization.

2 Institutional Background and Data

2.1 The Patent Application Process

Patents on mechanical, electronic, and chemical technologies are called *utility patents*. Such patents consist of a number of *claims*, which define the scope of the asserted invention. After submission, the application is given a technology classification and based on this classification is assigned to an *art unit*. These art units are organized by technology type of the patent application, each covering clusters of related technology subject-matter. By virtue of differences in technologies, art units differ in their size, the length of the review process and the rate of granted patents. The technological classification as well as the art unit assignment depends on the language used in an application's claims and specification. This allows for strategic use of specific language in applications with the goal of steering art unit assignment. Such behavior has limits, however, because an art unit assignment can be changed in the examination and search process after the initial assignment.

Within the art group, the applications are then assigned to individual patent examiners. How this assignment is is handled has changed recently, but is constant throughout our observation period. The procedure is codified in the Manual of Patent Examining Procedure (MPEP). The eighth edition of the MPEP from 2001 puts the assignment under the purview of the examiner supervisors, allowing them relatively free discretion in the assignment. This led to a debate in the literature on whether the assignment is random or which of the characteristics of the patent application are used in the assignment process. Lemley and Sampat (2012) and later Sampat and Williams (2019) argue that the assignment is more or less random, conditional on the current availability of examiners. They base this on their own conversations with examiner supervisors. While Lemley and Sampat (2012) concede that there may be some systematic aspect to the assignment process, they state that supervisors only take examiner familiarity with a technology into account, not the quality of an application.[2] A somewhat different opinion is taken by Righi and Simcoe (2019) who argue that supervisors pay more attention to examiner technical knowledge and specialization. They show some econometric evidence supporting their argument. However, it is clear that even if the individual examiner's knowledge base is taken into account, assignment has to take availability into account, as well.[3] They further show that proxies for patent quality and scope as well as the identity of the applicant are not randomly distributed across examiners. This is in line with our own results. However, while they claim that "there is no evidence that the broadest or most important applications are assigned to specific examiners" (p. 138), we show evidence to the contrary.

After assignment to an examiner, the examiners starts a process of search and examination. Based on this process, if they deem the patent sufficient, it is granted. If not, the patent is rejected

[2]See also Merges (1999) who argues that there is a tradition in the USPTO to treat all patents as equal.

[3]The most current version of the MPEP, instituted in February 2023, changes the assignment procedure and relegates it to an automated system. This system considers both examiner specialization and availability in the assignment, strengthening the hypothesis that both factors played a roll in the previous system, as well.

initially and the applicant, often represented through an attorney or agent, starts a negotiation process with the examiner through a series of submitted documents. This process can end with a granted patent that includes a reduced number of claims or a final rejection of the patent.

The examiners have the sole discretion on granting or rejecting the patent application. They work within relatively strict guidelines and are, in fact urged to grant the patent unless they can find a reason for refusal (Graham et al., 2018). Nevertheless, examiners differ quite substantially in their tendency to grant patents. This has lead to the patent system sometimes being called the patent *lottery* (Farre-Mensa et al., 2020). The argument is that the assignment of the patent examiner is decisively influential for the probability of getting a patent granted. Having a lenient examiner makes a grant more likely and constitutes a "win" in the lottery. This randomness in the patent application process has since then been used as an identification strategy for inferring the causal effect of a patent on various financial and societal outcomes (e.g., Sampat and Williams, 2019; Farre-Mensa et al., 2020; Melero et al., 2020).

2.2 Data

Our main source of data is the USPTO Patent Examination Research Dataset (PatEx) database provided directly from the USPTO (Graham et al., 2018, for details see). We use the 2022 release of these data that covers all patent applications up to the year 2022 inclusively. The data were drawn from the Public Patent Application Information Retrieval (PublicPAIR) system, a system that was put into place following the US Congress passing the American Inventors Protection Act (AIPA) in 2000.[4] AIPA is relevant because its passing drastically increased the number of observable patents and, crucially, made information about failed patent applications public starting November 2000. PatEx thus covers (almost) all patent applications from the year 2001 until 2022 inclusively.[5] The data does not cover those applications for which the applicants sought the opt-out provision under AIPA. How exactly such exceptions should be evaluated is discussed in Graham et al. (2018) and

[4] PublicPAIR has since been replaced by the USPTO Patent Center. For our data, however, it is the relevant system as PublicPAIR was retired in July 2022 and all data used in the analyses below stems from the period before its retirement.

[5] Another possible data source is PatentsView which also includes data on patent applications and extends to more recent years. Crucially, however, it does not list the assigned examiner before a patent is granted. Since this information is necessary for our analyses, we use the PatEx data, instead.

they constitute a inevitable omission for the data in our case (see Raffiee et al., 2023, for an analysis with the same data restrictions).

The initial data from PatEx contains information on slightly over 14 million applications. The data from the 2023 version of PatEx does not track a unique patent examiner ID, but merely identifies the examiner by name. The 2014 version of Patex was the last release that included a unique examiner ID. We use this release to connect the 2022 data to examiner IDs and then use the examiner name to associate the examiner names of the newer applications with an examiner ID.

A critical step in our data preparation involves linking patent applications to their representing law firms. The PatEx database contains attorney names as unstructured text strings with substantial variation in how firm names are recorded—a single law firm may appear under dozens of different text representations due to variations in punctuation, abbreviations, typos, OCR errors, and individual attorney listings. To address this challenge, we developed a comprehensive disambiguation pipeline that combines traditional text processing with large language model (LLM) assistance. The pipeline first applies fuzzy matching with inverse document frequency weighting to identify likely matches, then uses GPT-5-mini to normalize names, correct typos and OCR errors, and classify entities as law firms, in-house counsel, individuals, or government entities. A second LLM pass audits all proposed merges to ensure accuracy. This multi-stage approach successfully links over 6 million applications to identified law firms while maintaining high confidence in the matches. Full details of this attorney name cleaning and law firm disambiguation methodology are provided in Appendix D.

We then make a series of data selection steps. We only keep data with a valid filing date after 2000 and before 2021. With the former restriction, we consider the time period after AIPA and with the latter, we allow patent applications at least two years to obtain a grant decision. We further only consider data with a valid examiner name and art unit and focus on utility patents. After applying these restrictions and requiring successful law firm linkage as described above, we arrive at a base sample of 6,047,474 patent applications. The full data selection procedure is summarized in Table 1.

Table 1: Data Selection Procedure

Step	Description	Applications
0	Raw application data (PatEx)	14,100,378
1	Keep data with valid filing date	14,043,233
2	Keep applications filed after 2000	10,775,875
3	Keep data with valid examiner name	8,738,337
4	Keep data with valid art unit	7,995,530
5	Merge with attorney data, remove applications without law firm record	6,722,923
6	Keep only utility patents	6,258,185
7	Keep applications filed before 2021	6,047,474
Base Sample		**6,047,474**

Note: The table describes the steps in the data selection procedure and indicates the number of applications retained after every step.

To map noisy attorney-of-record strings to law-firm entities, we implement a multi-stage cleaning and disambiguation pipeline that (i) performs Unicode normalization and text standardization, (ii) tokenizes names and computes global token frequencies to obtain inverse-document-frequency (IDF) weights, and (iii) applies IDF-weighted token similarity to merge variants into canonical entities with stable, deterministic IDs. The pipeline relies on the USPTO attorney roster (monthly WebRoster files) and project-relative paths, uses sorted operations for determinism, and writes a one-to-many mapping from raw strings to canonical entities. We restrict the analysis to applications whose cleaned attorney string maps to a law-firm entity at high confidence; unmatched or non-firm entities are excluded. Full details are provided in Appendix D.

PublicPair and its successor system, the Patent Public Search Basic (PPUBS) are also central to the argument in this paper. While the PatEx data is a blanket retrieval from these portals at a later date, PPUBS allows accessing the data on granted patents and published applications on a daily basis. PPUBS can be searched for the latest information published by the USPTO such that patents granted on the previous day and, importantly, those patents' examiners are public knowledge almost in real time.

2.3 Examiner Leniency

The key variable analyzed in this study is the leniency for patent applications exhibited by the patent examiner. The advantage of this variable is that it is supposed to be random, conditional on the art unit and potentially other patent characteristics. This implies that the characteristics of the applicants or agents representing them should not influence the leniency. If it does, then the patent application process is either not impartial or can, in the least, be manipulated by strategic submission strategies.

For every observation in the data, we calculate the average leniency as a leave-out instrument. That is, we consider the average leniency of an examiner for all applications assigned to them in a given filing year except for the current observation. That is, the leniency of application i, analyzed by examiner e and filed in year t is calculated as

$$\text{Leniency}_{i,e,t} = \frac{1}{n_{e,t}-1}\sum_{j=1}^{n_{e,t}} \mathbf{1}(\text{Success}_{j,e,t}) - \mathbf{1}(\text{Success}_{i,e,t}). \tag{1}$$

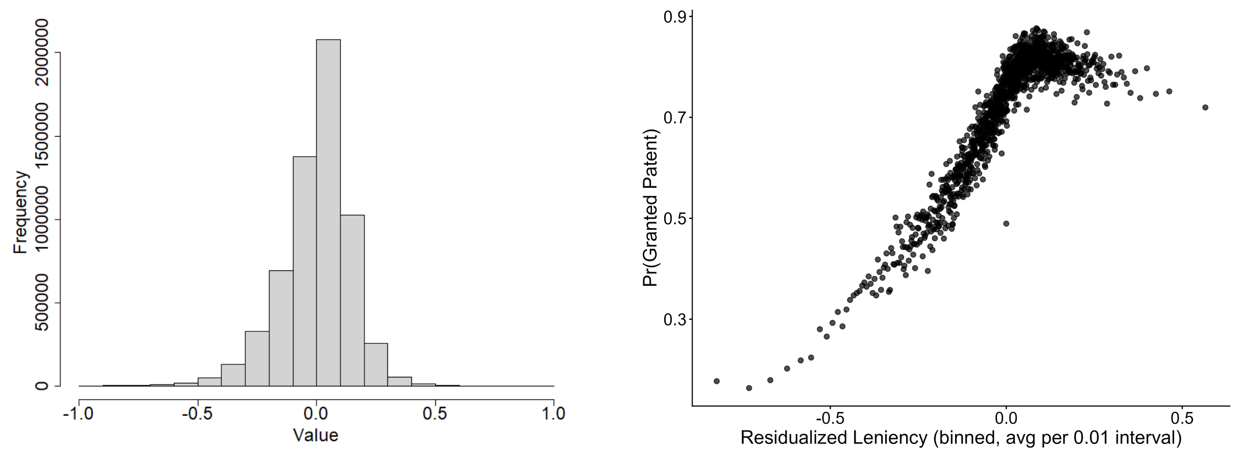
Here, $n_{e,t}$ is the number of patents examined by examiner e in year t and $\mathbf{1}(\text{Success})$ is an indicator function for whether an application is ultimately successful. Calculating the examiner leniency in this way removes data from examiners who only handled a single application. This likely includes data entries with misspelled examiner names. We can calculate an examiner leniency for 6,038,044 applications.

There are some notable aspects to the definition given in (1). First, we calculate the leniency for each examiner on an annual basis. This allows examiners to change their idiosyncratic leniency over time. Secondly, we calculate the examiners' annual leniency as an average over all art units the examiner operates in. Examiners operate in more than one art unit for slightly less than 40% of all examiner-year observations. In these cases, we make the assumption that the idiosyncratic leniency of the examiner does not change between art units. Our analyses nevertheless allow for an art-unit-level effect on the leniency. We simply assume that this effect is not examiner specific. To see the variation induced by the examiner leniency, we plot a histogram of the leniency for all patent applications considered in our main analysis in panel (a) of Figure 1. Data in the histogram is residualized by art-unit × year fixed effects. Even absent these effects, examiners vary significantly in their leniency. Moreover, this variation matters. Panel (b) plots the average success probability for binned values of examiner leniency. There is a clear positive relationship between both variables.

