Review of Economic Studies (2021) **0**, 1–48 doi:10.1093/restud/rdab073

Advance access publication 29 October 2021

Patent Screening, Innovation, and Welfare

MARK SCHANKERMAN
London School of Economics and CEPR

and

FLORIAN SCHUETT
Tilburg University

First version received January 2018*; Editorial decision June* 2021*; Accepted October* 2021 (*Eds.*)

Critics claim that patent screening is ineffective, granting low-quality patents that impose unnecessary social costs. We develop an integrated framework, involving patent office examination, fees, and endogenous validity challenges in the courts, to study patent screening both theoretically and quantitatively. In our model, some inventions require the patent incentive while others do not, and asymmetric information creates a need for screening. We show that the endogeneity of challenges implies that courts, even if perfect, cannot solve the screening problem. Simulations of the model, calibrated on U.S. data, indicate that screening is highly imperfect, with almost half of all patents issued on inventions that do not require the patent incentive. While we find that the current patent system generates positive social value, intensifying examination would yield large welfare gains. The social value of the patent system would also be larger if complemented by antitrust limits on licensing.

Key words: Innovation, Patent quality, Screening, Litigation, Courts, Patent fees, Licensing.

JEL Codes: D82, K41, L24, O31, O34, O38

1. INTRODUCTION

The patent system is one of the key policy instruments governments use to provide innovation incentives. However, there is growing concern among academic scholars and policymakers that patent rights are becoming an impediment, rather than an incentive, to innovation. They worry that the proliferation of patents, and their dispersed ownership, raise the transaction costs of doing R&D, and expose firms to rent extraction through patent litigation (Heller and Eisenberg, 1998; Lemley and Shapiro, 2005; Bessen and Maskin, 2009). These dangers have featured prominently in public debates (National Academy of Sciences, 2004; Federal Trade Commission, 2003; The Economist, 2015), U.S. Supreme Court decisions (eBay Inc. v. MercExchange, 547 U.S. 338 (2006); KSR Int'l Co. v. Teleflex Inc., 550 U.S. 398 (2007)), and the Leahy-Smith America Invents Act of 2011, the most significant patent legislation in half a century.

Critics claim that the main culprit is ineffective patent office screening, granting patents to inventions that do not represent a substantial inventive step, especially but not only in new

The editor in charge of this paper was Dirk Krueger.

areas such as business methods and software. Anecdotes of egregious cases buttress this view (Jaffe and Lerner, 2004). One high-profile example is the Amazon "one click" shopping patent, granted in 1999. The patent was licensed to Barnes & Noble to settle a patent suit and, despite widespread skepticism as to its validity, ran to full term without its validity ever being resolved by the courts.[1] Some scholars have gone as far as to argue that the patent system is unnecessary as an innovation incentive and actually reduces welfare (Boldrin and Levine, 2008). Of course, the real policy choice is not binary—it is not the current patent system or no system. A more constructive approach is to gauge the severity of the patent-quality problem and evaluate whether the current system is welfare enhancing. Perhaps most importantly, we need to assess whether a better-designed patent system would deliver greater welfare gains.

In this article, we develop an integrated framework designed to provide quantitative answers to these key policy questions. Specifically, we study how policymakers can improve screening through four policy instruments: the intensity of patent office examination, pre-grant (application) fees, post-grant (renewal) fees, and review by the courts of patents challenged by a competitor. For most of the analysis, we assume that courts are perfect. We use this benchmark to highlight the limitations of relying on courts, even if they are perfect. To our knowledge, we are the first to develop an integrative framework incorporating all four policy instruments.

We adopt a game-theoretic approach to studying patent-policy design. The advantage of this approach, compared to the alternative mechanism-design approach, is that it allows us to calibrate the model to evaluate the performance of the current patent system and the impact of policy reforms, which is the central objective of our framework.[2]

In our model, an inventor faces a competitor and is endowed with an idea for an invention. Inventions differ in two dimensions—the cost of developing the idea, which is private information, and the value of the invention, which is known both by the inventor and competitor. Some inventions require the patent incentive to be profitable, while others would be developed even without patents. We refer to the former as *high types* and to the latter as *low types*. Since patent protection increases the profit for both types, however, owners of low-type inventions also have a private incentive to seek a patent. Patents provide incentives to innovate but also cause social costs, so effective screening is important for welfare.

The inventor chooses whether to apply for a patent, in which case he pays a pre-grant fee, and the application is examined by the patent office, which detects low types at a rate that depends on the examination intensity. If subsequently approved, the inventor chooses whether to pay a post-grant fee in order for the patent to come into force. The inventor can then offer a license contract to the competitor, and the competitor chooses whether to challenge the validity of the patent in court. In the baseline model, the court is perfect: it always invalidates low-type patents and upholds high types. Formally, our model is a signalling game in which each decision by the inventor can reveal information about the invention type, and the competitor Bayesian updates.[3]

We augment the theoretical model with specific assumptions about the form of competition, licensing, and other features, and calibrate it to recover the key structural parameters. This is done

1. Although the one-click patent was upheld by the patent office on re-examination, the Court of Appeals for the Federal Circuit found "substantial questions" as to whether it was anticipated or rendered obvious by the prior art, leading the court to vacate the injunction Amazon got against Barnes and Noble (Amazon.com v. Barnesandnoble.com, 239 F.3d 1343 (Fed. Cir. 2001).

2. Examples of papers that take a mechanism design approach to study innovation incentives more generally include Hopenhayn *et al.* (2006), Weyl and Tirole (2012), Galasso *et al.* (2016), Akcigit *et al.* (2016), and Mitchell and Schuett (2020).

3. Reinganum and Wilde (1986) model a settlement game where the informed party makes an offer, as in our model. To our knowledge, Meurer (1989) was the first to analyse such a settlement game in the context of patent litigation, but his focus was not on patent screening.

by matching the equilibrium predictions of the theoretical model to the observed grant rate in the U.S. patent office, the litigation rate and patent validation rates in the courts, and R&D and total factor productivity growth per invention, which are constructed using disaggregated sector-level data and other information. The calibration allows us to infer the severity of the patent-quality problem, to assess whether or not the patent system is welfare-enhancing overall, and to evaluate the welfare impact of a number of policy reforms. The reforms we study fall into two categories: reforms to the patent office fee structure and examination intensity, and reforms to the litigation procedures for patent challenges and the antitrust treatment of patent licensing.

To begin, the estimates of the baseline model indicate that the patent-quality problem is real. Under the current patent policy, 60% of patent *applications* are on low-type inventions which, from an economic point of view, should not be granted. The patent office screens out 48% of these applications. Together, these two findings imply that the share of high-type inventions among patent *grants* is 56%, and hence that almost half of all patents are issued on low-type inventions, causing unnecessary social costs. Nonetheless, we find that even under current patent policy, the patent system is welfare-enhancing: the welfare it generates minus welfare from low-type inventions, that would anyway occur, is positive.

We turn next to a summary of our key results where we bring together insights from both the theoretical and the quantitative analysis. First, even if courts are perfect, they cannot eliminate all low-type patents. There are two reasons for this. Since litigation is costly, not all patents are sufficiently valuable to make them worth litigating. Thus, only a fraction of low types are potentially exposed to challenges. In addition, we show that for patents worth litigating, the equilibrium is semi-separating: low types play a mixed strategy whereby they sometimes mimic high types, and sometimes pre-empt challenges by setting a low license fee. Note that low *license fees* do not imply low *royalties*, as the competitor can be compensated for refraining from a challenge through other components of the license contract. With two-part tariffs, for example, low types set high royalties, enabling them to soften competition and raise industry profits, and ensure the competitor's participation through negative fixed fees (also known as "reverse payments"). Since it is royalties that determine deadweight loss, challenge-preempting low types cause the same social harm as those mimicking high types.

Critics claim that challenge preemption by means of license fees below litigation costs is a strategy adopted by so-called "patent trolls" (Federal Trade Commission, 2011). In our model, the extent to which this is done is endogenous to the design of patent policy. Our quantitative analysis indicates that most low-type patents escape scrutiny in the courts: only 17% of low-type patents (and 10% of all patents) are sufficiently valuable to even be at risk of a challenge; of those that are, about 67% strategically preempt challenges. This result raises serious doubts about over-reliance on the court system to weed out bad patents.

Second, the private incentives to challenge a patent can be either smaller or larger than the social incentives. Conventional wisdom is that the incentives to challenge are insufficient due to the positive externalities a successful challenge generates (Farrell and Merges, 2004; Choi, 2005; Farrell and Shapiro, 2008). While this point is valid, it is incomplete both because the challenger takes only his own litigation cost into account and because the private gains from a successful challenge can be larger than the deadweight loss avoided. Our quantitative analysis shows that, under current patent policy, the net social benefits from challenges are negative: litigation costs far exceed the deadweight loss they eliminate.

Third, whether the patent system is socially beneficial depends critically on patent policy design. We show theoretically that a well-designed patent system is welfare-improving relative to no patent rights. This result relies on the judicious use of both pre-grant fees and examination intensity. Our quantitative analysis complements this result by showing that the converse is also true: a poorly designed patent system reduces welfare. According to our simulations, a patent system with only registration (*i.e.* no patent office examination) is associated with 5.3% lower welfare than the current system, and its social value is *negative.* Thus, it is worse to have a registration system that gives temporary monopoly patent rights to all applicants than it is to have *no* patent rights at all, even if those patent rights are subject to review by perfect courts.

One design feature the theory identifies as desirable is to *frontload fees*, *i.e.*, rely on pre-grant rather than post-grant fees. In contrast, in the major patent offices around the world, fees are backloaded. The intuition for the optimality of frontloading is that low types prefer post-grant fees to pre-grant fees more strongly than high types because low types have a smaller chance of passing examination. Our simulation, however, shows that the impact of frontloading fees is small. This highlights how the quantitative analysis can help identify reforms that, though theoretically desirable, are not practically important.

Fourth, raising the examination intensity above the current level improves welfare. We show that accompanying frontloading by a policy that reinvests the additional fee revenue it generates into screening (a policy which is revenue-neutral), thereby enabling the patent office to raise its detection of low types from 48% to 56%, increases welfare by 0.9%. We also characterize the welfare-maximizing patent policy, where both examination intensity and fees are optimally set. We find an optimal examination intensity of 83%, much higher than the current level. Under this policy, welfare increases sharply, by 3%, and the share of high types among granted patents rises to 84%.

Our theoretical analysis sheds light on the channels through which an increase in examination intensity enhances welfare. We show that patent examination has three virtues: deterrence, detection, and limiting challenges (which is beneficial when net benefits of challenges are negative, as our estimates indicate is the case). More stringent examination deters low types whose inventions are not very valuable from applying and detects those that do apply more frequently. This leads to fewer challenges in equilibrium since the competitor infers that granted patents are more likely to be high types, and thus attaches a lower probability to winning the challenge.

Fifth, there is a tradeoff between screening and innovation. High fees help deter low-type applications, but they also discourage high-type investment. The optimal policy is a case in point. Optimal patent office fees are about fifteen times higher than current levels, leading to a 6% drop in high-type innovation. The reason such high fees are optimal nonetheless is that they curtail the number of applications, thus keeping examination costs under control. Moreover, the drop in high-type innovation is more than compensated by reductions in deadweight loss and litigation costs. The optimal policy thus trades losses in high-type innovation for savings in the social costs associated with patents.

If such large increases in fees seem politically infeasible, it is perhaps reassuring that much of the welfare gain from the optimal policy can be achieved through more modest increases in the patent office budget, accompanied by revenue-neutral increases in fees. We find that doubling (resp. tripling) examination resources per application would yield 50% (resp. 70%) of the gains from the optimal policy.

Sixth, we assess the impact of the Patent Trial and Appeal Board (PTAB), which was established by the U.S. Congress as part of the America Invents Act (2011). The PTAB is an administrative law body within the U.S. Patent and Trademark Office that conducts trials related to challenges of issued patents. It provides for a much cheaper way to challenge patents as compared to a court suit. It has become one of the most contentious aspects of the patent system. Critics claim that the PTAB has made it too easy to revoke patents—"patent death squads," as former Chief Judge Rader of the U.S. Court of Appeals for the Federal Circuit colorfully put it. Our theoretical analysis shows that the welfare effect of the PTAB is ambiguous: while the direct effect of lowering litigation costs is beneficial, there are also indirect, strategic effects which can work in the opposite direction. Our quantitative analysis, however, indicates that the direct effect dominates, with the introduction of PTAB resulting in a welfare increase of 0.8%.

Seventh, the social value of the patent system is much larger if complemented by antitrust limits on licensing. We show in the theory that prohibiting negative fixed fees in licensing contracts lowers the royalty rates charged by low-type inventors, but also increases challenges, so that the net effect on welfare is ambiguous. In 2013, the U.S. Supreme Court ruled that such "reverse payment" agreements may be illegal under the antitrust laws, while recognizing this tradeoff between litigation and higher prices (*FTC v. Actavis, Inc., 570 US 136*). Our quantitative analysis shows that the net welfare effect of prohibiting negative fixed fees is positive and very large—a 3.6% gain relative to the current patent system without the restriction.

Lastly, we study a reform that reallocates the burden of litigation costs. In the U.S., the default rule is for each party to pay their own costs. An alternative rule used in other countries is for the loser at trial to pay both parties' costs. Theoretically, a loser-pays rule raises high-type innovation but also the rate of challenges and thus has an ambiguous welfare impact. Our policy simulation indicates that a loser-pays rule reduces welfare.

