# Screening Property Rights for Innovation*

William Matcham[§] Mark Schankerman[§]

October 17, 2023

**Abstract**

We develop a dynamic structural model of patent screening incorporating incentives, intrinsic motivation, and multi-round negotiation. We use detailed data on examiner decisions and employ natural language processing to create a new measure of patent distance that enables us to study strategic decisions by applicants and examiners. We find that patent screening is moderately effective, given the existing standards for patentability. Examiners exhibit substantial intrinsic motivation that significantly improves the effectiveness of screening. A reform that limits negotiation rounds strongly increases screening quality. We quantify the annual net social costs of patent screening at $25.5bn, equivalent to 6.5% of U.S. private sector R&D.

**Keywords:** Patents, innovation, incentives, screening, intrinsic motivation
**JEL Classification:** D73, L32, O31, O34, O38

*We acknowledge Jakub Drabik for his excellent research assistance and contributions, especially on the distance metric. We thank Dietmar Harhoff, Marco Ottaviani, Florian Schuett, and seminar and conference participants at Tel Aviv University, Humboldt University of Berlin, Ben-Gurion University, CEMFI, University of Bologna, Royal Holloway, and Boston University for comments and suggestions on earlier drafts of the paper. We also thank Martin Rater, Daniel Ryman, and Andrew Toole in the U.S. Patent and Trademark Office for assistance in obtaining part of the data, and Janet Freilich and Michael Meurer for helpful discussions on procedural aspects of patent prosecution. This project was partly financed by a research grant from the Suntory-Toyota Centres for Economic and Related Disciplines at the London School of Economics. William Matcham also thanks the Edison Fellowship Program at George Mason University for financial assistance and the opportunity to discuss the project with legal scholars.

[§]Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. Emails: w.o.matcham@lse.ac.uk, m.schankerman@lse.ac.uk

# 1 Introduction

Public institutions play a central role in promoting innovation. The two most important channels are government support for public and private research, both in the form of direct funding and indirect fiscal subsidies, and the allocation of property rights, in the form of patents, to enhance innovation incentives for private sector R&D. To give a sense of the scale of investment, in 2015 the U.S. federal government financed 54.3% of overall R&D expenditures, or $151.5 billion (2023 U.S.D.), and 34.1% of university research. At the same time, the U.S. Patent and Trademark Office (hereafter, *Patent Office*) issued nearly 400,000 new patents. These property rights promote innovation by increasing the private returns to R&D, facilitating access to capital markets, and underpinning the market for technology, especially for small, high-technology firms (Hall and Lerner, 2010; Galasso and Schankerman, 2018). Moreover, the aggregate economic impact of these investments and property rights for innovation is magnified by the extensive knowledge spillovers they generate (Bloom, Schankerman, and Van Reenen, 2013).

Despite their evident importance, little is known about whether innovation-supporting public institutions allocate resources efficiently and how organizational changes affect agency performance. The aim of this paper, as part of a broader research program, is to show how structural models can be used to study and improve the efficiency of resource allocation by innovation-related public agencies. We study this topic in the context of the U.S. patent system, focusing on the quality of screening—that is, the allocation of property rights for innovation—by the Patent Office.

We develop a dynamic structural model of the patent screening process, which incorporates incentives, intrinsic motivation, and the actual structure of multi-round negotiation in the current system. We estimate the model using novel negotiation-round-level data on examiner decisions and text data from 20 million patent claims. From the claim text data, we use modern natural language processing (NLP) methods to develop a new measure of distance between patents, a key ingredient for characterizing strategic decisions by patent applicants and examiners. We conduct counterfactual analyses of how reforms to incentives, fees, and the structure of negotiations affect the quality and speed of patent screening, and we develop an approach to quantify these impacts and thus construct a "pseudo-welfare" measure of the quality of patent screening.

The effectiveness of patent screening and its implications for the quality of patents is a hotly debated policy issue. Academic scholars and policymakers have argued that patent rights have increasingly become an impediment to innovation rather than an incentive. These concerns have been prominently voiced in public debates (The Economist, 2015; Federal Trade Commission,

2011), recent U.S. Supreme Court decisions (eBay Inc. v. MercExchange L.L.C., 547 U.S. 338, 2006), and the major statutory reform of the patent system, the Leahy-Smith America Invents Act of 2011.

Critics of the patent system claim that the problems arise in large part from ineffective patent office screening, where patents are granted to inventions that do not represent a substantial inventive step—especially in emerging technology areas such as business methods and software (Jaffe and Lerner, 2004). The issue is important because granting "excessive" patent rights imposes static and dynamic social costs: higher prices and deadweight loss on patented goods, greater enforcement (litigation) costs, and higher transaction costs of R&D and the potential for retarding cumulative innovation (Galasso and Schankerman, 2015).

The patent prosecution process is an advantageous context to study the effects of incentives and motivation on screening for two primary reasons. First, the patent application process has a clear and well-documented structure that can be modeled. The multi-round negotiation between the applicant and examiner fits naturally into a dynamic game, which forms the basis of our model. The model involves an applicant who "pads" their patent application, attempting to extract more property rights than their invention truly entails. The examiner's role is to grant or reject the application based on the existing judicial interpretation of statutory criteria as applied to each claim in the patent application.

The fundamental trade-off for the applicant when choosing the level of padding is between the benefits of increased patent scope and the costs of engaging in a lengthy and costly negotiation with the examiner. The trade-off for the patent examiner for each specific application is between the incentives to grant patents quickly and the intrinsic utility cost of awarding an inappropriate degree of "patent scope"—i.e., granting only patent claims (after narrowing) that satisfy the patentability criteria.[1] The patent examiner searches prior art to estimate the appropriate scope of patent protection for the invention, but this estimate contains error. Allowing for examiner error is important because it implies that negotiation between the applicant and examiner, while costly, may not always be socially wasteful.

The second advantage of the patent context is the quality of data. The Patent Office collects detailed and extensive data on all *applications*, not just granted patents. We constructed a dataset covering around 55 million patent application decisions across 20 million patent claims between 2010–2015 and we observe the examiner's decisions on each patent claim over all rounds of the

[1]For a discussion of the economics and legal doctrines of patent scope, see Merges and Nelson (1990).

negotiation. These data allow us to formulate and estimate a structural model that reflects the actual patent application process.

Our estimates imply several key empirical findings. First, intrinsic motivation plays a significant role in contributing to the accuracy of patent screening. Junior examiners are more motivated than seniors on average, but both groups display substantial heterogeneity. Further, using the estimated parameters, counterfactual analysis shows that turning off intrinsic motivation increases the frequency of examiners granting invalid patents four-fold. This finding highlights the importance of designing human resource policies that effectively select examiners with high intrinsic motivation and ensure examiners sustain this motivation over their entire careers.

Second, we find that innovators substantially pad their patent applications, claiming (typically) greater property rights than are warranted by the true "inventive step" of their innovation. Moreover, there is a large degree of heterogeneity in the extent of padding across patent applications. This result highlights the importance of effective screening. An essential feature of our model is that the extent of padding is endogenous and thus is affected by various counterfactual policy reforms, which we detail later. We estimate the average level of padding at about 8%, rising to 10% when we weight by the value of the patent. This exaggerated scope of the patent applications is reflected in the fact that more than 80% of claims start below the distance threshold for patentability—as measured by the minimum required distance to claims in prior patents—and thus should be rejected.

However, the multi-round screening process is relatively effective at narrowing the scope of patent rights sought and, in so doing, reducing the number of invalid claims to about 7% among granted claims, but still, nearly one in five granted patents contains at least one claim that does not meet the threshold. One implication of this finding is that the proportion of patent applications that are granted—a commonly used indicator of the effectiveness of screening—is a misleading measure because it does not capture the extent to which granted property rights are narrowed during the screening process.

We evaluate counterfactual reforms involving changes to fees for the patent applicant, the structure of the negotiation process (e.g., limiting the number of rounds allowed), and the degree of intrinsic motivation of patent examiners. We quantify the effects of counterfactual reforms along three distinct dimensions. The first two relate to the accuracy of screening, meaning the degree of alignment between the scope of property rights granted and the scope justified by the invention. We assess accuracy in terms of granting claims that are not justified (false grants, or "type 1" error) and not granting claims that should be (false rejections, or "type 2" error).

Both errors carry their own social costs and benefits. Incorrect grants impose ex post welfare costs (deadweight loss) from higher prices and litigation costs associated with enforcing these patents, but at the same time may raise innovation incentives. False rejections dilute ex ante innovation incentives and discourage the development of new inventions that would contribute positive social value, but at the same time they reduce ex post deadweight loss. The last dimension is the speed of patent examination, measured by the number of negotiation rounds in equilibrium. We develop a method to quantify these impacts in terms of the associated net social costs and thus construct a "pseudo-welfare" measure of the quality of patent screening. We estimate the total net social cost of patent screening at $25.5bn per annual cohort of applications. This figure represents 6.5% of total R&D performed by business enterprises in the United States.

The counterfactual analysis highlights two key conclusions. First, restrictions on the number of allowable rounds of negotiation (currently absent in the U.S. patent system) significantly reduce the net social costs of screening, with a reduction of 45% in the case of allowing only one round. We show that these outcomes can be replicated through an equivalent fee per round for the applicant, but the required fees are too high to be politically feasible. Second, given the high levels of intrinsic motivation we estimate, extrinsic incentives are largely ineffective, leading to almost no change in net social costs. Extrinsic incentives do affect outcomes in a scenario with low intrinsic motivation, but they are counterproductive in that they raise the net social costs of screening.

The paper is organized as follows. Section 2 briefly summarizes the related literature. Section 3 describes the datasets and summarizes key descriptive features. The structural model is presented in Section 4. Section 5 describes our estimation methods. Section 6 presents the empirical estimates. Section 7 analyzes the impact of counterfactual reforms on the accuracy and speed of patent screening, and Section 8 describes our quantification of the net social costs and benefits associated with these counterfactual reforms.

## 2 Related Literature

*Intrinsic Motivation in Public Agencies*

We contribute to the literature that studies how intrinsic motivation affects the optimal design of incentives in mission-oriented agencies. On the theoretical side, Benabou and Tirole (2003; 2006) show conditions under which extrinsic rewards may crowd out intrinsic motivation. Particularly relevant to our paper, Besley and Ghatak (2005) emphasize how intrinsic motivation—which they define as the alignment between worker and agency objectives—induces welfare-improving sorting of workers across entities with different goals and also affects the optimal design of

incentives and authority.

Empirical studies use field experiments to analyze intrinsic motivation and public agency performance. These rely on various proxies for motivation. Leading examples include Ashraf, Bandiera, and Jack (2014), which evaluates the impact of extrinsic rewards on agents' performance in a public health organization in Zambia, and Ashraf, Bandiera, Davenport, and Lee (2020), which studies whether career benefits induce sorting at the expense of "pro-social" motivation. Both papers find that extrinsic rewards and intrinsic motivation are complementary.

Despite their interesting findings, these empirical studies cannot be used for counterfactual policy analysis, for which structural models are more appropriate. Our paper is the first structural model of a public agency that incorporates intrinsic motivation.[2] In doing this, we follow Besley and Ghatak's definition of intrinsic motivation—alignment of workers' objectives and the public agency mission. In our context, the Patent Office's mission is to award inventors property rights over their invention, consistent with statutory and judicial prescriptions. We model intrinsic motivation as an inherent disutility that examiners incur if they grant more intellectual property rights than they believe the inventor deserves, based on the information the examiners have. We show that patent examiners sometimes award patents to applications they believe are invalid due to strategic considerations and the extrinsic pay scheme they face.

Finally, recent papers study how screening mechanisms affect the performance of public agencies. Adda and Ottaviani (2023) develop a model of nonmarket allocation of resources, including but not limited to the award of grants to research projects. The authors study how the design of allocation rules and informational noise in the evaluation process affect the optimal design. In two empirical papers, Li and Agha (2015) and Li (2017) analyze the allocation of research grants at the National Institutes of Health (NIH) and show that peer review increases the effectiveness of grants in terms of post-grant citations. Azoulay, Graff Zivin, Li, and Sampat (2018) study the economic impact of these NIH grants, linking screening outcomes to publication citations and other innovation outcomes. Our contribution is to quantify some of the forces these papers identify and evaluate the equilibrium effects of various counterfactual reforms in the patent context.

*Patents and Innovation*

We also contribute to the limited empirical literature on patent screening. In a first paper on the

[2] Egan, Matvos, and Seru (2023) develop a structural model of consumer arbitration in which arbitrators differ in their idiosyncratic degrees of "slant" (or bias), which can be interpreted as a form of intrinsic motivation.

topic, Cockburn, Kortum, and Stern (2003) show that patent examiner characteristics affect the "quality" of issued patents, measured by subsequent citations and the frequency of litigation. Frakes and Wasserman (2017) use data on promotions of patent examiners (which are accompanied by lower extrinsic incentives) and show that promotions are associated with sharp increases in grant rates. They interpret this result as less rigorous screening and lower quality patents. While this is a striking finding, their analysis does not pin down whether it is driven by differences in extrinsic incentives, intrinsic motivation, or examiner opportunity costs, which our structural model will do.

Perhaps the most closely related paper is Schankerman and Schuett (2022), who develop an integrated framework to study patent screening, encompassing the patent application decision, examination, post-grant licensing, and litigation in the courts. They calibrate the model on data for the U.S. and use it to evaluate various counterfactual patent and court reforms. Their model estimates the effectiveness of patent examination, but they treat this as an exogenous parameter, but they do not model the prosecution process. In contrast, we develop and estimate the first equilibrium model of the patent examination process itself, which in turn allows us to investigate how various reforms to the incentives and design of patent screening affect the performance of this public agency.

Before turning to the data, we summarize a few key features of patents that guide our modeling choices. The critical feature of the patent document is the list of independent claims, which delineate the "metes and bounds," or scope, of the property right. The examination process involves assessing the patentability of *each* claim, not the patent as a single entity. In a departure from most existing literature, we treat a patent as a collection of claims that differ in both their private value and their similarity to previous patented claims. These two dimensions are a critical feature in the model. This heterogeneity is a first-order feature necessary to match the actual process of patent examination and to develop accurate statements about the potential effects of regime changes on the patent examination process.

## 3 Data and Descriptive Results

In this section, we describe our primary data sources, focusing on datasets not previously used in empirical studies of patents. We also present summary statistics and describe reduced-form evidence. Online Appendix B provides hyperlinks to all publicly available datasets and data sources we use in our empirical work.

*Distance Metric*

We construct a new measure of independent claim distance. To create this, we exploit the *U.S.PTO Patent Application Claims Full Text Dataset* and the *Granted Patent Claims Full Text Dataset*. The first dataset contains the full text for all U.S. patent application claims between 2001 and 2014 and an indicator for whether the claim is independent. The Granted Patent Claims Full Text Dataset records the full text for all U.S. patent claims granted between 1976 and 2014.

We summarize our approach to creating a distance measure here (see Online Appendix C for more detail). The approach calculates distances by representing a patent claim's text as a numerical vector and calculating a metric on that vector space.[3] We adopt the *Paragraph Vector* approach of Le and Mikolov (2014), which uses an unsupervised algorithm to "learn" the meaning of words by studying the context in which they appear and forming a vector representation for each word, picking up the meaning of paragraphs as a by-product.[4] As is common in the NLP literature, we measure distances between numerical vectors using the angular distance metric. To reflect distance to prior art, we compute the distance from each independent claim to every previously granted independent claim.[5]

*Rounds Data*

Since we estimate a model of the patent prosecution process over multiple rounds, comprehensive and reliable *round-level* data on the patent process are essential. We use the *Transactions History* data in the *Patent Examination (PatEx) Research Dataset* to create a dataset on the round-by-round evolution of utility patent applications between 2007 and 2014. In total, the transactions dataset includes 275.6 million observations covering 9.2 million unique applications. For every patent application, these data record examiner and applicant decisions at each round of the examination process.

*Sources Matched to Round Data*

We match the round-level data to three other datasets on patent applications. The first is the

[3] Kelly, Papanikolaou, Seru, and Taddy (2021) use similar methods to calculate patent similarity.

[4] The standard method (*bag-of-words*) for representing the patent claim text as a numerical vector has two significant weaknesses: it ignores the *ordering* and *semantics* of words.

[5] We conduct two falsification tests on our distance measure. First, we put independent claims into twenty, five-percentile bins of the distance measure and then calculate the proportion of claims rejected on novelty/obviousness grounds in each bin. We would expect that examiners are more likely to reject claims with a small distance to existing claims based on novelty/obviousness criteria. Thus the proportion of first-round rejections should be a declining function of the distance metric and the results confirm this prediction. Second, we conduct a similar test on the average number of examination rounds for each granted *patent*, by five-percentile bins of average distance of independent claims. Patents with higher average distance should be granted faster, and this is what we find.

*Application Data* in the *PatEx* Dataset, which contains features of the patent application, such as the identities of the applicant and examiner, the patent art unit (narrow technology classifications), and a binary indicator of the size of the applying firm (below or above 500 employees). Second, we match our data to renewal decisions by patent holders using the *U.S. PTO Maintenance Fee Events Dataset.* Third, since we focus on novelty/obviousness rejections, we require data on the *types* of rejections of each claim at each stage of the process. We obtain this from the *U.S.PTO Office Action Research Dataset for Patents.*

*Legal Fees*

For attorney fees, we use data from the *2017 American Intellectual Property Law Association (AIPLA) Report of the Economic Survey.* The survey reports means and percentiles of the distribution of hourly fees for different tasks, such as preparing and filing an application, issuing, paying renewal fees, and amending applications, split into three broad technology areas (biotechnology/chemical, electrical/computer, and mechanical). We use these moments to estimate the distributions of application and fighting costs for each patent application, adjusted for inflation.

*Seniority and Technology Complexity Credit Adjustments*

We obtained data on examiner seniority from Frakes and Wasserman (2017), who provide a panel of General Schedule (GS) grades for examiners, including each examiner's promotion dates. Using this, we work out the seniority of the examiner for each application. Finally, we received information on examiner extrinsic rewards from the Patent Office at the disaggregated U.S. Patent Classification level and then aggregated them to the technology center level.

*Descriptive Statistics*

Several features of the data are worth noting (Table A.1 in the Online Appendix provides details). First, 70% of applications resulted in the issuance of a patent. However, this is a misleading measure of the fraction of *content* granted because, as we will see, most applications are heavily narrowed during the examination process. Second, the prosecution time varies across applications—the mean duration is 2.96 years, and the mean number of rounds is 2.40. Third, the mean, median, and modal number of independent claims is three. Fourth, 24% of applications were by firms with fewer than 500 employees (a so-called "small entity"). Lastly, 46% of granted patents were renewed to the statutory limit, and only 13% were not renewed at the first renewal date.

Existing studies show that patent grant rates vary widely across technology centers and examiner seniority, with more senior examiners granting more frequently (Frakes and Wasserman, 2017; Sampat and Williams, 2019). In Online Appendix D, we confirm these findings about grant rates

using our data, and we also show that the likelihood of multi-round negotiation (lasting beyond one round) is much lower for senior examiners and varies substantially across technology centers. In addition, small entities are less likely to negotiate. We also analyze the variation in these outcomes for *each* examiner, decomposing variation in examiner-specific outcomes (such as their grant rate) within and between technology center-seniority pairs.[6] This decomposition shows that 80% of the variation in examiner grant rates and 81% of the variation in each examiner's average number of rounds is within-group variation.

Our model allows for several factors that can explain the substantial variation in examiner statistics even within seniority-technology-center dyads: we allow for a different distribution of intrinsic motivation for junior and senior examiners, we incorporate differences in the examiner credit structure across seniorities and technology centers, and we allow for heterogeneous legal (fighting) costs for applicants across technology centers. Our parameter estimates will enable us to disentangle the effects of these factors in explaining the variation in outcomes.