Figure 1: Examiner Leniency: Variation and Prediction Value

(a) Residual leniency: application level

(b) Relationship leniency and success probability



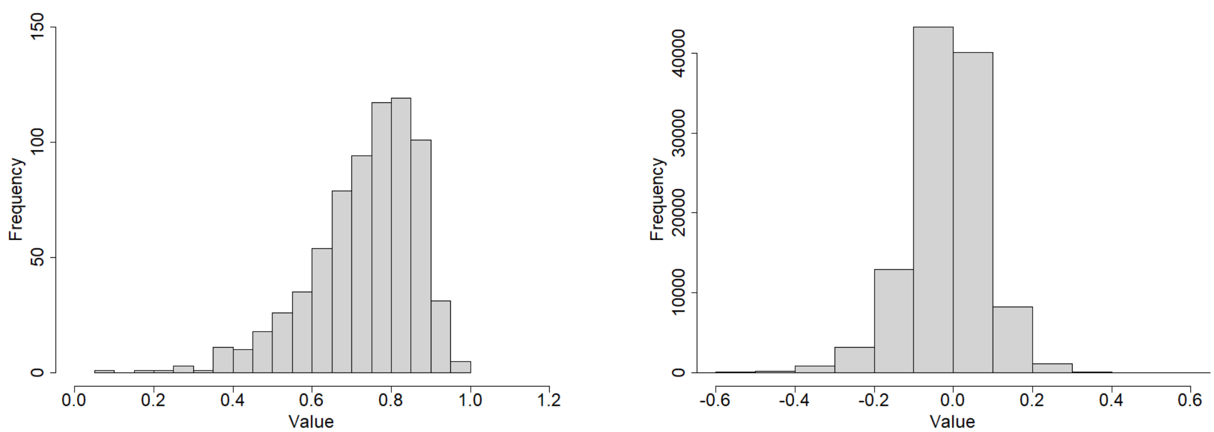
Note: Panel (a) shows a histogram of examiner leniency, calculated according to Equation (1), for n = 6,038,044 patent applications. The values are residualized using art-unit × year fixed effects. Panel (b) shows a binned scatter plot of the examiner leniency and the probability of getting the patent granted. There are 5000 observations per bin.

As mentioned above, it is generally accepted that patent grant rates differ by different technologies and thus differ by art units. Because examiner leniency is calculated from grant probabilities, this effect is reflected in the leniencies, as well. To visualize how much different art units differ in their leniency, we show a histogram of the average leniency per art unit (residualized by year fixed effects) in panel (a) of Figure 2. The strong variation across ar units makes clear why any analysis of examiner leniency has to control for art unit assignment. Righi and Simcoe (2019) argue that examiner specialization exists even below the art-unit level and the most recent version of the MPEP even codifies this. To visualize how much leniency varies by US patent classification (USPC) subclass, we show a histogram over the average residual leniency across the 14,214 subclasses in panel (b) of Figure 2. Data is residualized by art-unit × year fixed effects but nevertheless shows some variation on the subclass level. This variation is, however, significantly smaller than the variation on the examiner level, shown in panel (a) of Figure 1.

Figure 2: Leniency by Art Unit and and USPC Subclass

(a) Average leniency by art unit

(b) Average residual leniency by USPC subclass



Note: Panel (a) shows a histogram of average examiner leniency, calculated according to Equation (1), for n = 707 art units. The values are adjusted by year fixed effects. Panel (b) shows a histogram of average examiner leniency for n = 110,118 USPC subclasses. The values are residualized using art-unit $\times$ year fixed effects. USPC stands for United States Patent Classification.

3 The Effect of Legal Counsel

A first test of whether the entity providing legal counsel to a patent applicant can influence the patent lottery is to see whether the choice of counsel has an effect on the examiner leniency. By virtue of a random process, different entities will have different average examiner leniencies in their patent applications. We are concerned with the question of whether the idiosyncratic effects of the different entities exceed the level implied by a random process. For this purpose, we first derive the theoretical distribution of the entities' effects under the reference model of a random process and then compare this theoretical distribution with the distribution observed in the data. To have a sufficient number of observations per entity, we focus our analysis on those that provide counsel for at least 1,000 applications. Given that any one application can have up to three different entities providing counsel, this results in data on 5.6 million applications from 1,392 entities.

Under the reference model of a random process, we have n i.i.d. draws from a distribution of art-unit-by-year residualized patent examiner leniencies. Denote the i^{th} draw as $X_i \sim f$. We assume that f has finite variance, is not fat-tailed or overly skewed. These assumptions can be verified considering the empirical distribution of f in Panel (a) of Figure 1, which seems to be well behaved. We want to know the average residual examiner leniency of entity $l \in \{1, ..., L\}$. Under the reference model, these entities do not have an influence on the residual leniency of the patent examiner for any given application. Denote I_l as the set of applications with counsel from entity l. Then $\bar{X}_l = \frac{1}{|I_l|} \sum_{i \in I_l} X_i$ represents the average leniency of a given entity. Given our assumptions about f and the central limit theorem, we know that $\bar{X}_l \sim \mathcal{N}(0, \sigma_f^2/|I_l|)$, where σ_f^2 is the variance of the distribution of the examiner leniency.

We can now approach the calculation of the reference model's entity fixed effects in two ways. To achieve an analytical solution, we have to assume that all entities cover the same amount of patents, $|I_l| = c \; \forall \; l$. Under this assumption, we can state the mean of the k^{th} order statistic of the distribution of $\bar{X}$ as

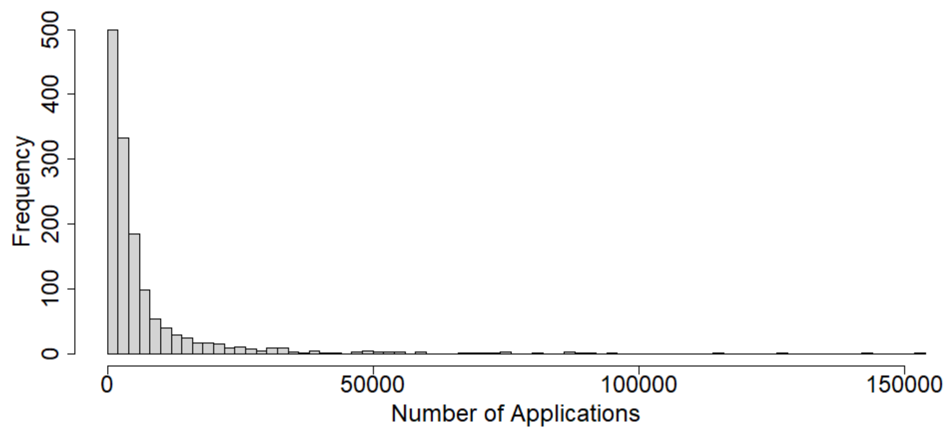
$$\mathbb{E}[\bar{X}_{(k)}] = \frac{\sigma_f}{\sqrt{c}} \mathbb{E}[Z_{(k)}]. \tag{2}$$

Here, $Z_{(k)}$ is the k^{th} order statistic of the standard normal distribution. Its expectation can be approximated as $\mathbb{E}[Z_{(k)}] \approx \Phi^{-1}\left(\frac{k}{L+1}\right)$ with Φ^{-1} being the cdf of the standard normal distribution. Our data dimensions would imply $c = 6,933$ and we can use (2) to calculate 1,392 analytical order statistics using the empirically observed value for σ_f over all observations in the data.

The issue with the analytical order statistics is that in practice, the number of patents per entity are not equal. A histogram of the $|I_l|$ in our data can be seen in Figure 3. The distribution has a mean of 6,933, a standard deviation of 12,679, and a median of 2,946. We thus have a strongly skewed number of patent applications per entity with a large amount of variation. The second approach to calculating the reference model fixed effects is thus numeric. For this, we randomly scramble the examiner leniencies across the 5.6 million patent applications. We then use this randomly allocated dataset to estimate fixed effects for the fictitious law firms. We repeat this process

1,000 times. The replications allow us to calculate the average value for all 1,392 ordered fixed effects.

Figure 3: Number of applications by counsel-providing entity



Note: Histogram of the number of applications our sample of 1,392 entities provided legal counsel for. One application can have up to three firms providing counsel.

To obtain the actual effects of entities on residual examiner leniency, we estimate

$$\text{Residual Examiner Leniency}_i = \sum_{\lambda=1}^{L} \beta_\lambda \mathbf{1}(\lambda \in \mathcal{L}(i)) + \varepsilon_i. \tag{3}$$

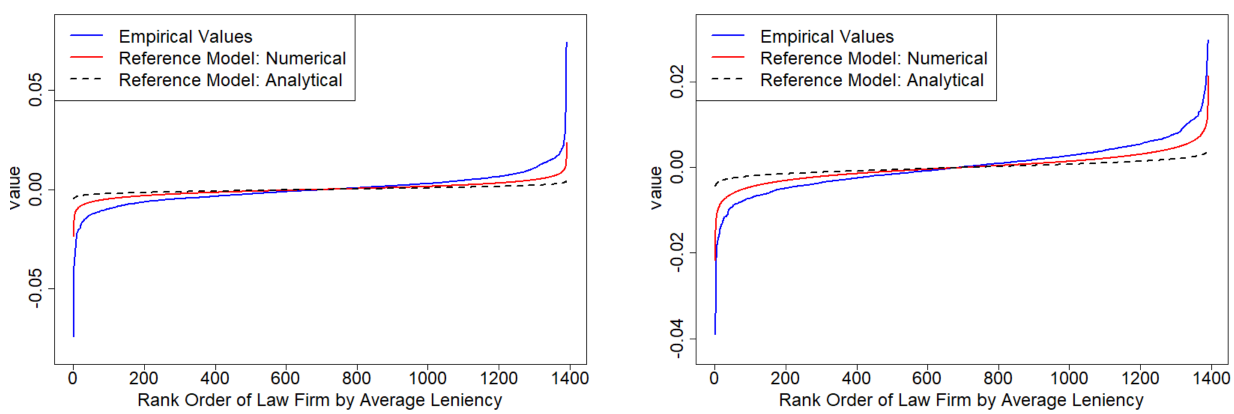
We start at the examiner leniency as calculated in equation (1) for patent application i. We then residualize it by projecting out art-unit $\times$ year fixed effects. We aim to observe the fixed effects for the different law firms, which are given by the different β_λ. $\mathcal{L}(i)$ represents the set of all entities that provide legal counsel for application i. Because multiple entities can be associated with the same patent, we estimate Equation (3) through a sparse matrix approach. For comparison with the analytical and numerical reference models, we order the fixed effects by their size.

The three different distributions of fixed effects are shown in panel (a) of Figure 4. Here we can see that the empirical differences in examiner leniency between the considered entities are considerably larger than would be predicted under the reference model where the allocation of examiners is random conditional on the art unit and the filing year. A Kolmogorov-Smirnoff test validates the visual assessment. It shows statistically significant differences between the empirical distribution and the two distributions derived from the reference model (against simulated values: $D = 0.16$, $p < 0.001$; against analytical values: $D = 0.31$, $p < 0.001$).

Figure 4: Comparison of Empirically Observed Law Firm Effects with Reference Model of Random Examiner Leniency

(a) Art-unit $\times$ Year Residual

(b) Art-unit $\times$ Year + USPC Subclass Residual



Note: The figure considers data from n = 5,605,011 observations from 1,392 entities proving legal counsel. We only consider patent applications associated entities associated with at least 1000 applications in the data. The analytical solution is calculated according to Equation (2). The numerical solution takes different numbers of applications per entity into account, using the empirical distribution. The empirical values come from estimating Equation (3).

Righi and Simcoe (2019) raise the point that patent examiners specialize into specific technological sub-classes of their art units and that these sub-classes have different grant rates associated with them (see panel (b) of Figure 2). In theory, it is possible that counseling entities also specialize on such sub-classes and that what we see in panel (a) of Figure 4 reflects this specialization. To understand whether such a specialization effect drives our results, we repeat the analysis using examiner leniency residualized by both art-unit $\times$ year effects and USPC subclass effects. The result can be seen in panel (b) of Figure 4. While the distribution of empirical fixed effects looks different, the main take-away of the analysis remains the same: entities seem to have a more extreme effect on examiner leniency than can be explained by simply random allocation of patent applications. These visual results are again validated by Kolmogorov-Smirnoff tests (against simulated values: $D = 0.13$, $p < 0.001$; against analytical values: $D = 0.27$, $p < 0.001$).