There is some economic analysis of patent screening in the literature. Caillaud and Duchêne (2011), Schuett (2013a,b), and Atal and Bar (2014) study models where the patent office sets examination intensity and pre-grant fees. Caillaud and Duchêne (2011) show that failure to screen rigorously has an adverse-selection effect, attracting low-type applications, which may cause a vicious circle. Schuett (2013b) also studies an adverse-selection model, focusing on incentive design for patent examiners. Schuett (2013a) examines a moral-hazard model in which patent screening affects the type of research projects that inventors select. Atal and Bar (2014) show that offering a menu of patents, including a "gold-plated" patent that is screened more rigorously, can improve patent quality. None of these papers models challenges as endogenous, a feature we show is crucial for many of our conclusions, and which allows us to study key policy reforms such as the introduction of the PTAB and antitrust restrictions on licensing.

Farrell and Shapiro (2008) and Llobet *et al.* (2021) study the welfare effects of patent examination in models that involve an explicit litigation stage. Like us, Farrell and Shapiro show that patents that are unlikely to survive a challenge may nevertheless command high royalties. In their symmetric-information setting, however, there is no litigation in equilibrium. Llobet *et al.* study a quality-ladder model of innovation with imperfect courts where judges endogenously choose effort provision, but challenges themselves are exogenous and information is symmetric. Neither paper considers fees as a policy instrument. What distinguishes our article is that we provide an integrated framework bringing together all relevant instruments—examination, fees, and challenges—and that we assess the welfare effects of patent policies quantitatively as well as qualitatively.

Finally, a number of recent papers in the macro literature adapt general-equilibrium, endogenous-growth models to study the welfare effects of innovation-related policies and institutions. Like our paper, they use micro evidence to calibrate the models and conduct counterfactual policy evaluation. Leading examples include Acemoglu *et al.* (2018), Akcigit *et al.* (2016), and Akcigit *et al.* (2021). While our article is a partial-equilibrium analysis, we share with this literature the objective of accounting for interlinkages between different components of the system of production and diffusion of innovation. In principle, one could consider embedding a stylized version of the patent screening process, which is the focus of this article, into an endogenous-growth framework to assess its impact on growth and welfare.

The article is organized as follows. Section 2 presents our model of patent screening. Section 3 derives the equilibrium and theoretically studies the welfare effects of patent policy. Section 4 sets up the model used for the quantitative analysis, describes its calibration, and presents the empirical results. Section 5 presents the counterfactual policy reforms. Section 6 concludes. All proofs are relegated to Appendix A.

2. A MODEL OF PATENT SCREENING

In the spirit of the ideas model of innovation developed by Scotchmer (2004), we consider an inventor endowed with an idea (v,κ), where κ is the R&D cost required to turn the idea into an invention and v is the social value of the invention. The social value v is observed by both the inventor and a competitor active in the same industry. The R&D cost κ is the inventor's private information. It is common knowledge that v is drawn from a distribution $F(\cdot)$ on $[\underline{v},\overline{v}]$ and that κ is drawn from a distribution $G_v(\cdot)$ on $[\underline{\kappa},\overline{\kappa}]$, both of which are continuously differentiable. Once an idea has been developed, the inventor can apply for a patent.[4]

Profits and welfare prior to invention are normalized to zero. If the invention is developed, profits and welfare depend on whether the inventor obtains a patent. In the absence of a patent, the inventor earns $\pi(v)\geq 0$ and welfare is $v\geq\pi(v)$ (both gross of R&D costs). Thus, $\pi(v)$ is a measure of the profit the inventor can appropriate without a patent. For this to be positive, competition must not be too fierce, or there must be an alternative appropriation mechanism such as lead time. The quantitative analysis in Section 4 uses imperfect competition.

2.1. *Licensing game*

To determine profits and welfare *with a patent*, we consider a licensing game between the inventor and competitor. The competitor observes v but not κ; assuming that the competitor receives a signal about κ would not change the qualitative results as long as the signal is imperfect, so that the inventor has better information than the competitor. To fix ideas, assume that the invention covers a cost-reducing technology and that, absent a patent, the competitor can copy the invention and earn $\pi(v)$.[5]

The timing of the licensing game is as follows. The inventor makes a take-it-or-leave-it offer to the competitor to license the invention through a royalty scheme that transfers an amount R from the competitor to the inventor (for brevity, we hereafter refer to R as the license fee). This should be seen as a reduced form for an explicit model of product-market competition in which the licensing contract takes the form of a two-part tariff.[6] The competitor decides whether to accept or reject; if she rejects, she further decides whether to challenge the patent in court, causing litigation costs $l_C(v)$ to the competitor and $l_I(v)$ to the inventor. The court either upholds or revokes the patent. If the patent is upheld, the inventor can offer a new license contract to the competitor; if the patent is revoked the competitor can use the invention for free.[7]

The payoff structure when the inventor holds a patent is as follows. Without a license agreement, only the inventor can use the new cost-reducing technology while the competitor

4. We assume that patent applications can only be submitted on inventions, not ideas. The basic patent ineligibility of abstract ideas was recently affirmed by the U.S. Supreme Court in *Bilski v. Kappos* (561 U.S. 593 (2010)).

5. Symmetry in profits is not important; what matters is that the competitor is made worse off by the inventor owning a patent.

6. Specifically, letting τ denote the fixed fee and ρ the per-unit royalty in the two-part tariff, if $\rho=0$ we have $R=\tau$. For $\rho>0$, the profit that is transferred from the competitor to the inventor, and thus the expression for R, depends on equilibrium behaviour. The inventor will choose ρ to maximize the firms' joint profits. See Section 4.1.

7. In practice, most validity challenges are filed as a response to an infringement suit. Our model abstracts from infringement suits and instead focuses on validity challenges. This implies that the competitor cannot use the patented invention without either taking a license or successfully challenging the validity of the patent. Incorporating infringement suits into the model would not fundamentally change the analysis as long as the competitor incurs a cost if she decides to challenge.

has to use the backstop technology. Assume that the invention generates a competitive advantage of $\Delta_I(v)>0$ for the inventor and a competitive disadvantage of $\Delta_C(v)>0$ for the competitor. That is, in the absence of a license agreement the inventor earns $\pi(v)+\Delta_I(v)$ while the competitor earns $\pi(v)-\Delta_C(v)$, which constitutes her outside option in the license negotiations.[8]

With a license agreement in place, both firms can use the invention in production. Moreover, the firms are able to (jointly) exercise market power, for example by using royalties to soften competition. We capture this in reduced form by assuming that the inventor earns $\pi(v)+m(v)+R$ and the competitor $\pi(v)-R$, where $m(v)\geq 0$ is the extra profit due to market power. That is, $m(v)$ is the difference in industry profit with a royalty-based license $(2\pi(v)+m(v))$ as compared to industry profit without the royalty $(2\pi(v))$. Welfare becomes $v-D(v)$, where $D(v)$ is the social cost of patents, including the deadweight loss from monopoly pricing. This setup captures in the simplest possible way the classic tradeoff whereby patents increase the incentives to innovate but at the same time create social cost.

We assume that $m(v)+\Delta_C(v)>\Delta_I(v)$, which ensures that the inventor prefers to license his invention rather than exclude the competitor and compete with asymmetric costs.[9] Let $\Delta^H(v)\equiv m(v)+\Delta_C(v)$ denote the extra profit the inventor earns thanks to the patent when R is set at the highest level the competitor would ever be willing to accept, namely, $R=\Delta_C(v)$.

Assumption 1. *$v-D(v)>\pi(v)+\Delta^H(v)$ for all v.*

This assumption says that the social returns exceed the private returns to R&D, which is consistent with the evidence in Bloom *et al.* (2013).

2.2. *Patentability*

An important modelling choice concerns the requirements for an invention to be patentable. We impose the patentability requirement $\kappa>\pi(v)$, *i.e.*, R&D costs must exceed the profits the inventor can appropriate without a patent. This requirement can be justified on both normative and descriptive grounds. From a normative perspective, this is the patentability requirement that a social planner would choose in our setup. The planner wants to give patents only to those inventions that would not be developed absent the patent incentive. Inventions with $\kappa\leq\pi(v)$ would be developed anyway, so giving them patents causes unnecessary social costs. (This argument abstracts from the potential benefits of disclosure.) The planner also does not want to give patents to ideas whose development would be privately valuable but socially harmful, *i.e.*, ideas such that $\pi(v)<\kappa\leq\pi(v)+\Delta^H(v)$ and $\kappa>v-D(v)$, but Assumption 1 rules this case out.

From a descriptive perspective, although statutory patent law makes no explicit reference to R&D costs and profits, this is likely because what matters is not *realized* costs and profits, but *expected* costs and profits at the time of investment, and these cannot be readily observed. Instead, lawmakers and courts are forced to resort to alternative standards that rely on empirical proxies to determine whether or not an invention would have been developed absent a patent. The most important such standard—non-obviousness—has been rationalized by courts and legal scholars in exactly this way. For an invention to meet the non-obviousness standard, it must not have been obvious to a "person having ordinary skill in the art" (35 US Code §103). In the landmark case of *Graham v. John Deere* (383 U.S. 1 (1966)) the U.S. Supreme Court stated: "The inherent problem

8. We do not specify welfare for this case because, under our assumptions, all patents will end up being licensed in equilibrium, except for those that are challenged and invalidated.

9. In a homogeneous-good Cournot model, which is what we use for the quantitative analysis in Section 4, a sufficient condition for this is that the invention is non-drastic.

was to develop some means of weeding out those inventions which would not be disclosed or devised but for the inducement of a patent." This rationale for non-obviousness was reaffirmed in *KSR v. Teleflex* (550 U.S. 398 (2007)), where the court held unpatentable "advances that would occur in the ordinary course." To draw the line, the court appealed to concepts closely related to the expected costs and benefits of an invention, such as predictability (implying less need for experimentation, and thus lower expected R&D costs *ceteris paribus*; see Merges, 1992) and the existence of clear market incentives to solve a problem. It is also the consensus view among legal scholars that non-obviousness is meant to distinguish inventions that would be made without patent protection from those that would not (Kitch, 1966; Eisenberg, 2004; Duffy, 2007; Meurer and Strandburg, 2008).[10]

It is important to make the distinction between what the courts are trying to achieve and how well they can achieve it with the legal standards they use as proxies. In our baseline model, these legal standards are assumed to be perfect proxies for the underlying economic standard of patentability. Later in the analysis, the legal standards are allowed to be imperfect proxies so that courts may make errors.

Assumption 2. *For all* $v \in (\underline{v}, \overline{v})$, $\underline{\kappa} < \pi(v)$, $\pi(v) + \Delta^H(v) < \overline{\kappa}$, *and* $g_v(\kappa) > 0$ *for* $\kappa \in (\underline{\kappa}, \overline{\kappa})$.

This assumption implies, in particular, that $0 < G_v(\pi(v)) < 1$ for all v. Thus, regardless of value, there is a fraction of ideas whose R&D cost exceeds $\pi(v)$ and a fraction whose R&D cost does not. In addition, it ensures that the cutoff on κ below which inventors invest in R&D is interior.

We will refer to inventors with ideas such that $\kappa > \pi(v)$ as type-H inventors or high types, and to those such that $\kappa \leq \pi(v)$ as type-L inventors or low types. Under our patentability criterion, type-H inventors should be given patents while type-L inventors should not.[11] The problem is that κ is privately observed by the inventor. Since low types also benefit from patent protection, society must put in place a screening mechanism. We model this mechanism as involving two stages: patent office examination and validity challenges in the courts.

2.3. *Obtaining patents*

To obtain a patent, the inventor must submit an application to the patent office and pay a *pre-grant fee* $\phi_A \geq 0$. The patent office then examines the application. We assume that type-H inventions always pass the examination, while type-L inventions pass the examination only with probability $1-e$, where $e \in [0,1)$ represents the patent office's examination intensity; with probability e type-L inventions are detected and refused patent protection.[12,13] Inventions that pass the examination

10. Economists have provided other rationales for non-obviousness; see Scotchmer and Green (1990), Scotchmer (1996), O'Donoghue (1998), Hunt (2004), and Kou *et al.* (2013).

11. This is not the same as saying that patents should be given to high social-value inventions, since both low and high-type inventions can be of either high or low (private and social) value. One possible objection to our patentability requirement is that, in an alternative environment where the inventor has multiple ideas but can pursue only one of them, it may encourage high-cost inventions. However, even in such an environment, our patentability requirement can be justified when inventors do not observe invention values. In that case, R&D cost can be a signal of value, and the planner may want to promote "ambitious" research. For a related argument, see O'Donoghue (1998).

12. Thus, the patent office is assumed to make only type II errors and no type I errors. This can be justified by the fact that rejections require examiners to come up with evidence showing that the invention is not patentable, whereas allowances do not. In addition, a relatively inexpensive appeals procedure is available to applicants whose application is rejected, so type I errors are likely to get corrected. Our results do not rely on this assumption.

13. The implicit assumption that the probability of detection depends only on the sign of $\pi(v) - \kappa$ and not on its magnitude keeps the model tractable for the licensing and challenge stages that follow.