# 4 Model of the Patent Screening Process

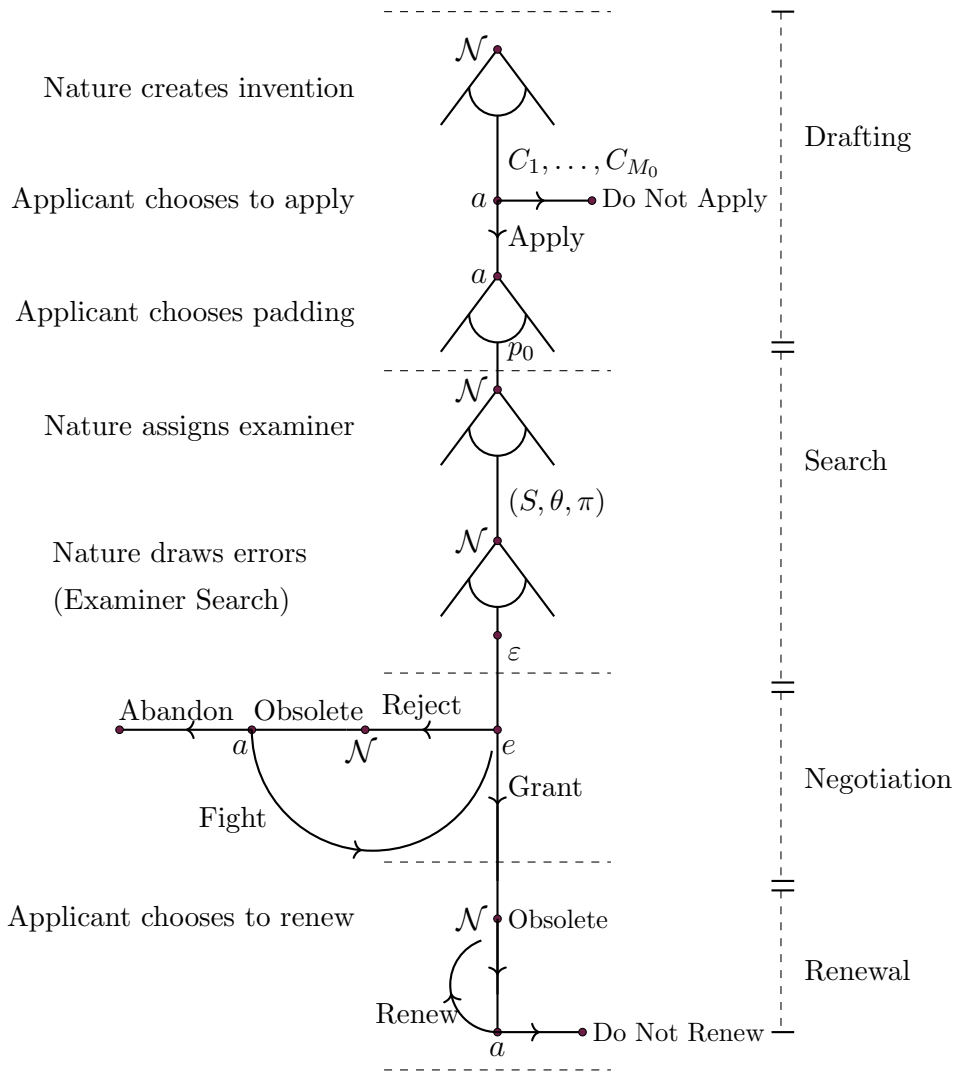
We model the patent screening process as a dynamic game in technology center $T$, between an inventor, $a$, and a realization of the examiner, $e$. There are four potential stages: (1) Application Decision and Patent Drafting, (2) Examiner Search, (3) Negotiation, and (4) Renewal. Figure 1 depicts the extensive form of the model.

In the baseline model, we analyze patent screening *conditional* on the invention being developed. For the validity of the structural model (and the counterfactual analysis), we do not need to model the potential inventor's decision whether to invest to develop their idea into an invention. However, to quantify the net social costs associated with these errors, we need to model the decision to develop (as well as how the patentee licenses their invention), which we do in Section 8.

Regarding the examiner, we present a model of how they act on *one specific application*. Therefore, we focus on intra-application incentives and costs for the examiner, rather than inter-application incentives induced by factors such as meeting their quarterly credit targets. A model in which examiners make decisions over time with consideration of the complete set of examinations in their docket would introduce significant complications and is not necessary to meet the

[6] Table D.2 provides more detail, along with the proportion of within-group variation for other dependent variables, such as mean examination length, mean number of rounds, etc.

Figure 1. Extensive Form of the Model




aims of our model.

## 4.1 Application Decision and Patent Drafting

### 4.1.1 Inventor Type

An inventor is endowed with a developed invention they are considering patenting. The patent application for the invention consists of $M_0$ initial independent claims $(C_1, ..., C_{M_0})$.[7] We characterize an independent claim $C_j$ by the pair $(D_j^*, v_j^*)$ where $D_j^* \sim G_D(\cdot)$ is the distance of the

[7] As we want to focus on the economic incentives for the applicant, we do not consider any agency issues between the inventor and the patent attorney who actually drafts the application.

true version of claim $j$ to the nearest claim in any existing invention and $v_j^* \sim G_v(\cdot)$ denotes the initial flow net returns generated by the true version of claim $j$ once it is commercialized.[8] We define the returns $v_j^*$ as relative to the inventor's outside option, e.g., protecting the invention by trade secrecy.[9]

#### 4.1.2 Application Decision

First, the inventor decides whether to apply. If they do not, the game ends, and their payoff is zero. If they do, they become an *applicant*, and the game continues. The inventor, a risk-neutral expected utility maximizer, chooses to apply if the expected utility of the game that follows applying is positive (because flow returns are defined relative to the next best alternative).

#### 4.1.3 Padding

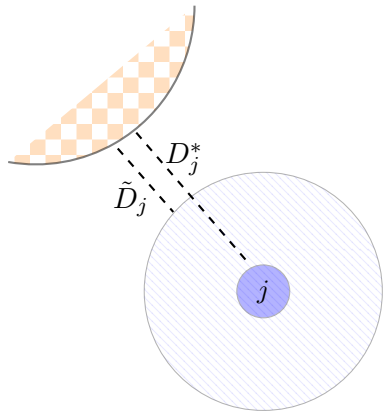
After deciding to apply, the applicant chooses the amount to exaggerate the claims on their patent application. We refer to this as the initial choice of *padding*, denoted $p$. Padding obfuscates the true "metes and bounds" of the invention, thereby concealing the inventive step and expanding the property rights claimed in the application. Padding allows the patent owner to extract potentially more revenue, by working it themselves or licensing it. However, greater padding also entails some obfuscation in defining the relationship between the actual invention and the boundaries of the patent rights claimed and necessarily moves the application closer to the prior art. Figure 2 illustrates the concepts of independent claims and padding.

There is a tradeoff for the applicant in the choice of padding. The advantage is that it increases the initial returns of claim $j$ for the applicant from $v_j^*$ to $\tilde{v}_j^0 = \mathcal{V}(v_j^*, p)$, where the padded value function $\mathcal{V}(\cdot,\cdot)$ is increasing in both arguments. On the other hand, padding increases the likelihood of examiner rejections during the examination process on the grounds of non-obviousness (closeness to existing patents) and indefiniteness. Padding shrinks independent claim distances from $D_j^*$ to $\tilde{D}_j^0 = \mathcal{D}(D_j^*, p)$, where the padded distance function $\mathcal{D}(\cdot,\cdot)$ is increasing in $D_j^*$ and decreasing in $p$. For simplicity, we assume that value (distance) is proportional (inversely

[8] We assume that distances and values are uncorrelated. Based on the theoretical literature on differentiated products, the relationship is ambiguous. Other things equal, being further from rivals (in product space, which we assume is correlated with claim distances) softens price competition and thus increases private value – implying a positive correlation between distance and value. However, the distribution of demand will typically vary with location, with firms endogenously locating (patenting) in areas of high demand. This implies a negative correlation between distance to rivals and values.

[9] Table A.2 provides our choices for parameterized distributions of distances $G_D(\cdot)$ and values $G_v(\cdot)$ (along with all other distributions in the model).

Figure 2. Distances and Padding




*Notes*: The orange semicircle in the top left corner represents the closest existing invention to the independent claim $j$, which is the small full blue circle in the bottom right corner. The applicant pads the true independent claim to create the larger cross-hatched circle. The distance between the true independent claim and the nearest existing invention is $D_j^*$, whereas the distance between the padded claim and nearest point is $\tilde{D}_j$.

proportional) to the degree of padding: $\tilde{v}_j^0 = p \cdot v_j^*$ and $\tilde{D}_j^0 = D_j^*/p$.

Finally, there is a direct cost of padding in the form of legal costs, which we assume is proportional to padding because heavily padded applications require more time to craft.[10] In particular, we specify legal costs as $F_{\text{app}} = f_{\text{app}} \cdot (1 + |p - 1|)$, where $f_{\text{app}}$ is the attorney fees associated with patent drafting (which is log-normally distributed across applicants). The motivation for this specification is that it takes additional time for the attorney either to under-pad ($p < 1$) or over-pad ($p > 1$); writing down the truth ($p = 1$) is quickest. We assume symmetry for simplicity.

#### 4.1.4 Applicant Expected Utility

The applicant decides the initial level of padding without knowing the identity of the examiner the Patent Office will assign. This feature is relevant because examiners differ in types (seniority, time cost, and intrinsic motivation) and, thus, in their strategies. As a result, applicants make initial padding decisions in light of the distribution of examiner types. The applicant chooses initial padding to maximize their expected utility less application legal costs, where the expectation is

[10]The applicant may choose to *understate* the true scope of the invention ($p < 1$) and thus earn lower returns, as it reduces the likelihood of rejection by the examiner (especially if there is a restriction of the number of rounds allowed). We find some evidence of such under-padding in the empirical results.

taken first over the roster of potential examiners $e = 1, \ldots \underline{E}$ (where the random assignment of applications implies an equally likely chance of each examiner in the relevant technology center), over the error of the examiner $\varepsilon \sim G_{e,\varepsilon}(\cdot)$, and potential obsolescence of their invention $\boldsymbol{\omega}$ (all described later).

Formally, the applicant's optimal padding choice $p_0$ maximizes the ex ante value of patent rights $\Gamma(p)$, defined[11]

$$\Gamma(p) = \mathbb{E}_{e,\varepsilon,\boldsymbol{\omega}}\left[U_a^0(e, \varepsilon, \boldsymbol{\omega}, p)\right] - F_{\text{app}}(p),$$

where

$$\mathbb{E}_{e,\varepsilon,\boldsymbol{\omega}}\left[U_a^0(e, \varepsilon, \boldsymbol{\omega}, p)\right] = \frac{1}{\underline{E}} \sum_{e=1}^{\underline{E}} \int \mathbb{E}_{\boldsymbol{\omega}} U_a^0(e, \varepsilon, \boldsymbol{\omega}, p)\, dG_{e,\varepsilon}(\varepsilon),$$

and we define $\mathbb{E}_{\boldsymbol{\omega}} U_a^0(e, \varepsilon, \boldsymbol{\omega}, p)$, the applicant expected utility (over the full vector of obsolescence) for a given examiner $e$ and error $\varepsilon$, later in Equation (4).[12] The applicant applies if

$$\Gamma^* \equiv \Gamma(p_0) \geq 0. \tag{1}$$

## 4.2 Examiner Search

### 4.2.1 Examiner Assignment

The patent office assigns the application randomly to an examiner within the relevant art unit of the technology center. We characterize an examiner by the tuple $(S, \theta, \pi)$. The first term $S$ represents examiner seniority. The type $\theta \sim G_{S,\theta}(\cdot)$ corresponds to the level of intrinsic motivation. Intrinsically motivated workers incur a disutility from awarding patent rights that do not meet the patentability standard, based on the information available to them (see Section 4.4.3 for formalization of how this enters the examiner's payoff). We let the distribution of $\theta$ depend on seniority $S$. Finally, $\pi \sim G_\pi(\cdot)$ corresponds to the examiner's cost of delay (i.e., the extra effort cost for going another round plus any pressure costs associated with timely docket management). The effort cost component will reflect the examiner's productivity.

[11] Throughout, we use the notation $\mathbb{E}_{\boldsymbol{\omega}}$ to denote expectations taken over the vector of obsolescence shocks that are not yet realized. Before applying, this is the full vector of 20 possible shocks that could occur, one each year after application. As the process continues, obsolescence shocks occur, and fewer shocks are left to be realized. With a slight abuse of notation, whenever we use $\mathbb{E}_{\boldsymbol{\omega}}$ with an emboldened $\boldsymbol{\omega}$, it refers to the sub-vector of $\boldsymbol{\omega}$ that have not yet occurred. The notation $\mathbb{E}_{\omega_r}$ refers to an expectation over $\omega$ only in round $r$.

[12] We simplify notation by using $e$ to denote both the random variable reflecting the (unknown) examiner prior to application its realization after applying. The same holds for examiner errors.

#### 4.2.2 Examiner Grounds for Rejection

Once assigned, the examiner learns the applicant's identity and thus their fighting costs. The examiner also knows the padded value of the application to the applicant. The examiner reads the application and independently searches the existing prior art to assess the grounds for rejection throughout the negotiation process. There are three main grounds for rejection: *novelty*, *non-obviousness*, and *indefiniteness*. Novelty requires that the claim has not been in use for one year before filing. Non-obviousness requires that the claim makes an inventive step beyond the closest existing invention that would not be immediate to anyone skilled in the relevant area. Indefiniteness requires that the claim is precise and clear on the exact boundaries of claimed property rights. In this paper, we focus on novelty/non-obviousness.[13]

After searching the prior art, the examiner assesses the obviousness/novelty of *each claim* $j$, with their assessment denoted by $\hat{D}_j$ and equal to

$$\hat{D}_j = \mathcal{D}(D_j^*, p) \cdot \varepsilon,$$

where $\varepsilon$ denotes the drawn examiner error in assessing obviousness/novelty, which is assumed to be independent of the true distances $D_j^*$. The distribution of search errors depends on the seniority of the examiner and may also depend on the technology center since the number and complexity of patents and other prior art vary across technology fields.

The distribution of search errors also depends on the intrinsic motivation of the examiner. We specify that the *mean* of the search error distribution satisfies two criteria. The first is that the mean of the error tends to one (the unbiased case) as $\theta \to \infty$. The second is that for all $\theta < \infty$, the mean of the search error distribution is greater than one. We specify the second feature because examiners who are not perfectly intrinsically motivated do not scour the literature so thoroughly, thereby missing relevant prior art. When they miss relevant prior art, they perceive distances to be larger than they are and hence have errors greater than one. However, these requirements do not force one-sided examiner error since some draws may still be below one, even if the mean is above one. Our functional form choice satisfying this assumption is $\mu_\varepsilon = 1 + \frac{1}{\theta}$.

We say the examiner has *grounds* for an obviousness rejection if $\hat{D}_j$ is less than an obviousness

[13] Using the Office Action Research Dataset described in Section 3, which identifies the reasons the examiner rejects claims in a patent at each round, we analyzed the overlap between novelty/non-obviousness (102/103) and indefiniteness (112) rejections. We find that 73% of office actions containing a 112 rejection also contain a 102/103 rejection. Thus, novelty/non-obviousness rejections cover most of the observed indefiniteness rejections, so omitting indefiniteness from the baseline model is a profitable abstraction.

threshold $\tau$. However, having grounds for rejection will not necessarily mean the examiner will reject the claim. The examiner's decision will be the one that maximizes their utility, taking into account their explicit incentives (credits) and intrinsic motivation. This is a crucial point as it implies that examiners' decisions in the data may not align with decisions made solely on legal grounds.

Finally, examiner errors are specified to be constant throughout the negotiation stage. In this sense, there is no updating of examiner error. However, the grounds for rejection will be recalculated at every negotiation round as the applicant narrows the extent of padding in response to a rejection by the examiner.

### 4.3 Information Structure

The information structure for the applicant and examiner is as follows. The inventor knows the set of claims covered by the invention (given by nature), their true distance to all prior art, and the private value of each claim. Before deciding whether to apply for a patent, the inventor does not know which examiner will be assigned to the application. After assignment, the applicant knows the characteristics of the examiner, including the level of intrinsic motivation, productivity, seniority, and structure of patent office incentives the examiner faces. The applicant also knows the structure of the process and the fees imposed by the patent office at each stage.

The assigned examiner does not observe the true claim distances or the applicant's extent of padding, only the padded distances, contaminated by examiner error, for each claim in the application. The examiner does not know the error she makes in determining the claim distances during the search of prior art. The examiner observes the fighting costs and padded private value of the applicant's claims.[14] Since the examiner reports their assessment of the padded distance to the applicant, the applicant knows the examiner's error.

### 4.4 Negotiation

The Negotiation Stage is a finitely repeated version of the stage game shown in the "Negotiation" section of Figure 1. At round $r$, if required to act, first, the examiner chooses whether to grant or abandon and, if rejected, the applicant chooses whether to abandon or fight. In between the examiner's and applicant's decision, the applicant's invention can become obsolete, in which case

[14] We could assume that the examiner does not perfectly observe the private value, but instead obtains an unbiased signal of the value. This feature would not deliver any additional insights and would increase computational burden.

the applicant abandons it immediately. The applicant and examiner discount each stage at rate $\beta$.

Let $\mathbf{x}_a$ and $\mathbf{x}_e$ be the vector of strategies of the applicant and examiner, respectively, if the invention is not obsolete.[15] We detail the actions and payoffs obtained at the two decision nodes, starting at the point at which the examiner has just rejected in round $r$ so that $x_e^r = \text{REJ}$.

### 4.4.1 Obsolescence and Credits

First, pre-grant obsolescence, denoted by $\omega_r$, is realized. If $\omega_r = 1$, the applicant's invention becomes obsolete. In this case, all returns shrink to zero permanently, and trivially, the applicant abandons and obtains a period payoff of zero.[16] In this case, the examiner obtains a period payoff of credits $g^r_{ABN}(S,T)$. If the invention does not become obsolete, then $\omega_r = 0$, and the applicant makes a non-trivial decision. Formally, obsolescence is a Markov process, where, for all $r$, if $\omega_r = 1$, then $\omega_{r+1} = 1$ (an absorbing state). Otherwise, if $\omega_r = 0$, $\omega_{r+1}$ is a Bernoulli random variable with parameter $P_{\omega,\text{pre}}$ if we are still in the application process, and parameter $P_{\omega,\text{post}}$ if a patent has been granted and we are in the renewals process.

We provide the full schedule of examiner credits in Appendix E. The most important feature to note is that credits weakly decline as the applicant enters subsequent requests for continued examination, which make early granting more attractive to the examiner.

### 4.4.2 Applicant Decision

Upon receiving a rejection, if the invention has not become obsolete, the applicant has two choices. They can abandon ($x_a^r = \text{ABN}$), in which case the applicant's and examiner's payoffs are as described in the event of obsolescence. Instead of abandoning, the applicant can continue the application ($x_a^r = \text{FIGHT}$). Continuing involves narrowing rejected claims, which we model as a reduction in padding $p$ by proportion $\eta$.[17] Hence for all *rejected* claims $j$, the padding

[15] Of course, the vectors include a rejection/acceptance decision and abandonment/fight decision for *every* round. To check whether a strategy is optimal, we must specify what each player would do in every round, even if the prior parts of the strategy dictate that this round will not be reached on the equilibrium path.

[16] The applicant obtains a period payoff of zero because the Patent Office reveals all applications (after 18 months), so their potential for appropriation of innovation returns (e.g., by trade secrecy as an alternative) has essentially vanished.

[17] We could extend the model to allow the applicant to choose whether to narrow by proportion $\eta$ with some probability or respond by arguing that the examiner is in error and not narrow at all. However, our data on patent word counts imply that this extension is empirically unimportant. To see this, we look at word counts on patents granted with one rejection after publication and calculate the proportion of cases whether the applicant resubmits

becomes $p_{j,r+1} = \eta p_{j,r}$. The padding level remains the same for all accepted claims.