Who are these entities that produce particularly good (or bad) draws from the patent lottery? To answer this question, we consider the observable characteristics of the 1,392 entities with 1000 or more applications in our data. Specifically, we estimate the equation

$$\text{Rank}_i = \beta_0 + \beta_1 \cdot \ln(\#\text{Apps.}_i) + \beta_2 \cdot \mathbf{1}(\text{IP Stars}_i) + \beta_3 \cdot \text{HHI}_i + \beta_4 \cdot \mathbf{1}(\text{In-House}_i) + \beta_5 \cdot \mathbf{1}(\text{Other}_i) + \varepsilon_i. \quad (4)$$

The rank is the result of the fixed effect analysis displayed in Figure 4. We consider both the fixed effects from leniency residualized with Art-unit $\times$ Year fixed effects and that with additional USPC subclass fixed effects. Entities are variable with regard to their number of applications, which we reflect by considering the logarithm of this number. 61 of the 974 law firms in our sample are listed in the 2020 ranking of the IP Stars specialists guide. According to the guides own information, the ranking is "based on a weighted review of information submitted by firms, publicly available information, and market feedback."[6] The Art Unit HHI is a concentration measure for the applications of a given entity. A higher value implies that the entities focus on a more narrow set of art units, implying greater familiarity with both the subject matter and the bureaucratic idiosyncrasies of the art units in question. In order to make the results more interpretable, we normalize the variable to mean zero and a standard deviation of one. When all entities are considered, law firms (n=974) are the reference category in the estimation. In addition, there are 363 in-house counsels in our data. Other entities includes individuals (n=34), the government (n=9) and unclassified entities (n=12).

Results of the estimation are given in Table 2. Throughout all specifications, we can see that the influence of the number of applications is limited. When only law firms are considered and the stricter resiualization is applied, the influence is negative. This could stem from the fact that law firms which are more adapt at manipulating the lottery are also more expensive and thus take on fewer applications. However, given the lacking significance in other specifications, this result should not be overinterpreted. In contrast, the IP Stars ranking is always positive and highly significant, irrespective of the specification. The ranking either already takes the positive

[6] See https://www.ipstars.com/ for futher details. We accessed the 2020 data of the ranking using the wayback machine (https://web.archive.org/). It was the earliest ranking accessible in this fashion.

Table 2: Determinants of Entity Fixed Effect Rank

	Dependent Variable: Rank of Entity Fixed Effect			
	(1)	(2)	(3)	(4)
Constant	610.8***	668.8***	829.1***	919.7***
	(90.07)	(103.4)	(91.10)	(106.6)
Log(No. Applications)	7.595	1.865	-17.93*	-29.56**
	(10.66)	(12.43)	(10.76)	(12.82)
IP Stars Ranked (2020)	167.4***	184.4***	183.8***	197.5***
	(45.64)	(46.60)	(46.49)	(47.64)
Art Unit HHI	56.88***	108.9***	28.76**	6.805
	(16.87)	(37.46)	(14.62)	(37.68)
In-House Counsel	60.74**		26.15	
	(28.21)		(27.66)	
Other Entity	9.275		-17.81	
	(60.73)		(62.68)	
Entity Types	All	Law Firms	All	Law Firms
Residualized	Art Unit × Year	Art Unit × Year	Art Unit × Year + Subclass	Art Unit × Year + Subclass
R^2	0.03516	0.02043	0.01541	0.01474
Observations	1,392	974	1,392	974

Note: Table shows the results of estimating Equation (4) either with data on all entities (columns (1) and (3)) or with data for law firms only (columns (2) and (4)). The rank is the result of the fixed effect analysis displayed in Figure 4. For columns (1) and (2) we consider the fixed effects from leniency residualized with Art-unit × Year fixed effects. For columns (3) and (4) we additionally add USPC subclass fixed effects. 61 of the 974 law firms are listed in the 2020 IP Stars ranking. The Art Unit HHI is a concentration measure for the applications of a given entity. This variable is normalized to mean 0 and standard deviation 1. When all entities are considered, law firms are the reference category. Other entities includes individuals (n=34), the government (n=9) and unclassified entities (n=12). Standard errors are heteroscedasticity robust. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

outcomes into account or better reputed firms are better at influencing the patent lottery.

The concentration measure is interesting. When only residualizing by art unit and year, it is highly significant and positive, even more so when only law firms are considered. This shows that concentration on a few technological fields increases the effectiveness at influencing the patent lottery. This effect becomes smaller and even insignificant among law firms, when the USPC subclass is added to the residualization process of the leniency. We know from Righi and Simcoe (2019) that examiners (and thus their leniency) differ by these subclasses, at least for some art units. Strategic use of language when writing claims in the patent application can influence the subclass allocation of the application. Comparing the coefficient on the Art Unit HHI between the first two and the last two columns indicates that the institutional knowledge necessary for such strategic behavior might be what drives the effect of art unit concentration on entity rank. Lastly, we can see that in-house counsels are better than law firms at gaining an advantage in the patent lottery. This effect, however, is not stable across different residualizations. Additionally, it is not homogeneous. Because the IP Stars coefficient is so much higher than that for in-house counsel, we can see that highly reputable law firms are actually better at manipulating the patent lottery than the average in-house counsel.

There are, of course, caveats to our analysis. First, we do not know whether potential specialization effects of counseling entities and patent examiners are adequately controlled for using the USPC subclass.[7] Even though there are over 110000 subclasses in the data, the differentiation might not be fine enough (at least for some art units). Further, even if we interpret the analysis in the way that patent applications represented by certain law firms get assigned more lenient examiners, we still do not know how these firms achieve this outcome. To address the former issue, we turn to application-level evidence that does not use information on law firms below. We shed light on how law firms manipulate the patent lottery in Section **??**.

4 Application-Level Evidence

The previous section provided evidence that law firms and in-house counsel can manipulate the patent lottery. We now consider which applicants are particularly likely to invest into such entities. We first provide theoretical predictions for this question using a simple game-theoretic model in Section 4.1. The comparative statics from this model provide hypotheses which are then analyzed empirically in Section 4.2.

[7] See the results in Righi and Simcoe (2019) for some support of this hypothesis.

4.1 Model

We use a simplified version of the Nordhaus (1969) model in the notation of Budish et al. (2016). The model considers the societal trade-off between incentives given to firms for innovating through a period of time in which their innovations are protected by patents and the welfare losses due to the monopoly rents induced by patents in that time period. For this, the model considers a collection of firms that differ on their cost of innovation, the chance that innovation succeeds and the monopoly rents they can obtain from a patent. Other than is typically the case, we are not interested in finding a societal equilibrium for the optimal patent protection length. Rather, we use the model's setting to analyze the effect of a random element in the patent system and the possibility to manipulate this element.

To focus on this aspect, we use a series of simplifying assumptions to keep the analysis tractable and the exposition intuitive. First, we assume that there are only two firms that innovate, $i \in \{1, 2\}$, and each firm has a single possible innovation to patent. Because it is inessential for the purpose of the model, we assume that if a firm attempts to innovate, it is always successful. If a patent application is successful, the firm receives increased profits due to the patent protection. If the application fails, there is free firm entry to the technology and competitive pressure results in reduced profits for the firm. We denote the net difference between these two profits as π_i. When using this term, we abstract from the exact length of the patent protection period. π_i can be seen as the discounted total net benefit to the firm of a patent protected innovation.[8]

Cost of innovation for a firm are given by c_i. We consider firms with innovations which can differ in costs and potential benefits, but do not differ in their *ex-ante* probability of succeeding in the patent lottery. Absent manipulation, both patent applications have probability $p_i = \bar{p}$ of succeeding. Because there is no manipulation, the probability of succeeding in the patent lottery is not influenced by firm or application characteristics. While independence from application characteristics is a simplification in the model, it is not material for the empirical test, in which we

[8]Note that this benefit is private for the firm and does not represent the societal benefit of the innovation. It appears because of monopoly protection for the firm and thus comes at the expense of the rest of society. Given the deadweight loss inherent to monopolies, total welfare is larger if the application fails than if it succeeds. Nevertheless, a probability of success has to be present, because otherwise the firms would have no incentive to innovate.

simply use examiner leniency to measure p_i.[9] Given the probability of succeeding in the manipulation free patent lottery, a risk neutral firm i innovates if $\bar{p}\pi_i \geq c_i$. Since the model uses $p_i = \bar{p}$, there are no comparative statics of p_i with regard to firm characteristics.

Comparative statics do appear if firms can manipulate the patent lottery. We model this by letting them spend expenses e_i at marginal cost k_i. A higher level of e_i increases the probability of a granted application, while spending of the other firm (e_{-i}) decreases it. We assume that $p(e_i, e_{-i})$ is twice continuously differentiable and that expenses have decreasing benefits as well as a negative cross derivative such that $p_1 > 0, p_2 < 0, p_{11} < 0, p_{22} > 0$ and $p_{12} = p_{21} < 0$. As argued in the introduction, the nature of the manipulation is to spend ressources such that the application is assigned to a more lenient patent examiner. Because the pool of examiners is fixed (at least in the short term), manipulating the lottery in this form is a zero-sum game. As such, it has to hold that

$$p(e_1, e_2) + p(e_2, e_1) = 2\bar{p}. \tag{5}$$

If a firm expends money to get a more lenient patent examiner, the other firm must get the less lenient examiner. Since the expenses are targeted at the selection of the examiner rather than at influencing the examiners themselves (for example through bribes), the average leniency in the system does not change.

We can now see that firms innovate if

$$\pi_i p(e_i^*, e_{-i}^*) - k_i e_i^* \geq c_i \tag{6}$$

where asterisks indicate the equilibrium solution to the manipulation game played by the two firms. The equilibrium of the game is given in the following proposition.

Proposition 1. *Conditional on both firms innovating,* $p_1(0,0) > \max\left\{\frac{k_1}{\pi_1}, \frac{k_2}{\pi_2}\right\}$ *and* p_{12} *not being too*

[9] Alternative model approaches in which $\bar{p}$ is a baseline probability that is then adjusted based on the characteristics of the innovation are possible but will only complicate the exposition without any further information. The advantage of using examiner leniency as a measure for the probability of patent lottery success is that it is independent from the characteristics of applicant and application unless there is manipulation, which is the exact hypothesis our empirical analysis aims to study.

negative, the unique Nash equilibrium of the patent lottery manipulation game is given by

$$p_1(e_1^*, e_2^*) = \frac{k_1}{\pi_1} \qquad \textit{and} \qquad p_1(e_2^*, e_1^*) = \frac{k_2}{\pi_2}. \tag{7}$$

All proofs are provided in the appendix. An interior solution implies that both firms spend money on manipulating the patent lottery. While we cannot show this claim explicitly in the empirical data, we can show evidence that is consistent with it. For this, we derive the comparative statics of the Nash equilibrium. Specifically, we are interested in how the grant probability of a patent changes when the parameters of the firm or its application change. Showing support for these results in the data will give credibility to the claim that manipulation does indeed exist.

Proposition 2. *Under the conditions of the interior Nash equilibrium,* $\frac{\partial p(e_i^*, e_{-i}^*)}{\partial k_i} < 0$ *and* $\frac{\partial p(e_i^*, e_{-i}^*)}{\partial \pi_i} > 0$.

This proposition provides us with two testable hypotheses. $\frac{\partial p(e_i^*, e_{-i}^*)}{\partial k_i} < 0$ states that firms with higher marginal costs of influencing the patent lottery will invest less into the patent lottery and are thus less likely to have a patent examiner with a low leniency assigned to them. In the real world, differences in marginal costs will likely appear due to differences in liquidity or access to informal networks. Both are more likely to be present in large firms. Proposition 2 thus implies that smaller firms will, on average, have less lenient examiners. $\frac{\partial p(e_i^*, e_{-i}^*)}{\partial \pi_i} > 0$ states that firms with patent applications that have a higher value once they are granted will invest more in the patent lottery and are thus more likely to have a lenient patent examiner. We test this hypothesis using the number of claims in a patent as a proxy for the patent's quality.

4.2 Empirical Evidence

We test the hypotheses derived from the theoretical model using proxies of firm size and patent quality that are included in the USPTO data. For the cost of capital, we already argued that individual inventors or smaller firms will have a harder time raising sufficient funds to influence the patent lottery. The USPTO tracks the size of application through a small entity indicator. When an individual, a nonprofit organization or a firm with fewer than 500 employees applies for a patent, the application fee is reduced by 50%. We use this indicator as a proxy for cost of capital, building on the well-established result that the cost of capital is decreasing in firm size (e.g. Petersen and Rajan, 1994; Fama and French, 1995; Bao et al., 2011). The USPTO grants an even greater discount of 75% if applicants meet more strict requirements on the number of previous applications and, when applicable, the revenue of a company. In this case, they distinguish between "small" and "micro" applicants. However, since only around 70,000 applications meet the criteria for "micro" in our data, we report results both for a summarizing category of discounted applicants and for the full differentiation.

For the quality of the patent, we follow Sampat and Williams (2019). They argue that the quality measure needs to be known at the time of the patent application. This disqualifies *ex-post* measures such as forward citations (Trajtenberg, 1990), patent renewals (Schankerman and Pakes, 1986), patent litigation (Harhoff et al., 2003), or excess stock returns (Kogan et al., 2017). They settle on two quality indicators, one of which is immediately available from the USPTO: the number of claims made in a patent. This is also the quality indicator utilized by Farre-Mensa et al. (2020).