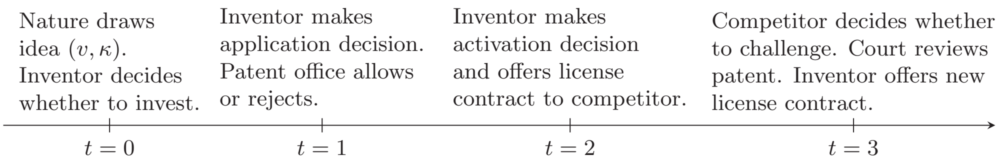

FIGURE 1
Timing of the game

must pay a *post-grant fee* $\phi_P \geq 0$ in order to be issued a patent. This payment thus occurs after the patent office has decided whether to allow or reject the application, and has to be paid only in case of allowance. We will refer to payment of ϕ_P as the inventor *activating* the patent. One can think of the post-grant fee ϕ_P as a renewal fee paid in lump sum, whereby the inventor chooses to maintain his patent for some duration in exchange for payment of ϕ_P. If the inventor does not apply, does not pass the examination, or does not pay ϕ_P, the invention falls into the public domain.

The cost of examining a patent application with intensity e is $\gamma(e)$, which we take to be increasing and convex and satisfy $\lim_{e\to 1}\gamma'(e) = \infty$. We refer to a combination of examination intensity, pre-grant fee, and post-grant fee $(e, \phi_A, \phi_P) \in [0,1) \times \mathbb{R}^2_+$ as a *patent policy*.

2.4. *Court challenges*

We model courts as differing from the patent office in two ways. First, while the patent office examines all applications, the courts only review patents that are challenged. Second, the courts have a different propensity to make mistakes than the patent office. To begin, we adopt the simplifying assumption that the courts do not make any mistakes: they always uphold type-H patents and revoke type-L patents.[14] In Section 3.2.3 and Supplementary Appendix C, we examine a generalization of the model in which courts can make mistakes.

2.5. *Timing*

Figure 1 summarizes the timing of the game, from the investment stage to the licensing stage. The game we study is a signalling game in which the inventor's decisions to apply for and activate a patent, as well as which license fee to propose, can convey information about his type to the competitor. Our solution concept is Perfect Bayesian Equilibrium. To deal with multiplicity, we use the D1 refinement.

3. THEORETICAL ANALYSIS

We start with the simpler case in which there is a single invention value and ideas differ only in R&D cost (Section 3.1). We then move to the full-fledged model in which ideas differ in both value and R&D cost (Section 3.2). The results obtained for the different ranges of invention value considered in Section 3.1 form the basis of the analysis in Section 3.2.

14. Court cases have more time to gather evidence and hear arguments than the typical time patent examiners can allot to screening, and also have the patent office's prior art search as input. Thus there is reason to believe that courts (especially bench trials) may be more accurate.

3.1. *Homogeneous invention values*

Consider an invention of value v. We first examine the case where a patent on the invention is not worth litigating, $\Delta_C(v) < l_C(v)$, so that challenging the patent is not a credible threat regardless of the competitor's beliefs, and then turn to the case where patents are worth litigating, $\Delta_C(v) \geq l_C(v)$, so that challenges are potentially credible, depending on the competitor's beliefs. To frame these two cases informally, patents worth litigating correspond to relatively high-value inventions, while patents not worth litigating correspond to relatively low-value inventions. This relationship between value and litigation emerges under plausible conditions that we specify formally in Section 3.2, where we analyze heterogeneous values.

3.1.1. Patents not worth litigating. Suppose $\Delta_C(v) < l_C(v)$, so that challenges are never credible. Then, the competitor accepts any license fee $R \leq \Delta_C(v)$, and both types of inventors propose $R = \Delta_C(v)$. An inventor of type H invests in R&D, applies for a patent, and activates the patent if and only if

$$\kappa \leq \pi(v) + \Delta^H(v) - \phi_P - \phi_A \equiv \Pi_{nc}(v). \tag{1}$$

An inventor of type L always invests as $\kappa \leq \pi(v)$ by definition. Deterring him from applying for a patent requires

$$(1-e)(\Delta^H(v) - \phi_P) - \phi_A \leq 0. \tag{2}$$

Note that this condition does not depend on κ.

Consider a patent policy (e, ϕ_A, ϕ_P). The following proposition combines (1) and (2) to examine how patent policy affects innovation by high types and deterrence of low types.

Proposition 1. *When patents are not worth litigating:*

(i) Deterrence of low types without suppressing high-type innovation cannot be achieved if either $e = 0$ or $\phi_A = 0$.
(ii) Conditional on deterrence of low types, the maximum level of high-type innovation that can be achieved is increasing in e and decreasing in ϕ_P.
(iii) Fixing the sum of fees, and thus the level of high-type innovation, the examination intensity required to achieve deterrence of low types is increasing in ϕ_P.

Since, by definition, for high types $\kappa > \pi(v)$, high-type innovation is suppressed when $\Pi_{nc}(v)$ falls below $\pi(v)$. Result (i) of Proposition 1 says that the set of policy combinations that achieve deterrence without suppressing high-type innovation, given by

$$S_{nc} = \{(e, \phi_A, \phi_P) \in [0,1) \times \mathbb{R}^2_+ : (1-e)(\Delta^H(v) - \phi_P) \leq \phi_A < \Delta^H(v) - \phi_P\}, \tag{3}$$

obtained by combining (1) and (2), includes only policies with strictly positive pre-grant fee and strictly positive examination intensity. The intuition is related to the fact that, conditional on passing examination, a patent is worth the same to high and low types (namely, $\Delta^H(v)$), but high types have higher R&D costs. This implies that if $e = 0$, low types' payoff from applying strictly exceeds high types' for any (ϕ_A, ϕ_P), so setting fees high enough to deter low types also deters high types. Similarly, if $\phi_A = 0$, low types will apply regardless of e as long as $\phi_P \leq \Delta^H(v)$; higher post-grant fees also deter high types.

Result (ii) highlights a tradeoff between deterrence and innovation. Patent fees hamper high-type innovation but at the same time facilitate deterrence. Patent policies that involve higher examination intensity can accommodate lower fees without jeopardizing deterrence, and are thus associated with more high-type innovation.

Results (ii) and (iii) both reflect the insight that pre-grant fees are a more effective screening device than post-grant fees. This is because low types prefer fees to be backloaded more strongly than high types. Keeping the sum of pre-grant and post-grant fees constant at $\phi_A+\phi_P=\phi$, type H is indifferent over all combinations of fees, whereas type L's expected total fee payment, $\phi_A+(1-e)\phi_P=\phi_A+(1-e)(\phi-\phi_A)$, is increasing in ϕ_A. Intuitively, pre-grant fees must be paid whether or not the application passes examination, whereas post-grant fees are only paid conditional on passing examination.[15]

Welfare. We now derive expressions for expected welfare as a function of patent policy, which we will use later in the analysis. There are three cases. If $\phi_A \geq \Delta^H(v)-\phi_P$, high-type innovation is suppressed, and neither high nor low types apply for patents. Low types still invest, so welfare is

$$w^0(v)=\int_{\underline{\kappa}}^{\pi(v)}(v-\kappa)dG_v(\kappa).$$

If $(1-e)(\Delta^H(v)-\phi_P)\leq\phi_A<\Delta^H(v)-\phi_P$, low types are deterred but some high-type innovation occurs, so welfare is

$$w^d(v)=\int_{\underline{\kappa}}^{\pi(v)}(v-\kappa)dG_v(\kappa)+\int_{\pi(v)}^{\Pi_{nc}(v)}(v-D(v)-\kappa-\gamma(e))dG_v(\kappa).$$

If $\phi_A<(1-e)(\Delta^H(v)-\phi_P)$, low types apply, and welfare is

$$w^{nd}(v)=\int_{\underline{\kappa}}^{\pi(v)}(v-(1-e)D(v)-\kappa-\gamma(e))dG_v(\kappa)+\int_{\pi(v)}^{\Pi_{nc}(v)}(v-D(v)-\kappa-\gamma(e))dG_v(\kappa).$$

3.1.2. Patents worth litigating. Now suppose $\Delta_C(v)\geq l_C(v)$, so that challenges are potentially credible. To derive the condition for challenge credibility, suppose high-type inventors charge a license fee $R^H\leq\Delta_C(v)$ and expect the competitor to challenge the patent with probability x. Moreover, suppose low-type inventors mimic high types. Then the competitor's belief that the inventor is of type H is

$$\underline{\lambda}\equiv\frac{G_v(\pi(v)+(1-x)(R^H+m(v))+x(\Delta^H(v)-l_I(v))-\phi_A-\phi_P)-G_v(\pi(v))}{G_v(\pi(v)+(1-x)(R^H+m(v))+x(\Delta^H(v)-l_I(v))-\phi_A-\phi_P)-eG_v(\pi(v))}.$$

The numerator is the probability that the inventor is issued a patent and is of high type, i.e., that κ is greater than $\pi(v)$ and less than $\pi(v)+(1-x)(R^H+m(v))+x(\Delta^H(v)-l_I(v))-\phi_A-\phi_P$, which is the cutoff below which high types invest. The denominator is the probability that the inventor is issued a patent being of either high or low type, which equals the numerator plus the probability that κ is less than $\pi(v)$ times the probability that low types pass examination, $1-e$.

Given any belief λ, the competitor prefers challenging to not challenging if and only if $(1-\lambda)R^H-\lambda(\Delta_C(v)-R^H)\geq l_C(v)$: the benefit from a challenge, given by the expected change in the

15. In a more general setup in which type-H inventors also face a risk of rejection, both types prefer fees to be backloaded, but as long as a type-L inventor is less likely to pass examination he still prefers backloading more strongly than a type-H inventor.

license fee (a decrease by R^H if the patent is revoked; an increase by $\Delta_C(v)-R^H$ if the patent is upheld), must exceed the cost of litigation. Because $\underline{\lambda}$ is based on *all* low types mimicking high types, it is the lowest belief the competitor can hold.[16] Hence, the condition for challenges to be credible is

$$\left(1-\underline{\lambda}\right)R^H-\underline{\lambda}(\Delta_C(v)-R^H)\geq l_C(v). \tag{4}$$

Since $\Delta_C(v)\geq l_C(v)$, (4) holds for some patent policy (e,ϕ_A,ϕ_P) when $R^H=\Delta_C(v)$.[17] Because low types are less likely to survive a challenge than high types, holding a patent is worth less to low types than to high types for any given probability of being challenged. One might thus expect that challenge credibility would make deterrence easier. In particular, one might expect that it would be possible to deter low types even in the absence of patent examination, by setting (pre-grant or post-grant) fees just above the expected value of a patent to low types. As the following proposition shows, however, this is not the case.

Proposition 2. *The set of patent policies that achieve deterrence of low types without suppressing innovation by high types when patents are worth litigating is contained in the set of policies that do so when patents are not worth litigating, S_{nc}.*

The intuition is simple. The competitor is Bayesian and updates her beliefs based on the inventor's equilibrium strategy. If patent policy induces sorting, so that low types are deterred but (some) high types are not, then it takes away the competitor's incentive to challenge, because with perfect courts, any challenge would fail. It follows that challenge credibility cannot help achieve sorting, highlighting the importance of modelling challenges as endogenous. In particular, Proposition 2 implies that the result from Proposition 1—that policies with $e=0$ or $\phi_A=0$ cannot achieve sorting—carries over to the case where patents are worth litigating. The condition for deterrence is exactly the same as when patents are not worth litigating, namely (2).

Although challenge credibility has no effect on the conditions for deterrence, it *does* change the equilibrium in the absence of deterrence, which we now derive. In particular, challenge credibility can reduce the license fees that low types ask for. Fix a patent policy (e,ϕ_A,ϕ_P) for which challenges are credible. To streamline the exposition, we restrict attention to patent policies such that $\phi_A<(1-e)(\Delta^L(v)-\phi_P)$, where $\Delta^L(v)\equiv l_C(v)+m(v)$.[18] (Because challenge credibility requires $l_C(v)\leq\Delta_C(v)$, this implies that the patent policy does not deter low types.) As the following proposition shows, the equilibrium is semi-separating: high-type inventors charge high license fees, low-type inventors randomize between high and low license fees, and the competitor randomizes over the challenge decision when observing high license fees.[19]

Proposition 3. *Consider a patent policy (e,ϕ_A,ϕ_P) such that (4) holds for $R^H=\Delta_C(v)$ and $x=0$, and suppose $\phi_A<(1-e)(\Delta^L(v)-\phi_P)$. In the unique equilibrium,*

(i) high-type inventors invest if and only if $\kappa\leq\pi(v)+\Delta^H(v)-\tilde{x}(v)l_I(v)-\phi_A-\phi_P\equiv\Pi_{cc}(v)$, where $\tilde{x}(v)$ is defined below; if they invest, they apply, activate, and propose $R=\Delta_C(v)$;

16. This applies for a given x. There is an interdependence between the competitor's belief $\underline{\lambda}$ and the rate of challenges x, which we discuss in Section 3.2.

17. For $\phi_A=0$, and $e\in[0,1)$, $\underline{\lambda}\to 0$ as $\phi_P\to\Delta^H(v)$.

18. As we argue below, this is the empirically relevant range of patent policies. Supplementary Appendix C derives the equilibrium when this condition does not hold.

19. The equilibrium is unique up to the randomization strategies of low types. See the proof of Proposition 9 in Supplementary Appendix C.