Continuing involves a fighting cost to the applicant. The applicant must pay the attorney the fee for amending the application, $F_{\text{amend}}$. In the case of a Request for Continued Examination, the applicant must pay the associated patent office fee, $F^r_{\text{round}}$. Continuation involves delay costs for the examiner, denoted by $\pi$. After narrowing occurs, the applicant pays fighting costs, and we move to round $r+1$.

Formally, let the value function for the applicant *upon being rejected in round* $r$ be $U^r_a(\omega_r, \mathbf{x}_e)$. Clearly, the value function for the applicant is a function of the future strategies of the examiner. Further, because $\omega$ is a Markov process, the value function for the applicant only depends on the realization of $\omega$ in period $r$. The term $U^r_a(\omega_r, \mathbf{x}_e)$ is defined as follows. If the invention becomes obsolete, so that $\omega_r = 1$, we have (for all $\mathbf{x}_e$)

$$U^r_a(1, \mathbf{x}_e) = 0. \tag{2}$$

Otherwise,

$$\begin{aligned} U^r_a(0, \mathbf{x}_e) = \max\Bigg\{0, -F_{\text{amend}} \quad &- \quad F^{r+1}_{\text{round}} + \beta\Big(1(x^{r+1}_e = \text{GR})[V^{r+1} - \phi] \\ &+ \quad 1(x^{r+1}_e = \text{REJ})\mathbb{E}_{\omega_{r+1}} U^{r+1}_a(\omega_{r+1}, \mathbf{x}_e)\Big)\Bigg\}, \end{aligned} \tag{3}$$

where $1(A)$ is the indicator function, equal to one if statement $A$ is true and zero otherwise, $V^{r+1}$ defines the ex post net expected benefits from patent rights if granted in round $r+1$, as given in Equation (9) in Section 4.5, and $\phi$ is the finalizing fee. Equation (3) says that the value for the applicant in round $r$, provided they are not obsolete, is either zero if it is optimal for them to abandon or the sum of fighting costs, plus either the payoff of being granted in the next round (if the examiner will grant them) or the expected value from round $r+1$ if the examiner will reject them in round $r+1$ (both discounted by $\beta$).

If $x^{r+1}_e = \text{GR}$, and $\omega_r = 0$, the applicant abandons in round $r$ if

$$F_{\text{amend}} + F^{r+1}_{\text{round}} > \beta[V^{r+1} - \phi]$$

and if $x^{r+1}_e = \text{REJ}$, the applicant abandons in round $r$ if

$$F_{\text{amend}} + F^{r+1}_{\text{round}} > \beta\mathbb{E}_{\omega_{r+1}} U^{r+1}_a(\omega_{r+1}, \mathbf{x}_e).$$

an application with the same word count. This happens only 7% of the time, so we view the choice to ignore the possibility of no narrowing as a profitable abstraction in the baseline.

At this point, we can define the expected utility for the applicant before applying, for a given choice of padding, as

$$\mathbb{E}_{\boldsymbol{\omega}} U_a^0(e, \varepsilon, \boldsymbol{\omega}, p) = 1(x_e^1 = \text{GR})[V^1 - \phi] + 1(x_e^1 = \text{REJ})\mathbb{E}_{\omega_1} U_a^1(\omega_1, \mathbf{x}_e), \tag{4}$$

where all four terms on the right-hand side are (implicitly) functions of the level of padding.

#### 4.4.3 Examiner Grant/Rejection

If the applicant fights ($x_a^r = \text{FIGHT}$), we move to a new round $r+1$, and the examiner obtains updated assessments $\hat{D}_j^{r+1} = \mathcal{D}(D_j^*, p_0\eta^r)$ on previously rejected claims. Based on their updated assessment, the examiner recalculates the grounds for rejection and decides whether to grant the patent.

*Granting*

Granting a patent in round $r+1$ ($x_e^{r+1} = \text{GR}$) ends the negotiation game and moves the applicant into the renewal stage. Let $\mathcal{R}^{r+1} \in [0, 1]$ denote the proportion of claims the examiner thinks they should reject on obviousness/novelty grounds. Then the immediate payoff to the examiner from granting is

$$\mathcal{G}^{r+1} = g_{GR}^{r+1}(S, T) - \theta\mathcal{R}^{r+1}.$$

Here $g_{GR}^{r+1}(S, T)$ is the credit received by the examiner for granting at stage $r+1$. The term $\theta\mathcal{R}^{r+1}$ captures the intrinsic utility cost for the examiner. For intuition on this term, consider the extreme cases. When $\mathcal{R}^{r+1} = 0$, the examiner believes there are no independent claims on which they have grounds to reject and therefore feels no intrinsic disutility in granting the application. On the other hand, when $\mathcal{R}^{r+1} = 1$, the examiner believes that they should reject every independent claim, so the examiner is going against the organization's mission statement in granting a patent. The examiner's intrinsic penalty from premature granting is the product of the proportion of strategically incorrect claim acceptances and their intrinsic motivation parameter.

One might be concerned that our specification of intrinsic motivation also captures examiner career concerns within the Patent Office. Even if the examiner were not intrinsically motivated, their internal career prospects may depend on the frequency with which they grant invalid claims. However, while the Office does have a "random review" of examiners' decisions by a senior panel, these reviews are very rare, they do not come with explicit punishments, and Patent Office data confirm that decisions are frequently successfully appealed by the head examiner in the art unit.

*Rejecting*

If the examiner chooses not to grant in round $r+1$ ($x_e^{r+1} = \text{REJ}$) they get credits $g_{REJ}^{r+1}(S, T)$, and the stage game continues. The examiner follows this choice by rejecting any claim on which

they believe there are grounds to reject. Hence, the examiner rejects any independent claim $j$ if $\hat{D}_j^{r+1} < \tau$. After this, the application moves back into the hands of the applicant, at which point another obsolescence realization occurs, and then the applicant decides again whether to abandon or continue.

Formally, we define the value function for *the examiner after rejecting in round* $r$ as $W_e^r(\omega_r, \mathbf{x}_a)$. The value function for the examiner satisfies

$$W_e^r(\omega_r, \mathbf{x}_a) = \begin{cases} g_{ABN}^r & \text{if} \quad x_a^r = \text{ABN or } \omega_r = 1 \\ -\pi + \beta \max\left\{\mathcal{G}^{r+1}, g_{REJ}^{r+1} + \mathbb{E}_{\omega_{r+1}} W_e^{r+1}(\omega_{r+1}, \mathbf{x}_a)\right\} & \text{if} \quad x_a^r = \text{FIGHT} \end{cases} \tag{5}$$

In the bottom branch of Equation (5), where the applicant fights, the value to the examiner of rejecting in round $r$ is the cost $\pi$ plus either the (discounted) benefits of granting in round $r+1$ or the net benefits of rejecting in round $r+1$, whichever is larger.

Given the applicant's strategy $\mathbf{x}_a$ the examiner grants in round $r$ if

$$\mathcal{G}^r > g_{REJ}^r + \mathbb{E}_{\omega_r} W_e^r(\omega_r, \mathbf{x}_a).$$

This says that the examiner grants if the period payoff from granting exceeds the credits from rejecting plus the expected continuation value from the point of having rejected in round $r$, with expectation taken over obsolescence outcomes.

### 4.5 Renewal

We enter the renewals stage if the examiner grants the patent and the applicant pays the finalizing fee. Our renewal model adapts Schankerman and Pakes (1986) to the United States context, adding a probability of post-grant obsolescence in addition to deterministic depreciation. Suppose the patent is granted in round $r$. The returns for each granted claim $j$ start at $\tilde{v}_{j,r} = v_j^* \cdot p_r$ and depreciate at rate $\delta$ each period after grant. With probability $P_{\omega,\text{post}}$, the invention becomes obsolete, at which point the returns shrink to zero permanently. To keep the patent rights, the applicant must pay renewal fees $F_4$, $F_8$, and $F_{12}$ at years four, eight, and twelve after grant. The patent life ends at $L = 20$ years after submission of the patent application, at which point the invention enters the public domain.

The renewal decisions by the applicant are those that maximize their expected utility from retaining patent rights. Formally, define the expected returns from years $t_1$ to $t_2$ as

$$\mathbb{E}_{\boldsymbol{\omega}} V_{t_1,t_2} = \sum_{t=t_1}^{t_2} \left[\beta(1-\delta)(1-P_{\omega,\text{post}})\right]^{t-t_1} \sum_j \tilde{v}_{j,r}$$

and let $I_t$ be equal to one if the applicant will renew at year $t$ (provided the patent is not obsolete) and zero otherwise. Then, the applicant will renew at year four if the net expected benefit after year four is positive:

$$V_4^{N,r} \equiv \mathbb{E}_{\boldsymbol{\omega}} V_{4,7} - F_4 + I_8 \beta^4 V_8^{N,r} > 0, \tag{6}$$

where $V_8^{N,r}$ is the net returns from patent rights after year eight, which is defined analogously. The renewal decision at year eight is analogous, and the decision at year 12 is similar, except there is no future renewal decision post year 12.[18] Finally, we define the ex post net expected benefits from patent rights, when granted in round $r$, denoted as $V^r$ (as in Equation (3), as

$$V^r = \mathbb{E}_{\boldsymbol{\omega}} V_{1,3} + I_4 \beta^4 V_4^{N,r}. \tag{9}$$

*Characterizing the Equilibrium*

For every given parameter vector and choice of padding, the negotiation game is a finite game of perfect information, and hence has a subgame-perfect equilibrium that can be found through backward induction. The equilibrium strategies ($\mathbf{x}_a^*$ and $\mathbf{x}_e^*$) are characterized by (for all $r$):[19]

1. $x_e^{r,*} = \text{GR}$ if and only if
$$\mathcal{G}^r > g_{REJ}^r + \mathbb{E}_{\boldsymbol{\omega}} W_e^r(\omega_r, \mathbf{x}_a^*).$$

2. If $x_e^{r+1,*} = \text{GR}$, $x_a^{r,*} = \text{ABN}$ if and only if
$$F_{\text{amend}} + F_{\text{round}}^{r+1} > \beta[V^{r+1} - \phi].$$

3. If $x_e^{r+1,*} = \text{REJ}$, $x_a^{r,*} = \text{ABN}$ if and only if
$$F_{\text{amend}} + F_{\text{round}}^{r+1} > \beta \mathbb{E}_{\omega_{r+1}} U_a^{r+1}(\omega_{r+1}, \mathbf{x}_e^*).$$

[18] To be precise, conditional on not becoming obsolete, the applicant renews at year eight if

$$V_8^{N,r} \equiv \mathbb{E}_{\boldsymbol{\omega}} V_{8,11} - F_8 + I_{12} \beta^4 V_{12}^{N,r} > 0, \tag{7}$$

and, conditional on not becoming obsolete, the applicant renews at year 12 if

$$V_{12}^{N,r} \equiv \mathbb{E}_{\boldsymbol{\omega}} V_{12,20-r} - F_{12} > 0. \tag{8}$$

[19] In practice, we limit the process to six rounds (around 95% of applications last at most three rounds of negotiation and the modal number is two) so the characterization holds for $r < 6$. In the sixth round, if rejected, we force the applicant to abandon. The examiner's continuation value is therefore only $g_{ABN}^6(S, T)$.

4. The terms $U_a^r(1, \mathbf{x}_e^*)$, $U_a^r(0, \mathbf{x}_e^*)$, and $W_e^r(\omega_r, \mathbf{x}_a^*)$, and $W_e^r(0, \mathbf{x}_a^*)$ satisfy Equations (2), (3), and (5), respectively.

5. $I_4$, $I_8$, and $I_{12}$ are equal to one if and only if inequalities (6), (7), and (8), respectively, are satisfied.

We want to highlight the important advantages of modelling patents as comprised of multiple claims rather than as a single object. First, a model with multiple claims allows a more realistic description of the patent prosecution system and thus a tighter link to the data on which the model parameters will be estimated. Second, endowing applicants with multiple claims allows for specific claims to be narrowed only up to the round at which they are granted. Third, a multiple claim model enables us to specify the examiners' intrinsic motivation disutility as a function of the proportion of granted claims they judge invalid.

# 5 Estimation

Our primary estimation method is simulated method of moments (SMM), though we estimate some parameters outside the model. For reference, Online Appendix Table A.2 summarizes all parameters and their associated distributional assumptions.

## 5.1 External Estimation

*Discount Rate* ($\beta$)
The data lack some detailed information for identifying all parameters. Specifically, discount rates are traditionally difficult to identify. Hence, we set $\beta = 0.95$—as in most of the literature (Pakes, 1986).

*Distance Threshold* ($\tau$)
We estimate the distance threshold externally using observations on claim distances and examiners' grant decisions. For every examiner $e$, we calculate the minimum value of the distances among claims they grant. This number corresponds to their "personal distance threshold," denoted as $\tau_e = \min_{j \in J_e} \tilde{D}_j$, where $J_e$ is the set of claims granted across all applications by examiner $e$. Since examiners are not perfectly intrinsically motivated, some examiners' personal thresholds are below the true threshold $\tau$ in cases where they knowingly grant patents with relatively small distances. However, we assume that the most intrinsically motivated examiner will have a personal threshold $\tau_e$ equal to the true threshold $\tau$. This assumption allows us to estimate the distance threshold as the maximum of the distribution of examiners' individual thresholds. The validity of this approach relies on $\max_{e=1,\ldots,\underline{E}} \tau_e \to \tau$ as the number of examiners in the tech-

nology center $\underline{E} \to \infty$.[20] We calculate these thresholds separately for each technology center to create technology-center-specific thresholds. Notably, our estimates of the threshold in different technology centers are very similar, ranging from 0.48 to 0.52.

*Applicant Fighting Costs* ($f$.)
We have data on the quantiles of the distributions of *amendment*, *maintenance*, and *issuance* hourly fees charged by lawyers. We assume these three costs are log-normally distributed. Since these moments directly correspond to the elements of applicant fighting costs and do not identify any other parameters in the model, we estimate the mean and variances of the log of fighting costs using the optimal two-step generalized method of moments estimation procedure for each of the three negotiation-based fighting costs. The data allow us to estimate different negotiation fighting cost distributions for simple applications (less than ten claims) and complex applications in chemical, electrical, and mechanical fields.[21]

*Depreciation of patent returns* ($\delta$)
Bessen (2008) estimates the combined effect of depreciation and the probability of obsolescence at 0.14, using U.S. renewal data. In our context, this corresponds to $(1 - P_{\omega,\text{post}}) \cdot \delta + P_{\omega,\text{post}} \cdot 1$. Hence, for each parameter guess of $P_{\omega,\text{post}}$, we extract the implied pure depreciation rate from this relationship.

## 5.2 Simulated Method of Moments

We estimate the remaining set of model parameters using SMM. The model does not admit an analytic solution for endogenous variables as a function of all the model primitives. Hence, the goal is to choose the parameters that best match the moments of the data with the corresponding moments computed from the model's numerical solution. We estimate the model using moments from the data described in Section 3, assuming the model's equilibrium generates the data.

We denote the full vector of parameters to estimate as $\boldsymbol{\psi} = (\boldsymbol{\psi}_e, \boldsymbol{\psi}_a)$. The vector of applicant parameters is $\boldsymbol{\psi}_a = \left(\eta, P_{\omega,\text{pre}}, P_{\omega,\text{post}}, \alpha_D, \beta_D, \mu_v, \sigma_v, \boldsymbol{\mu}_{\boldsymbol{f}_{\text{app}}}, \boldsymbol{\sigma}_{\boldsymbol{f}_{\text{app}}}\right)$. We described the narrowing

[20] In practice, we experiment with the first and fifth percentiles as robustness checks. We also remove examiners who have conducted fewer than a threshold number of examinations. We experiment with values of 50 and 100 for this threshold and find only minor differences in all of these cases.

[21] On application fighting costs, though we have similar moments on lawyers' application drafting fees, because application fighting cost is proportional to padding in the model, its distribution is contaminated by the endogenous choice of padding (which is a function of all model parameters). This feature means that we cannot estimate the distribution of application fighting costs outside the model: we must estimate these parameters as part of the simulated method of moments procedure described in the next subsection.

and obsolescence parameters in Section 4. For distances, we assume $D_j^*$ is Beta distributed with parameters $(\alpha_D, \beta_D)$. The Beta distribution is a natural choice as it provides a flexible distribution on the interval $[0, 1]$, which coincides with the interval of our distance metric. Further, we use a multivariate normal distribution copula to correlate claim distances within an application.[22] Motivated by Schankerman and Pakes (1986), the log of initial *claim* flow returns is normally distributed with mean $\mu_v$ and variance $\sigma_v^2$. Finally, we assume that the log of application drafting legal fees per unit padding, $f_{app}$, are normally distributed with mean $\mu_{f_{app}}$ and variance $\sigma^2_{f_{app}}$, with different parameters for simple and complex applications in chemical, electrical, and mechanical fields.

The vector of examiner parameters is $\boldsymbol{\psi}_e = (\mu_{\theta,\text{junior}}, \mu_{\theta,\text{senior}}, \sigma_\theta, \mu_\pi, \sigma_\pi, \sigma_\varepsilon)$. The first three parameters $(\mu_{\theta,\text{junior}}, \mu_{\theta,\text{senior}}, \sigma_\theta)$ correspond to log-normal parameters for the distribution of examiner intrinsic motivation. We estimate different $\mu$ parameters for "junior" (pre-GS-14 grade) and "senior" examiners. Though we constrain the $\sigma$ parameter to be the same for juniors and seniors, given the log-normal specification, this does not force the variances (or even the variance relative to the mean) to be the same for juniors and seniors. The log of examiner delay costs, $\pi$, are normally distributed with mean $\mu_\pi$ and variance $\sigma_\pi^2$. Finally, examiner errors are normally distributed, with mean $1 + \dfrac{1}{\theta}$ and variance $\sigma_\varepsilon^2$.

We estimate $\boldsymbol{\psi}$ using a minimum-distance estimator that matches moments of the data with the corresponding moments implied by the model. More specifically, for any value of $\boldsymbol{\psi}$, we solve the model for several simulated draws from the distributions of exogenous variables. Then, we calculate moments of the endogenous variables across the simulated observations. The minimum-distance estimator minimizes the SMM objective function:

$$\hat{\boldsymbol{\psi}} = \arg\min_{\psi} \left(\boldsymbol{m}(\psi) - \boldsymbol{m}_{\mathcal{S}}\right)' \Omega \left(\boldsymbol{m}(\psi) - \boldsymbol{m}_{\mathcal{S}}\right),$$

where $\boldsymbol{m}(\psi)$ is the vector of simulated moments computed from the model when the parameter vector is $\psi$, $\boldsymbol{m}_{\mathcal{S}}$ is the vector of corresponding sample moments, and $\Omega$ is a symmetric, positive-definite weighting matrix.[23]

[22] Specifically, in the simulation, for each application, we draw a vector of size $M_0$ from a standard multivariate normal with correlation coefficient $\rho$. We apply the quantile function of the normal to the draws to create correlated uniform random variables. Then for the estimation guess $(\tilde{\alpha}_D, \tilde{\beta}_D)$, we apply the inverse CDF of a Beta distribution with these parameters to the uniform draws to generate correlated beta distributed initial distances. For $\rho$, we use the empirical correlation of granted distances. Simulations confirm that the correlation of the multivariate copula is very close to the correlation of the distances. See Nelsen (2007) for details.

[23] For the weighting matrix we use a diagonal matrix that scales moments to a uniform scale. We cannot use the optimal two-step weight matrix because we do not have application-specific data on fighting costs that can

### 5.3 Choice of Moments

We now briefly describe our choice of moments for the SMM estimation. In Online Appendix F, we provide some intuition about how these moments aid in identifying the parameters we estimate.