Another commonly considered characteristic of patent applications is the applications' scope. Kuhn and Thompson (2019) propose the length of the first claim in a patent application as an inverse measure of scope. The argument is that shorter claims imply fewer qualifiers or conditional statements, thus broadening the scope of the potential patent. Note that both in their own work and in the works of others, it is argued that this measure is different from application quality (e.g. Righi and Simcoe, 2019). We thus expect the parameter π_i from the model to be captured by the number of claims, not by the scope measure.

We estimate the equation

$$\text{Leniency}_{i,a,t} = \beta_1 \cdot \mathbf{1}(\text{Discounted}_{i,a,t}) + \beta_2 \cdot \text{No. Claims}_{i,a,t} + \beta_3 \cdot \text{Scope}_{i,a,t} + \nu_{a,t} + \varepsilon_{i,a,t}. \tag{8}$$

Here, the leniency of the examiner of patent application i in art unit a and year t is a function of the small entity indicator, the number of claims and the applications scope, measured inversely by the length of the application's first claim. The estimation features art-unit-by-year fixed effects and the standard errors are clustered on the level of the art unit. Summary statistics for the dependent and independent variables are provided in Table 3. We analyze data on over 5.7 million patent applications. The applications show an average leniency of 71.4% and about one fifth of them are made by applicants qualifying for a discounted fee. The median application has 20 claims and while the average is slightly below the median, the range of this variable is between one claim and over 8,000 claims. The scope is also fairly symmetric with an average number of 774 characters in the first claim of the patent.

Table 3: Summary Statistics on the Application-level

	n	Mean	St.Dev.	1. Quartile	Median	3. Quartile
Examiner Leniency (in 100%)	5,708,617	71.41	20.91	59.52	76.71	87.69
Discounted Application	5,708,617	0.216	0.412	-	-	-
No. Claims	5,708,617	19.46	26.28	12	20	21
Scope	5,708,617	774.60	831.55	413	658	979

Note: Examiner leniency is defined as the average probability of an application's examiner to grant a patent in the applications' year and art unit. The application is discounted if it comes from an individual, a nonprofit organization or a company with fewer than 500 employees. No. Claims is the number of claims listed at the time of the application. Scope is the number of characters in the first claim of the application at the initial time of application.

Table 4 shows the results of the estimation. The estimates show clear and highly statistically significant support for the hypotheses derived in the theoretical model. On average, an application made by a small entity gets assigned a 1.13 percentage point less lenient patent examiner. Similarly, the number of claims increases the leniency of the examiner such that an additional claim leads to a 0.003 percentage point higher leniency. While both point estimates are small, it has to be kept in mind that we are using one-dimensional proxies for both quality of the patent and the marginal cost of influencing the lottery. Both concepts are, however, highly complex and multi-dimensional. Attenuation bias thus likely decreases the point estimates provided here. The fact that the coefficients' test statistics are nevertheless so high points towards the validity of our model predictions. On the contrary, the scope of the application is not statistically significantly associated with the leniency of the examiner. This is in line with our model that does not offer any predictions on this application characteristic.

Table 4: Application-level Evidence for Model Hypotheses

	Dependent Variable: Examiner Leniency (in 100%)					
	(1)	(2)	(3)	(4)	(5)	(6)
Discounted Appl.	-1.130^{***}				-1.132^{***}	
	(0.104)				(0.104)	
Small Entity		-1.090^{***}				-1.092^{***}
		(0.096)				(0.096)
Micro Entity		-1.838^{***}				-1.835^{***}
		(0.312)				(0.312)
No. Claims			0.003^{***}		0.003^{***}	0.003^{***}
			(0.001)		(0.001)	(0.001)
Application Scope				0.00001	0.00001	0.00001
				(0.00003)	(0.00003)	(0.00003)
Fixed effects	Art unit × Year	Art unit × Year	Art unit × Year	Art unit × Year	Art unit × Year	Art unit × Year
Clustered st. err.	Art unit	Art unit	Art unit	Art unit	Art unit	Art unit
Observations	5,708,588	5,708,588	5,708,588	5,708,588	5,708,588	5,708,588

Note: Table shows the results of estimating Equation (8) either in full (columns (5) and (6)) or with each explanatory variable in isolation (columns (1) through (4)). Examiner leniency is defined as the average probability of an application's examiner to grant a patent in the applications' year and art unit. The application is discounted if it comes from an individual, a nonprofit organization or a company with fewer than 500 employees. Micro entities get a larger discount than small entities. No. Claims is the number of claims listed at the time of the application. Scope is the number of characters in the first claim of the application at the initial time of application. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Less than 7% of the applicants who receive a discount are micro entities by the definition of the USPTO. These micro entities are even smaller than the typical small entities receiving a discount. Amon others, one condition to qualify is that the gross income of the entity cannot be more than three times the median household in the previous year. Such entities likely have even larger costs of capital than mere small entities. Our model would thus predict a lower leniency for their application. This prediction is supported in the data, both in the univariate and the multivariate analysis, as can be seen in columns (2) and (6) of Table 4. This lends further support for our model and the ultimate conclusion that the patent lottery can be manipulated by applicants (at a given cost).

From the argument of Righi and Simcoe (2019) and the descriptive analyses of Section 2, we know that the USPC subclass explains variability in the examiner leniency. For the purpose of our analysis, however, it is unclear whether this is variability we aim to explain using application characteristics or whether it should be projected out beforehand. If applicants (or their counsel) manipulate the patent lottery by strategically choosing claim language such that the patent application is given a USPC subclass with more lenient examiners, then we should not control for the effect of the subclass in our analysis. If the manipulation uses a different channel, then controlling for subclasses should not affect the results. As can be seen in Table B1 of the Appendix, including subclass fixed effects decreases the coefficients of our hypothesis variables, but does not leave them insignificant. We can thus conclude that likely both strategic subclass assignment and alternative channels are at work.

5 Predicting success of a patent application

We now turn to the question how the patent lottery can be manipulated. Several channels are possible. An immediate thought is that entities might use illicit means and unofficial channels to influence the patent examiner assignment. However, the nature of the patent lottery and the (almost) real time publishing of patent decisions in PublicPair and PPUBS allows for a different, completely legal way to get a favorable assignment. Because there is a limited number of examiners in any art unit and because applications are assigned taking examiner availability into account, an application can be timed strategically so that only more lenient examiners are available at the time of submission. Below, we use a relatively simple prediction model to show that such strategic submission timing is in fact possible.

We approach the issue as follows. If strategic submission timing was not feasible, we would be unable to predict the success of an application based on timing alone. The null hypothesis would thus be that conditional on the art unit, the success of an application on time t could not be predicted based on previously publicly available information. We thus test whether a simple time-t conditioning variable predicts the subsequent success of a newly filed patent application. Our predictor is expected leniency, which is constructed strictly prior to filing, uses no information of the focal application other than its art unit and summarizes contemporaneous leniency. Evidence of predictability rejects the null hypothesis and speaks against independence of the patent process. The setup mirrors market timing: an investor uses a time-t signal (e.g., the dividend yield) observed today to forecast future returns. Likewise, a law firm or in-house counsel observes expected leniency before filing and forecasts the application's success. Applications with higher expected leniency should have higher success probabilities.

Let i index applications, e examiners, a art units, and t_i the filing date (day) of application i. Outcomes resolve on application status dates s. All "as-of" quantities use end-of-day logic, that is events on the filing day t_i are excluded. We write $x(s^-)$ for "strictly before day s" and t^- for "strictly before the filing day."

Examiner leniency. Define a rolling grant-rate ("leniency") step function for each (e, a) based only on prior days as

$$\ell_{e,a}(s^-) \;=\; \frac{\#\{\text{grants for } (e,a) \text{ with status date } < s\}}{\#\{\text{disposals for } (e,a) \text{ with status date } < s\}}\,, \tag{9}$$

updated only on status dates s. We require at least $K_{\min} = 10$ prior disposals for $\ell_{e,a}(s^-)$ to be defined; otherwise it is missing. If an examiner has multiple disposals on a day, that day yields the same $\ell_{e,a}$ for all disposals that day. Note that the leniency of an examiner is calculated for every art unit separately. Thus, the approximately 40% of examiners who handle applications in multiple art units have more than one leniency value.

Expected leniency. For application i in art unit $a(i)$ with filing day t_i, we fix a ten-day lookback window, so the conditioning interval is $[t_i - 10,\, t_i)$, left–inclusive at $t_i - 10$ and excluding the filing day t_i. For each examiner e, let $\ell_{e,a(i)}(s^-)$ denote the examiner–art–unit leniency computed strictly before day s. Define the set of distinct examiners $\mathcal{E}_i(10)$ as those who, within $[t_i - 10,\, t_i)$, have at least one disposal in $a(i)$ for which $\ell_{e,a(i)}(s^-)$ is defined. For brevity, we henceforth write $\mathcal{E}_i \equiv \mathcal{E}_i(10)$. For each $e \in \mathcal{E}_i$, let $s_e^*(i)$ be e's latest eligible day in this window (the most recent $s \in [t_i - 10,\, t_i)$ for which $\ell_{e,a(i)}(s^-)$ is defined). The expected leniency for application i with filing day t_i is the cross–examiner mean of these latest leniencies:

$$\text{ExpLen}_{i,10} \;=\; \frac{1}{|\mathcal{E}_i|}\sum_{e\in\mathcal{E}_i} \ell_{e,a(i)}\big(s_e^*(i)^-\big). \tag{10}$$

This quantity is defined only if $|\mathcal{E}_i| \geq 1$; otherwise it is set to missing.

Out-of-sample estimations. Let $Y_i \in \{0,1\}$ indicate whether application i is ultimately granted. As described in Section 2, we limit our dataset to all applications whose last application status is prior to 2021 to ensure that the application process has been completed. Our baseline predictive relation treats expected leniency as the independent variable:

$$\Pr(Y_i = 1 \mid \mathcal{F}_{t_i^-}) \;=\; \beta_0 + \beta_1\,\text{ExpLen}_i \;+\; \varepsilon_i, \tag{11}$$

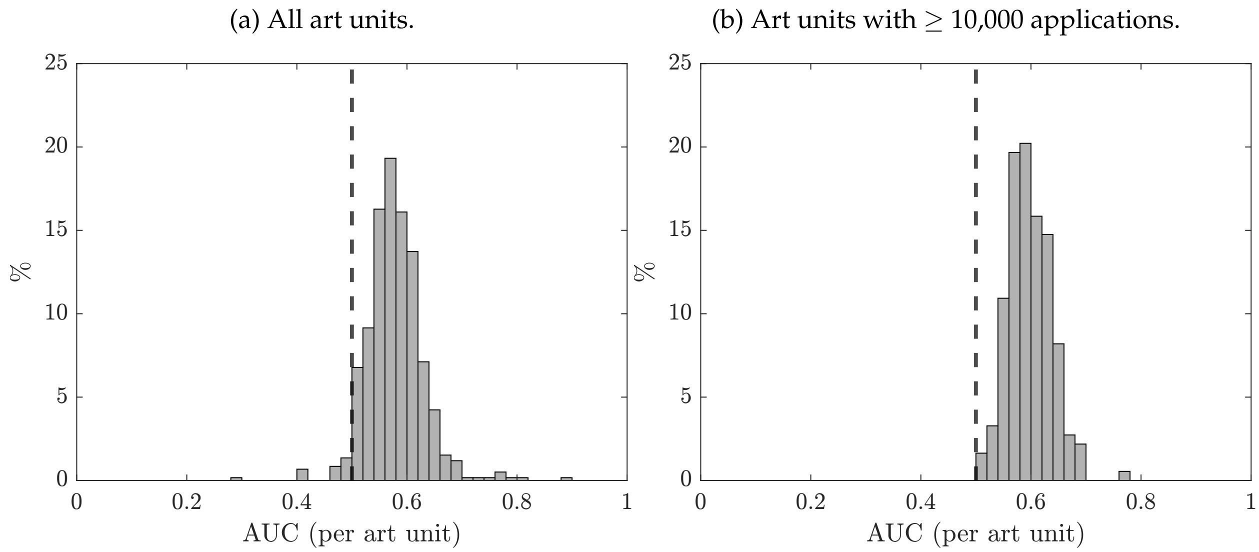
where $\mathcal{F}_{t_i^-}$ denotes the information set strictly prior to the filing day. The coefficient β_1 captures how strongly expected leniency maps into subsequent success.[10]

To simulate a real world prediction problem, in which a law firm would use all available information to predict the subsequent success of a patent application, we implement an out-of-sample (OOS) procedure. For each day D in the time series (with a 252-day training window), we estimate a linear probability model on the training set containing all filings strictly before D and record the estimated parameters $(\hat{\beta}_0, \hat{\beta}_1)$. The training set must contain at least one success and one failure. We then apply these estimates to the test set containing all filings on day D to obtain predictions $\hat{Y}_i$. This rolling-window exercise is repeated day by day until the end of the time series. A pooled ROC area-under-the-curve (AUC) is computed from all $(Y_i, \hat{Y}_i)$ pairs across OOS days.