(ii) low-type inventors always invest, apply and activate; they propose

$$R=\begin{cases}\Delta_C(v) & \textit{with probability } \tilde{y}(v)\\ l_C(v) & \textit{with probability } 1-\tilde{y}(v),\end{cases}$$

where

$$\tilde{y}(v)\equiv\left(\frac{l_C(v)}{\Delta_C(v)-l_C(v)}\right)\left(\frac{G_v(\Pi_{cc}(v))-G_v(\pi(v))}{(1-e)G_v(\pi(v))}\right); \tag{5}$$

(iii) the competitor never challenges the patent if offered $R=l_C(v)$ and challenges with probability $\tilde{x}(v)$ if offered $R=\Delta_C(v)$, where

$$\tilde{x}(v)\equiv\frac{\Delta_C(v)-l_C(v)}{\Delta^H(v)+l_I(v)}.$$

In equilibrium, high types charge the same license fee as when challenges are not credible, namely $R=\Delta_C(v)$, which triggers a challenge with probability $\tilde{x}(v)$. High types invest if and only if

$$\kappa\le\pi(v)+\Delta^H(v)-\tilde{x}(v)l_I(v)-\phi_A-\phi_P=\Pi_{cc}(v). \tag{6}$$

Low types sometimes reveal themselves by charging a low license fee $l_C(v)$, which prevents the competitor from challenging even though she knows she would win for sure. At other times, low types instead mimic high types by charging a high license fee $\Delta_C(v)$. The probability of challenges $\tilde{x}(v)$ is such that low types are indifferent between high and low fees. Hence, low types' equilibrium payoff is $\pi(v)-\kappa+(1-e)(\Delta^L(v)-\phi_P)-\phi_A$.

The rate at which low types charge the high license fee, $\tilde{y}(v)$, is chosen so as to make the competitor indifferent between challenging and not challenging. Formally, $\tilde{y}(v)$ solves $(1-\tilde{\lambda})\Delta_C(v)=l_C(v)$, where

$$\tilde{\lambda}=\frac{G_v(\Pi_{cc}(v))-G_v(\pi(v))}{G_v(\Pi_{cc}(v))-G_v(\pi(v))+(1-e)\tilde{y}(v)G_v(\pi(v))}$$

is the competitor's belief that the patent is of high type conditional on observing the high fee. Note that the rate of challenges $\tilde{x}$ does not depend on patent policy. Hence, patent policy affects high types' profit only through the sum of fees, $\phi_A+\phi_P$.

Corollary. *The courts do not eliminate all low-type patents in equilibrium.*

The corollary is a direct implication of Proposition 3 and is again driven by the endogeneity of challenges. Even though courts are perfect at discriminating between high and low types, endogenously some patents are not litigated. Low types sometimes preempt challenges by charging a low license fee (equal to the competitor's litigation costs). And even when low types charge high license fees, the competitor challenges only a fraction of them.

These results underline the importance of the patent office. The distinctive feature of patent office review is that all applications are examined. By contrast, court review only occurs if the competitor challenges, and that depends on the inventor's equilibrium strategy. This is the fundamental drawback of a registration system, relying entirely on the courts for screening.[20]

20. We implicitly assume that the patent office can commit to examining all patent applications, even if in equilibrium no low types apply. This assumption can be justified by appealing to the richer setting we analyse in Section 3.2, where inventions differ in value. In that setting, it is difficult to completely deter low-type inventors, as some have very valuable inventions. Under the plausible assumption that the patent office does not observe the value of an invention, from the examiner's perspective there will thus always be a strictly positive probability of the application being a low type. By contrast, the richer setting leaves our results on private challenges unaffected.

Welfare. Welfare in the equilibrium of Proposition 3 is

$$w^{cc}(v)=\int_{\underline{\kappa}}^{\pi(v)}(v-\kappa-(1-e)[D(v)+\tilde{x}(v)\tilde{y}(v)(l_C(v)+l_I(v)-D(v))]-\gamma(e))dG_v(\kappa)$$
$$+\int_{\pi(v)}^{\Pi_{cc}(v)}(v-\kappa-D(v)-\tilde{x}(v)(l_C(v)+l_I(v))-\gamma(e))dG_v(\kappa),$$

where the first line is welfare from low types and the second line welfare from high types. Regrouping terms differently, we can rewrite this as

$$w^{cc}(v)=\int_{\underline{\kappa}}^{\Pi_{cc}(v)}(v-D(v)-\kappa)dG_v(\kappa)+G_v(\pi(v))eD(v)-G_v(\Pi_{cc}(v))\gamma(e)+\chi(v), \tag{7}$$

where

$$\chi(v)\equiv(1-e)G_v(\pi(v))\tilde{x}(v)\tilde{y}(v)[D(v)-l_C(v)-l_I(v)]$$
$$-(G_v(\Pi_{cc}(v))-G_v(\pi(v)))\tilde{x}(v)(l_C(v)+l_I(v)). \tag{8}$$

The first term in (7) is welfare before examination and challenges, the second is the deadweight loss avoided thanks to *patent examination*, the third is the cost of examination, and χ is the deadweight loss avoided thanks to *challenges*, net of the associated litigation costs. In what follows, we will refer to χ as the *net benefits from challenges*.[21] Whether these are positive or negative is *a priori* ambiguous. On the one hand, challenges help rid society of low-type patents, which raises welfare (provided deadweight loss exceeds litigation costs). On the other hand, challenges create wasteful litigation of high-type patents.

The next proposition examines the sign of χ under the conditions of the equilibrium in Proposition 3.

Proposition 4. *Consider (e,ϕ_A,ϕ_P) such that (4) holds for $R^H=\Delta_C(v)$ and $x=0$, and suppose $\phi_A<(1-e)(\Delta^L(v)-\phi_P)$. Then,*

$$\chi(v)=(G_v(\Pi_{cc}(v))-G_v(\pi(v)))\left(\frac{l_C(v)D(v)-(l_C(v)+l_I(v))\Delta_C(v)}{\Delta^H(v)+l_I(v)}\right).$$

The net benefits from challenges are positive if and only if

$$\Delta_C(v)\leq\left(\frac{l_C(v)}{l_C(v)+l_I(v)}\right)D(v). \tag{9}$$

This proposition reinforces the previous results about the perils of relying on the courts for screening. Challenges may do more harm than good: if (9) does not hold, the net benefits from challenges are negative. This condition can be understood by examining the private and social

21. Note that χ captures only the *ex post* welfare effects of challenges (after the patent is granted) and not the *ex ante* effects that operate through investment and application decisions.

incentives to initiate a challenge. Society is better off challenging than not challenging, given a posterior λ that the inventor is of high type, if and only if

$$(1-\lambda)D(v) \geq l_C(v)+l_I(v). \tag{10}$$

The left-hand side is the expected social benefit from a challenge, given by the probability of invalidation times the deadweight loss that is saved if the patent is invalidated; the right-hand side is the social cost, equal to the sum of litigation costs. Privately, the competitor finds it optimal to challenge if and only if $(1-\lambda)\Delta_C(v) \geq l_C(v)$. In equilibrium, $\lambda=\tilde{\lambda}$ is such that expected private benefits and costs of a challenge are equalized, as the competitor must be indifferent between challenging and not when the inventor charges the high license fee. Evaluating (10) at $\lambda=\tilde{\lambda}=(\Delta_C(v)-l_C(v))/\Delta_C(v)$ yields (9).

Are the incentives to challenge likely to be excessive or insufficient? In a homogeneous-good Cournot model with linear demand and constant marginal cost (which is what we use for the quantitative analysis in Section 4), we have $\Delta_C(v)>D(v)$, *i.e.*, the competitive disadvantage that the patent inflicts on the competitor is greater than the deadweight loss it causes (see Supplementary Appendix F). This alone is enough to imply that condition (9) is violated, and thus that challenges are socially excessive.[22] We conjecture that the condition holds much more generally, though, because there is a second force that also works in the direction of excessive challenges: the competitor does not take into account the litigation cost $l_I(v)$ it imposes on the inventor (and society) by challenging.

It is perhaps surprising that the result in Proposition 4 does not depend on examination intensity. In particular, one might have expected challenges to be more beneficial when e is low. Inspection of (8) reveals that, for a given $\tilde{y}$, the deadweight loss eliminated by challenges is indeed larger when e is small. It turns out, however, that in equilibrium $\tilde{y}$ adjusts to exactly cancel out the effect of a change in e. Formally, this is because $1-e$ is in the denominator of $\tilde{y}$; see (5). Intuitively, as e increases and more low types are eliminated by patent office examination, the rate at which low types mimic high types $\tilde{y}$ increases to keep the competitor's beliefs constant at $\tilde{\lambda}=(\Delta_C(v)-l_C(v))/\Delta_C(v)$. Notice also that a proportional change in litigation costs does not affect (9). This implies in particular that a policy reducing litigation costs (such as the introduction of a cheaper administrative procedure for patent challenges) does not change the *sign* of the net benefits from challenges.

3.2. *Heterogeneous invention values*

To study how the design of patent policy affects innovation and welfare in a richer and more general environment, we now extend the analysis of the previous subsection to the case where inventions differ in value as well as R&D costs. Introducing both dimensions of heterogeneity enables us to account for the impact of patent policy on the mix of inventions that are developed and litigated.

22. The assumption that there is a single competitor plays a role here. The conventional view, derived from symmetric-information models, is that introducing multiple competitors generates free riding and thus too few challenges. Note, however, that it is not obvious how the presence of multiple competitors would change the condition in our asymmetric-information model. On the one hand, an individual competitor's benefit from invalidation is smaller. On the other hand, the cost of a challenge may also be smaller. To see this, consider an equilibrium in which all competitors randomize over the challenge decision, and suppose that, if several of them challenge the patent, one of them is drawn at random (say, whoever arrives at the courthouse first). Then, a competitor who challenges only incurs the litigation cost with some probability.

Assume that all relevant functions of v (namely, $\pi(v)$, $\Delta_C(v)$, $m(v)$, $l_C(v)$, $l_I(v)$, and $D(v)$) are continuously differentiable and strictly increasing. Then, there exist two unique cutoffs, v^* and $\hat{v}$, defined by

$$\Delta^H(v^*) = \phi_A + \phi_P \tag{11}$$

$$\Delta^H(\hat{v}) = \frac{\phi_A}{1-e} + \phi_P, \tag{12}$$

with $v^* \leq \hat{v}$, such that no high types apply for $v \leq v^*$ and no low types apply for $v \leq \hat{v}$.[23]

To begin, we show that, under a mild condition, a well-designed patent system is socially desirable.

Proposition 5. *If $\Delta^H(\overline{v}) > \gamma(0)$, there is a non-empty set of patent policies for which welfare with a patent system strictly exceeds welfare without a patent system.*

This does not mean that *any* patent system is better than no patent system. In the quantitative analysis in Section 4, we show that a patent system that is poorly designed (one with inadequate screening) can be worse than no patent system at all.

The idea of the proof of Proposition 5 is as follows. In the absence of a patent system, there is no high-type innovation. Consider a patent policy constructed in such a way that no low types apply, namely, $(e, \phi_A, \phi_P) = (\varepsilon, (1-\varepsilon)\Delta^H(\overline{v}), 0)$, with $\varepsilon \in [0,1)$. Note that challenges are not credible, and for $\varepsilon > 0$, some high types invest. Thus, it suffices to show that for an appropriately chosen $\varepsilon > 0$, the welfare gains from high-type innovation are larger than the cost of examining their applications. Now choose ε such that $(1-\varepsilon)\Delta^H(\overline{v}) = \gamma(\varepsilon)$. Because high types invest if and only if $\kappa \leq \pi(v) + \Delta^H(v) - (1-\varepsilon)\Delta^H(\overline{v})$, the welfare gains for all those that invest satisfy $v - D(v) - \kappa - \gamma(\varepsilon) \geq v - D(v) - \pi(v) - \Delta^H(v)$, which—because social returns exceed private returns by Assumption 1—is positive. Intuitively, the fee is chosen in such a way that only those high types apply for which the welfare gains outweigh the cost of examination.[24]

We now study how equilibrium behaviour depends on patent policy. As a first step, we derive a cutoff on v for challenges to be credible. The lowest belief the competitor can hold is

$$\underline{\lambda}(v) = \frac{G_v(\Pi_{cc}(v)) - G_v(\pi(v))}{G_v(\Pi_{cc}(v)) - eG_v(\pi(v))}. \tag{13}$$

Using $R^H = \Delta_C(v)$, the condition for challenge credibility becomes

$$1 - \underline{\lambda}(v) \geq \frac{l_C(v)}{\Delta_C(v)}. \tag{14}$$

Assumption 3. *$\Delta_C(v)$ and $l_C(v)$ satisfy:*

23. We are applying the convention that $v^* = \underline{v}$ if $\Delta^H(\underline{v}) > \phi_A + \phi_P$ and $v^* = \overline{v}$ if $\Delta^H(\overline{v}) < \phi_A + \phi_P$, and similarly for the other cutoffs we define.

24. It is worth pointing out that this result relies on both instruments—fees and examination intensity—being suitably combined. This contrasts with Schuett (2013b), where the patent system is socially desirable even if fees are the only available instrument (*i.e.*, a registration system in which fees are optimally chosen is better than no patent system). The reason for the difference is that, in Schuett (2013b), it is assumed that the probability of challenges is exogenous, so that sorting is possible even if $e = 0$.

(i) $\Delta_C(\underline{v}) < l_C(\underline{v})$
(ii) $\Delta_C(\overline{v}) > l_C(\overline{v})$
(iii) $l_C(v)/\Delta_C(v)$ *is strictly decreasing in* v.