The number of moments we can calculate on endogenous variables in the model far exceeds the number of model parameters. To select a subset of moments for our estimation procedure, we followed a rigorous, data-driven methodology, based on the sensitivity matrix of parameter estimates to the inclusion of particular moments (Andrews, Gentzkow, and Shapiro, 2017), along with plots of how estimated model moments (and separately, the value of the SMM objective) vary with parameter values. We provide details on the complete set of moments we considered and our pruning procedure in Appendix F. Through this procedure, we pruned the set of moments down to 40 that clearly assist in estimating the parameters.

The selected moments corresponding to outcomes for examiners are the proportion of applications granted by round and seniority, the standard deviation of examiner rejection rates by seniority, and the proportion of patents granted containing an invalid claim (again, by seniority and round). The selected moments corresponding to outcomes for applicants are the proportion of abandonments by round and examiner seniority, patent renewal rates, means and standard deviations of granted claim distances by round granted, and means and medians of legal application fees by technology class.

## 6 Empirical Results

In this section, we present and interpret our parameter estimates and briefly discuss model fit and robustness. For model estimates, we bootstrap standard errors. Standard errors are negligible for all parameters, which is unsurprising since we calculate data moments using millions of observations.

### 6.1 Applicant Parameters

Table 1 presents the estimates for parameters relating to the applicant. First, we estimate the proportion of narrowing per round as $1 - \eta = 0.25$. This estimate indicates that screening substantially narrows over-claiming by the applicant. Second, we estimate two probabilities of

allow us to compute the correlation between these moments and others. Details on computation and numerical optimization are available on request.

Table 1. Applicant Parameter Estimates

| Parameter | Symbol | Estimate | S.E. |
|---|---|---|---|
| Per-round narrowing | $1-\eta$ | 0.25 | 0.000 |
| Pre-grant obsolescence | $P_{\omega,\text{pre}}$ | 0.14 | 0.001 |
| Post-grant obsolescence | $P_{\omega,\text{post}}$ | 0.04 | 0.000 |
| Initial returns log-mean | $\mu_v$ | 10.55 | 0.077 |
| Initial returns log-sigma | $\sigma_v$ | 1.32 | 0.022 |
| Initial distance alpha | $\alpha_D$ | 4.57 | 0.003 |
| Initial distance beta | $\beta_D$ | 7.74 | 0.004 |
| Simple application fighting cost log-mean | $\mu_{f,\text{simple}}$ | 8.53 | 0.011 |
| Simple application fighting cost log-sigma | $\sigma_{f,\text{simple}}$ | 0.87 | 0.054 |

*Notes*: This table provides the applicant's model parameters. Standard errors are bootstrapped. Table A.3 provides fighting cost parameters by technology area.

obsolescence: a pre-grant probability during the application process and post-grant obsolescence during the patent's life. The estimated pre-grant obsolescence probability is 14% for each negotiation round. The post-grant rate is 4% per year, which is broadly similar to other estimates in the literature.[24] The probability of obsolescence is higher during the application process for two reasons. First, applicants are more likely to discover their invention to be obsolete earlier in its life cycle (e.g., discovering that commercialization costs make the project unviable). Second, the prosecution stage contains applications that are eventually granted and those who abandon, and many of those who abandon do so precisely because they become obsolete.

Third, the distribution of initial returns from an unpadded independent claim is highly skewed. Though the mean is \$91,046, the median is \$38,069, and the modal value of initial returns for an unpadded independent claim is \$6,656. To understand the distribution of unpadded initial returns on the *application*, we take the distribution of the number of independent claims and use it to construct sums of draws from the distribution of claim returns. For example, the first patent application in our dataset has two independent claims. Hence, we draw two values from the distribution of claim initial unpadded returns and add them to get the total initial unpadded

[24]Using data for three European countries, pooled across technology fields, Pakes (1986) calculates values of 6%, 4% and 1% for the likelihood of obsolescence in the first, second and third year after grant, respectively. Using German data, disaggregated by four technology areas, Lanjouw (1998) estimates a range of 7-12%.

returns on that application. The median initial unpadded returns from a patent application are \$129,659.

It is difficult to compare our estimates of initial returns to existing estimates in the literature on total patent returns since we estimate the distribution of initial returns for (a) all *applications* (not just granted ones) and (b) *unpadded* claims. Nonetheless, it is worth noting that Bessen (2008) estimates the mean net present value of patents (adjusted to 2018 U.S.D) for all U.S. patentees as \$78,168 and \$113,067 for just U.S. public firms in manufacturing.

Next, we discuss the implied distribution of initial unpadded distances and fighting costs. The mean distance is 0.37, and the distribution is approximately symmetric. Given that our estimated thresholds are between 0.48 and 0.52, these estimates imply that about 83% of application claims have distances below the threshold. Despite this, many applications are eventually granted because of extensive narrowing and examiners granting invalid claims. Fighting costs for simple applications are lower than all other categories. Recall that legal costs per application are specified as $F_{\text{app}} = f_{\text{app}} \cdot (1 + |p - 1|)$, where $f_{\text{app}}$ is the attorney fees associated with patent drafting. Evaluated at the mean levels of $p$ and $f_{\text{app}}$, we estimate these transaction costs at \$7,920 for simple applications and \$12,333 for electrical applications.

*Padding (overclaiming property rights)*
We compute statistics on the model's endogenous variables by simulating the model at our estimates. Relating to the applicant, we calculate the distribution of optimal initial padding for those who apply. The mean padding level is 8%, with $70^{th}$ and $90^{th}$ percentiles equalling 18% and 31%, respectively. These results suggest that many applicants substantially exaggerate the true extent of their invention when they apply for patent rights.

We also compute two weighted averages of padding, where our weights are either the mean (over claims) of initial unpadded distances ($\bar{D}_s^*$) or initial unpadded values ($\bar{v}_s^*$). The weighted average of padding rises to 10% when weighted by values and 9% when weighted by distances, indicating that inventors increase padding for applications with claims that are more valuable and distant from the prior art (where such padding is less likely to induce the examiner to reject).

## 6.2 Examiner Parameters

Table 2 presents the estimates of the examiner parameters. To understand examiner costs and intrinsic motivation, we provide a slight digression on the units of examiner payoffs in the model,

Table 2. Examiner Parameters Estimates

| Parameter | Symbol | Estimate | Standard Error |
|---|---|---|---|
| Junior intrinsic motivation log-mean | $\mu_{\theta,\text{junior}}$ | 3.92 | 0.004 |
| Senior intrinsic motivation log-mean | $\mu_{\theta,\text{senior}}$ | 3.38 | 0.005 |
| Intrinsic motivation log-sigma | $\sigma_\theta$ | 0.77 | 0.055 |
| Delay cost log-mean | $\mu_\pi$ | 0.19 | 0.006 |
| Delay cost log-sigma | $\sigma_\pi$ | 0.27 | 0.015 |
| Error standard deviation | $\sigma_\varepsilon$ | 0.02 | 0.000 |

*Notes*: This table provides the model parameters relating to the examiner. Standard errors are bootstrapped.

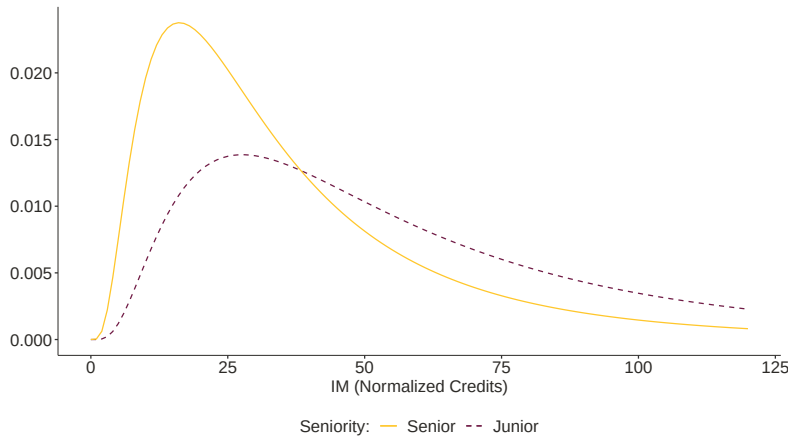
which we call "normalized credits."[25] The Office adjusts each examiner's credits based on their seniority and the technological complexity of applications. We use the same adjustments when we model payoffs for examiners.[26] These normalized credits are the unit of examiner payoffs. This ensures that payoffs are in the same units for all examiners, regardless of their seniority and technology center.

We start by interpreting the parameters of intrinsic motivation. To our knowledge, these are the first structural estimates of intrinsic motivation in a public agency. We estimate $\sigma_\theta$ as 0.77, which implies, by the properties of the log-normal distribution, a coefficient of variation of 0.82 (82%). This estimate implies substantial variation in intrinsic motivation across examiners, even within seniority category. We estimate $\mu_{\theta,\text{junior}} = 3.92$ and $\mu_{\theta,\text{senior}} = 3.38$, which implies that, on average, junior examiners are more intrinsically motivated than senior examiners. Figure 3 plots the distribution of intrinsic motivation for junior and senior examiners as implied by the log-normal assumptions. It is clear that the distribution of senior examiners' intrinsic motivation (yellow solid) is generally lower than that of junior examiners (maroon dashed). At least two countervailing forces influence the relationship between seniority and intrinsic motivation. In-

[25] Online Appendix Section E provides a detailed derivation of the examiners' credit structure.

[26] For example, an examiner receives two credits for granting a patent in the first negotiation round. We adjust these credits by dividing by a seniority factor (for example, by 1.25 for a senior GS-14 examiner) and multiplying by a technology correction (say, 29 for the relatively complex category of computer networks). Therefore, a GS-14 examiner in technology center "computer networks" receives 46.4 normalized credits for granting a patent in the first round. Tables E.1 and E.2 report the values of seniority and technology corrections across all seniorities and technology centers, respectively.

Figure 3. Density of examiner intrinsic motivation




*Notes*: Orange solid curve represents the distribution for senior examiners; maroon dashed curve represents the distribution for junior examiners. To interpret the x-axis, consider an examiner in technology center 36, where the technology correction is 22.4. Dividing the values on the x-axis by 22.4 yields the number of credits the examiner pays as an intrinsic motivation cost to a GS-12 examiner for granting a patent for an application on which every claim is invalid.

trinsic motivation will fall with seniority if examiners become "jaded" with experience. However, selection cuts the other way since the least intrinsically motivated examiners are likelier to move to the private sector with higher remuneration. The evidence thus indicates that the jading effect dominates the selection effect.

To interpret the magnitude of intrinsic motivation, we calculate the associated cost for a median intrinsically motivated GS-12 (junior) examiner in a selected technology center 36 ("Miscellaneous" category). For this examiner, the seniority correction is one and the technology correction is 22.4. Recall that intrinsic motivation cost (in terms of normalized credits) is $C_{IM} = \theta\mathcal{R}$, where $\theta$ is the intrinsic motivation parameter, and $\mathcal{R}$ is the proportion of claims the examiner believes invalid. We divide $C_{IM}$ by 22.4 to change the units back to pure credits. Hence, in terms of raw credits, this examiner's intrinsic motivation cost is $2.25R$, which means that the examiner faces a cost of 2.25 credits for knowingly granting a patent with *100%* of its claims as invalid. This cost is equivalent to the credits the examiner obtains for making *three* final rejections. This example is only an illustration, but our estimates generally imply that intrinsic motivation costs are sizeable relative to extrinsic rewards.

Next, we consider examiner delay costs. The coefficient of variation of examiner costs is 0.08, ten times smaller than examiner intrinsic motivation. Moreover, delay costs are estimated to be

small, with the median cost for a GS-12 (junior) examiner in technology center 36 paying an equivalent of 0.05 credits to go an extra round on this particular application. The fact that these costs are so small suggests that examiners are not pressured explicitly to finish applications fast and that the opportunity cost of devoting more time to this application relative to the next on their desk is small. This finding is intuitive since the most time-consuming activity for the examiner is their initial literature search. Hence, continuing to make decisions on an application they have already reviewed is less time-intensive than starting a new application (though also less compensated).

Finally, we discuss examiner error parameters. Recall that examiner errors are normally distributed, with an estimated standard deviation and mean equal to $\mu_\varepsilon = 1 + \frac{1}{\theta}$, where $\theta$ is the examiner's intrinsic motivation. The error that an examiner draws multiplies padded distances to create the examiner's distance assessment. Since junior examiners are more intrinsically motivated on average, the mean of the junior examiners' error distribution is closer to one. We estimate the standard deviation of examiner errors to be 0.02, indicating that errors are modest, typically within 4% of the examiner-specific mean.

*Calculating Examiner Errors*

We compute statistics on two kinds of examiner errors by simulating the model with our baseline estimates. The first error occurs when an examiner grants a patent with invalid claims. We refer to this as a "type 1" error. We calculate that this happens for 19% of grants, suggesting that while examiners are screening out some invalid patents, nearly one in five applications contain some claims that should not have been granted. The last statistic represents the "extensive" margin of this type of examiner error; we can also calculate an "intensive margin" error. Among all granted claims, 7% are invalid (compared to 83% of claims whose unpadded distance is below the threshold), implying that most invalid patents contain only a few invalid claims.

We also calculate the weighted errors (focusing on the intensive margin), where weights reflect the distance of the claim from the patentability threshold. Among simulations, indexed by $s$, let $S_G$ be the set granted and $j$ represent a claim. We calculate the measure

$$\sum_{s \in S_G} \sum_j \frac{w_{sj}}{\sum_{s' \in S_G} \sum_{j'} w_{s'j'}} E_{1sj,\text{int}}, \tag{10}$$

where $E_{1sj,\text{int}}$ is equal to one if claim $j$ on simulation $s$ is invalid (has a distance below the threshold) and zero otherwise, and the weight $w_{sj} = |\tilde{D}_{sj} - \tau_s|$, where $\tau_s$ is the threshold relevant to simulation $s$. The idea is to put more weight on errors where claims are further away from the threshold (making the error more "egregious"). If the weighted average is lower than the unweighted average, it implies that errors occur in marginal cases in which it is not obvious

whether the patent is valid. Indeed, the weighted error is 2%, much lower than the unweighted value of 7%, suggesting that most errors occur in cases of marginal validity.[27]

The other kind of "error" (or "undesirable" outcome) occurs when an applicant abandons an application that contains valid claims. We refer to these as "type 2" errors. Approximately 36% of abandonments have at least one valid claim. Strictly speaking, these are not a mistake by the examiner since they should only grant patents to applications on which *all* claims are valid. At the intensive margin, among all claims the applicant abandons, 18% are valid.

Similar to Equation (10), we calculate a weighted average of type 2 intensive margin errors, where the weights are the same as before—the distance of a claim from the threshold. In this case, the weighted error falls to 6%, again implying that abandonments occur on marginally valid claims rather than clearly valid ones. When we compute the extensive margin weighted error, the proportion of abandoned applications with at least one valid claim is 11%.

*Model Fit and Robustness*

We compare the values of simulated moments, calculated at the estimated parameters, with moments in the data.[28] As expected, we match most of the internal moments well (two exceptions are described in Online Appendix F). The real test of model fit, however, is how well we match moments that are not used in the estimation procedure. Online Appendix Figure A.2 displays these comparisons for the excluded moments described in Appendix F, which include percentiles on granted distances in each round, the mean of distances for latter rounds, and means and percentiles of round one rejection rates across seniority categories. We match all of these moments well.

We run a series of robustness checks on our baseline model (Online Appendix Table A.5 summarizes the results). First, we examine changes to how we define the distance threshold for patentability. In the baseline, we define each examiner's "revealed" threshold as the minimum distance they grant and then take the threshold as the maximum of those values over examiners. We experiment with using the first and fifth percentile of distances granted for each examiner, which allows for measurement error in their personal threshold. We also check robustness to a discount factor of 0.99 and a broader definition of examiner seniority. In all cases, the parameter estimates are generally robust.

[27] We also calculate weighted averages for the extensive margin type 1 errors (details available on request). The weighted error in this case is 5.1%, similarly suggesting that most errors occur in marginal cases.

[28] Online Appendix Figure A.1 displays the full set of comparisons.

Table 3. Counterfactual Experiments

| **Counterfactual** | **Not Apply** (%) | **Pad** (%) | **Rounds** | **R1 Gr** (%) | **T1** (%) | **T2** (%) |
|---|---|---|---|---|---|---|
| Baseline | 6.3 | 8.0 | 2.5 | 11.5 | 18.8 | 36.5 |
| 25K Round Fee | 8.4 | 6.6 | 2.4 | 12.8 | 18.0 | 37.9 |
| 50K Round Fee | 12.2 | 5.9 | 2.4 | 14.2 | 17.7 | 39.8 |
| Three Rounds | 27.2 | 3.6 | 2.1 | 15.1 | 15.9 | 49.2 |
| Two Rounds | 51.1 | 0.8 | 1.6 | 25.8 | 11.9 | 56.1 |
| One Round | 79.6 | -2.3 | 1.0 | 98.4 | 0.5 | 91.7 |
| 15% IM | 3.9 | 8.0 | 2.1 | 30.2 | 89.3 | 22.4 |
| Credit↘ | 6.3 | 7.9 | 2.5 | 11.5 | 18.7 | 36.3 |
| Credit↘ + 15% IM | 3.4 | 18.1 | 2.1 | 32.8 | 88.9 | 17.7 |

*Notes*: "Not Apply' is the percent of inventors who do not apply for a patent; Pad is the mean level of padding. Rounds is the mean number of rounds. "R1 Gr" is the percent of applications granted in Round 1. T1 represents the proportion of granted patents with some invalid claims. T2 represents the proportion of abandoned applications with some valid claims.

# 7 Counterfactual Analysis

We use the estimated model to conduct a series of counterfactual analyses to examine the impacts of various reforms on the speed and quality of the screening process and the degree of padding in patent applications. The counterfactual scenarios we examine include removing intrinsic motivation, and changing the level of patent office fees, the number of allowable rounds in the process, and examiner extrinsic incentives (credits).

Table 3 presents the results. We focus on four endogenous outcomes. The first is the proportion of applicants who choose not to apply for a patent on their developed invention. The second is the applicant's choice of how much to pad the application. The third set of outcomes is the proportion of grants in round one and the average number of rounds (speed of resolution). The fourth set, relating to screening quality, is the proportion of granted patents with at least one invalid claim (type 1 error at the extensive margin) and abandoned applications with at least one valid claim (type 2 error at the extensive margin). We note but do not report that the changes in these errors at the intensive margin errors are similar.

*Fees*

In the baseline, there are relatively low fees for applicants throughout the prosecution process.

In the first counterfactual, we introduce a substantial per-round fee that the applicant must pay for *each* negotiation (not just for an RCE).[29] This fee acts as a marginal cost per round of negotiation. Since each round is now more expensive, applicants have increased incentives to exit the patent process as soon as possible, and less incentive to apply in the first place. A substantial $50,000 fee for every extra round reduces padding by a quarter (from 8.0% to 5.9%) and slightly reduces the mean number of rounds, from 2.5 to 2.4. The proportion of grants in round one increases from 11.5% to 14.2%, reflecting the reduced padding, and the fraction of granted patents with some invalid claims (type 1 error) falls slightly. However, the rounds fee increases type 2 error—rising from 36.5% to 39.8%. The trade-off between these two types of errors is a feature of many of the counterfactuals we analyze.

It is at first surprising that per-round fees as high as $50,000 do not substantially change the speed or quality of patent prosecution. The explanation is that the private value of patent rights is large enough to make applying for a patent on many of these inventions worthwhile, even with high per-round fees. Further, the applicant has the option to apply and then abandon after the first rejection, without paying any negotiation fee. Fees would have to be much higher to substantially impact outcomes.[30]

*Restricting the Number of Rounds*

Instead of using fees, we consider limiting the maximum number of rounds of negotiation between the applicant and examiner. We consider a maximum of three rounds, then a maximum of two rounds (equivalent to removing all RCEs, allowing only one round of interaction between applicant and examiner), and finally, we allow for only one round (that is, no negotiation between applicant and examiner so that the examiner's first decision is final). These counterfactuals are motivated by a 2007 U.S.PTO proposal to restrict the number of RCEs. The proposed rulemaking was challenged in federal court, which judged the restrictions as an overreach of Patent Office

[29] We also consider substantially increasing the *application* fees to as much as $50,000. However, because this is a fixed fee paid upon application, provided it is still profitable to apply, applicants will not change their padding decision. Even at this level, the fee does not materially alter average padding and, since there is practically no change to the proportion of inventors who choose to apply, introducing an application fee acts mainly as a transfer from applicants to the Patent Office, with minimal changes to quality or speed of prosecution. If the additional resources from the higher application fee were reinvested in patent office examination, there would be improvements. This finding—that application fees only really help if they are reinvested—is similar to the findings in Schankerman and Schuett (2022), who use a completely different theoretical model and data.