Figure 5 reports the results. We run the prediction exercise separately by art unit to avoid a pooled model that would mainly identify the application's art unit rather than predict success. The figure shows AUCs for all art units meeting the minimum data requirements (588 of 720).

[10]By specifying expected leniency as the sole explanatory variable, our approach intentionally preserves a parsimonious and transparent structure, thereby enhancing interpretability. Naturally, a commercially used model could have improved predictive performance by incorporating additional covariates or employing more sophisticated machine learning methods.

Figure 5: Distribution of Predictive Quality across Art Units



The figure shows the out-of-sample area-under-the-curve (AUC) per art unit. In Panel (a), the sample includes all art units meeting the minimum data requirements described above, whereas in Panel (B), the sample includes all art units with at least 10,000 applications. The time horizon of the sample is from 2014 to 2020.

For about 97% of art units, the AUC exceeds the random-assignment benchmark of 0.5, indicating that even across all art units success can be predicted with relatively little uncertainty. Restricting to art units with at least 10,000 applications yields 183 units (nearly 60% of all applications) of which all exceed an AUC of 0.5. These findings suggest that the patent process is not independent. Instead, strategic timing of applications allows for manipulating the patent lottery. This possibility is inherent in the fact that the assignment of the examiner to a patent application has to take availability of the examiner into account for the sake of efficiency. As long as public information systems provide information about recently completed assignments of examiners, such information is likely to be exploited.

It has to be noted that the model employed here is deceptively simple and uses very little data. A law firm with access to more extensive data can use this model or more sophisticated methods to predict application success even better. Since our model is already able to provide meaningful predictions for almost all art units, such better models are likely able to make even better predictions for an even larger range of art units.

6 Trading based on examiner leniency

Is there value in getting assigned a more lenient patent examiner? Previous research has established that obtaining a patent has value (Farre-Mensa et al., 2020) and that successful patent filings are associated with higher market valuations (Pakes, 1985). At the time that the application is published, information on the examiner should thus be valuable to the market. A more lenient examiner promises a higher chance at a patent and thus a higher market valuation. Motivated by this insight, we construct a trading strategy that exploits cross-sectional variation in examiner leniency at the time a patent application becomes public. Our objective is to test whether firms whose patent applications are reviewed by more lenient examiners subsequently earn higher stock returns, consistent with markets valuing the increased likelihood of successful patent outcomes.

We begin by aligning each patent application with the firm's trading record. We merge the application data with firm identifiers using the matching tables provided by Arora et al. (2017) and Arora et al. (2021). Because their matching table is available only through 2021, we restrict the patent sample to applications whose status dates are no later than 2021. When multiple applications are published on the same day for the same firm, we aggregate them to a single firm-day observation by taking the mean leniency measure and the maximum of the implied holding windows. For every application, we define the opening date of the investment position as the next trading day after the publication date. The position is held until the next trading day following the end of the application's resolution window: the patent issue date for successful applications or the application status date for unsuccessful ones. Thus, all investment decisions are made using information available only with a one-day lag ($t + 1$), ensuring a conservative and implementable trading rule.

For the corresponding stock returns, we use CRSP and apply standard filters. We restrict the sample to common equities (share codes 10 and 11) listed on NYSE, NYSE MKT, NASDAQ, or ARCA. Returns are adjusted for delistings following Shumway (1997). We further remove firm–day observations with arithmetic returns below –100%, trading volume below 100 shares, or stock prices below $1.

On each publication day, we sort all firms with published applications into five quintile portfolios based on the examiner leniency. Quintile 1 (Q1) contains firms exposed to the strictest examiners (bottom 20th percentile), and Quintile 5 (Q5) contains firms exposed to the most lenient examiners (top 20th percentile). For each quintile, we compute daily value-weighted returns across all stocks that are active, which are all whose open and close dates bracket the current trading day, and we require at least 100 stocks per quintile-day for a return to be included in our analysis. To capture the cross-sectional spread in returns associated with examiner leniency, we also construct a long-short (LS) portfolio that buys Q5 and sells Q1, inheriting the same timing and holding-period rules as the quintile portfolios. Because the LS portfolio contrasts two groups of firms that both experience the publication of patent applications, it nets out the mere effect of publication and isolates the return differential associated with examiner leniency. Statistical significance of this return spread therefore provides evidence that variation in examiner leniency is reflected in expected stock returns.

We evaluate risk-adjusted performance by regressing value-weighted portfolio excess returns on the Fama-French market return factor as well as 17 value-weighted Fama-French industry portfolios.[11] The industry controls are important given the strong alignment between technology fields (art units) and industry classifications, which may otherwise confound the estimated leniency effect. We report the resulting intercepts (alphas) in basis points per day.

Table 5 reports the regression results. All quintile alphas are positive and statistically significant. Notably, the first quintile exhibits an alpha that is higher than expected given its relatively low leniency, whereas the remaining portfolios display a clear monotonic increase in alphas across quintiles. This non-monotonicity at the lower end implies that the long–short portfolio does not capture the maximal possible spread between high- and low-leniency firms, although the resulting long–short alpha remains positive and statistically significant.

We list the full result of the regression, including the industry controls, in Table B2 of the Appendix. It shows that the *Machinery* industry contributes the strongest positive exposure, indicating that leniency-based return differences are particularly pronounced among machinery firms.

[11] We use the returns of the market factor and industry portfolios on Kenneth French's website.

Table 5: Value-weighted alphas of trading strategy

	Q1	Q2	Q3	Q4	Q5	Q5-Q1
Alpha (bps)	1.880^{***}	1.477^{***}	1.776^{***}	2.148^{***}	2.803^{***}	0.924^{**}
	(0.250)	(0.316)	(0.401)	(0.422)	(0.346)	(0.365)
Mkt − rf	0.855^{***}	1.112^{***}	1.013^{***}	1.340^{***}	1.373^{***}	0.517^{***}
	(0.106)	(0.143)	(0.126)	(0.141)	(0.109)	(0.124)

Note: Table displays the daily alphas (in bps) of portfolios sorted on examiner leniency. Alphas are the intercepts from time-series regressions of the value-weighted portfolio (excess) returns against the market factor and the 17 value-weighted Fama-French industry portfolio returns. Industries include Automobiles, Consumption, Construction, Finance, Food, Chemicals, Consumer Durables, Fabricated Products, Machinery, Mining, Oil, Retail Stores, Steel, Textiles, Transportation, Utilities, and Other. The coefficients of the industry portfolios are not shown due to brevity. Table B2 in the Appendix shows the table including the estimated coefficients for the industries. Standard errors are calculated following Newey and West (1986), where we choose the optimal truncation lag as suggested by Newey and West (1994). The standard errors are provided below the respective alphas. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

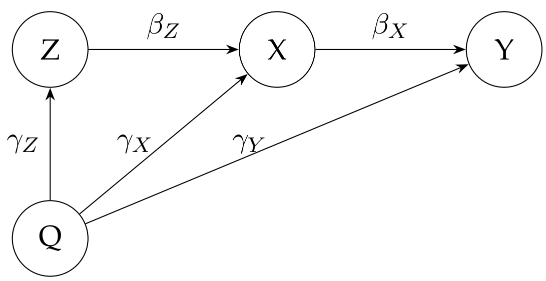
In contrast, the *Consumption* industry exhibits the most negative loading, suggesting that cross-sectional return differences driven by examiner leniency are substantially reduced - or even reversed - within this sector.

Findings on the trading strategy indicate that examiner leniency contains economically and statistically meaningful information about firms' expected stock returns. Statistically significant alphas suggest that financial markets only partially incorporate the effect of examiner assignment on the expected success of patent applications, creating systematic return differences that can be exploited by a simple trading strategy.

7 Revisiting Sampat and Williams (2019)

What do the results of this study imply for research using the patent lottery as a source of exogenous variation? In their paper, Sampat and Williams (2019, henceforth SW) consider the influence of patents on follow-up innovation. They use the patent examiner leniency (henceforth Z) as an instrument for the probability of a granted patent (henceforth X). In light of our findings, this seems questionable. Interestingly, in Section IV.D, SW acknowledge the existence of a quality dimension to the invention underlying the patent application, which we denote Q. They provide a test of the

Figure 6: Causal Influence Diagram for Patent Examiner Leniency IVs



Note: The diagram shows the causal influences considered in a typical patent examiner randomness designs, but adds the quality dimension Q and its potential influence on examiner leniency (Z), the probability of a granted patent (X) and the final outcome (Y).

quality dimension and argue that based on this test, its importance can be discarded – a conclusion at odds with our results. Below, we reconsider their empirical set-up, their test of the quality dimension, and their empirical identification strategy. We argue that given the values provided in the paper, one can construct data in which there is no influence of a patent grant on follow-up innovation but the identification strategy nevertheless results in a statistically significant result.

SW provide three key values. Specifically, in a regression $X = \tilde{\beta}_z Z + \eta$, they obtain $\tilde{\beta}_Z = 0.876$ with a standard error of 0.037 (p. 222). In a regression of X on the quality dimension, they obtain a statistically significant influence, not specified further than the p-value of the F statistic (p. 225). However, from this regression, they take the predicted values of X, denote them $\hat{X}^Q$, and regressing Z on $\hat{X}^Q$, they obtain a coefficient (that we denote $\tilde{\beta}_{\hat{X}}$) of 0.013 with a standard error of 0.003 (p. 225). Lastly, in a 2SLS regression of Y on X, instrumented with Z, they obtain $\tilde{\beta}_X^{2SLS} = -0.023$ with a standard error of 0.010 (Table 3, Panel A, Column (1), p. 226). Together, they take these numbers as evidence for a statistically significant, negative and causal influence of a granted patent (X) on follow-up innovation (Y). We summarize the entire empirical set-up in Figure 6. Their argument boils down to the claim that a small influence between Q, X and Z lets them ignore the potential impact of Q on Y and thus, β_X is correctly identified by their coefficient $\tilde{\beta}_X^{2SLS}$.

Using Figure 6 as our DGP, we now consider the following system of equations

$$Y = \beta_X X + \gamma_Y Q + \varepsilon \tag{12}$$

$$X = \beta_Z Z + \gamma_X Q + \eta + \mu \tag{13}$$

$$Z = \gamma_Z Q + \nu \tag{14}$$

From the values in SW, we know relatively little about the underlying values. From Table 1 (p. 212), we obtain $E[x] = 0.3043$ which implies $Var(X) = 0.2117$. From own analysis of the USPTO data, we know that Art-unit-by-year demeaned examiner leniency has $Var(Z) = 0.0225$. Further, from their reported estimates, we can obtain some additional information about the coefficients.[12] $\tilde{\beta}_Z = 0.876$ implies that

$$\tilde{\beta}_Z = \frac{\gamma_Z(\beta_Z\gamma_Z + \gamma_X)Var(Q) + \beta_Z Var(\nu)}{\gamma_Z^2 Var(Q) + Var(\nu)} = 0.876. \tag{15}$$

Further, the coefficient estimate from regressing Z on $\hat{X}^Q$ implies

$$\tilde{\beta}_{\hat{X}} = \frac{\gamma_Z}{\beta_Z\gamma_Z + \gamma_x} = 0.013. \tag{16}$$

Since no coefficient for the quality dimension is reported[13], we assume $\gamma_X = 0.2$.[14] We further assume that Z and Q explain exactly $\varphi_Z = 50\%$ of the variance of X. This implies

$$Var(Q) = \frac{\varphi_Z Var(X) - \beta_Z^2 Var(Z)}{\gamma_X^2 + 2\beta_Z\gamma_X\gamma_Z}. \tag{17}$$

Lastly, equation (14) implies that

$$Var(\nu) = Var(Z) - \gamma_Z^2 Var(Q). \tag{18}$$

[12]For derivations of these expressions, see Appendix C.

[13]And, in fact, their quality scale has two variables, so reporting a single coefficient is not possible.