Parts (i) and (ii) ensure that patents on sufficiently low-value inventions are not worth litigating while patents on sufficiently high-value inventions are. Part (iii) says that the right-hand side of (14) is decreasing. The next assumption ensures, as we show in Lemma 1, that the left-hand side of (14) is non-decreasing.

Assumption 4. *The distribution of R&D costs satisfies*

$$\frac{\Pi_{cc}' g_v(\Pi_{cc}) + \partial G_v(\Pi_{cc})/\partial v}{G_v(\Pi_{cc})} \leq \frac{\pi' g_v(\pi) + \partial G_v(\pi)/\partial v}{G_v(\pi)} \quad \textit{for all } v. \tag{15}$$

This assumption is more likely to hold if G_v is log-concave, so that g_v/G_v is decreasing. Log-concavity is a characteristic of most commonly used distributions (Bagnoli and Bergstrom, 2005). In Supplementary Appendix F, we show that Assumption 4 is satisfied when R&D costs are exponentially distributed, which is the formulation we use for the quantitative analysis in Section 4.

Lemma 1. *If Assumption 4 holds,* $\underline{\lambda}$ *is non-increasing in* v.

Lemma 1 implies that, under Assumptions 3 and 4, there exists a unique $v^{cc} \in (\underline{v}, \overline{v}]$, defined by $1 - \underline{\lambda}(v^{cc}) = l_C(v^{cc})/\Delta_C(v^{cc})$, such that challenges are credible only if $v > v^{cc}$. Note that $v > v^{cc}$ is a necessary but not sufficient condition for challenge credibility. The reason is that the competitor's beliefs depend on the inventor's expectations about the rate of challenges x. This naturally leads to multiplicity of equilibria for a range of invention values $v \in (v^{cc}, v^{nc})$, where $v^{nc} \geq v^{cc}$ is defined analogously to v^{cc}, replacing Π_{cc} by Π_{nc} in $\underline{\lambda}$. Within this range, if high types do not expect to be challenged, sufficiently many of them invest for challenges not to be credible, so $x = 0$ (fulfilling expectations). If high types expect to be challenged at rate $\tilde{x}(v) > 0$, sufficiently few of them invest for challenges to be credible, so $x = \tilde{x}(v)$ (again fulfilling expectations).[25] For ease of exposition, we select the equilibrium with $x = \tilde{x}(v)$ for $v > v^{cc}$. The following proposition describes this equilibrium.

Proposition 6. *Consider a patent policy* (e, ϕ_A, ϕ_P) *such that* $\phi_A < (1-e)(\Delta^L(v^{cc}) - \phi_P)$. *In equilibrium:*

(i) *for* $v \in [\underline{v}, v^*]$, *nobody applies;*
(ii) *for* $v \in (v^*, \hat{v}]$, *only high types apply, charging* $R = \Delta_C(v)$; *there are no challenges;*
(iii) *for* $v \in (\hat{v}, v^{cc}]$, *both low and high types apply, charging* $R = \Delta_C(v)$; *there are no challenges;*
(iv) *for* $v \in (v^{cc}, \overline{v}]$, *both low and high types apply; high types charge* $R = \Delta_C(v)$; *low types charge*

$$R = \begin{cases} \Delta_C(v) & \textit{with probability } \tilde{y}(v) \\ l_C(v) & \textit{with probability } 1 - \tilde{y}(v); \end{cases}$$

25. To be precise, there is a third equilibrium with an intermediate rate of challenges, but this equilibrium does not always exist; see Proposition 9 in Supplementary Appendix C.

the competitor never challenges the patent if offered $R = l_C(v)$ and challenges with probability $\tilde{x}(v)$ if offered $R = \Delta_C(v)$.

Low types always invest. High types invest if and only if $v > v^*$ and $\kappa \leq \Pi(v)$, where

$$\Pi(v) \equiv \begin{cases} \Pi_{nc}(v) & \text{for } v \leq v^{cc} \\ \Pi_{cc}(v) & \text{for } v > v^{cc}, \end{cases}$$

with $\Pi_{nc}(v) > \pi(v)$ for all $v > v^*$ and $\Pi_{cc}(v) > \pi(v)$ for all $v > v^{cc}$.

Recalling that $\Delta^L(v) = l_C(v) + m(v)$, the assumption that $\phi_A < (1-e)(\Delta^L(v^{cc}) - \phi_P)$ essentially says that the sum of examination-adjusted fees, $\phi_A/(1-e) + \phi_P$, should be smaller than the litigation costs of patents valuable enough to be challenged. By all estimates, even for relatively low-value patents, the cost of litigation runs into the hundreds of thousands of dollars, so this is the empirically relevant case. The assumption implies that the conditions of Proposition 3 are satisfied whenever challenges are credible, and hence that the equilibrium is as described in that proposition for all $v > v^{cc}$.[26]

Proposition 6 implies that, if $\phi_A < (1-e)(\Delta^L(v^{cc}) - \phi_P)$, the application decision of low-type inventors does not depend on the availability of court challenges. Put differently, challenges do not have any deterrence effect. This is because the condition that $\phi_A < (1-e)(\Delta^L(v^{cc}) - \phi_P)$ ensures that the payoff from the challenge-preempting license fee, Δ^L, which they can always guarantee themselves, is large enough to make applying worthwhile for all low types above v^{cc}, and thus also for a range of values below v^{cc}, where they are not exposed to challenges and their payoff is $(1-e)(\Delta^H(v) - \phi_P) - \phi_A$. This means that the cutoff for low types to apply is given by $\hat{v}$. The availability of challenges affects the license contracts low types propose but not their application behavior.[27]

Welfare. Given the cutoffs identified in Proposition 6, welfare is given by

$$W = \int_{\underline{v}}^{v^*} w^0(v)dF(v) + \int_{v^*}^{\hat{v}} w^d(v)dF(v) + \int_{\hat{v}}^{v^{cc}} w^{nd}(v)dF(v) + \int_{v^{cc}}^{\overline{v}} w^{cc}(v)dF(v).$$

After rearranging, the welfare function can be rewritten as

$$\begin{aligned} W = & \int_{\underline{v}}^{v^*} \int_{\underline{\kappa}}^{\pi(v)} (v-\kappa)dG_v(\kappa)dF(v) + \int_{v^*}^{\overline{v}} \int_{\underline{\kappa}}^{\Pi_{nc}(v)} (v - D(v) - \kappa - \gamma(e))dG_v(\kappa)dF(v) \\ & + \int_{v^*}^{\hat{v}} \int_{\underline{\kappa}}^{\pi(v)} (D(v)+\gamma(e))dG_v(\kappa)dF(v) + \int_{\hat{v}}^{\overline{v}} \int_{\underline{\kappa}}^{\pi(v)} eD(v)dG_v(\kappa)dF(v) \\ & + \int_{v^{cc}}^{\overline{v}} \left[\chi(v) - \int_{\Pi_{cc}(v)}^{\Pi_{nc}(v)} (v - D(v) - \kappa - \gamma(e))dG_v(\kappa) \right] dF(v). \end{aligned} \tag{16}$$

Patent policy affects both the cutoffs on v and κ (i.e., integration bounds) and the integrands.

26. Supplementary Appendix C describes the equilibrium when the assumption does not hold. The main difference is that, for some invention values, rather than randomize over high and low license fees, low types randomize over the decision whether to apply for a patent.

27. The absence of a deterrence effect from challenges is a result that holds more generally. Proposition 2 implies that even if $\phi_A \geq (1-e)(\Delta^L(v^{cc}) - \phi_P)$, challenges cannot achieve full deterrence for any v. As we show in Supplementary Appendix C, challenges can only achieve partial deterrence: the availability of challenges can lead some low types to randomize over the application decision where they would otherwise apply with probability one.

3.2.1. Patent policy tradeoffs. We now study the tradeoffs involved in setting a patent policy (e,ϕ_A,ϕ_P). The following proposition first shows that there is no tradeoff in choosing the structure of patent office fees: they should be frontloaded.

Proposition 7. *Starting from any patent policy* $(e,\phi_A,\phi_P)\in[0,1)\times\mathbb{R}_+^2$ *such that* $\phi_A+\phi_P>0$ *and* $\phi_A+\phi_P<(1-e)\Delta^L(v^{cc})$*, frontloading fees by switching to the policy* $(e,\phi_A+\phi_P,0)$ *raises welfare.*

The proof of Proposition 7 shows that, holding the sum of pre-grant and post-grant fees fixed, $\hat{v}$ is decreasing in ϕ_P while all other welfare components are invariant to ϕ_P. Recall that because the rate of challenges is independent of patent policy, high types' profit depends only on the sum of fees and not their distribution. By contrast, the low-type cutoff depends on $\phi_A/(1-e)+\phi_P$, so that shifting fees from post-grant to pre-grant amplifies their deterrence effect. As a result, frontloading raises the threshold above which low types apply, reducing deadweight loss and examination costs, without affecting anything else. This is a generalization of the result in Proposition 1 that pre-grant fees are a more effective screening device than post-grant fees.

It is worth noting that there are reasons outside the model that may justify post-grant fees. First, a schedule of renewal fees can ensure that high-value inventions, or high-productivity inventors, receive longer patents, which can be optimal under some conditions (Scotchmer, 1999; Cornelli and Schankerman, 1999). Second, liquidity-constrained firms would be hurt by frontloading, as they are likely to have better access to outside funding once the patent is granted.[28] Third, when inventors learn about the value of their inventions over time, renewal fees provide less of a disincentive to innovation than same-sized application fees, as they are only paid when the invention turns out to be sufficiently valuable. In Section 3.2.2, we show how frontloading can be adjusted to preserve its beneficial effects in the presence of uncertainty about value.

In light of Proposition 7, in what follows we focus on patent policies (e,ϕ_A,ϕ_P) with $\phi_P=0$ and $\phi_A=\phi\geq 0$. Starting from any patent policy such that $\Delta^H(\underline{v})<\phi<(1-e)\Delta^L(v^{cc})$, so that Proposition 6 applies and $\underline{v}<v^*\leq\hat{v}$, differentiating W in (16) with respect to e and ϕ, and rearranging, the welfare effects of marginal increases in examination intensity and (pre-grant) fees can be expressed as follows:

$$\frac{\partial W}{\partial e}=\overbrace{\int_{\hat{v}}^{\overline{v}}\int_{\underline{\kappa}}^{\pi(v)}D(v)dG_v(\kappa)dF(v)}^{\text{detection}}+\overbrace{\frac{\partial\hat{v}}{\partial e}\int_{\underline{\kappa}}^{\pi(\hat{v})}((1-e)D(\hat{v})+\gamma(e))dG_{\hat{v}}(\kappa)f(\hat{v})}^{\text{deterrence}}$$
$$+\overbrace{\frac{\partial v^{cc}}{\partial e}\left[\int_{\Pi_{cc}(v^{cc})}^{\Pi_{nc}(v^{cc})}(v^{cc}-D(v^{cc})-\kappa)dG_{v^{cc}}(\kappa)-\chi(v^{cc})\right]f(v^{cc})}^{\text{challenge discouragement}}$$
$$-\underbrace{\left[\gamma'(e)(A^L+A^H)+\frac{\partial v^{cc}}{\partial e}\int_{\Pi_{cc}(v^{cc})}^{\Pi_{nc}(v^{cc})}\gamma(e)dG_{v^{cc}}(\kappa)f(v^{cc})\right]}_{\text{examination cost increase}},\qquad(17)$$

28. For a theoretical analysis of how patent policy affects the financing problem of liquidity-constrained firms, see Schuett (2013a, Section VI). In practice the U.S. patent office addresses this concern by giving discounts to small firms.

where A^L and A^H denote the number of applications by low and high types, given by

$$A^L \equiv \int_{\hat{v}}^{\overline{v}} G_v(\pi(v))dF(v) \qquad A^H \equiv \int_{v^*}^{\overline{v}} (G_v(\Pi(v))-G_v(\pi(v)))dF(v),$$

and

$$\begin{aligned}
\frac{\partial W}{\partial \phi} = & \overbrace{\frac{\partial \hat{v}}{\partial \phi}\int_{\underline{\kappa}}^{\pi(\hat{v})}((1-e)D(\hat{v})+\gamma(e))dG_{\hat{v}}(\kappa)f(\hat{v})}^{\text{deterrence}} \\
& \underbrace{-\int_{v^*}^{v^{cc}}(v-D(v)-\Pi_{nc}(v))g_v(\Pi_{nc}(v))dF(v)-\int_{v^{cc}}^{\overline{v}}(v-D(v)-\Pi_{cc}(v))g_v(\Pi_{cc}(v))dF(v)}_{\text{innovation discouragement I}} \\
& \underbrace{-\frac{\partial v^*}{\partial \phi}\int_{\pi(v^*)}^{\Pi_{nc}(v^*)}(v^*-D(v^*)-\kappa)dG_v^*(\kappa)f(v^*)}_{\text{innovation discouragement II}} \\
& \underbrace{-\int_{v^{cc}}^{\overline{v}}\frac{l_C(v)D(v)-(l_C(v)+l_I(v))\Delta_C(v)}{\Delta^H(v)+l_I(v)}g_v(\Pi_{cc})dF(v)}_{\text{challenge discouragement (intensive margin)}} \\
& \underbrace{+\frac{\partial v^{cc}}{\partial \phi}\left[\int_{\Pi_{cc}(v^{cc})}^{\Pi_{nc}(v^{cc})}(v^{cc}-D(v^{cc})-\kappa)dG_{v^{cc}}(\kappa)-\chi(v^{cc})\right]f(v^{cc})}_{\text{challenge encouragement (extensive margin)}} \\
& \underbrace{+\int_{v^*}^{v^{cc}}\gamma(e)g_v(\Pi_{nc}(v))dF(v)+\int_{v^{cc}}^{\overline{v}}\gamma(e)g_v(\Pi_{cc}(v))dF(v)}_{\text{examination cost decrease I}} \\
& \underbrace{+\frac{\partial v^*}{\partial \phi}\int_{\pi(v^*)}^{\Pi_{nc}(v^*)}\gamma(e)dG_v^*(\kappa)f(v^*)-\frac{\partial v^{cc}}{\partial \phi}\int_{\Pi_{cc}(v^{cc})}^{\Pi_{nc}(v^{cc})}\gamma(e)dG_{v^{cc}}(\kappa)f(v^{cc})}_{\text{examination cost decrease II}}. \qquad (18)
\end{aligned}$$