[30] Of course, these fees would be significant for small firms or single inventors who may be cash constrained. However, round fees for small and micro entities could be reduced, as the Patent Office already does for other types of fees.

authority.[31] The court decision did not consider the quantitative impact of such changes on patent office screening quality or its welfare effects, which our paper makes possible.

Round restrictions have material consequences on screening outcomes. Removing all RCEs (allowing only two rounds) would lead to half of all inventors not applying for a patent and would virtually eliminate padding. In this case, 25.8% of applicants are granted in the first round, and because applicants respond to the restriction by reducing padding, the proportion of patents granted with invalid claims falls. In particular, with only one opportunity for negotiation, type 1 error falls sharply, from 18.8% to 11.9%.

The disadvantage of limiting the scope for negotiation is that it increases the proportion of abandoned applications with valid claims. With no RCEs, this proportion rises from 36.5% to 56.1%. As with fees, making the process tougher for applicants through fewer allowable rounds generally reduces the granting of invalid claims and speeds up the process but leads to the abandonment of valid claims. As we discuss in the next section, granting invalid claims and not granting valid claims each imposes social costs, and we need to measure these to evaluate the overall impact of the reforms.

Finally, we compare the effectiveness of fees and round restrictions ("price versus quantity" instruments) by computing the equivalent per-round fee—in the sense of equalizing the mean number of rounds in equilibrium—to restrictions on the number of RCEs. The simulations show that the fee equivalent to removing all RCEs is a massive $600,000 per round. Using fees generally produces lower type 1 and type 2 errors than their rounds equivalents, but such fee levels are politically unpalatable.

*Removing Intrinsic Motivation*

Next, we evaluate the impact of removing intrinsic motivation by reducing it for every examiner to 15% of its original value.[32] Knowing that examiners will be more unwilling to grant invalid patents, only 3.9% of inventors do not apply, the number of rounds falls from 2.5 to 2.1, and the

[31] The proposed changes are in *U.S.PTO Changes to Practice for Continued Examination Filings, Patent Applications Containing Patentably Indistinct Claims, and Examination of Claims in Patent Applications—the "New Rules"* (SmithKline Beecham Corp. v. Dudas, 541 F. Supp. 2d 805, 2008). The court decided that the "New Rules" were substantive and that the Patent Office did not have the rulemaking authority to make substantive changes, though the Court noted that the Patent Office could make procedural changes, such as fees. As we will show, one can achieve the same equilibrium number of rounds with an "equivalent" fee, so from an economic point of view, this distinction is problematic.

[32] We cannot fully remove intrinsic motivation because our specification of mean error being inversely related to IM implies that IM cannot be exactly zero. We provide the reason behind our choice of 15% in Section 8.

proportion of applications granted in round one almost triples, increasing from 11.5 to 30.2. Not surprisingly, type 1 error jumps sharply to 89.3%, while type 2 error declines. This counterfactual highlights the quantitative importance of intrinsic motivation on the quality of patent screening and confirms its potential salience for economic analyses of other public agencies.[33]

*Removing Credits*

Finally, we consider changes to the structure of credits for examiners. We remove all credits for the examiner after the first round. If it is the case that examiner costs of delay represent the marginal cost of an extra examination round, then such a policy change could be justified on efficiency grounds of "marginal cost pricing" since we estimated examiner delay costs to be small.[34] When we remove all credits after the first round in the baseline model, there are minimal, if any, impacts on any of the outcome variables. This result suggests that our baseline estimates of intrinsic motivation are sufficiently large for examiners to want to avoid granting invalid patents even in a context where they will receive no further extrinsic reward if they do so. This striking finding reflects the extent to which patent office examiners are intrinsically motivated. The results are not consistent with extrinsic incentives crowding out intrinsic incentives.

To complement this exercise, we also analyze the effect of removing all credits after the first round alongside reducing intrinsic motivation to 15% of its value (at any higher value of intrinsic motivation, removing credits has no material effect). In this case, we find non-trivial impacts of credits consistent with economic intuition. First, padding doubles, up to 18% (relative to 8.0% when only intrinsic motivation is changed) and first-round grants increase from 30.2% to 32.8%. Type 2 error declines because the increased padding means that abandonments are less likely to include valid claims. These results indicate that extrinsic incentives and intrinsic motivation are *substitutes*, not complements, as sometimes found in the experimental literature (see Section 2 for citations): credits only work as an effective device to incentivize examiners when examiners are not intrinsically motivated (and even then, as we show in the next section, credits do *not* reduce social costs of screening).

[33]Interestingly, increasing intrinsic motivation (not reported) does not have much impact in reducing padding or type-1 error. The explanation for this outcome is that examiners are already sufficiently intrinsically motivated to get most of the benefits, so further increases do not have much bite.

[34]This counterfactual has limitations that the others do not because our model is focused on optimal decisions on a given patent application. It does not incorporate any interactions between different applications the examiner faces, such as optimizing docket management across applications (including meeting quarterly or annual targets). This counterfactual is best thought of informing an examiner that for one of the new applications in their docket, they will only receive credits for the first round.

In summary, these counterfactual experiments show that no reform we consider unambiguously improves both prosecution speed and quality. There is typically a trade-off: policies that make prosecution stricter lead to fewer grants of invalid patents but increased abandonments of valid applications. Evaluating reforms requires converting these outcomes into social costs, which we do in the following section.

# 8 Quantifying the Social Costs of Patent Screening

We classify net social costs into three categories: *type 1*, *type 2*, and *prosecution* costs. Type 1 costs refer to the costs induced by granting invalid claims. Type 2 costs refer to the social value of inventions that are not developed ex ante because of the potential threat of not being granted valid claims. Type 1 benefits refer to the social value of inventions that would not be developed ex ante without type 1 error. Type 2 benefits refer to the ex post deadweight loss *not* incurred when inventors abandon valid claims. Prosecution costs are the Patent Office's costs of examining applications plus the legal fees incurred by the applicants during the negotiation process. In what follows, we summarize our quantification approach; full details are in Online Appendix G. We start with the costs of each type of error and then discuss the benefits.

## 8.1 Type 1 Costs

There are two sources of costs from type 1 error: the deadweight loss associated with the royalties extracted by the patentee and the litigation costs associated with legal challenges against invalid patents that are granted (and that are valuable enough to warrant a challenge).

*Deadweight loss from royalties*

We assume that the patentee charges the Arrow royalty equal to the unit cost savings due to the invention, $\Delta c$. The deadweight loss from royalties depends on the market structure of licensees. Our baseline specification is perfect competition among licensees, with a linear demand and constant unit cost.[35] In this case, the deadweight loss is

$$DWL = \frac{1}{2}\Delta\wp\Delta q = \frac{\lambda}{2}\frac{\Delta\wp}{\wp}\tilde{V},$$

where $\wp$ is the initial price (without the royalty associated with the claim), $\Delta\wp = \Delta c$ with perfect competition, $\tilde{V} = q\Delta\wp$ denotes total *royalty payments*, and $\lambda$ is the elasticity of product demand

[35] In Online Appendix G, we extend the approach to Cournot competition. Our calibration indicates that this extension yields quantitatively very similar results.

(in absolute value).[36] To calibrate this expression, we follow Schankerman and Schuett (2022), who estimate the ratio of corporate licensing revenue from intangible industrial property to R&D at 39.3%. Multiplying this ratio by the ratio of R&D to sales in manufacturing in 2002 (4.1%), we take $\frac{\Delta\wp}{\wp} = 1.61\%$. We do the computation for values of the demand elasticity $\lambda \in (1, 3)$ and report $\lambda = 2$ in the main analysis (qualitative conclusions hold for the other values).

*Cost of litigation on invalid patents*

The social cost of type 1 error also involves litigation costs on invalid patents. Not all invalid patents are "exposed" to litigation because their private value is not large enough to justify the litigation expense. Letting $G_{\tilde{V}}(\cdot)$ denote the distribution of the value at stake $\tilde{V}$, we take the proportion of patents not exposed to litigation from Schankerman and Schuett (2022) ($\check{v} = 89.6\%$) and calculate the $\check{v}^{st}$ percentile of the value at stake distribution, $\check{V} = G^{-1}_{\tilde{V}}(\check{v})$. Then, all patents with $\tilde{V}$ exceeding the threshold $\check{V}$ are exposed to litigation.

The social cost for invalid patents not exposed to litigation is only the deadweight loss from royalties. From Schankerman and Schuett (2022), exposed invalid patents have a 16.3% probability of being litigated, in which case, we assume that courts are perfect and thus always invalidate wrongly granted claims. In this case, the social cost is the sum of litigation costs for the patentee and challenger, each denoted $\mathcal{C}(\tilde{V})$.[37] The remaining 83.7% of exposed invalid patents are not litigated and only impose the deadweight loss.[38]

In summary, the expected social cost of granting an invalid patent of value $\tilde{V}_s$ is

$$S_{1s} = I_s DWL_s + (1 - I_s)\left[0.837 \cdot DWL_s + 0.163 \cdot 2\mathcal{C}(\tilde{V}_s)\right], \tag{11}$$

where $I_s = 1(\tilde{V}_s \leq \check{V})$ is an indicator equal to one if the patent is not exposed to litigation. Then, the total type 1 cost is

$$T_1 = \sum_{s \in S_G} E_{1s} S_{1s} \tag{12}$$

where $E_{1s}$ is equal to one if a granted application $s \in S_G$ is invalid and zero otherwise.

[36] For invalid patents, we cannot use the model estimates of values of patent rights $V$ to represent royalty payments for invalid patents $\tilde{V}$, since our estimates of $V$ are contaminated with potential legal costs (explained in the next subsection). In Online Appendix G, we explain how we overcome this challenge to calculate type 1 costs.

[37] We take $\mathcal{C}(\tilde{V})$ as linear in $\tilde{V}$ and calibrate the coefficients using AIPLA data.

[38] Patentees with invalid patents can pre-empt a challenge by charging a royalty payment (typically a lump sum) equal to the cost of litigation for the challenger (this is commonly referred to as "trolling" behavior). For these cases, the social cost is only the deadweight loss associated with the patent, since the payment is a pure transfer from the licensee to the patentee (we ignore possible R&D incentive effects of the transfer). See Schankerman and Schuett (2022) for more discussion.

## 8.2 Type 2 Costs

From the ex post perspective, there is *no social cost* from type 2 errors because the innovation has already been produced and the R&D cost is sunk (this is essentially ex post hold-up). Therefore, it only makes sense to analyze the social cost of type 2 errors from the ex ante (incentive) perspective. Type 2 error reduces the expected value of patent protection for the inventor and, thus, the ex ante decision of inventors to develop their (exogenous) ideas. We want to calculate the social value of the set of socially valuable inventions that are *not* developed when there is the possibility of type 2 error but which *would be* developed in the absence of type 2 error. This task requires us to construct a simple model of development. We emphasize that we do not require this extension to estimate the screening model, nor to calculate padding, the number of equilibrium rounds, type 1 and type 2 errors, type 1 social costs, and prosecution costs.

The decision to develop an idea into an invention depends on three things: the ex ante value of patent rights ($\Gamma^*$), the value of the invention without patent rights ($\pi$), and the development cost ($\kappa$). To compute $\Gamma^*$, we use our model to calculate the ex ante value of patent rights (net of all costs), as in Equation (1). To calculate the private value of the invention without patent rights, we define the patent premium ($\xi$) as the percentage increase in private value due to patent protection. Hence, for positive $\Gamma^*$, by definition $\Gamma^* = \xi\pi$, implying a set of values of $\pi$. We assume that the patent premium is constant across inventions and calibrate it based on existing estimates from the literature on patent renewal models (Schankerman, 1998).[39] For the cost of developing an idea into an invention, $\kappa$, we draw values from the distribution estimated by Schankerman and Schuett (2022).[40]

An inventor *does not* invest to develop an idea $i$ if

$$ND_i \equiv \mathcal{B}_i - \kappa_i \leq 0,$$

where $\mathcal{B}_i \equiv \pi_i + \max\{\Gamma_i^*, 0\}$ is the private benefit of development. An idea is socially valuable to develop if the net social benefit of development,

$$S_{2i} \equiv \frac{\rho_{\text{soc}}}{\rho_{\text{priv}}}\mathcal{B}_i - \kappa_i,$$

[39] This is a strong assumption, but it is not feasible to identify $\pi_s$ if we allow the patent premium to vary. The reason is that we do not have any information on who develops their ideas, which might allow us to back out $\pi$ from the decision to develop and our estimated value of $\Gamma^*$. Furthermore, we must specify $\pi$ for inventions with negative ex ante value of patent rights. To do this, we draw from the distribution of $\pi$ created from positive values of patent rights.

[40] An alternative approach is to assume that inventors do not know their development cost, and thus use the mean cost $\bar{\kappa}$. We experimented with this approach and qualitative conclusions are robust.

is positive (where $\rho_{\text{priv}}$ and $\rho_{\text{soc}}$ denote the private and social rates of return). We use a conservative estimate of $\frac{\rho_{\text{soc}}}{\rho_{\text{priv}}} = 2$ from Bloom, Schankerman, and Van Reenen (2013).

Let $\Upsilon_0$ denote the set of ideas that are socially beneficial to develop ($S_{2i} > 0$) but which are not developed ($ND_i \leq 0$). To calculate type 2 social cost, we compute the subset of $\Upsilon_0$, which we denote $\Upsilon_1$, that *would* develop in the absence of type 2 error. To do this, we simulate the outcome from a "counterfactual" patent prosecution where, at the point of patent abandonment, the inventor obtains the value of all valid claims in that patent. By definition, in this scenario, all abandoned claims are invalid, so there is no type 2 error. Let $\Gamma'$ denote the expected value of patent rights in this new scenario. The idea $i$ would be developed in this scenario if

$$ND'_i \equiv \pi_i + \max\{\Gamma'_i, 0\} - \kappa_i > 0.$$

We then compute type 2 costs as

$$T_2 = \sum_{i \in \Upsilon_1} S_{2i}, \tag{13}$$

where $\Upsilon_1$ is the subset of $\Upsilon_0$ with $ND'_i > 0$. This is the set of ideas that are socially beneficial to develop, that are not developed in the scenario with type 2 error, but that would be developed in the absence of any type 2 error.

### 8.3 Patent Prosecution Costs

The social cost of patent prosecution for each application $s$ consists of two components: applicant legal costs of amending the application each round and Patent Office administrative costs. The amendment cost is the per-negotiation cost $F_{\text{amend},s}$ drawn from the estimated distribution, multiplied by the equilibrium number of negotiations for application $s$ (equal to the number of rounds $r_s$ minus 1). For the administrative cost, we calculate the patent operations budget per application as \$4,117 (in 2018 dollars). This value excludes patent office fees, as these are transfers from the applicant to the patent office, as well as loss in patent value associated with pre-grant obsolescence since that, too, is a transfer from the applicant to the owner of the invention that superseded it. We divide the operations budget per application by the average number of rounds across all simulations and by the average number of independent claims in an application, to create the average patent office cost per round and claim, denoted by $RCC$. Then, the total social cost of patent prosecution is

$$T_3 = \underbrace{\sum_s (r_s - 1) F_{\text{amend},s}}_{\text{Applicant Fighting Costs}} + \underbrace{\sum_s M_{0,s} r_s RCC}_{\text{Office Costs}}, \tag{14}$$

where $M_{0,s}$ is the initial number of claims in application $s$.

## 8.4 Benefits of Type 1 and Type 2 Errors

There are also benefits from errors. In the type 1 case, when invalid patents are incorrectly granted, the ex ante incentives for inventors to develop and patent their ideas are increased. This is analogous to the *costs* of type *2* error. We compute these benefits as the sum of social development benefits from welfare-enhancing projects that would not be developed without type 1 error but that are developed with type 1 error.[41] The method is similar to the approach described in Section 8.2.

Further, there are benefits from type 2 errors. Not granting valid patents saves the deadweight loss on those patents. We compute these benefits as described in Section 8.1. Note that there is no benefit associated with litigation cost savings since, under our assumption of costly but perfect courts (always upholding valid patents and overturning invalid ones), valid patents that are granted would not be challenged.

One important point to note is that the quantification of net social costs in this section is based on the presumption that the patentability threshold used by the Patent Office corresponds to the social optimum, that is, the threshold that only grants patents to inventions that are welfare-enhancing but would not be developed without patent rights. To see this, suppose the threshold is too low (the conventional wisdom) so that some patents are considered "valid" and granted despite not being welfare-enhancing. We would incorrectly not count these as a type 1 error, so they would not contribute to our measure of type 1 social costs. Thus, we would understate type 1 costs (and type 2 benefits). By an analogous argument, we would overstate type 2 costs (and type 1 benefits). Therefore, if the threshold is too low, the consequence is that we would understate net type 1 social costs and overstate net type 2 social costs. It remains an open and important research question to determine the "optimal" distance threshold, that is, the one that grants patents only to inventions that are welfare-enhancing and not otherwise developed (Schankerman and Schuett, 2022).

## 8.5 Social Costs in Counterfactual Reforms

Table 4 summarizes the three components of net social costs for the baseline model and the set of counterfactual reforms.[42] The baseline row approximates the net social costs associated with

[41] The "counterfactual" patent prosecution in this case is one where, at the point of patent grant, the inventor only obtains the value of the *valid* claims in the patent.

[42] The table presents the values of net social costs for $\lambda = 2$, $\frac{\rho_{\text{soc}}}{\rho_{\text{priv}}} = 2$, $\xi = 0.1$, and development costs drawn. The qualitative conclusions are similar for a range of other parameter values. In Appendix A, we provide results for the cases of a 5% patent premium with $\frac{\rho_{\text{soc}}}{\rho_{\text{priv}}}$ equal to 1.5 and 2, and a 10% patent premium with $\frac{\rho_{\text{soc}}}{\rho_{\text{priv}}}$ equal

a yearly cohort of ideas, averaged over 2011–2013 (Appendix G explains how we calibrate the annual number of ideas). Subsequent rows provide the net social costs in that counterfactual scenario. All values are adjusted for inflation, presented in 2023 U.S. dollars.

In the baseline, total type 1 net costs equal \$6.4bn, total type 2 net costs are \$1.5bn, and prosecution costs equate to \$17.6bn. In the final column, we sum these three net costs and estimate the total net social cost of patent screening at \$25.5bn. This total constitutes 6.5% of total R&D performed by business enterprises in the U.S. in 2011.[43]

Introducing a per-round fee lowers type 1 and prosecution costs because it discourages applications and lowers padding for those that do apply. This, in turn, implies that fewer grants are invalid and that grants occur in fewer rounds. However, a round fee increases type 2 costs as applicants are more likely to abandon with some valid claims in a scenario with high negotiation fees. With a \$25,000 round fee, the latter effect dominates, so the total net social cost increases by a very modest 1.9%. As mentioned earlier, for sufficiently large rounds fees (likely to be politically infeasible), the reductions in type 1 and prosecution costs eventually dominate. Further, in these counterfactuals, the extra revenues generated by the fees are *not* reinvested in more intensive or faster examinations. If they were reinvested, social costs from introducing fees would be mitigated or even converted to social gains.