[14]Note that pretty much any assumption that we make is innocuous. We simply provide a counter example to their identification strategy. So if we can find *any* combination of parameters that assumes $\beta_Y = 0$ and can recreate the reported results, we have shown our point.

Taken together, we can solve these equations to obtain $\beta_Z = 0.8239$, $\gamma_Z = 0.0026$, $Var(Q) = 2.2164$, and $Var(\nu) = 0.0224$. From the fact that Z and Q explain exactly 50% of the variance of X, we know that $Var(\eta) = 0.5Var(X) = 0.1058$.

Given the full DGP, we can write

$$\tilde{\beta}_X^{2SLS} = \frac{\beta_X \beta_Z Var(Z) + (\beta_X \gamma_X + \gamma_Y)\gamma_Z Var(Q)}{\beta_Z Var(Z) + \gamma_X \gamma_Z Var(Q)}. \tag{19}$$

Our critical assumption now is that $\beta_X = 0$. When applying it, rearranging leaves us with

$$\gamma_Y = \tilde{\beta}_X^{2SLS} \frac{\beta_Z Var(Z) + \gamma_X \gamma_Z Var(Q)}{\gamma_Z Var(Q)} \tag{20}$$

to which we know all values such that we obtain $\gamma_Y = -0.0778$.

Lastly, we need to make some further assumptions. Similar to the variance of X, we assume that Q explains 50% of the variance of Y, setting $\varphi_Y = 50\%$. ε, η, ν and Q are generated as normally distributed with mean zero and their respective standard deviations. The μ in Equation (13) is set to the empirically reported value of 0.3043.

We run 10,000 Monte Carlo simulations with $n = 15,000$ observations each. We use the aforementioned assumptions and parameters derived above and, crucially, set $\beta_X = 0$. Our setting has some slight differences to the correct empirical setting. In the real data, there are two quality dimensions (family size and claims count) rather than one. We abstract from this and treat our variable Q as an aggregated measure. The variable X is binary in SW and thus bimodal, while it is normally distributed and unimodal in our simulations. Working with the binary variable adds unnecessary complication to the derivations above and does not meaningfully alter our results. The sample size in SW is slightly more intricate than assumed here. $\tilde{\beta}_Z$ and $\tilde{\beta}_{\hat{X}}$ are reported based on 14,476 observations, which is what we match our n to. $\tilde{\beta}_X^{2SLS}$ is reported based on a panel with 293,652 observations. We do not emulate this panel, as it would have required significant additional assumptions. We simply treat the analysis as cross-sectional. Given that SW cluster their standard errors on the level of the 14,476 patent applications, this seems appropriate.

Results of the simulation can be found in Table 6. We can see that our simulation design very closely matches the results of Sampat and Williams (2019). Even though we have set β_X, and thus the causal influence of X on Y, to 0, we still find a statistically significant coefficient for $\tilde{\beta}_X^{2SLS}$ of equivalent size as in the original paper. Note also that our average simulated value of $\tilde{\beta}_{\hat{X}}$ is the same as the one reported in the paper, with its standard error being a little higher. Even though our DGP leads to the same small coefficient and its statistical significance is less impressive than reported in the paper, we still show the same value for $\tilde{\beta}_X^{2SLS}$ as SW *with the explicit assumption* of $\beta_X = 0$ in the DGP. The last line of Table 6 shows the share of simulations with statistical significance at the 5% level (the same as in SW). We can see that our simulations render this significance level in 63.7% of all replications.

Table 6: Reported and simulated results for the SW IV Design

	$\tilde{\beta}_X^{2SLS}$	$\tilde{\beta}_Z$	$\tilde{\beta}_{\hat{X}}$
Reported Coefficient	-0.023	0.876	0.013
Reported Standard Error	0.010	0.037	0.003
Average Simulated Coefficient	-0.023	0.876	0.013
Average Simulated Standard Error	0.010	0.020	0.004
Share of coefficients significant at 5%	0.637	1.000	0.889

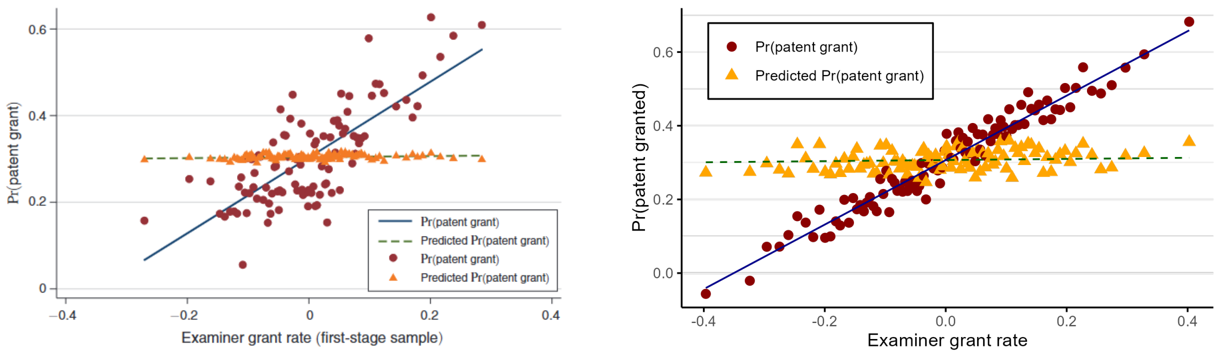
Note: The first two rows report the values from Sampat and Williams (2019), the last two rows report the values obtained in our simulation using the DGP displayed in Figure 6 with the assumption that $\beta_X = 0$. The table is based on 10,000 Monte Carlo simulations.

The argument provided by SW against the importance of the quality dimension is both a quantitative one and a graphical one. Quantitatively, they argue that the small coefficient size for $\tilde{\beta}_{\hat{X}}$ implies a low importance of the quality dimension in their data. We show above that this argument is misleading. Our simulated data lead to a similarly small coefficient for $\tilde{\beta}_{\hat{X}}$, but the entire influence of X on Y that is captured in $\tilde{\beta}_X^{2SLS}$ is due to the quality of the invention underlying the patent application rather than due to the grant of the patent. The second argument put forward by SW is a graphical one. In their Figure 3, which is reprinted in panel (a) of our Figure 7, they argue that the statistical relationship of Z and $\hat{X}^Q$ is difficult to determine. The graph puts this into an (indirect) comparison with the statistical relationship of Z and X. However, there is no reason why one should be informative about the other. Also, panel (b) of Figure 7 plots the same picture for our last simulation run. The message is very comparable. The statistical relationship of the yellow triangles looks indiscernible and definitely weaker than that implied by the red dots. Nevertheless, we estimate a $\tilde{\beta}_X^{2SLS} = -0.186$ in this simulation, as well.

Figure 7: Probability of Patent Grant by Examiner Leniency: Empirical v Simulated Data

(a) Figure as it appears in SW (2019)

(b) Figure from a sample simulation run



Note: Panel (a) is a reprint of Figure 3 from Sampat and Williams (2019). Panel (b) is the equivalent graph from the first simulation run of our Monte Carlo analysis. Red dots and blue line are the first stage of IV regression, while the orange triangles and green dashed line are the raw data and regression of leniency on the predicted probability of a patent grant from the quality dimension.

In total, we make assumptions about three values in the simulation: γ_X, φ_Z, and φ_Y. In appendix C we show that changing these values has no bearings about the average estimated values for $\tilde{\beta}_X^{2SLS}$, $\tilde{\beta}_Z$, or $\tilde{\beta}_{\hat{X}}$. Changing γ_X and φ_Z does, however, change the calculated values for γ_Z, γ_Y, and β_Z. Whether or not these values are more or less realistic is of secondary to the question at hand: can the identification strategy of SW provide credible evidence of a causal influence of a patent grant on follow-up innovation? The answer seems to be negative.

The most obvious criticism of the analysis presented here are the relatively extreme values assumed for φ_Z and φ_Y. Decreasing them makes the statistical relationships between Q and X and Q and Y less precise and thus leads to less commonly statistically significant coefficients in the 2SLS estimation. However, as we show in Table 7, even reducing both φ_Y and φ_X to 10% still does not allow us to reject the hypothesis that the statistical significance of $\tilde{\beta}_X^{2SLS}$ stems from the chance and the connection through Q rather than from $\beta_X \neq 0$. Note also that the values in the table below refer to the true quality dimension, not the imprecise measurement of SW. In fact, even if quality is measured as $\tilde{Q} = Q + \omega$ with ω iid, Equations (15) through (18) still hold.[15] This makes the assumed values of φ_Z and φ_Y more realistic because they pertain to the true quality dimension and not just the one measured by SW.

Table 7: Joint robustness to φ_X and φ_Y

	$\varphi_Y = 0.1$	$\varphi_Y = 0.2$	$\varphi_Y = 0.35$	$\varphi_Y = 0.5$	$\varphi_Y = 0.65$
$\varphi_X = 10\%$	0.052	0.063	0.067	0.077	0.085
$\varphi_X = 20\%$	0.087	0.120	0.172	0.237	0.277
$\varphi_X = 35\%$	0.129	0.217	0.340	0.444	0.562
$\varphi_X = 50\%$	0.178	0.313	0.478	0.634	0.755
$\varphi_X = 65\%$	0.231	0.399	0.614	0.769	0.861

Note: Table displays the share of Simulations with statistically significant $\tilde{\beta}_X^{2SLS}$ as a function of φ_X and φ_Y. γ_X is set to the default value of 0.2. The table is based on 10,000 Monte Carlo simulations for each parameter pair.

The results of the present study call into question whether the patent lottery can be treated as exogenous variation and used for causal identification. If firms can invest into manipulating the patent lottery, then any empirically identified effect of a granted patent could instead by a driver of such investment or be caused by some omitted underlying variable. Our reanalysis of the results by Sampat and Williams (2019) show that published results are not immune to this critique.

8 Conclusions

We examine whether the ostensibly random assignment of patent examiners—the "patent lottery"—can be gamed and what this implies for innovation and empirical research. Using 6 million U.S. applications from 2001–2020, we construct the null distribution of examiner leniency and show that assignment patterns at the law firm level deviate significantly from randomization. Larger entities receive examiners who are, on average, 1.13 percentage points more lenient, applications

[15]The measured quality dimension only appears in the determination of $\tilde{\beta}_{\hat{X}}$, the only coefficient reported by SW for which their measure of quality is actually used. In the notation above, if $\tilde{\beta}_{\hat{X}} = \frac{Cov(Z,\hat{X})}{Var(X)}$, and we express $Var(\omega) = \theta Var(Q)$, then $Cov(Z, \hat{X}) = (1+\theta)^{-1}\gamma_Z(\gamma_X + \beta_Z\gamma_Z)Var(Q)$ and $Var(\hat{X}) = (1+\theta)^{-1}(\gamma_X + \beta_Z\gamma_Z)^2 Var(Q)$ such that the influence of ω cancels out.

with more claims are matched to more lenient examiners, and publicly available information enables predictive timing. A simple trading strategy based on examiner leniency earns positive abnormal returns. A model with heterogeneous examiner quality and costly manipulation rationalizes these facts.

These findings matter for both policy and research. Fair and independent assignment is a cornerstone of the patent system; systematic deviations shift approval probabilities toward sophisticated repeat players and away from small innovators, altering the distribution of innovative rents. Our results also challenge a common identification strategy that instruments patent grants with examiner leniency under the assumption of random assignment, highlighting the risk of spurious inference when assignment is partly endogenous.

Our analysis has limitations. While we document patterns consistent with manipulation and provide strong predictive evidence, we do not directly observe all channels of influence or the assignment algorithm's internal logic. Measures of examiner leniency and attorney identity, while carefully constructed, are subject to measurement error. The return predictability we document is based on historical backtests and may be attenuated in real time.

The broader implications are nevertheless important. If assignment can be influenced at scale, then the patent system may inadvertently reward influence rather than invention, with consequences for entry, competition, and the geography of innovation. Transparent, auditable randomization and periodic assignment audits could restore confidence in the system and improve the external validity of research designs that rely on examiner heterogeneity.

Several directions merit future work. First, evaluate policy reforms—for example, stricter automated randomization, throttling of filing-time discretion, or routine anomaly detection—using pilot programs or staggered rollouts. Second, develop designs that exploit plausibly exogenous shocks (e.g., staffing or procedural changes) to pin down causal channels. Third, quantify welfare and distributional effects across firm size, technology, and regions. Fourth, extend the analysis to other jurisdictions and to downstream outcomes such as litigation, licensing, and market entry. Finally, explore assignment mechanisms that preserve workload balancing while hardening randomness against strategic behavior.