To interpret the various terms in (17) and (18), note that, by the implicit function theorem, $\partial v^*/\partial e = 0$, $\partial v^*/\partial \phi > 0$, $\partial \hat{v}/\partial e \geq 0$, and $\partial \hat{v}/\partial \phi > 0$. In addition, $\partial v^{cc}/\partial e \geq 0$ and $\partial v^{cc}/\partial \phi \leq 0$.[29]

29. Formally,

$$\begin{aligned}
&\frac{\partial v^*}{\partial e}=0 && \frac{\partial \hat{v}}{\partial e}=\frac{\phi}{(1-e)^2(d\Delta^H/dv)}\geq 0 && \frac{\partial v^{cc}}{\partial e}=-\frac{\partial \underline{\lambda}/\partial e}{\partial \underline{\lambda}/\partial v+d(l_C/\Delta_C)/dv}\geq 0 \\
&\frac{\partial v^*}{\partial \phi}=\frac{1}{d\Delta^H/dv}>0 && \frac{\partial \hat{v}}{\partial \phi}=\frac{1}{(1-e)(d\Delta^H/dv)}>0 && \frac{\partial v^{cc}}{\partial \phi}=-\frac{\partial \underline{\lambda}/\partial \phi}{\partial \underline{\lambda}/\partial v+d(l_C/\Delta_C)/dv}\leq 0,
\end{aligned} \qquad (19)$$

where the inequalities for the derivatives of v^{cc} follow from the fact that by Assumptions 3 and 4, combined with Lemma 1, the denominator is negative, and

$$\frac{\partial \underline{\lambda}}{\partial e}=\frac{G_v(\pi(v))(G_v(\Pi_{cc}(v))-G_v(\pi(v)))}{(G_v(\Pi_{cc}(v))-eG_v(\pi(v)))^2}\geq 0 \qquad \frac{\partial \underline{\lambda}}{\partial \phi}=-\frac{g_v(\Pi_{cc}(v))(1-e)G_v(\pi(v))}{(G_v(\Pi_{cc}(v))-eG_v(\pi(v)))^2}\leq 0.$$

The intuition for the sign of the derivatives of v^{cc} is as follows. An increase in e makes it more likely that a granted patent originates with a high-type inventor, which renders challenges less attractive and hence raises the threshold for challenge credibility. An increase in ϕ reduces high-type innovation and thus the share of patents that originate with high types; this reduces the threshold for challenge credibility.

Raising the examination intensity has four distinct effects on welfare. First, it increases *detection* of low-type applications, which eliminates deadweight loss. Second, it increases *deterrence* by raising the threshold above which low types apply, $\hat{v}$; this eliminates deadweight loss and saves on examination costs. Third, it *discourages challenges* by raising the threshold above which challenges are credible, v^{cc}. This increases high-type innovation by suppressing litigation for inventions that are at the borderline of challenge credibility.[30]

At the same time, for inventions with borderline challenge credibility, raising e also takes away $\chi(v)$, *i.e.*, the net benefits from challenges. By Proposition 4, $\chi(v)$ may be positive or negative, depending on (9). If $\chi(v) < 0$, as will be the case in the quantitative analysis in Section 4, then the challenge discouragement effect is unambiguously positive. Fourth, it *increases examination costs* because $\gamma' > 0$ and because the increase in v^{cc} means that more high-type applications need to be examined.

Raising pre-grant fees has five distinct welfare effects, some of which mirror those of an increase in e. First, it increases *deterrence* by raising $\hat{v}$. Second, by reducing the profitability of patents it directly *discourages high-type innovation* in two ways: it lowers the thresholds on R&D costs below which high types invest, Π_{nc} and Π_{cc}, across all invention values, and it pushes up the minimum invention value for which high types invest, v^*. Third, it *discourages challenges at the intensive margin* for both high and low-type inventors by reducing Π_{cc}, implying both fewer high-type inventors and fewer low-type inventors charging high license fees (*i.e.*, lower $\tilde{y}$) across all invention values. Formally, this effect is $\int_{v^{cc}}^{\overline{v}}(\partial\chi(v)/\partial\phi)dF(v)$; its sign depends on (9).

Fourth, raising ϕ *encourages challenges at the extensive margin* by lowering v^{cc}. By triggering litigation for inventions with borderline challenge credibility, this indirectly decreases high-type innovation and adds $\chi(v)$; the sign of the latter effect again depends on (9). Fifth, by directly and indirectly discouraging innovation it *decreases examination costs*, as fewer high-type applications need to be examined.

This analysis reveals the various channels through which examination intensity and fees affect welfare. It highlights the usefulness of our framework, which integrates the different stages of the patent-screening process. It also helps us shed light on the mechanisms underlying the effects of the policy reforms analyzed in the quantitative analysis in Section 5.

The next proposition fixes fees at a level that covers at least the cost of *registering* patents (without examination), $\gamma(0)$, and examines the welfare effect of raising examination intensity above zero.

Proposition 8. *Consider a patent policy* $(e,\phi_A,\phi_P)=(0,\phi,0)$ *with* $\phi<\Delta^L(v^{cc})$*, and suppose* $\Delta_C(v^{cc})\geq D(v^{cc})l_C(v^{cc})/(l_C(v^{cc})+l_I(v^{cc}))$*. If* $\phi\geq\gamma(0)$ *and* $\gamma'(0)=0$*, an increase in e enhances welfare.*

This proposition implies that, when fees are set to cover the costs of registering patents and the net benefits from challenges are negative, a registration system cannot be optimal. The condition

30. To see that $\int_{\Pi_{cc}(v^{cc}))}^{\Pi_{nc}(v^{cc}))}(v^{cc}-D(v^{cc})-\kappa)dG_{v^{cc}}(\kappa)$ is positive, notice that, for $\kappa\leq\Pi_{nc}(v)$, we have $v-D(v)-\kappa\geq v-D(v)-(\pi(v)+\Delta^H(v))>0$ by Assumption 1.

that fees cover costs is of particular relevance when the patent office is self-funded, as is currently the case for the USPTO.

3.2.2. Uncertainty about invention value. So far, we assumed that inventors know the value of their invention upfront. In practice, there is uncertainty: some inventions whose expected value justifies the costs of a patent application ex ante turn out to have little value ex post, and their inventors allow them to lapse, thus saving on renewal fees. To see how uncertainty about invention value changes the analysis, we now consider a variation on the model whereby the inventor learns the value of his invention only after investing and paying the pre-grant fee, but before paying the post-grant fee. Specifically, assume that an invention (v,κ) has value v with probability b and value zero with probability $1-b$; we will refer to $1-b$ as the probability of *obsolescence*.

We modify the timing in Figure 1 as follows. At $t=0$, the inventor learns (v,κ) but is uncertain whether the invention will turn out to have value. Thus, v is now interpreted as the potential value of the invention, conditional on having value at all. Between $t=1$ and $t=2$—*i.e.*, after application and examination, but before activation—the uncertainty about value is resolved.

This is a simple way to capture learning about invention value. Although in practice the resolution of uncertainty takes place more gradually and renewal fees are paid throughout the patent life, the empirical literature suggests that learning is completed within a few years, and thus before the bulk of renewal fees is due (Pakes, 1986; Lanjouw, 1998). In the absence of the patent incentive, the inventor's expected payoff from investing becomes $b\pi(v)-\kappa$. We thus adapt our previous terminology by defining inventors as low types if $\kappa \leq b\pi(v)$ and as high types if $\kappa > b\pi(v)$.

At $t=2$, an inventor whose invention turns out to have value pays the post-grant fee only if $v > v_P$, where v_P is implicitly defined by $\Delta^H(v_P)=\phi_P$. An inventor whose invention turns out to have zero value never pays the post-grant fee. At $t=1$, an inventor with $v \leq v_P$ anticipates that he will not activate the patent even if the invention has value. Thus he does not apply at $t=1$ and does not invest at $t=0$ if of high type. For $v > v_P$, a high type's maximum payoff from investing and applying is $b(\pi(v)+\Delta^H(v)-\phi_P)-\phi_A-\kappa$, while a low type's maximum payoff from applying is $b(1-e)(\Delta^H(v)-\phi_P)-\phi_A$. Hence, the thresholds v^* and $\hat{v}$ below which no high types invest and no low types apply are now implicitly defined by

$$\Delta^H(v^*)=\frac{\phi_A}{b}+\phi_P$$

$$\Delta^H(\hat{v})=\frac{\phi_A}{b(1-e)}+\phi_P.$$

As before, we restrict attention to the case where $b(1-e)(\Delta^L(v^{cc})-\phi_P) > \phi_A$, which ensures in particular that $\hat{v} < v^{cc}$. Because both v^* and $\hat{v}$ are greater than v_P, all inventors who pay the pre-grant fee and whose invention turns out to have value will thus also pay the post-grant fee. High types invest, apply, and activate if $v > v^*$ and $\kappa \leq \Pi(v)$, where

$$\Pi(v)=\begin{cases} \Pi_{nc}(v)=b[\pi(v)+\Delta^H(v)-\phi_P]-\phi_A & \text{for } v \leq v^{cc} \\ \Pi_{cc}(v)=b[\pi(v)+\Delta^H(v)-\tilde{x}(v)l_I(v)-\phi_P]-\phi_A & \text{for } v > v^{cc}. \end{cases}$$

Here, $\tilde{x}(v)$ is the same as before, and v^{cc} is implicitly defined by $(1-\underline{\lambda}(v^{cc}))\Delta_C(v^{cc})=l_C(v^{cc})$, where

$$\underline{\lambda}(v)=\frac{G_v(\Pi_{cc}(v))-G_v(b\pi(v))}{G_v(\Pi_{cc}(v))-eG_v(b\pi(v))}.$$

Finally, the probability that low types charge high license fees when challenges are credible becomes

$$\tilde{y}(v)=\left(\frac{l_C(v)}{\Delta_C(v)-l_C(v)}\right)\left(\frac{G_v(\Pi_{cc}(v))-G_v(b\pi(v))}{(1-e)G_v(b\pi(v))}\right).$$

Uncertainty on value implies that our result on frontloading in Proposition 7 needs to be modified. Keeping the sum of pre-grant and post-grant fees constant and shifting fees forward to the application stage no longer leaves high types unaffected; instead, it raises v^*. The intuition is that, unlike post-grant fees, which do not have to be paid if the invention turns out to have low value, pre-grant fees are always due, so frontloading raises the expected cost of obtaining patent rights. To restore the result, frontloading needs to be done in such a way that the *expected* fee payment for high types remains the same. In effect, any combination of fees that keeps $\phi_A+b\phi_P$ constant leads to the same cutoff v^* and the same $\Pi(v)$. By contrast, low types still prefer fees to be backloaded. In the quantitative analysis below, we calibrate the parameter b based on obsolescence rates estimated in the empirical literature and implement frontloading by shifting only a fraction b of post-grant fees forward.

3.2.3. Administrative patent review. One of the most significant reforms of the 2011 America Invents Act was the creation of an administrative law body for patent review through the establishment of the Patent Trial and Appeal Board (PTAB). The likely effects of such a reform are twofold. First, the procedure is cheaper than judicial review, so it reduces litigation cost. Second, the procedure is likely to be less thorough than judicial review, and thus to result in more errors, *i.e.*, high-type patents being invalidated or low-type patents being upheld.[31] To analyze the effects of introducing administrative patent review, we now extend our model to allow for erroneous decisions at the litigation stage.

Assume that high-type patents are upheld with probability q^H and low-type patents with probability $q^L<q^H$. To make sure that it is optimal *not to challenge* if the competitor is certain of facing a high type, we assume for the purposes of this section that $(1-q^H)\Delta_C(\bar{v})\leq l_C(\bar{v})$, which together with Assumption 3 implies that $1-q^H\leq l_C(v)/\Delta_C(v)$ for all $v\geq v^{cc}$. For the quantitative analysis below, we do not impose this assumption, and in fact for some of the counterfactuals and some values of v, it turns out not to be satisfied. Supplementary Appendix C derives the equilibrium for the full range of (q^L,q^H), including for the case where $(1-q^H)\Delta_C(\bar{v})>l_C(\bar{v})$.