Restrictions on the allowable number of negotiation rounds have qualitatively similar effects on social costs as rounds fees, but the impacts are much larger. Removing all RCEs (two rounds) yields a 10.4% fall in total social costs relative to the baseline. Restricting the process to one round reduces net social costs by 45%.

Removing intrinsic motivation (down to 15% of its original level) increases the total social cost by 68.6%. When examiners have almost no intrinsic motivation, they are willing to grant applications fast, even if they are padded. As a result, administrative costs fall when intrinsic motivation is removed.[44] However, the grants of patents with invalid claims cause type 1 net costs to triple and consequently lead to an overall rise in net social costs. This finding confirms the critical role

to 1.5. We do not present results for different values of $\lambda$ because quantitative values in this case are very similar to the baseline.

[43]It is worth noting that this is at the lower end of estimates of the private value of patent rights (Pakes, 1986; Schankerman, 1998). This suggests that the patent system, as it is currently configured, generates net positive social value. For similar findings in a different framework, see Schankerman and Schuett (2022).

[44]The decrease in prosecution costs is countervailed by the fact that when intrinsic motivation is low, there is an extensive margin increase in the number of inventors applying for patent rights.

Table 4. Net Social Costs of Patent Prosecution

| **Counterfactual** | $T_1$ | $T_2$ | $T_3$ | **Total** |
|---|---|---|---|---|
| Baseline ($Bn) | 6.4 | 1.5 | 17.6 | 25.5 |
| 25K Round Fee | 5.9 | 3.7 | 16.4 | 26.0 |
| 50K Round Fee | 6.1 | 6.3 | 15.1 | 27.5 |
| Three Rounds | 4.9 | 10.1 | 10.2 | 25.1 |
| Two Rounds | 2.9 | 15.6 | 4.7 | 23.1 |
| One Round | 0.0 | 13.4 | 0.7 | 14.1 |
| 15% IM | 25.8 | 2.3 | 15.0 | 43.0 |
| Credit$\searrow$ | 6.4 | 1.5 | 17.6 | 25.5 |
| Credit$\searrow$ + 15% IM | 15.9 | 4.0 | 15.8 | 35.7 |

*Notes*: Equation (12) defines $T_1$; Equation (13) defines $T_2$, respectively; Equation (14) defines $T_3$; Total sums the three kinds of costs. The "baseline" row provides the total social costs in billions of 2023 U.S. dollars.

that intrinsic motivation plays in this public agency.

Finally, with the baseline level of intrinsic motivation, removing all examiner credits after the first round for one examination has almost no effect on social costs – precisely as we would expect, given the negligible changes to any endogenous variables. In fact, examiners' intrinsic motivation must be as low as 15% of original values for credits to have any effect on net social costs. With 15% intrinsic motivation, type 2 gross (and net) costs, prosecution costs, and type 1 *gross* costs all increase when credits are removed. As a result, when intrinsic motivation is lowered by 85%, removing credits increases total *gross* social costs. Yet, total *net* social costs *decrease*, suggesting that credits are counter-productive even when intrinsic motivation is low. This finding is driven by a large increase in type 1 *benefits* (and hence a decrease in type 1 *net* social costs) from removing credits. This result highlights the importance of accounting for the increased development from relaxed patent granting, as opposed to just the ex post social costs that arise through deadweight losses and litigation.

# 9 Conclusion

In this paper, we develop and estimate a structural model of the patent screening process. The model incorporates incentives, intrinsic motivation, and multi-round negotiation between the examiner and applicant. The paper shows how structural modeling of the incentives and

organization of innovation-supporting public agencies can be used to design reforms to improve agency performance. Our paper highlights the fact that, to analyze the impact of reforms on the effectiveness of screening, it is critical to incorporate *both* the agency's decision-making and the endogenous responses of applicants being screened.

Our findings show that patent screening is moderately effective *given* the statutory and judicial standards for patentability within which the Patent Office is required to operate. This effectiveness is driven by substantial intrinsic motivation of examiners. We find that restrictions on the number of allowable rounds of negotiation reduce the social costs of screening. This outcome can be replicated through an equivalent round fee for the applicant, but the required fees are too high to be politically feasible. Finally, we estimate the total net social cost of patent screening at $25.5bn per annual cohort of applications. This figure represents 6.5% of R&D in the United States performed by business enterprises.

This paper studies patent screening and instruments to improve its effectiveness at the *pooled* technology level. A fruitful extension would be to estimate the model for individual technology fields, such as biotechnology and software, which would allow for the evaluation of the differential effectiveness of various instruments in different areas. More generally, we hope this paper illustrates the value of using structural models to inform decisions on how to reform public agencies, particularly those that affect the allocation of R&D resources, including leading institutions like the National Institutes of Health and National Science Foundation, and similar institutions in other countries.

## References

Adda, J. and M. Ottaviani (2023): "Grantmaking, Grading on a Curve, and the Paradox of Relative Evaluation in Nonmarkets," *forthcoming in The Quarterly Journal of Economics*.

Andrews, I., M. Gentzkow, and J. M. Shapiro (2017): "Measuring the Sensitivity of Parameter Estimates to Estimation Moments," *The Quarterly Journal of Economics*, 132, 1553–1592.

Ashraf, N., O. Bandiera, E. Davenport, and S. S. Lee (2020): "Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the Delivery of Public Services," *American Economic Review*, 110, 1355–94.

Ashraf, N., O. Bandiera, and K. Jack (2014): "No margin, no mission? A field experiment on incentives for public service delivery," *Journal of Public Economics*, 120, 1 – 17.

Azoulay, P., J. S. Graff Zivin, D. Li, and B. N. Sampat (2018): "Public R&D Investments and Private-sector Patenting: Evidence from NIH Funding Rules," *The Review of Economic Studies*, 86, 117–152.

Benabou, R. and J. Tirole (2003): "Intrinsic and Extrinsic Motivation," *The Review of Economic Studies*, 70, 489–520.

——— (2006): "Incentives and Prosocial Behavior," *American Economic Review*, 96, 1652–1678.

Besley, T. and M. Ghatak (2005): "Competition and Incentives with Motivated Agents," *American Economic Review*, 95, 616–636.

Bessen, J. (2008): "The value of U.S. patents by owner and patent characteristics," *Research Policy*, 37, 932–945.

Bloom, N., M. Schankerman, and J. Van Reenen (2013): "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, 81, 1347–1393.

Cockburn, I., S. Kortum, and S. Stern (2003): *Are All Patent Examiners Equal? Examiners, Patent Characteristics, and Litigation Outcomes*, Washington, DC: The National Academies Press.

Egan, M. L., G. Matvos, and A. Seru (2023): "Arbitration with Uninformed Consumers," *conditionally accepted in The Review of Economic Studies.*

Federal Trade Commission (2011): *The Evolving IP Marketplace: Aligning Patent Notice and Remedies with Competition*, Washington D.C.: Government Printing Office.

Frakes, M. D. and M. F. Wasserman (2017): "Is the Time Allocated to Review Patent Applications Inducing Examiners to Grant Invalid Patents? Evidence from Microlevel Application Data," *The Review of Economics and Statistics*, 99, 550–563.

Galasso, A. and M. Schankerman (2015): " Patents and Cumulative Innovation: Causal Evidence from the Courts," *The Quarterly Journal of Economics*, 130, 317–369.

——— (2018): "Patent rights, innovation, and firm exit," *The RAND Journal of Economics*, 49, 64–86.

Hall, B. and J. Lerner (2010): *The Financing of R&D and Innovation*, vol. 1, Elsevier.

Jaffe, A. and J. Lerner (2004): *Innovation and Its Discontents: How Our Broken Patent System is Endangering Innovation and Progress, and What to Do About It*, Princeton University Press.

Kelly, B., D. Papanikolaou, A. Seru, and M. Taddy (2021): "Measuring Technological Innovation over the Long Run," *American Economic Review: Insights*, 3, 303–20.

Lanjouw, J. O. (1998): "Patent Protection in the Shadow of Infringement: Simulation Estimations of Patent Value," *The Review of Economic Studies*, 65, 671–710.

Le, Q. and T. Mikolov (2014): "Distributed representations of sentences and documents," in *International conference on machine learning*, PMLR, 1188–1196.

Li, D. (2017): "Expertise versus Bias in Evaluation: Evidence from the NIH," *American Economic Journal: Applied Economics*, 9, 60–92.

Li, D. and L. Agha (2015): "Big names or big ideas: Do peer-review panels select the best science proposals?" *Science*, 348, 434–438.

Merges, R. P. and R. R. Nelson (1990): "On the Complex Economics of Patent Scope," *Columbia Law Review*, 90, 839–916.

Nelsen, R. B. (2007): *An introduction to copulas*, Springer Science & Business Media.

Pakes, A. (1986): "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, 54, 755–784.

Sampat, B. and H. L. Williams (2019): "How Do Patents Affect Follow-On Innovation? Evidence from the Human Genome," *American Economic Review*, 109, 203–36.

Schankerman, M. (1998): "How Valuable is Patent Protection? Estimates by Technology Field," *The RAND Journal of Economics*, 29, 77–107.

Schankerman, M. and A. Pakes (1986): "Estimates of the Value of Patent Rights in European Countries during the Post-1950 Period," *Economic Journal*, 96, 1052–1076.

Schankerman, M. and F. Schuett (2022): "Patent Screening, Innovation, and Welfare," *The Review of Economic Studies*, 89, 2101–2148.

The Economist (2015): "Time to fix patents," *The Economist Group*, August 8th-14th, 9.

# Screening Property Rights for Innovation

## Online Appendices

William Matcham and Mark Schankerman

## A Additional Tables and Figures

Table A.1. Summary Statistics

| Variable | Observations | Mean | Median | Std. Dev. |
|---|---|---|---|---|
| Issued | 4,846,053 | 0.70 | 1.00 | 0.46 |
| Duration of Prosecution (years) | 4,846,053 | 2.96 | 2.67 | 1.57 |
| Number of Rounds | 4,608,833 | 2.40 | 2.00 | 1.45 |
| Independent Claims | 3,838,553 | 2.99 | 3.00 | 2.94 |
| Small Entity | 4,781,012 | 0.24 | 0.00 | 0.43 |
| Not Renewed at 4 | 410,667 | 0.13 | 0.00 | 0.33 |
| Renewed at 4, not at 8 | 410,667 | 0.19 | 0.00 | 0.39 |
| Renewed at 8, not at 12 | 410,667 | 0.23 | 0.00 | 0.42 |
| Renewed at 12 | 410,667 | 0.46 | 0.00 | 0.50 |

*Notes*: Sample sizes are lower for rounds, claims, and examiner variables since the datasets containing these variables cover a subset of the years 2001-2017. On renewal variables, we restrict attention to patents granted before 2006 to ensure that we have full renewal data on all granted patents. Categorical variables may not sum to one due to rounding.

## Table A.2. Estimated and Assigned Parameters

| Estimated Parameters | | | |
|---|---|---|---|
| **Variable** | **Notation** | **Distribution** | **Parameters** |
| ***Examiner*** | | | |
| Intrinsic motivation | $\theta \sim G_{S,\theta}(\cdot)$ | Log-normal | $\sigma_\theta$, $\mu_{\theta,\text{junior}}$ or $\mu_{\theta,\text{senior}}$ |
| Examiner Delay Cost | $\pi \sim G_\pi(\cdot)$ | Log-normal | $\mu_\pi, \sigma_\pi$ |
| Error | $\varepsilon \sim G_{e,\varepsilon}(\cdot)$ | Normal | $\sigma_\varepsilon$ |
| ***Applicant*** | | | |
| Initial claim returns | $v_j^* \sim G_v(\cdot)$ | Log-normal | $\mu_v, \sigma_v$ |
| Initial claim distances | $D_j^* \sim G_D(\cdot)$ | Beta | $\alpha_D, \beta_D$ |
| Obsolescence | $\boldsymbol{\omega}$ | Bernoulli | $P_{\omega,\text{pre}}$ or $P_{\omega,\text{post}}$ |
| Application legal costs | $f_{\text{app}}$ | Log-normal | $\mu_{f,\text{app}}, \sigma_{f,\text{app}}$ |
| Issuance legal costs | $f_{\text{iss}}$ | Log-normal | $\mu_{f,\text{iss}}, \sigma_{f,\text{iss}}$ |
| Maintenance legal costs | $f_{\text{main}}$ | Log-normal | $\mu_{f,\text{main}}, \sigma_{f,\text{main}}$ |
| Amendment legal costs | $f_{\text{amend}}$ | Log-normal | $\mu_{f,\text{amend}}, \sigma_{f,\text{amend}}$ |
| Narrowing | $\eta$ | - | - |

| Assigned Parameters | | |
|---|---|---|
| **Variable** | **Notation** | **Values** |
| Discount rate | $\beta$ | 0.95 |
| Depreciation | $\delta$ | $\frac{0.14-P_{\omega,\text{post}}}{1-P_{\omega,\text{post}}}$ |
| Threshold by technology center | $\tau$ | Range from 0.48 to 0.52 |
| Credits | $g^r(S,T)$ | - |
| Finalizing fee | $\phi$ | \$2,268 |
| RCE fees | $F^3_{\text{round}} = F^5_{\text{round}}$ | \$1,034 |
| | $F_4$ | \$1,685 |
| Renewal fees | $F_8$ | \$3,791 |
| | $F_{12}$ | \$7,792 |

## Table A.3. Application Fighting Costs by Technology Area

| **Parameter** | **Symbol** | **Estimate** | **S.E.** |
|---|---|---|---|
| Chemical application fighting cost log-mean | $\mu_{f,\text{chem}}$ | 9.15 | 0.008 |
| Chemical application fighting cost log-sigma | $\sigma_{f,\text{chem}}$ | 0.38 | 0.010 |
| Electrical application fighting cost log-mean | $\mu_{f,\text{elec}}$ | 9.18 | 0.010 |
| Electrical application fighting cost log-sigma | $\sigma_{f,\text{elec}}$ | 0.57 | 0.014 |
| Mechanical application fighting cost log-mean | $\mu_{f,\text{mech}}$ | 9.02 | 0.008 |
| Mechanical application fighting cost log-sigma | $\sigma_{f,\text{mech}}$ | 0.47 | 0.011 |

*Notes*: Standard errors are bootstrapped.

Table A.4. Applicant Fighting Costs by Technology Area

| **Parameter** | **Symbol** | **Estimate** |
|---|---|---|
| Simple amendment fighting cost log-mean | $\mu_{f,\text{amend,simp}}$ | 7.60 |
| Simple amendment fighting cost log-sigma | $\sigma_{f,\text{amend,simp}}$ | 0.37 |
| Chemical amendment fighting cost log-mean | $\mu_{f,\text{amend,chem}}$ | 8.13 |
| Chemical amendment fighting cost log-sigma | $\sigma_{f,\text{amend,chem}}$ | 0.45 |
| Electrical amendment fighting cost log-mean | $\mu_{f,\text{amend,elec}}$ | 8.07 |
| Electrical amendment fighting cost log-sigma | $\sigma_{f,\text{amend,elec}}$ | 0.38 |
| Mechanical amendment fighting cost log-mean | $\mu_{f,\text{amend,mech}}$ | 7.95 |
| Mechanical amendment fighting cost log-sigma | $\sigma_{f,\text{amend,mech}}$ | 0.43 |
| Issuance cost log-mean | $\mu_{f,\text{iss}}$ | 6.54 |
| Issuance cost log-sigma | $\sigma_{f,\text{iss}}$ | 0.62 |
| Maintenance cost log-mean | $\mu_{f,\text{main}}$ | 5.67 |
| Maintenance cost log-sigma | $\sigma_{f,\text{main}}$ | 0.46 |

Table A.5. Robustness of Estimates

| **Parameter** | **Symbol** | **Baseline** | 1% $\boldsymbol{\tau}$ | 5% $\boldsymbol{\tau}$ | $\boldsymbol{\beta = 0.99}$ | **Definition of Seniority (GS13 + GS14)** |
|---|---|---|---|---|---|---|
| Junior intrinsic motivation log-mean | $\mu_{\theta,\mathrm{j}}$ | 3.92 | 3.96 | 3.96 | 3.90 | 4.16 |
| Senior intrinsic motivation log-mean | $\mu_{\theta,\mathrm{s}}$ | 3.38 | 2.90 | 2.73 | 3.18 | 2.93 |
| Intrinsic motivation log-sigma | $\sigma_\theta$ | 0.77 | 0.82 | 0.79 | 0.90 | 0.99 |
| Examiner delay cost log-mean | $\mu_\pi$ | 0.19 | 0.16 | 0.18 | 0.49 | 0.12 |
| Examiner delay cost log-sigma | $\sigma_\pi$ | 0.27 | 0.37 | 0.42 | 0.10 | 0.60 |
| Error standard deviation | $\sigma_\varepsilon$ | 0.02 | 0.02 | 0.02 | 0.03 | 0.02 |
| Initial returns log-mean | $\mu_v$ | 10.55 | 10.59 | 10.88 | 10.07 | 10.28 |
| Initial returns log-sigma | $\sigma_v$ | 1.32 | 1.13 | 1.61 | 2.94 | 0.57 |
| Initial distance alpha | $\alpha_D$ | 4.57 | 3.92 | 3.90 | 4.56 | 3.75 |
| Initial distance beta | $\beta_D$ | 7.74 | 6.72 | 6.22 | 7.79 | 7.15 |
| Narrowing probability | $\eta$ | 0.75 | 0.73 | 0.74 | 0.75 | 0.72 |
| Application obsolescence probability | $P_{\omega,\mathrm{pre}}$ | 0.14 | 0.13 | 0.13 | 0.12 | 0.14 |
| Renewal obsolescence probability | $P_{\omega,\mathrm{post}}$ | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 |
| Simple application fighting cost log-mean | $\mu_{f,\mathrm{simple}}$ | 8.53 | 8.43 | 8.56 | 8.60 | 8.53 |
| Simple application fighting cost log-sigma | $\sigma_{f,\mathrm{simple}}$ | 0.87 | 0.97 | 0.79 | 0.74 | 0.95 |
| SMM Objective | | 1.23 | 1.47 | 1.29 | 1.25 | 1.33 |

*Notes*: This table provides estimates of the model parameters across various model alternatives. The baseline model defines senior examiners as those at the GS14 level. The last column expands this to include GS13 and GS14.

Table A.6. Net Social Costs of Patent Prosecution: Robustness

| | Patent Premium ($\xi$) = 0.10 | | | | Patent Premium ($\xi$) = 0.05 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Counterfactual** | $\boldsymbol{T_1}$ | $\boldsymbol{T_2}$ **(1.5)** | $\boldsymbol{T_3}$ | **Total** | $\boldsymbol{T_1}$ | $\boldsymbol{T_2}$ **(1.5)** | $\boldsymbol{T_2}$ **(2.0)** | $\boldsymbol{T_3}$ | **Total (1.5)** | **Total (2.0)** |
| Baseline ($Bn) | 6.4 | 0.7 | 17.6 | 24.7 | 6.6 | 0.0 | 0.2 | 20.6 | 27.2 | 27.4 |
| 25K Round Fee | 5.9 | 1.8 | 16.4 | 24.1 | 6.3 | 0.7 | 1.4 | 19.1 | 26.1 | 26.8 |
| 50K Round Fee | 6.1 | 3.1 | 15.1 | 24.2 | 5.5 | 1.7 | 3.5 | 17.1 | 24.7 | 26.1 |
| Three Rounds | 4.9 | 4.8 | 10.2 | 19.8 | 5.4 | 1.9 | 3.9 | 11.5 | 18.8 | 20.8 |
| Two Rounds | 2.9 | 7.4 | 4.7 | 14.9 | 2.9 | 3.2 | 6.6 | 5.2 | 11.4 | 14.8 |
| One Round | 0.0 | 6.3 | 0.7 | 7.0 | 0.0 | 1.6 | 3.3 | 0.8 | 2.4 | 4.1 |
| 15% IM | 29.0 | 1.1 | 15.0 | 45.1 | 31.6 | 0.4 | 0.8 | 17.3 | 50.1 | 49.8 |
| Credit$\searrow$ | 6.4 | 0.7 | 17.6 | 24.7 | 6.5 | 0.0 | 0.2 | 20.6 | 27.2 | 27.3 |
| Credit$\searrow$ + 15% IM | 24.3 | 1.9 | 15.8 | 42.0 | 23.7 | 0.7 | 1.5 | 18.2 | 47.8 | 43.3 |

*Notes*: This table provides the values of net social costs for alternative values of the patent premium and social multiplier. Columns denoted $\boldsymbol{T_2}$ **(1.5)** and $\boldsymbol{T_2}$ **(2.0)** provide values of type 2 net social costs when $\frac{\rho_{\text{soc}}}{\rho_{\text{priv}}}$ is equal to 1.5 and 2.0, respectively. Columns **Total (1.5)** and **Total (2.0)** provide the total net social costs when $\frac{\rho_{\text{soc}}}{\rho_{\text{priv}}}$ is equal to 1.5 and 2.0, respectively.