References

Arora, A., S. Belenzon, and L. Sheer (2017). Back to basics: why do firms invest in research? Technical report, National Bureau of Economic Research.

Arora, A., S. Belenzon, and L. Sheer (2021). Matching patents to compustat firms, 1980–2015: Dynamic reassignment, name changes, and ownership structures. Research Policy 50(5), 104217.

Bao, J., J. Pan, and J. Wang (2011). The illiquidity of corporate bonds. The Journal of Finance 66(3), 911–946.

Budish, E., B. N. Roin, and H. Williams (2016). Patents and research investments: Assessing the empirical evidence. American Economic Review 106(5), 183–187.

Fama, E. F. and K. R. French (1995). Size and book-to-market factors in earnings and returns. The Journal of Finance 50(1), 131–155.

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

Goldsmith-Pinkham, P., P. Hull, and M. Kolesár (2025). Leniency designs: An operator's manual.

Graham, S. J., A. C. Marco, and R. Miller (2018). The uspto patent examination research dataset: A window on patent processing. Journal of Economics & Management Strategy 27(3), 554–578.

Harhoff, D., F. M. Scherer, and K. Vopel (2003). Citations, family size, opposition and the value of patent rights. Research policy 32(8), 1343–1363.

Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman (2017). Technological innovation, resource allocation, and growth. The quarterly journal of economics 132(2), 665–712.

Kuhn, J. M. and N. C. Thompson (2019). How to measure and draw causal inferences with patent scope. International Journal of the Economics of Business 26(1), 5–38.

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

Melero, E., N. Palomeras, and D. Wehrheim (2020). The effect of patent protection on inventor mobility. Management Science 66(12), 5485–5504.

Merges, R. P. (1999). As many as six impossible patent before breakfast: Property rights for business concepts and patent system reform. Berkeley Tech. LJ 14, 577.

Newey, W. K. and K. D. West (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. NBER Technical Paper No. 55, National Bureau of Economic Research.

Newey, W. K. and K. D. West (1994). Automatic lag selection in covariance matrix estimation. The Review of Economic Studies 61(4), 631–653.

Nordhaus, W. D. (1969). Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change. MIT Press.

Pakes, A. (1985). On patents, r & d, and the stock market rate of return. Journal of Political Economy 93(2), 390–409.

Petersen, M. A. and R. G. Rajan (1994). The benefits of lending relationships: Evidence from small business data. The Journal of Finance 49(1), 3–37.

Raffiee, J., F. Teodoridis, and D. Fehder (2023). Partisan patent examiners? exploring the link between the political ideology of patent examiners and patent office outcomes. Research Policy 52(9), 104853.

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

Rosen, J. B. (1965). Existence and uniqueness of equilibrium points for concave n-person games. Econometrica: Journal of the Econometric Society, 520–534.

Sampat, B. and H. L. Williams (2019). How do patents affect follow-on innovation? Evidence from the human genome. American Economic Review 109(1), 203–236.

Schankerman, M. and A. Pakes (1986). Estimates of the value of patent rights in european countries during the post-1950 period. The economic journal 96(384), 1052–1076.

Shumway, T. (1997). The delisting bias in crsp data. The Journal of Finance 52(1), 327–340.

Trajtenberg, M. (1990). Economic analysis of product innovation: The case of CT scanners, Volume 160. Harvard University Press.

Appendix A Proofs

A.1 Proof of Proposition 1

Proof. We can collect the first order conditions in a system $F(e_1, e_2) = 0$ with $F_1(e_1, e_2) = p_1(e_1, e_2) - \frac{k_1}{\pi_1} = 0$ and $F_2(e_1, e_2) = p_1(e_2, e_1) - \frac{k_2}{\pi_2} = 0$. The Jacobian is then given by

$$J = \begin{bmatrix} p_{11}(e_1, e_2) & p_{12}(e_1, e_2) \\ p_{12}(e_2, e_1) & p_{11}(e_2, e_1) \end{bmatrix}. \tag{A1}$$

Defining

$$S := \frac{1}{2}(J + J^\top) = \begin{bmatrix} p_{11}(e_1, e_2) & \frac{1}{2}(p_{12}(e_1, e_2) + p_{12}(e_2, e_1)) \\ \frac{1}{2}(p_{12}(e_1, e_2) + p_{12}(e_2, e_1)) & p_{11}(e_2, e_1) \end{bmatrix}. \tag{A2}$$

From this we can see that S is own-concave by $p_{11} < 0$ and that it has a positive determinant if

$$p_{11}(e_1, e_2)p_{11}(e_2, e_1) - \frac{1}{4}(p_{12}(e_1, e_2) + p_{12}(e_2, e_1))^2 > 0. \tag{A3}$$

Thus, by Rosen (1965), (7) identifies the unique Nash equilibrium as long as $|p_{12}|$ is not too large and optimal expenses are not given by the corner solution of $e_i^* = 0$. The latter can be guaranteed by $p_1(0, 0) > \max\left\{\frac{k_1}{\pi_1}, \frac{k_2}{\pi_2}\right\}$. □

A.2 Proof of Proposition 2

Proof. Due to Proposition 1, we know that we study an interior solution with the first order conditions given by (7). Define the system of first-order conditions as a mapping $F : \mathbb{R}^2 \times \mathbb{R}^4_{++} \to \mathbb{R}^2$:

$$F_1(e_1, e_2; k_1, k_2, \pi_1, \pi_2) := p_1(e_1, e_2) - \frac{k_1}{\pi_1} = 0, \tag{A4}$$

$$F_2(e_1, e_2; k_1, k_2, \pi_1, \pi_2) := p_1(e_2, e_1) - \frac{k_2}{\pi_2} = 0. \tag{A5}$$

Let $A := \partial F/\partial(e_1, e_2)$ evaluated at (e_1^*, e_2^*). This is equivalent to $J(e_1^*, e_2^*)$ from (A1). Under the assumption from Proposition 1 that p_{12} is not too negative, we can see that

$$\det(A) = p_{11}(e_1^*, e_2^*)\, p_{11}(e_2^*, e_1^*) - p_{12}(e_1^*, e_2^*)\, p_{12}(e_2^*, e_1^*) > 0. \tag{A6}$$

Let $E = (e_1^*, e_2^*)^\top$. The Implicit Function Theorem (IFT) implies that, in a neighborhood where A is nonsingular, E is a continuously differentiable function of (k_1, k_2) and

$$\frac{\partial E}{\partial k_i} = -A^{-1}\frac{\partial F}{\partial k_i}, \qquad i \in \{1, 2\}. \tag{A7}$$

Focusing on k_1, the relevant parameter derivative is $\frac{\partial F}{\partial k_1} = (-1/\pi_1, 0)^\top$. Solving the system of linear equations in (A7) renders

$$\frac{\partial e_1^*}{\partial k_1} = \frac{p_{11}(e_2^*, e_1^*)}{\pi_1 \det A} < 0 \qquad \text{and} \tag{A8}$$

$$\frac{\partial e_2^*}{\partial k_1} = -\frac{p_{12}(e_2^*, e_1^*)}{\pi_1 \det A} > 0. \tag{A9}$$

Analogously, we can find $\frac{\partial e_1^*}{\partial k_2} > 0$ and $\frac{\partial e_2^*}{\partial k_2} < 0$.

Treating the benefit parameter analogously renders $\frac{\partial E}{\partial \pi_i} = -A^{-1}\frac{\partial F}{\partial \pi_i}$ for $i \in \{1, 2\}$. The relevant parameter derivative is $\frac{\partial F}{\partial \pi_1} = \left(\frac{k_1}{\pi_1^2}, 0\right)^\top$. Solving the system of linear equations again renders

$$\frac{\partial e_1^*}{\partial \pi_1} = -\frac{k_1 p_{11}(e_2^*, e_1^*)}{\pi_1^2 \det A} > 0 \qquad \text{and} \tag{A10}$$

$$\frac{\partial e_2^*}{\partial k_1} = \frac{k_1 p_{12}(e_2^*, e_1^*)}{\pi_1^2 \det A} < 0. \tag{A11}$$

And analogously $\frac{\partial e_1^*}{\partial \pi_2} < 0$ and $\frac{\partial e_2^*}{\partial \pi_2} > 0$.

The proposition follows from $\frac{\partial p(e_i, e_{-i})}{k_i} = p_1(e_i, e_{-i})\frac{\partial e_i}{\partial k_i} + p_2(e_i, e_{-i})\frac{\partial e_{-i}}{\partial k_i} < 0$ and $\frac{\partial p(e_i, e_{-i})}{\pi_i} = p_1(e_i, e_{-i})\frac{\partial e_i}{\partial \pi_i} + p_2(e_i, e_{-i})\frac{\partial e_{-i}}{\partial \pi_i} > 0$. □

Appendix B Additional Empirical Results

Table B1: Application-level Evidence with USPC Subclass Fixed Effects

	Dependent Variable: Examiner Leniency					
	(1)	(2)	(3)	(4)	(5)	(6)
Discounted Appl.	-0.518^{***}				-0.518^{***}	
	(0.033)				(0.033)	
Small Entity		-0.513^{***}				-0.513^{***}
		(0.033)				(0.033)
Micro Entity		-0.612^{***}				-0.606^{***}
		(0.120)				(0.120)
No. Claims			0.002^{**}		0.002^{**}	0.002^{**}
			(0.001)		(0.001)	(0.001)
Application Scope				-0.00005^{**}	-0.00004^{**}	-0.00004^{**}
				(0.00002)	(0.00002)	(0.00002)
Fixed effects	Art unit $\times$ Year + Subcl.	Art unit $\times$ Year + Subcl.	Art unit $\times$ Year + Subcl.	Art unit $\times$ Year + Subcl.	Art unit $\times$ Year + Subcl.	Art unit $\times$ Year + Subcl.
Clustered st. err.	Art unit	Art unit	Art unit	Art unit	Art unit	Art unit
Observations	5,708,617	5,708,617	5,708,617	5,708,617	5,708,617	5,708,617

Note: Table shows the results of estimating Equation (8) either in full (columns (5) and (6)) or with each explanatory variable in isolation (columns (1) through (4)). Examiner leniency is defined as the average probability of an application's examiner to grant a patent in the applications' year and art unit. The application is discounted if it comes from an individual, a nonprofit organization or a company with fewer than 500 employees. Micro entities get a larger discount than small entities. No. Claims is the number of claims listed at the time of the application. Scope is the number of characters in the first claim of the application at the initial time of application. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Table B2: Value-weighted alphas of trading strategy

	Q1	Q2	Q3	Q4	Q5	Q5-Q1
Alpha (bps)	1.880^{***}	1.477^{***}	1.776^{***}	2.148^{***}	2.803^{***}	0.924^{**}
	(0.250)	(0.316)	(0.401)	(0.422)	(0.346)	(0.365)
Mkt − rf	0.855^{***}	1.112^{***}	1.013^{***}	1.340^{***}	1.373^{***}	0.517^{***}
	(0.106)	(0.143)	(0.126)	(0.141)	(0.109)	(0.124)
Cars	−0.002	−0.006	-0.015^{*}	−0.008	0.007	0.010
	(0.004)	(0.006)	(0.008)	(0.010)	(0.007)	(0.008)
Chems	0.001	0.005	−0.009	-0.028^{***}	-0.024^{***}	-0.024^{***}
	(0.005)	(0.008)	(0.009)	(0.009)	(0.008)	(0.008)
Clths	-0.008^{**}	-0.023^{***}	-0.041^{***}	-0.054^{***}	-0.048^{***}	-0.040^{***}
	(0.004)	(0.006)	(0.008)	(0.008)	(0.006)	(0.006)
Cnstr	-0.024^{***}	-0.056^{***}	-0.051^{***}	-0.053^{***}	-0.055^{***}	-0.031^{***}
	(0.005)	(0.008)	(0.009)	(0.010)	(0.009)	(0.010)
Cnsum	0.474^{***}	0.275^{***}	-0.052^{***}	-0.152^{***}	-0.125^{***}	-0.599^{***}
	(0.015)	(0.020)	(0.017)	(0.019)	(0.015)	(0.018)
Durbl	-0.020^{***}	-0.039^{***}	-0.073^{***}	-0.064^{***}	-0.032^{***}	−0.012
	(0.007)	(0.008)	(0.010)	(0.011)	(0.007)	(0.008)
FabPr	−0.004	-0.016^{**}	-0.028^{***}	-0.039^{***}	-0.021^{**}	-0.016^{*}
	(0.005)	(0.008)	(0.010)	(0.012)	(0.009)	(0.009)
Finan	-0.148^{***}	-0.182^{***}	-0.151^{***}	-0.195^{***}	-0.219^{***}	-0.071^{***}
	(0.018)	(0.022)	(0.021)	(0.023)	(0.018)	(0.019)
Food	0.006	-0.023^{*}	−0.012	-0.043^{***}	-0.055^{***}	-0.061^{***}
	(0.009)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Machn	-0.054^{***}	−0.015	0.101^{***}	0.232^{***}	0.346^{***}	0.400^{***}
	(0.012)	(0.018)	(0.017)	(0.019)	(0.017)	(0.017)
Mines	-0.012^{***}	-0.031^{***}	-0.015^{***}	-0.016^{***}	-0.012^{**}	−0.000
	(0.003)	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)
Oil	0.000	0.005	0.006	−0.014	−0.012	−0.012
	(0.009)	(0.011)	(0.009)	(0.011)	(0.008)	(0.011)
Other	0.075^{**}	0.087^{*}	0.426^{***}	0.188^{***}	−0.015	-0.090^{**}
	(0.034)	(0.046)	(0.046)	(0.049)	(0.039)	(0.045)
Rtail	-0.017^{*}	0.005	0.008	0.008	0.011	0.029^{**}
	(0.010)	(0.014)	(0.014)	(0.013)	(0.013)	(0.014)
Steel	-0.015^{***}	-0.020^{***}	-0.029^{***}	-0.028^{***}	-0.020^{***}	−0.005
	(0.003)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)
Trans	-0.031^{***}	-0.022^{**}	-0.019^{*}	-0.023^{*}	-0.037^{***}	−0.007
	(0.008)	(0.011)	(0.010)	(0.012)	(0.010)	(0.010)
Utils	-0.054^{***}	-0.046^{***}	-0.034^{***}	-0.041^{***}	-0.045^{***}	0.009
	(0.006)	(0.008)	(0.009)	(0.009)	(0.008)	(0.008)