The low type can now preempt challenges by setting $R=l_C(v)+q^L\Delta_C(v)$. We thus let $\Delta^L(v)\equiv l_C(v)+m(v)+q^L\Delta_C(v)$, and we maintain the assumption that $(1-e)(\Delta^L(v^{cc})-\phi_P)>\phi_A$, so that the cutoffs described in Proposition 6 remain relevant. Accordingly, the welfare function given in (16) remains valid if we replace Π_{cc} by

$$\Pi_{cc}(v)=\pi(v)+(1-\tilde{x}(v))\Delta^H(v)+\tilde{x}(v)[q^H\Delta^H(v)-l_I(v)]-\phi_A-\phi_P, \tag{20}$$

the equilibrium values of $\tilde{x}(v)$ and $\tilde{y}(v)$ by

$$\tilde{x}(v)=\frac{(1-q^L)\Delta_C(v)-l_C(v)}{(1-q^L)\Delta^H(v)+l_I(v)} \tag{21}$$

$$\tilde{y}(v)=\left(\frac{(q^H-q^L)\Delta_C(v)}{(1-q^L)\Delta_C(v)-l_C(v)}-1\right)\left(\frac{G_v(\Pi_{cc}(v))-G_v(\pi(v))}{(1-e)G_v(\pi(v))}\right), \tag{22}$$

31. See Section 5.2 for a discussion.

and the net benefits from challenges, χ, by

$$\chi(v)=(G_v(\Pi_{cc}(v))-G_v(\pi(v)))\left(\frac{(q^H-q^L)[l_C(v)D(v)-(l_C(v)+l_I(v))\Delta_C(v)]}{(1-q^L)\Delta^H(v)+l_I(v)}\right). \tag{23}$$

Note also that the challenge-credibility threshold v^{cc} is now implicitly defined by

$$\underline{\lambda}(v^{cc})(1-q^H)+(1-\underline{\lambda}(v^{cc}))(1-q^L)=\frac{l_C(v^{cc})}{\Delta_C(v^{cc})}. \tag{24}$$

We now discuss the welfare effects of changes in the costs of litigation and the accuracy of adjudication. Conventional wisdom has it that a reduction in litigation costs is a good thing, while a reduction in adjudication accuracy is a bad thing. On both accounts, our results, details of which can be found in Supplementary Appendix D, are more nuanced. First, reducing litigation costs is not unambiguously good. One of the reasons is that a decrease in either l_C or l_I raises the rate of challenges $\tilde{x}$, and thus the incidence of litigation. Second, it matters whether the inventor's or the competitor's litigation costs are reduced. As it turns out, many of the welfare effects are of opposite sign. Lowering the competitor's litigation cost decreases Π_{cc}, whereas lowering the inventor's increases it.

Third, reducing the accuracy of adjudication is not unambiguously bad. A marginal decrease in q^H reduces Π_{cc} but also enhances the net benefits from challenges (if $\chi(v)<0$). While the high-type innovations that are lost as a result are only marginally profitable (and thus add little net social value), the deadweight loss that is eliminated applies to all inframarginal high-type innovations that are successfully challenged. Even an increase in q^L can be welfare-enhancing, because it raises the challenge-credibility threshold and thereby encourages innovation.

In short, the welfare effects of introducing administrative patent review are theoretically ambiguous. This lends further value to the quantitative analysis below, the results of which suggest that administrative review generates substantial welfare gains.

3.2.4. Antitrust limits on licensing contracts. So far we have assumed that the extra profits from licensing, $m(v)$, and the associated deadweight loss, $D(v)$, are the same regardless of whether the high or the low license fee is charged. This makes sense if there are no restrictions on licensing. As Section 4.1 shows, when license contracts take the form of unrestricted two-part tariffs consisting of a fixed fee and a per-unit royalty, inventors will structure the contract so as to maximize industry profits by softening competition through royalties, and then share the surplus through the fixed fee. This implies that royalties, and thus deadweight loss, are the same with both the high-type and low-type contract. The fixed fees in the contracts differ, however: the high-type contract involves no fixed fees, while the low-type contract involves negative fixed fees to compensate the competitor for foregoing a challenge. The negative fixed fees in the low-type contract bear some resemblance to so-called reverse payment settlements, which have recently come under scrutiny. In *FTC v. Actavis, Inc.* (570 US 136 (2013)), the U.S. Supreme Court effectively introduced limits on the use of negative fixed fees in licensing contracts. We now examine how such a policy affects welfare in our model. We return to this issue in Section 5, where we analyse its impact quantitatively.

A restriction on negative fixed fees has no effect when challenges are not credible, as low types always offer the high-type contract in that case. When challenges are credible, its effects are twofold. First, it introduces an additional social benefit from challenge credibility: the threat of a challenge now leads low types to sometimes reduce their royalty, rather than the fixed fee, thereby eliminating deadweight loss. Second, it raises the rate of challenges and lowers the threshold for challenge credibility.

To see this formally, suppose that the extra profit from licensing and the deadweight loss are $m^H(v)$ and $D^H(v)$ with the high-type contract and $m^L(v) \leq m^H(v)$ and $D^L(v) \leq D^H(v)$ with the low-type contract. Then, w^{cc} becomes

$$w^{cc}(v) = \int_{\underline{\kappa}}^{\Pi_{cc}(v)} \Big(v - D^H(v) - \kappa - \gamma(e)\Big) dG_v(\kappa) + G_v(\pi(v)) e D^H(v) + \chi(v),$$

where

$$\chi(v) = (1-e)G_v(\pi(v)) \Big[\tilde{x}(v)\tilde{y}(v)[D^H(v) - l_C(v) - l_I(v)] + (1-\tilde{y}(v))(D^H(v) - D^L(v))\Big] \\ - (G_v(\Pi_{cc}(v)) - G_v(\pi(v)))\tilde{x}(v)(l_C(v) + l_I(v)) \quad (25)$$

are the adjusted net benefits from challenges. The equilibrium rate of challenges, $\tilde{x}$, and of charging high license fees, $\tilde{y}$, become

$$\tilde{x}(v) = \frac{\Delta_C(v) - l_C(v) + m^H(v) - m^L(v)}{\Delta^C(v) + m^H(v) + l_I(v)}$$

$$\tilde{y}(v) = \left(\frac{l_C(v)}{\Delta_C(v) - l_C(v)}\right)\left(\frac{G_v(\Pi_{cc}(v)) - G_v(\pi(v))}{(1-e)G_v(\pi(v))}\right).$$

In addition to its effect on $w^{cc}(v)$, a restriction on negative fixed fees also impacts welfare by changing the challenge-credibility threshold v^{cc}.

Compared to the expression for χ in (8), the one in (25) has an extra term, namely, $(1-e)G_v(\pi(v))(1-\tilde{y}(v))(D^H(v) - D^L(v)) \geq 0$. This corresponds to deadweight loss that challenge credibility avoids, in the presence of restrictions on negative fixed fees, without any challenges taking place: the *threat* of a challenge, rather than an actual challenge, leads low types to offer lower royalties a fraction $1-\tilde{y}(v)$ of the time, reducing deadweight loss from $D^H(v)$ to $D^L(v)$.

In the absence of other changes, this extra benefit from challenges would make restrictions on negative fixed fees unambiguously welfare enhancing. The restrictions also induce an increase in the rate of challenges and a decrease in the challenge-credibility threshold, however. Using the expression of $\tilde{y}$, we obtain

$$\chi(v) = \tilde{x}(v)(G_v(\Pi_{cc}(v)) - G_v(\pi(v)))\left[\frac{l_C(v)D^H(v) - (l_C(v) + l_I(v))\Delta_C(v)}{\Delta_C(v) - l_C(v)}\right] \\ + \left((1-e)G_v(\pi(v)) - \frac{l_C(v)(G_v(\Pi_{cc}(v)) - G_v(\pi(v)))}{\Delta_C(v) - l_C(v)}\right)(D^H(v) - D^L(v)).$$

We know that the expression on the second line is positive, and the expression in square brackets is the same as the one that determines the desirability of challenges in the absence of restrictions on licensing. The impact of licensing restrictions jointly depends on the sign of the latter expression and on whether $\tilde{x}(v)(G_v(\Pi_{cc}(v)) - G_v(\pi(v)))$ increases or decreases, which is indeterminate: $\tilde{x}$ increases (as $m^H(v) - m^L(v) \geq 0$), but $\Pi_{cc}(v)$ decreases (due to the increase in $\tilde{x}$). The decrease in $\Pi_{cc}(v)$ also implies that the share of high types goes down, thus lowering v^{cc}.

In summary, although a restriction on negative fixed fees leads to a reduction in deadweight loss, it also leads to an increase in litigation, making its overall welfare effect theoretically ambiguous. In the quantitative analysis that follows, we find that, when the model is evaluated at the calibrated parameters, the welfare effect of such a restriction is large and positive.

3.2.5. Changing the allocation of legal costs. The baseline model is based on each party paying its own legal costs, regardless of the outcome of trial. This is the standard rule in the U.S., but other parts of the world use loser-pays rules (sometimes called the English rule), under which legal costs are shifted from the winner to the loser. To study the impact of a change in the allocation of legal costs, we now extend the model by assuming that a share $\zeta \in [0,1]$ of the winner's litigation costs must be paid by the loser. To preempt challenges the low type then has to set $R=(1-\zeta)l_C(v)$, and accordingly we let $\Delta^L(v) \equiv (1-\zeta)l_C(v)+m(v)$. We maintain the assumption that $(1-e)(\Delta^L(v^{cc})-\phi_P) > \phi_A$, so that the cutoffs described in Proposition 6 and the welfare function given in (16) continue to apply after appropriately adjusting the definitions of Π_{cc}, χ, $\underline{\lambda}$, and v^{cc}.[32] Supplementary Appendix E provides the corresponding expressions and the details of the analysis. We now briefly discuss the results.

Changing the allocation of legal costs has ambiguous welfare effects. On the one hand, it decreases high types' litigation costs by $\zeta l_I(v)$. Holding the rate of challenges fixed, this raises high-type innovation. On the other hand, it increases the rate of challenges $\tilde{x}$, which can potentially more than offset the direct effect on high types' litigation costs (except, under perfect courts, when $\zeta = 1$). The intuition is that shifting legal costs reduces the payoff to low types from being involved in a challenge, but also the payoff from pre-empting challenges as the license fee that can be extracted in this way shrinks. Because both payoffs are reduced by the same amount, $\zeta l_C(v)$, while the payoff from charging the high license fee and escaping a challenge is unchanged, the rate of challenges required to make the low type indifferent has to go up. For similar reasons, cost shifting also has ambiguous effects on the challenge-credibility threshold v^{cc} and on the net benefits from challenges χ. In the quantitative analysis, we simulate the effects of cost shifting and find that it reduces welfare.

4. QUANTITATIVE ANALYSIS

In this section, we set up the model used for the quantitative analysis and describe how we recover its key structural parameters. These allow us to evaluate the effectiveness of patent screening and the welfare generated by the current patent system. This will then lay the foundation for Section 5 in which we use the estimated parameters to assess the impact of various counterfactual policy reforms on innovation and welfare.

4.1. *Summary of the empirical model and computation of equilibrium*

This section specifies the empirical model for the quantitative analysis and the computation of equilibrium, for any given set of parameters. We now describe how we specify the key elements in the theoretical model that are left in general form—including profits, gains from patenting, and deadweight loss—as well as the functional forms we adopt for litigation costs, examination costs, and the distributions of R&D costs and invention values. Supplementary Appendix F provides the

32. This is now a stronger assumption as $\Delta^L(v)$ is lower. Nevertheless, it holds with the functional forms and calibrated parameters of the quantitative analysis below.

full derivations of the empirical model and verifies that the assumptions on which the theoretical analysis is based are satisfied.

We embed the theoretical model into a Cournot product market setting. The two firms—an inventor I and competitor C—produce a homogeneous good and compete à la Cournot in continuous time. Inverse demand is given by $p=a-Q$ and firms (initially) have symmetric unit costs c. The inventor has the opportunity to invest in R&D to develop a cost-reducing invention. If he invests, the invention occurs at time $t=0$. In each subsequent time both firms choose their output levels. The invention reduces unit production cost from c to $c'=(1-s)c$, where $s\in[0,1]$.[33] Flow profits are cumulated at discount rate r.

In the theoretical analysis, inventions are indexed by their social value, v. In the empirical model, we microfound invention value. The primitive is taken to be a distribution of invention sizes (in terms of the cost reduction the invention generates), $F(s)$, specified below. Given a random draw from this distribution, and the licensing game and product market competition, there is an implied flow value in equilibrium for the invention which will depend on (s,a,c).

If the invention is not patented, the competitor can copy it without cost or delay and both firms earn the present value of the infinite sum of flow profits, given by:

$$\pi(s)=\frac{(a-(1-s)c)^2-(a-c)^2}{9r}.$$

The social value of the invention is $v(s)=4\pi(s)$.[34]

If the invention is patented, the inventor can exclude the competitor from its use, in which case the competitor's flow profit is diminished. The present value of the reduction in the competitor's flow profit during the patent life, T, is:

$$\Delta_C(s)=\frac{(1-\exp(-rT))4(a-c)sc}{9r}.$$

Alternatively, the inventor can make a take-it-or-leave-it offer of a license contract to the competitor. We assume that the contract takes the form of a two-part tariff with fixed fee τ and per-unit royalty ρ. The present value of the deadweight loss associated with royalty ρ is given by

$$D(s,\rho)=\frac{(1-\exp(-rT))(a-(1-s)c+\rho/2)\rho}{9r}.$$

The inventor chooses the royalty ρ to maximize the firms' joint profits, while he uses the fixed fee τ to share the surplus. Our theoretical analysis implies that, in equilibrium, high types will offer a license contract (τ^H,ρ^H) that holds the competitor down to her outside option. As Supplementary Appendix F shows, the joint-profit maximizing royalty is equal to the (absolute) cost reduction, so $\rho^H=sc$. The fixed fee that holds the competitor down to $\pi(s)-\Delta_C(s)$ then is $\tau^H=0$, and the present value of the gains from market power that the optimal royalty allows the firms to achieve, by softening competition, is:

$$m(s)=\frac{(1-\exp(-rT))(a-c)sc}{9r}.$$

33. We assume $a/2\geq c$ which ensures the invention is non-drastic, so that the competitor does not shut down post-invention regardless of the value of s. The assumption is satisfied by our calibrated values of a and c.