Figure A.1. Match of internal data and model moments

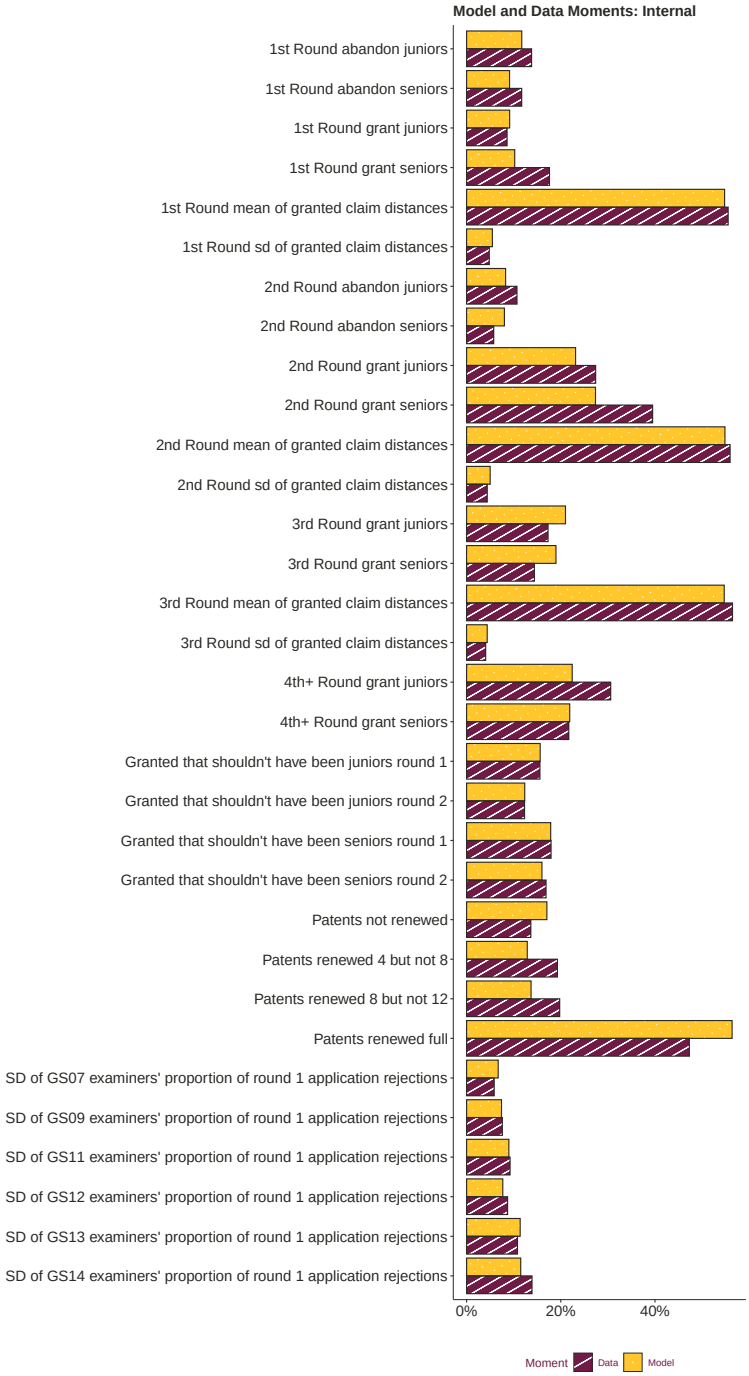
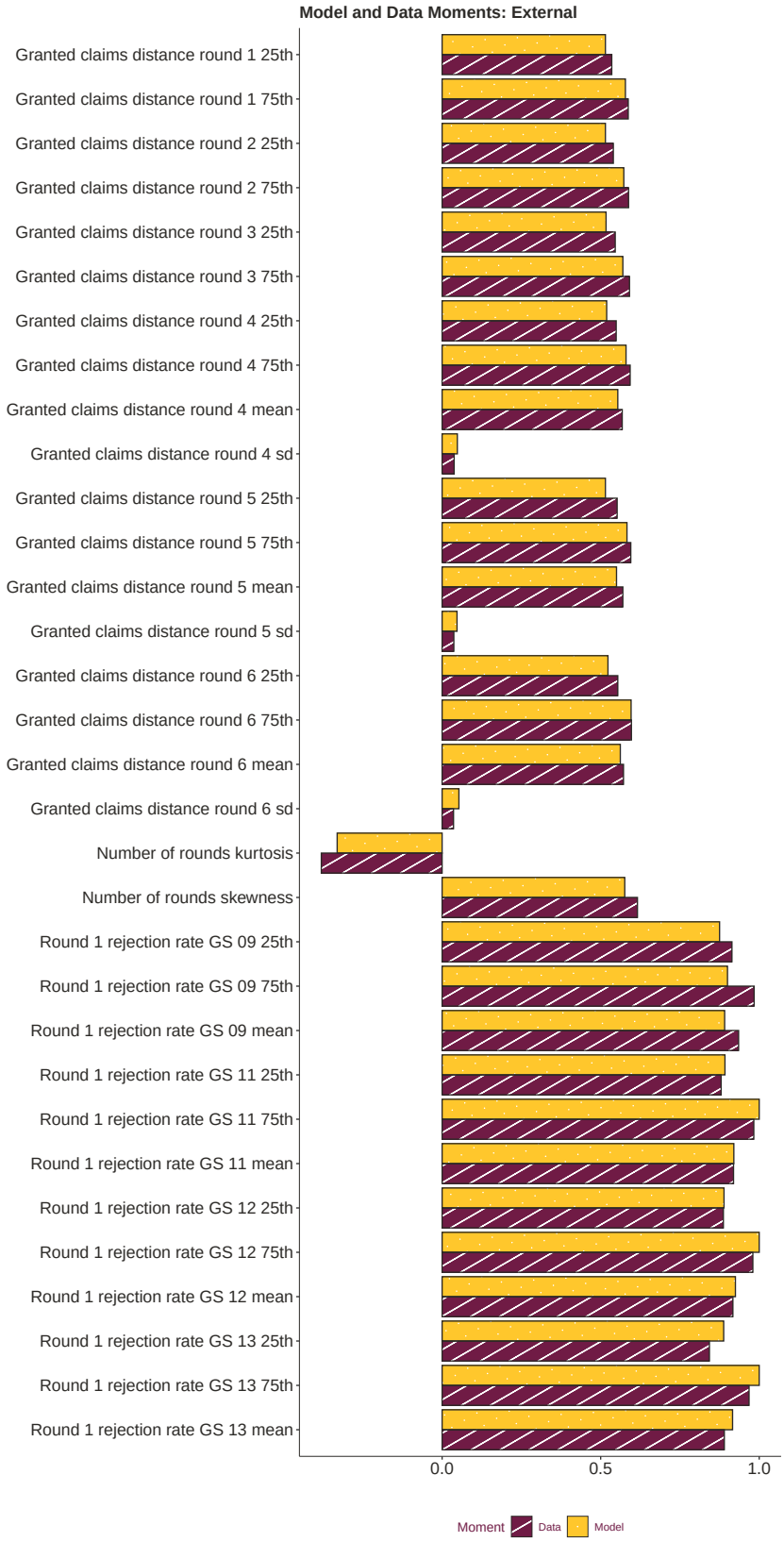


Figure A.2. Match of external data and model moments



# B Data Sources

If the links are broken, the documents are available upon request.

## B.1 Publicly Available Datasets

1. *U.S.PTO Patent Application Claims Full Text Dataset* and *U.S. PTO Patent Claims Full Text Dataset*: https://www.uspto.gov/learning-and-resources/electronic-data-products/patent-claims-research-dataset
2. *Patent Examination Research Dataset*: https://www.uspto.gov/ip-policy/economic-research/research-datasets/patent-examination-research-dataset-public-pair
3. *U.S.PTO Maintenance Fee Events Dataset*: https://developer.uspto.gov/product/patent-maintenance-fee-events-and-description-files
4. *U.S.PTO Office Action Research Dataset*: https://www.uspto.gov/ip-policy/economic-research/research-datasets/office-action-research-dataset-patents
5. *Frakes and Wasserman (2019)*: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ABE7VS

## B.2 Data from Public Documents

6. GDP Deflator: https://fred.stlouisfed.org/series/GDPDEF.
7. *AIPLA Report of the Economic Survey*: See https://www.aipla.org/detail/journal-issue/economic-survey-2017 for 2017.
8. Industry concentration: https://www.census.gov/content/dam/Census/programs-surveys/economic-census/data/archived_tables/2007/sector31/2007_31-33_Con_Ratios_US.zip.
9. Patent Office fees: https://www.govinfo.gov/content/pkg/CFR-2011-title37-vol1/pdf/CFR-2011-title37-vol1.pdf or from https://www.uspto.gov/sites/default/files/aia_implementation/AC54_Final_Table_of_Patent_Fee_Changes.pdf.
10. Patent operations costs:

    **2005:** https://www.uspto.gov/sites/default/files/about/stratplan/ar/USPTOFY2005PAR.pdf
    **2010:** https://www.uspto.gov/sites/default/files/about/stratplan/ar/USPTOFY2010PAR.pdf

**2015:** https://www.uspto.gov/sites/default/files/documents/USPTOFY15PAR.pdf

11. Patent applications: https://www.uspto.gov/web/offices/ac/ido/oeip/taf/us_stat.htm.

12. R&D expenditures: https://www.nsf.gov/statistics/infbrief/nsf14307/.

# C Distance Measure

This section provides details on how we construct our patent distance metric. We describe our preferred choice, the paragraph vector approach.[1] The method consists of four steps: (1) standardizing the independent claim text, (2) turning the text into a numerical vector, (3) calculating the distances between a focal patent claim on an application to all existing granted patent claims and (4) calculating the distance to the closest existing independent claim.

The first step before converting text into a numerical vector is text standardization. We perform basic changes to the content of the text and remove words that carry no informational content. Once we standardize the text, we drop any claims with fewer than two words or illegible text.

We use the paragraph vector approach to represent the text of a patent claim as a numerical vector. The paragraph vector approach is an improvement of the word vector approach. We implement the Paragraph Vector approach using Gensim's `Doc2Vec` Python model (Řehůřek and Sojka, 2010).

The step above converts all patent claims, including those on applications and those granted, into a numerical vector. The next step involves taking every focal application patent claim vector and calculating its distance to every *existing* granted claim at the point of application. After representing a patent claim's text as a numerical vector, we use cosine similarity and angular distance, both of which are standard in the text matching and the NLP literature. We compute the cosine similarity (CS) between claim text vectors $x$ and $y$ as

$$cs(x, y) = \frac{\sum_i x_i y_i}{\sqrt{\sum_j x_j^2 \sum_j y_j^2}}.$$

Then, we calculate the angular distance (AD) metric, $AD(x, y) = \arccos(cs(x, y))/\pi$ and then double $AD$ to obtain a normalized distance in the interval $[0, 1]$.

[1] At the time of writing this paper, we used the state-of-the-art approach, but there is a fast-moving frontier. The most recent approaches use GPT-4 or BERT word embeddings integrated directly into Neural Networks. See Elliot and Hansen (2023) for details on text algorithms.

With all distances computed, it is a simple step to find the closest 50 claims to each application. We experiment with different choices on which percentile of the closest 50 distances to use. We also experimented with taking an average of the five closest distances for example, and the resulting distances were similar.

# D Descriptive Results

We show how patent application outcomes vary with technology center and examiner seniority. First, we regress a binary variable equal to one if the application process lasts more than one round against fixed effects for examiner seniority grade, technology center, year of application, and a small entity indicator (applying firm having fewer than 500 employees). The results in Column (1) of Table D.1 reveal substantial variation across technology centers; e.g., Computer Networks (TC-24) has a 12 percentage point higher likelihood of multi-round negotiation than the reference category, Biotechnology (TC-16). Further, the likelihood of any negotiation decreases with the seniority of the examiner, with senior (GS-14) examiners nine percentage points less likely to require negotiation relative to the most junior, holding technology center and application year fixed. Further, small entities are 12 percentage points less likely to negotiate (all else fixed).

In Column (2), we do the same analysis for the dependent variable equal to one if the examiner grants a patent. We match the findings of Frakes and Wasserman (2017) – senior examiners are more likely to grant and grant rates vary substantially across technology centers. In our model, we explain this variation by letting the distribution of intrinsic motivation vary with seniority level, by incorporating differences in the credit structure for examiners that vary across seniority and technology centers, and by allowing fighting costs to differ for applicants, with technology category-specific distributions. Our parameter estimates enable us to disentangle the effects of these factors in explaining the variation in outcomes, as we discuss in the text.

Table D.1. Regression Results

| Variable | (1) Negotiation | (2) Grant |
|---|---|---|
| Intercept | 0.7433 (0.006) | 0.542 (0.005) |
| GS-7 | -0.002 (0.004) | 0.003 (0.005) |
| GS-9 | -0.016 (0.004) | 0.035 (0.004) |
| GS-11 | -0.020 (0.004) | 0.066 (0.004) |
| GS-12 | -0.034 (0.004) | 0.092 (0.004) |
| GS-13 | -0.045 (0.004) | 0.126 (0.004) |
| GS-14 | -0.091 (0.004) | 0.178 (0.004) |
| Chemicals (17) | 0.064 (0.002) | 0.067 (0.002) |
| Comp. Software (21) | 0.105 (0.002) | 0.196 (0.002) |
| Comp. Networks (24) | 0.123 (0.002) | 0.192 (0.002) |
| Communications (26) | 0.047 (0.002) | 0.198 (0.002) |
| Electronics (28) | -0.010 (0.001) | 0.244 (0.001) |
| Other (36) | 0.065 (0.002) | 0.136 (0.002) |
| Mech Engineering (37) | 0.042 (0.002) | 0.139 (0.001) |
| Small Entity | -0.120 (0.001) | -0.170 (0.001) |
| Year FE | Yes | Yes |
| N | 1,641,333 | 1,759,313 |

*Notes*: Omitted grade is GS-5 and omitted technology center is Biotechnology and Organic Fields (16). Technology center "Other" refers to Center 3600, which is "Transportation, Electronic Commerce, Construction, Agriculture, Licensing and Review." Following Frakes and Wasserman (2017), we omit GS-15 grade examiners. We report heteroskedasticity robust (HC1) standard errors in parentheses.

These results show stark differences in average grant rates and likelihood of negotiation across technology centers and examiner seniority grades. Next, we investigate the variation in *examiner-specific* decisions within and between seniority grades and technology center pairs. To do this, we calculate examiner-specific outcomes (average grant rates, number of rounds, length of examination period, probability of negotiation, etc.) within each seniority grade examiners are in at the time. We decompose the variation in these examiner averages into within and between seniority grade-technology center pairs by introducing dummies for each seniority-grade-technology-center dyad in Table D.2. The proportion of within-group variation in examiner grant rates is 80%, im-

Table D.2. ANOVA Results

| Variable | Grade × TC Fixed Effects |
|---|---|
| Grant rate | 79.84 |
| Duration of examination (years) | 75.79 |
| Number of rounds | 80.89 |
| No negotiation (one round) | 89.53 |
| Independent claims granted | 74.93 |

*Notes*: For each variable $y$, and an examiner $e$ when they are in seniority grade $S$ and technology center $T$, we calculate $\bar{y}_{eST}$. Then we regress $\bar{y}_{eST}$ on a set of interactive dummies for seniority grade and technology center. We report $1 - R^2$ (as a percentage) for these regressions, thereby providing the proportion of within group variation.

plying substantial variation in examiner grant rates not explained by seniority and technology centers. Our model explains this variation in examiner-specific grant rates within the technology center and seniority groups by incorporating group-specific *distributions* of examiner intrinsic motivation and costs of delay.

# E Examiner Credit Structure

Here we provide expressions for $g^r_{GR}(S,T)$, $g^r_{ABN}(S,T)$, $g^r_{RCE}(S,T)$ and $g^r_{REJ}(S,T)$. For $y \in \{GR,\ ABN,\ REJ,\ RCE\}$, we write $g^r_y(S,T) = \nu^r_y \cdot c(S,T)$, and give expressions for $\nu^r_y$ and $c_(S,T)$ separately.

## E.1 Credits

Granting in the first round gives the examiner a payoff of $\nu^1_{GR} = 2$ credits. Rejecting in the first round gives $\nu^1_{REJ} = 1.25$. If the applicant abandons in round one, the examiner obtains $\nu^1_{ABN} = 0.75$. Granting in the second round gives $\nu^2_{GR} = 0.75$ credits. Rejecting in the second round gives $\nu^2_{REJ} = 0.25$ credits, with an extra $\nu^2_{ABN} = \nu^2_{RCE} = 0.5$ credits whether the applicant abandons or continues to an RCE. Ultimately, the examiner obtains two credits irrespective of what happens in the first two rounds. The only difference is whether they obtain the credits immediately (say, from an immediate grant) or spread out over two rounds.

The structure of the payoffs in the first RCE are the same, except $\nu^3_{REJ} = 1$ and $\nu^3_{GR} = 1.75$. In this case, irrespective of what happens in the RCE, the examiner will obtain 1.75 credits. The

Table E.1. Seniority Corrections

| Seniority Grade | Signatory Authority | $c_{SEN}(S)$ |
|---|---|---|
| GS-5 | None | 0.55 |
| GS-7 | None | 0.7 |
| GS-9 | None | 0.8 |
| GS-11 | None | 0.9 |
| GS-12 | None | 1.0 |
| GS-13 | None | 1.15 |
| GS-13 | Partial | 1.25 |
| GS-14 | Partial | 1.25 |
| GS-14 | Full (primary examiner) | 1.35 |

*Notes*: This table provides the seniority factors for credit adjustment. In the empirical work, we use 1.15 for GS-13 and 1.25 for GS-14.

difference comes from whether they receive all 1.75 credits at once by granting, or 1 credit from their non-final rejection and $\nu^4_{REJ} = 0.25$ plus $\nu^4_{ABN} = \nu^4_{RCE} = 0.5$ credits from the applicant's response.

In the second and any subsequent RCEs, the structure of the payoffs is still the same, except $\nu^{2r+1}_{REJ} = 0.75$ and $\nu^{2r+1}_{GR} = 1.5$ ($r > 1$). As before, the examiner will receive 1.5 credits from second and subsequent RCEs. The difference comes from whether they receive all 1.5 credits at once from granting, or 0.75 credits from their non-final rejection and $\nu^{2r+2}_{REJ} = 0.25$ plus $\nu^{2r+2}_{ABN} = \nu^{2r+2}_{RCE} = 0.5$ credits from the applicant's response.

## E.2 Seniority and Technology Complexity Adjustments

The seniority and technology complexity adjustment term is

$$c(S,T) = \frac{c_{TECH}(T)}{c_{SEN}(S)}.$$

Table E.1 gives the values of $c_{SEN}(S)$ across the GS categories. Higher seniority factors imply larger values of $c_{SEN}$, and therefore lower values of credits. Table E.2 gives the values of $c_{TECH}(T)$ we created for the different technology centers and use in the estimation of the model. The Patent Office does not have adjustments at the technology center level, but rather at the more detailed U.S. Patent Class (USPC) level. We obtained the adjustments at the USPC level from the Patent Office and constructed a patent-application weighted average for each technology center.