Note: Table displays the daily alphas (in bps) of portfolios sorted on examiner leniency. Alphas are the intercepts from time-series regressions of the value-weighted portfolio (excess) returns against the market factor and the 17 value-weighted Fama-French industry portfolio returns. Standard errors are calculated following Newey and West (1986), where we choose the optimal truncation lag as suggested by Newey and West (1994). The standard errors are provided below the respective alphas. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Appendix C Details on the Reassessment of Sampat and Williams (2019)

C.1 Derivation of Coefficients

$\tilde{\beta}_Z$ is the coefficient resulting from estimating $X = \alpha_1 + \tilde{\beta}_Z Z + \tilde{\eta}$. We know that

$$\tilde{\beta}_Z = \frac{Cov(X,Z)}{Var(Z)}. \tag{A12}$$

From the DGP, we know that $Z = \gamma_Z Q + \nu$ and $X = \beta_Z Z + \gamma_X Q + \eta + \mu$. Substituting renders

$$X = (\beta_Z \gamma_Z + \gamma_X)Q + \beta_Z \nu + \alpha_1 + \eta \tag{A13}$$

Because $Cov(\nu, Q) = 0$ and $Cov(\nu, \eta) = 0$ and because α_1 is a constant, we can write $Cov(X,Z) = \gamma_Z(\beta_Z\gamma_Z + \gamma_X)Var(Q) + \beta_z Var(\nu)$. By a similar argument, $Var(Z) = \gamma_Z^2 Var(Q) + Var(\nu)$. Equation (15) follows.

$\tilde{\beta}_{\hat{X}}$ is the coefficient resulting from estimating $Z = \alpha_2 + \tilde{\beta}_{\hat{X}}\hat{X} + \tilde{\nu}$. Define δ such that $\hat{X} = \alpha_3 + \delta Q$. Then, by the DGP, we have $\delta = \gamma_X + \beta_Z\gamma_Z$. Substituting renders

$$\tilde{\beta}_{\hat{X}} = \frac{Cov(Z,\hat{X})}{Var(\hat{X})} = \frac{\gamma_Z(\gamma_X + \beta_Z\gamma_Z)Var(Q)}{(\gamma_X + \beta_Z\gamma_Z)^2 Var(Q)} = \frac{\gamma_Z}{\gamma_X + \beta_Z\gamma_Z}. \tag{A14}$$

$Var(Q)$ can be defined from

$$\varphi_Z Var(X) = \beta_Z^2 Var(Z) + \gamma_X^2 Var(Q) + 2\beta_Z\gamma_Z Cov(Z,Q). \tag{A15}$$

By the DGP, $Cov(Z,Q) = \gamma_Z Var(Q)$. Substituting and rearranging renders (17).

$\tilde{\beta}_X^{2SLS}$ is the coefficient of regressing Y on X when X is instrumented by Z. Thus

$$\tilde{\beta}_X^{2SLS} = \frac{Cov(Z,Y)}{Cov(Z,X)} = \frac{Cov(Z, \beta_X\beta_Z Z + (\beta_X\gamma_X + \gamma_Y)Q + \beta_X\eta + \varepsilon)}{Cov(Z, \beta_Z Z + \gamma_X Q + \eta)}. \tag{A16}$$

Because $Cov(Z,\eta) = 0$, $Cov(Z,\varepsilon) = 0$, and $Cov(Z,Q) = \gamma_Z Var(Q)$, (19) follows.

C.2 Numerical Solution

For the numerical solution, we rearrange (15) to

$$\beta_Z(\gamma_Z) = \frac{\gamma_Z - 0.013\gamma_X}{0.013\gamma_X}. \tag{A17}$$

Note that realistically, β_Z has to lie in the unit interval. Thus, $\gamma_Z \in \left[0.013\gamma_X, \frac{0.013\gamma_X}{1-0.013}\right]$.

We then substitute (18) into (15) and obtain $\tilde{\beta}_Z = \gamma_X\gamma_Z\frac{Var(Q)}{Var(Z)} + \beta_Z(\gamma_Z)$. Further substituting (17) renders

$$\tilde{\beta}_Z(\gamma_Z) = \gamma_X\gamma_Z\frac{\varphi_Z\frac{Var(X)}{Var(Z)} - (\beta_Z(\gamma_Z))^2}{\gamma_X^2 + 2\beta_Z(\gamma_Z)\gamma_X\gamma_Z} + \beta_Z(\gamma_Z). \tag{A18}$$

This expression can be used to numerically solve $\tilde{\beta}_Z(\gamma_Z) - 0.876 = 0$ in the interval $\left[0.013\gamma_X, \frac{0.013\gamma_X}{1-0.013}\right]$.

C.3 Robustness to Parameter Choices

Table C3: Robustness of Simulation Result to Parameter Assumptions

(a) Changing γ_X				
	$\gamma_X = 0.05$	$\gamma_X = 0.1$	$\gamma_X = 0.3$	$\gamma_X = 0.5$
$\tilde{\beta}_X^{2SLS}$	-0.023	-0.023	-0.023	-0.023
share sign.	*0.634*	*0.635*	*0.626*	*0.640*
$\tilde{\beta}_Z$	0.876	0.876	0.876	0.876
share sign.	*1.000*	*1.000*	*1.000*	*1.000*
$\tilde{\beta}_{\hat{X}}$	0.013	0.013	0.013	0.013
share sign.	*0.891*	*0.888*	*0.886*	*0.891*
γ_Z	0.001	0.001	0.004	0.007
γ_Y	-0.019	-0.039	-0.117	-0.194
β_Z	0.824	0.824	0.824	0.824
(b) Changing φ_X: the share of $Var(X)$ explained by Z and Q				
	$\varphi_X = 0.2$	$\varphi_X = 0.35$	$\varphi_X = 0.65$	$\varphi_Z = 0.8$
$\tilde{\beta}_X^{2SLS}$	-0.023	-0.023	-0.023	-0.023
share sign.	*0.235*	*0.454*	*0.772*	*0.854*
$\tilde{\beta}_Z$	0.876	0.876	0.877	0.876
share sign.	*1.000*	*1.000*	*1.000*	*1.000*
$\tilde{\beta}_{\hat{X}}$	0.013	0.013	0.013	0.013
share sign.	*0.392*	*0.723*	*0.963*	*0.990*
γ_Z	0.003	0.003	0.003	0.003
γ_Y	-0.275	-0.121	-0.057	-0.045
β_Z	0.862	0.843	0.806	0.787
(b) Changing φ_Y: the share of $Var(Y)$ explained by Q				
	$\varphi_Y = 0.2$	$\varphi_Y = 0.35$	$\varphi_Y = 0.65$	$\varphi_Y = 0.8$
$\tilde{\beta}_X^{2SLS}$	-0.023	-0.023	-0.023	-0.023
share sign.	*0.308*	*0.494*	*0.742*	*0.825*
$\tilde{\beta}_Z$	0.876	0.876	0.875	0.876
share sign.	*1.000*	*1.000*	*1.000*	*1.000*
$\tilde{\beta}_{\hat{X}}$	0.013	0.013	0.013	0.013
share sign.	*0.884*	*0.894*	*0.888*	*0.885*
γ_Z	0.003	0.003	0.003	0.003
γ_Y	-0.078	-0.078	-0.078	-0.078
β_Z	0.824	0.824	0.824	0.824

Note: The table shows the results of the simulation when input values for γ_X, φ_X and φ_Y are changed. Unless indicated otherwise, the simulations use $\gamma_X = 0.2$, $\varphi_Z = 0.5$ and $\varphi_Y = 0.5$. The table is based on 10,000 Monte Carlo simulations for each combination of parameters.

Appendix D Attorney Name Cleaning and Law Firm Disambiguation

The PatEx dataset contains attorney names as free-text strings without standardized formatting or unique identifiers. To link patent applications to law firms for our analysis, we developed a multi-stage disambiguation pipeline that combines traditional natural language processing techniques with large language model (LLM) assistance to address the substantial variation in attorney name representations.

The raw attorney names in PatEx appear in various formats with inconsistent punctuation, spacing, character encoding, and frequent typos or OCR errors. A single law firm may be represented by dozens of different text strings due to variations in how individual attorneys list their firm affiliation, changes in firm names over time, data entry inconsistencies, and transcription errors. For example, the same firm might appear as "Smith & Jones LLP," "Smith and Jones, L.L.P.," "Smth & Jonnes Law Firm," or with individual attorney names followed by the firm name. These variations, particularly when compounded by typos and OCR errors, make traditional string matching insufficient.

Our disambiguation pipeline operates in five main stages. First, we perform initial fuzzy matching using token-based similarity with inverse document frequency (IDF) weighting. Each attorney name is tokenized, and tokens are weighted by their rarity across the corpus—common tokens like "Law" or "LLP" receive low weights while distinctive firm names receive high weights. This identifies obvious matches while flagging ambiguous cases for further processing.

Second, we employ GPT-5-mini to process every unique attorney name in the dataset. The model performs multiple tasks simultaneously: correcting typos and OCR errors (e.g., "Jhonson" to "Johnson," "rn" to "m"), expanding abbreviations (e.g., "Intl" to "International"), standardizing legal suffixes, removing address components, and classifying each entity as a law firm, in-house counsel, individual attorney, or government entity. The model also identifies "anchor tokens"—the distinctive words that uniquely identify each entity. This unified approach leverages the LLM's contextual understanding to handle complex variations that rule-based systems would miss.

Third, we perform additional fuzzy matching on the LLM-corrected names to identify matches that were obscured by typos or formatting variations in the original data. Fourth, we consolidate all matches using graph-based clustering, where each connected component represents a single law firm entity. We assign stable canonical identifiers using deterministic hashing to ensure reproducibility.

Finally, as a quality control measure, we send every merged cluster back to the LLM for a second-pass audit. The model reviews all proposed merges and flags any obvious errors where distinct entities may have been incorrectly combined. This audit focuses on identifying clear mistakes rather than being overly cautious about minor variations. The LLM approves merges with minor differences (typos, abbreviations, punctuation variations) while rejecting only those that are clearly different entities.

This multi-stage approach, combining traditional NLP techniques with LLM capabilities, successfully processes over 14 million attorney name instances and links them to canonical law firm entities. The use of LLMs is particularly crucial for handling the pervasive typos and OCR errors in the data, which traditional fuzzy matching alone cannot reliably correct. By leveraging both algorithmic precision and LLM contextual understanding, we achieve high-confidence entity resolution while maintaining the conservative approach necessary for rigorous empirical analysis.

The resulting dataset links over 6 million patent applications to their representing law firms, enabling the analysis of how legal representation influences patent examination outcomes presented in the main text.