Table E.2. Technology Center Adjustments

| **Technology Center** $T$ | **U.S.PTO Number** | **Correction** ($c_{TECH}(T)$) |
|---|---|---|
| Chemical and Materials Engineering | 17 | 22.2 |
| Computer Architecture Software and Information Security | 21 | 31 |
| Computer Networks, Multiplex, Cable and Cryptography/Security | 24 | 29 |
| Communications | 26 | 26.5 |
| Semiconductors, Electrical and Optical Systems and Components | 28 | 21.4 |
| Transportation, Electronic Commerce, Construction, Agriculture... | 36 | 22.4 |
| Mechanical Engineering, Manufacturing and Products | 37 | 19.9 |

# F Moment Selection and Identification Intuition

First, we provide further details on the possible moments we could use to estimate our model. Then, we provide some information on our methods to prune moments from the full set. Finally, we provide some intuition on how the moments identify the model parameters.

## F.1 Available Moments

We have seven sets of moments available, which we describe in turn.

Our first group of moments corresponds to examiners' issuance and applicants' abandonment decisions. For each round in the model and each seniority level, we calculate the proportion of applications examiners grant and the proportion that applicants abandon. Since there are nine seniority grade-signatory authority pairs, and we observe at least six rounds, this implies at least 108 moments on grants and abandonments.

Second, we observe the distribution of the proportion of claims rejected, both by round (six) and by seniority grade-signatory authority pair (nine). These observations generate another 54 moments. Third, we observe the proportion of granted patents that renew at four, eight, and twelve years after issuance. These observations generate four moments on patent renewals (don't renew at four, renew at four but not eight, renew at eight but not twelve and renew at twelve).

Fourth, we calculate the distribution of claim distances by round. We calculate the mean and standard deviation of the distance distribution by round for at least six rounds, implying at least 12 moments on distance. Another moment comes from the within-application distance correlation. Fifth, at each of the nine seniority grades, we calculate each examiner's *leniency*, which is their average rejection rate across all the applications they examine. Hence for each seniority grade-signatory authority pair, we obtain a distribution of examiner rejection rates, for

which we can calculate the mean and standard deviation of the distribution of examiner fixed effects. From this we obtain another 18 moments.

Next, given that we can identify the distance threshold externally, we calculate the proportion of granted patents containing at least one invalid claim (that is, a claim whose distance is below the distance threshold). Hence, for each round and each seniority level, we calculate the proportion of patents granted containing an invalid claim, implying another 54 moments.

Finally, we observe the distribution of application fighting costs. We have six moments on the distribution of legal application fees for four technology categories (simple, chemical, electrical and mechanical), which we match to the technology centers on which we estimate the model. This implies another 24 moments.

### F.2 Choosing Moments

We have more than two hundred data moments that we can calculate from endogenous variables in the model. Since we have 21 model parameters to estimate with simulated method of moments, in principle, we are over-identified. However, not all moments will aid the estimation procedure in identifying the parameters, so we begin by pruning the set of moments for estimation.

We follow a rigorous, data-driven methodology to create a subset of the moments that best estimate the parameters. To do this, we calculate the sensitivity matrix described in Andrews, Gentzkow, and Shapiro (2017). As the authors explain, "sensitivity gives a formal, quantitative language in which to describe the relative importance of different moments for determining the value of specific parameters." If a moment had a small value in the sensitivity matrix for all parameters, we considered it as not useful in estimating our model. Further, as described in Jalali, Rahmandad, and Ghoddusi (2015), for each parameter and moment, we plot the value of the moment for different values of the parameter, fixing the other parameters at their estimates. If this curve is flat, this parameter does not influence on the value of the moment. For a given moment, if the curve is flat across *all* parameters, it suggests that the moment offers no useful variation to identify the parameters.

For each parameter, we also plot the value of the SMM objective across all values of the parameter, fixing other parameters at their estimates. Ideally, the SMM will be U-shaped in each parameter to ensure a well-defined global minimum exists. By doing this, we learn how well we pin down parameters based on the set of moments we have available.

By combining the sensitivity matrix with moment and SMM plots, we pruned the set of moments

down to those that offer some assistance in estimating the parameters. Since we split many parameters into two seniority groups (junior and senior), we split some of our moments into the same seniority categories.

### F.3 Full Set of Moments

The full set of moments we use for estimation is as follows. The selected moments corresponding to outcomes for examiners are:

(i) The proportion of applications granted in each round for juniors and seniors, for rounds one, two, three, and all rounds after four combined [eight moments]

(ii) The standard deviation of the distribution of examiner rejection rates for the six seniority categories used by the Patent Office (GS levels 7, 9, 11, 12, 13, and 14) [six moments]

(iii) The proportion of patents granted containing an invalid claim (for juniors and seniors) for rounds one and two [four moments]

The moments corresponding to outcomes for applicants are:

(i) The proportion of abandonments in each round, when the assigned examiner is junior and senior, for rounds one and two [four moments]

(ii) The proportion of granted patents not renewed, renewed at year four but not eight, renewed at year eight but not twelve, and renewed at year twelve [four moments]

(iii) The mean and standard deviation of the distribution of granted claim distances for rounds one, two, and three [six moments]

(iv) Mean and median of legal application fees for simple applications and complex applications in electrical, mechanical, and chemical technologies [eight moments]

### F.4 Identification

A model is either point identified or not, and technical conditions on the required variation in exogenous variables determine whether a model is identified (Andrews, Gentzkow, and Shapiro, 2017). Due to our model's complicated and nonlinear nature, we cannot calculate these conditions. Identification with simulated method of moments is based on how different moments are affected by specific parameters. While we cannot identify this link exactly, we provide some intuition of how moments aid in pinning down specific parameters of the model.

We start with the parameters relating to the applicant. The renewal rates, together with first-round abandonment decisions, aid in identifying the parameters of the distribution of flow returns,

i.e., $\mu_v$ and $\sigma_v$. This is because, all else equal, an applicant with higher returns is less likely to abandon after learning their examiner and more likely to renew their patent, conditional on being granted. The renewal moments also aid in identifying the post-grant obsolescence probability $P_{\omega,\text{post}}$. Similarly, the ex post claim distribution of padded distances, as calculated using the distance between text vectors, aids in identifying the parameters of the distribution of ex ante unpadded distance, i.e., $\alpha_D$ and $\beta_D$. Moments on application fighting costs directly pin down the distribution of application fighting costs, $\mu_{f_{app}}$, and $\sigma_{f_{app}}$.

Regarding pre-grant obsolescence $P_{\omega,\text{pre}}$, the only case in which an applicant abandons in interim rounds two to four is when they become obsolete. If an applicant, upon learning their examiner calculates that they will want to abandon in any round after the first, they will abandon immediately in round one. Therefore, interim round abandonments offer substantial assistance in identifying the obsolescence probability in the application process.

Intuition for examiner parameters is more complicated. Observing that examiners grant several invalid patents could result from low intrinsic motivation, high examiner error, or high examiner delay costs. Three factors make this challenge less formidable. First, since we assume that only intrinsic motivation varies by seniority, differences in grant rates and examiner errors by seniority pick up the value of intrinsic motivation, $\mu_{\text{im}}$ by seniority, and differences in the variation in examiner-specific grant rates by seniority capture the variation in intrinsic motivation, $\sigma_\theta$ by seniority.

Second, we assume that each examiner has the same delay cost across all applications and rounds but faces varying intrinsic motivation costs at each round of every application (because $\mathcal{R}^r$, the proportion of invalid independent claims varies across rounds and applications). This implies that the proportion of invalid patents granted in rounds one and two offer the best assistance in identifying the mean examiner intrinsic motivation and mean examiner delay costs. Third, examiner error is two-sided and symmetric. This feature creates cases where examiners do not grant valid patents, whereas intrinsic motivation and delay costs only incentivize examiners to grant when they should not. Otherwise, we know that an examiner, making no mistake, and facing a fully valid patent, will always issue it. Together, this implies that we can use the residual variation in grant rates (valid and invalid) by round and seniority to learn about the distribution of examiner error.

### F.5 Details on Model Fit

As shown in Figure A.1, we match most of the internal moments well, though there are two exceptions. The first is the proportion of fully renewed patents, which we overestimate. The

other exception is the second-round grant rate. This moment is difficult to match with our model because examiners have incentives to wait until the third round and obtain RCE credits if they do not choose to grant in the first round. Since examiners have incentives and targets across applications on their desks (docket management), they are more likely to grant in the second round than our baseline model predicts.

# G Quantification of Social Costs

## G.1 Implementing Type 1 Social Cost Calculation

As indicated in the text, a key challenge in implementing our calculation of type 1 social costs comes from the fact that the estimates of the value of patent rights for invalid patents include potential litigation costs. To impute the "value at stake" in litigation for these patents, we need to adjust our methodology to exclude these costs.

To do this, we make two assumptions:

A1: Valid patents are not litigated. This assumption holds in a model with perfect courts, where a competitor knows (or can pay a fee to discover) whether a patent is valid or not, and then choose whether to litigate based on the result.[2] This assumption allows us to calculate the value of patent rights for valid patents, $\tilde{V}$, as equal to the observed value since there are no litigation costs to net out.

A2: The *distribution* of the value at stake, $G_{\tilde{V}}(\cdot)$, is the same for valid patents as invalid patents. The basis for this assumption is that initial distances and values are uncorrelated in the model. This assumption allows us to draw values from the observed distribution of $\tilde{V} = V$ for valid patents and use them as draws from the distribution of $\tilde{V}$ for invalid patents.

Given A1 and A2, the procedure for calculating type 1 social costs is as follows:

1. Estimate the parameters of a log-normal distribution for the value at stake for *valid* patents.[3] Let the estimated distribution be denoted as $\hat{G}_{\tilde{V}}(\cdot)$.

[2]This assumption is *not* at odds with Schankerman and Schuett (2022), where *high types* are litigated with some probability even though they will not be invalidated. The important point is that high types in their model (patents that would not be developed without patent rights) are not the same as valid patents in our model, which are defined as those with distance larger than the threshold.

[3]The sum of log-normal distributions is approximately log-normal (Dufresne, 2004), which our simulation here exhibits.

2. Let $\bar{P}$ be the total number of *invalid* patent grants for the given period we simulate. Then, for each $p = 1, \ldots, \bar{P}$:

   (a) Take a draw from the estimated distribution of *valid* patents' value at stake (ex post value), $\hat{G}_{\tilde{V}}(\cdot)$, to represent the value at stake for the invalid patent $p$

   (b) Using the draw, calculate $S_{1p}$ from Equation (11).

3. Calculate the total social cost of type 1 error as $\sum_{p=1}^{\bar{P}} S_{1p}$.

Finally, note that we calculate the threshold for exposure to litigation from the *empirical* distribution of the value at stake for valid patents, $\hat{G}_{\tilde{V}}(\cdot)$.

## G.2 Implementing Type 2 Social Cost Calculation

The primary challenge in implementing our calculation of type 2 social costs comes from calibrating the value of the invention without patent rights ($\pi$), particularly for inventions with $\Gamma^* \leq 0$, where we cannot use the patent premium. In a similar vein to our approach to type 1 social costs, we assume that the distributions of $\pi$ for those with positive and negative $\Gamma^*$ are the same and then draw values of $\pi$ from this distribution for those inventions.

To be precise, our specific implementation is as follows:

1. Draw a pilot set of potential inventions, used to calculate a distribution of $\pi$. Run these set of potential inventions through the model and calculate $\Gamma^*$. For those with positive $\Gamma^*$, create a distribution of $\pi$ using the relationship $\Gamma = \xi\pi$.

2. Now start the simulation for type 2 social costs by drawing a new set of potential inventions (returns, distances, number of claims, fighting costs, examiner etc.). For each potential invention $i$, calculate $\Gamma_i^*$. If $\Gamma_i^* > 0$, calculate $\pi_i = \frac{\Gamma_i^*}{\xi}$. If $\Gamma_i^* \leq 0$, draw a value of $\pi_i$ from the distribution calculated in 1. Also, draw a development cost $\kappa_i$.

3. For each of the potential inventions $i$, work out the set $i = 1, \ldots, \mathcal{I}_{\text{no dev}}$ that do not develop as those with $\max\{\Gamma_i^*, 0\} + \pi_i < \kappa_i$

4. For $i = 1 \ldots, \mathcal{I}_{\text{no dev}}$, run the potential invention through a model where, at the point of abandonment, the inventor obtains all valid claims they have, and so obtains the patent value of their valid claims, instead of a payoff of 0. By definition, this scenario has the property that all abandoned claims are invalid, so that there is no type 2 error. Let $\Gamma_i'$ denote the expected value of patent rights in this new scenario.

5. For $i = 1 \ldots, \mathcal{I}_{\text{no dev}}$, calculate the set $i = 1, \ldots, \mathcal{I}_{\text{now dev}}$ who have $\max\{0, \Gamma_i'\} + \pi_i \geq \kappa_i$. This is the set who do not develop because of type 2 error but do develop in the absence of type 2 error.

6. For $i = 1, \ldots, \mathcal{I}_{\text{now dev}}$, calculate $S_{2i} = \frac{\rho_{\text{soc}}}{\rho_{\text{priv}}}\Big(\max\{0, \Gamma_i\} + \pi_i\Big) - \kappa_i$ and calculate the total type 2 social cost as
$$T_2 = \sum_{i=1}^{\mathcal{I}_{\text{now dev}}} S_{2i}.$$

## G.3 Calibrating Deadweight Loss

In the derivation of deadweight loss, note that
$$DWL = \frac{1}{2}\Delta\wp\Delta q = \frac{1}{2}\frac{\Delta q}{q} q\Delta\wp = \frac{\lambda}{2}\frac{\Delta\wp}{\wp}\tilde{V},$$
by the definitions of $\tilde{V}$ and $\lambda$. Further, note that
$$\frac{\Delta\wp}{\wp} = \frac{q\Delta\wp}{q\wp} = \frac{\text{lic. rev}}{\text{sales}} = \frac{\text{lic. rev}}{\text{R\&D}} \cdot \frac{\text{R\&D}}{\text{sales}}$$
As described in the text, we use Schankerman and Schuett (2022) for the ratio of licensing revenue to R&D, and data from the Bureau of Economic Analysis for the ratio of R&D to sales.

## G.4 Deadweight Loss Under Cournot Competition

In the main text, we compute deadweight loss from a patented invention assuming symmetric licensees operate in a perfectly competitive industry. Suppose instead that the licensees compete in a Cournot setting. By standard calculations, the equilibrium price-cost margin is $\frac{\wp - c}{\wp} = \frac{m^*}{\lambda}$ where $m^* = \frac{1}{N}$ is the average market share and $\lambda$ is the demand elasticity. We write this as $\frac{\wp - c}{\wp} = \frac{H^e}{\eta}$ where $H^e$ is the symmetric-equivalent Herfindahl index of concentration. Thus for $H^e < 1$
$$\wp = \frac{c}{1 - \frac{H^e}{\lambda}}.$$
With imperfect competition, the change in equilibrium price is larger than the Arrow royalty due to double marginalization: $\Delta\wp = \frac{\Delta c}{1 - \frac{H^e}{\lambda}} > \Delta c$. The associated deadweight loss with Cournot competition is
$$DWL_{\text{cournot}} = \frac{1}{2}\Delta\wp\Delta q = \frac{1}{2}\frac{\Delta c}{1 - \frac{H^e}{\lambda}}\Delta q = DWL_{\text{pc}} \cdot \frac{1}{1 - \frac{H^e}{\lambda}},$$
where it should be noted that in this case $\tilde{V} = q\Delta c$ denotes total *royalty payments.* Since $H^e \in (0, 1)$ and we require that $|\lambda| > 1$, deadweight loss in this imperfect competition setting is larger than in perfect competition case.

Using U.S. Census data for 2007, the value added weighted-average Herfindahl index for manufacturing industries (based on the 50 largest firms), $H$, for manufacturing sectors is 0.05. As is well-known,the Herfindahl index can be decomposed as $H = \frac{1}{N} + N \cdot \text{Var}(m) = H^e + N \cdot \text{Var}(m)$, where $m$ is the market share of each firm. Thus, the observed $H$ overstates the unobserved $H^e$, so the computed deadweight loss will be an upper bound to the true value of $DWL$. Despite this, the upper bound for the Cournot setting is not materially different from the competitive case in the text.

The value of $H$ varies widely across industries. We do not compute deadweight loss using industry-specific values because it is difficult to assign patents in different patent classes to industries, and the existing Patent Office concordance is problematic (e.g., the mapping is not unique).

### G.5 Calibrating Litigation Costs

To calibrate litigation costs, $\mathcal{C}(\tilde{V})$, we use data from the American Intellectual Property Law Association (AIPLA) surveys on litigation costs as a function of (intervals) of the value at stake, which we assume is the same for the patentee and challenger. We use the linear specification

$$\mathcal{C}(\tilde{V}) = \ell_0 + \ell_1 \tilde{V}$$

Using this same specification, Schankerman and Schuett (2022) estimate $\ell_0 = \$624,000$ and $\ell_1 = 0.162$ (2018 USD). Note that this calibration of legal costs is at the patent, not claim, level.

### G.6 Calibrating Development Costs

We apply the estimates from Schankerman and Schuett (2022) to our context. They assume that development costs $\kappa$ are exponential, with mean equal to $k_0 + k_1 s$, where $s$ is the size reduction of the invention and $k_0$ and $k_1$ are estimated as $254.6 \times 10^3$ and $2.33 \times 10^{10}$, respectively. Regarding the size reduction, they assume that $s$ is log-logistic distributed with parameters $\beta_0 = 1.02$ and $\beta_1 = 1.14 \times 10^{-6}$. We use the mean value of $s$ in our calibration.

In the baseline quantification, we draw values of $\kappa$ from the distribution described above, which assumes that development costs are independent of $\Gamma^*$ and $\pi$. In this model, inventors know their development costs prior to their decision to develop their idea. We also experiment with another model, which makes the opposite assumption that inventors do not know their development costs and thus use the mean value, $\bar{\kappa} = k_0 + k_1 \bar{s}$, to make their development decision. Both models produce similar conclusions; results are available upon request.

### G.7 Calibrating the Number of Ideas

To compute the number of ideas, we start with the average annual number of utility patent applications in the period 2011–2013. We convert this number into the number of ideas in two steps. First, we use the estimates from Schankerman and Schuett (2022) that about two-thirds of applications are "low type" inventions (defined by them as those that would have been developed even without patent protection), and second, that one-third of ideas become a low type patent application. Together, this implies about one million ideas for potential inventions for each cohort of applications.

## References


Andrews, I., M. Gentzkow, and J. M. Shapiro (2017): "Measuring the Sensitivity of Parameter Estimates to Estimation Moments," *The Quarterly Journal of Economics*, 132, 1553–1592.

Dufresne, D. (2004): "The Log-Normal Approximation in Financial and Other Computations," *Advances in Applied Probability*, 36, 747–773.

Elliot, A. and S. Hansen (2023): "Text Algorithms in Economics," *Unpublished Working Paper*.

Frakes, M. D. and M. F. Wasserman (2017): "Is the Time Allocated to Review Patent Applications Inducing Examiners to Grant Invalid Patents? Evidence from Microlevel Application Data," *The Review of Economics and Statistics*, 99, 550–563.

Jalali, M., H. Rahmandad, and H. Ghoddusi (2015): *Using the method of simulated moments for system identification*, MIT Press.

Řehůřek, R. and P. Sojka (2010): "Software Framework for Topic Modelling with Large Corpora," *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45–50.

Schankerman, M. and F. Schuett (2022): "Patent Screening, Innovation, and Welfare," *The Review of Economic Studies*, 89, 2101–2148.