34. As a robustness check, we also use a specification that allows for a lag in copying the unpatented invention. Based on a survey of 100 U.S. firms, Mansfield (1985) finds that rival firms learn about the "nature and operation" of new inventions relatively quickly, most within 18 months. We experiment with imitation lags of 1–3 years and the results are similar.

As in the theoretical analysis, the (present value of the) gain to the inventor if the competitor accepts the contract (τ^H, ρ^H) is $\Delta^H(s) = \Delta_C(s) + m(s)$.

Low types propose the same contract as high types when challenges are not credible. When challenges are credible, the theoretical analysis implies that low types propose the high-type contract with probability y and a different contract (τ^L, ρ^L) with probability $1-y$, while the competitor challenges with probability x when offered the high-type contract (see Proposition 6).[35] We show in Supplementary Appendix F that $\rho^L = sc$ – the joint-profit maximizing royalty—and $\tau^L < 0$. By setting a negative fixed fee, the low type compensates the competitor for refraining from a challenge. If negative fixed fees are prohibited by antitrust authorities or the courts, we set $\tau^L = 0$ and solve for the value of ρ^L that achieves the same (preempting a challenge). Of course, in this case ρ^L is lower than when negative fees are allowed. We return to this point in the counterfactual simulations, where we assess the welfare impact of a recent Supreme Court decision that puts limits on negative fixed fees in licensing agreements.

Througout the quantitative analysis, we adopt the formulation described in Section 3.2.2 whereby inventors face uncertainty regarding the value of their invention, so that an idea (s, κ) leads to a cost reduction of size s with probability b and to no cost reduction (obsolescence) with probability $1-b$. For the calibration and most counterfactual experiments, we assume the courts do not make mistakes. As in the theoretical model, there is a threshold invention size s^* such that high types never invest if $s \le s^*$. For $s > s^*$ they invest and apply if and only if

$$\kappa \le \Pi(s) = \begin{cases} \Pi_{nc}(s) & \text{for } s \le s^{cc} \\ \Pi_{cc}(s) & \text{for } s > s^{cc}, \end{cases}$$

where $\Pi_{nc}(s) = b[\pi(s) + \Delta^H(s) - \phi_P] - \phi_A$ and $\Pi_{cc}(s) = b[\pi(s) + \Delta^H(s) - \tilde{x}(s)l(s) - \phi_P] - \phi_A$. Similarly, there is a threshold invention size $\hat{s}$ such that low types apply if and only if $s > \hat{s}$. Finally, challenges are credible if and only if $s > s^{cc}$.[36]

4.1.1. Distribution of invention size (cost reduction), $F(s)$.

We assume that cost reductions from inventions follow a log-logistic distribution $F(s; \beta_0, \beta_1) = \frac{s^{\beta_0}}{s^{\beta_0} + \beta_1^{\beta_0}}$ where $\beta_0 > 0, \beta_1 > 0$. The log-logistic distribution is similar in shape to the log-normal (which the patent renewal literature shows provides the best fit for the distribution of the value of patent rights; *e.g.* Schankerman and Pakes, 1986). The advantage of the log-logistic is that it has a closed-form expression for the probability density, which is convenient for computational purposes.

4.1.2. Distribution of development costs, $G_s(\kappa)$.

We assume that invention development cost, κ, follows an exponential distribution with mean $k(s)$, which is allowed to depend on the size of the cost reduction generated by the invention, s. Specifically, we use $G_s(\kappa; k_0, k_1) = 1 - \exp(-\kappa / k(s))$ where $k(s) = k_0 + k_1 s$, $k_0 \ge 0$ and we expect $k_1 > 0$, in which case $G_s(\kappa; k_0, k_1)$ is stochastically increasing in s. The exponential distribution is parsimonious (having a single parameter) and admits a simple analytic expression for the conditional expectation of R&D costs, which is convenient for computational purposes.

35. This is true only for a certain parameter range, but the estimated parameters satisfy the required conditions. Details are in Supplementary Appendix C.

36. This description of the thresholds characterizing the equilibrium needs to be modified when we introduce imperfect courts, as done in Supplementary Appendix C. There we show that in the presence of imperfect courts, there is an additional threshold $s_1 > s^{cc}$ above which low types always offer the challenge-preempting contract, high types charge a lower license fee, and the competitor randomizes over challenges. The quantitative analysis of the counterfactuals that use imperfect courts accounts for this.

4.1.3. Examination cost function, $\gamma(e)$. We specify the examination cost function per patent application as $\gamma(e;\gamma_0,\gamma_1)=(\gamma_0+g(e))^{\gamma_1}$, where $\gamma_0 \geq 0$ and $\gamma_1 > 0$ are parameters and $g(e)=e/(1-e)$. This formulation allows for a fixed component in examination costs (so we can estimate what the cost of a registration system, with $e=0$, would be) as well as a variable component $g(e)$, and has the flexibility to match the empirical targets with only two parameters. The function $g(e)$ is chosen so as to ensure that examination costs go to infinity as e tends to 1.[37]

4.1.4. Litigation cost function, $l(s)$. We assume that litigation costs are symmetric for the inventor and competitor, given by $l(s)=l_0+l_1\Delta^H(s)$, where $\Delta^H(s)$ is the present value of the patent premium for an invention of size (cost reduction) s. This formulation reflects the idea that the willingness to invest in a patent challenge for each party depends on the potential gains from winning, and is in line with the observation that survey estimates of litigation costs increase with the value at stake.

4.2. *Calibration and estimation of parameters*

Table 1 lists all the parameters of the model, as well as the empirical targets used in its calibration. There are two sets of parameters: seven parameters that we estimate by matching on a set of calibration targets, and nine parameters that we assign on the basis of external information. The estimated parameters are: examination intensity, e; two parameters of the log-logistic distribution of invention size, $F(s;\beta_0,\beta_1)$; two for the exponential distribution of development costs, $G_s(\kappa;k_0,k_1)$; and two for the examination cost function, $\gamma(e;\gamma_0,\gamma_1)$.

The set of assigned parameters include: demand and cost parameters, (a,c); obsolescence (cumulated over the patent life), $1-b$; discount factor, r; patent life, T; litigation cost parameters, (l_0,l_1); and pre-grant and post-grant patent fees, ϕ_A and ϕ_P, respectively. To assign values for (a,c), we first compute values at the most disaggregated manufacturing sectors available (6-digit level, North American Industry Classification System (NAICS); for details see Supplementary Appendix G). We use data on sales, price-cost markups, and Herfindahl concentration index to infer demand and cost parameters under the assumption that firms in the sector play equilibrium strategies in quantities. Then we weight each sector to construct an estimate of (a,c) for a randomly drawn patent. We set the rate of obsolescence (cumulated over the patent life) and the rate of depreciation of the returns to patents based on Lanjouw (1998) and Bessen (2008). Litigation cost parameters are based on survey information from the American Intellectual Property Law Association (2005; 2011; 2017). Patent life is set at the statutory value, 20 years. Pre- and post-grant fees are set using USPTO data (plus an external estimate of the legal cost of preparing a patent application). Supplementary Appendix G provides details of how we assign these parameter values.

Estimation of parameters is done by matching the empirical targets to their theoretical equivalents from the model. This is conducted in two stages. First, we choose the five parameters $(e,\beta_0,\beta_1,k_0,k_1)$ to match the following five targets: grant rate, *GR*, litigation rate, *LR*, validation rate, *VR*, R&D per invention, *RPI*, and TFP growth per invention, *TFPI*. Based on the first-stage estimates, we then choose the parameters (γ_0,γ_1) to match the USPTO budget per application, B, and the elasticity of the examination cost function, E. Construction of the empirical targets is described below.

The number of estimated parameters in the model is equal to the number of empirical targets, so the system is exactly identified. With no overidentifying restrictions, we cannot conduct internal validation tests. However, we conduct four different checks for external validation, which are described in Section 4.4.

37. Quantitative estimates are similar with an alternative specification $g(e)=-\ln(1-e)$.

TABLE 1
List of calibrated parameters and calibration targets

Assigned parameters		
Demand scale	a	\$550,929
Unit cost	c	\$84,460
Obsolescence	$1-b$	0.427
Discount rate	r	0.078
of which:		
- Depreciation: 0.048		
- Real interest rate: 0.03		
Patent duration	T	20
Minimum litigation cost	l_0	\$623,487
Variable litigation cost	l_1	0.162
Pre-grant fees	ϕ_A	\$16,282
of which:		
- Filing, search and examination fees: \$1,282		
- Patent drafting cost: \$15,000		
Post-grant fees	ϕ_P	\$10,791

Estimated parameters	
Examination intensity	e
Distribution of R&D cost, $G_s(\kappa)$	k_0, k_1
Distribution of invention size, $F(s)$	β_0, β_1
Examination cost function, $\gamma(e)$	γ_0, γ_1

Empirical targets	
Grant rate (GR, %)	71.2
Litigation rate (LR, %)	1.71
Validation rate (VR, %)	57.6
R&D per invention (RPI)	\$721,818
TFP growth per invention ($TFPI$, $\times 10^{-5}$)	2.52
Patent office budget/application (B)	\$4,117
Elasticity of examination costs (E)	2.10

4.2.1. Empirical targets and their theoretical equivalents. We summarize below the equations for the five first-stage empirical targets, and briefly describe how they are constructed. For more details, see Supplementary Appendix F on the derivation of these expressions and Supplementary Appendix G on the construction of the targets.

1. **Grant rate**: The grant rate is equal to the number of granted patents divided by the number of applications,

$$GR = \frac{(1-e)A^L + A^H}{A^L + A^H}$$

where A^L and A^H denote the number of applications by low and high types:

$$A^L = \int_{\hat{s}}^{1} G_s(b\pi(s))dF(s) \qquad A^H = \int_{s^*}^{1} (G_s(\Pi(s)) - G_s(b\pi(s)))dF(s).$$

Empirical target: The grant rate is taken from Carley *et al.* (2015), based on 2.15 million patent applications covering cohorts 1996-2005.

2. **Litigation rate**: The litigation rate is equal to the number of litigated patents divided by the number of grants,

$$LR = \frac{X}{(1-e)A^L + A^H}$$

where

$$X = b\int_{s^{cc}}^{1} \tilde{x}(s)\Big[\Big(G_s(\Pi_{cc}(s)) - G_s(b\pi(s))\Big) + (1-e)\tilde{y}(s)G_s(b\pi(s))\Big]dF(s).$$

Litigated patents are given by the probability that the invention has value, b, times the equilibrium probability of being challenged given that the high license fee is charged, $\tilde{x}(s)$, times the number of high types above the challenge credibility threshold plus the number of low types that are not screened out, times the probability they charge the high license fee, $\tilde{y}(s)$.

Empirical target: The litigation rate is the percentage of granted patents for domestic corporate entities in the U.S., over the period 1978–99, that are involved in at least one suit. This is taken from Lanjouw and Schankerman (2001) and corresponds to the probability that a randomly drawn patent is sued at least once.

3. **Validation rate**: The validation rate is equal to the number of challenges won by the patentee divided by the number of litigated patents,

$$VR = \frac{PW}{X},$$

where

$$PW = b\int_{s^{cc}}^{1} \tilde{x}(s)(G_s(\Pi_{cc}(s)) - G_s(b\pi(s)))dF(s).$$

In the baseline model with perfect courts, only high types win validity challenges.[38]

Empirical target: We use the fraction of patent challenge cases in which the validity of (all claims in) the patent is upheld by the court, taken from Allison *et al.* (2014).[39] Data cover all cases filed in U.S. district courts for 2008–09.

4. **R&D cost per invention**: R&D cost per invention is given by

$$RPI = \frac{K}{A^L + A^H}$$

where K denotes total R&D expenditure by patent applicants:

$$K = \int_{\hat{s}}^{1}\int_{0}^{b\pi(s)} \kappa\, dG_s(\kappa)dF(s) + \int_{s^*}^{1}\int_{b\pi(s)}^{\Pi(s)} \kappa\, dG_s(\kappa)dF(s).$$

Empirical target: R&D cost per invention is constructed for each 3-digit NAICS manufacturing industry, and then aggregated based on the number of patent grants. For each sector, we use R&D spending by private firms averaged over the 1999–2004 period and divide it by an estimate of the number of inventions. For the latter, we use the number of patent grants at the sector level (constructed by the USPTO) and divide by the grant rate to estimate patent applications by sector. We then adjust by estimates of the patent propensity for each sector, based on the large survey of U.S. corporations by Cohen *et al.* (2000).

38. With imperfect courts, the expression for PW is modified; see Supplementary Appendix F.

39. We use validation rates rather than patentee win rates more generally because the latter also depend on whether the court finds infringement. The idea behind our calibration strategy is to draw inferences about the underlying validity of issued patents from the outcomes in patent suits, among other things. Rulings concerning infringement are unrelated to this question. It is only the validity component which is informative for our purposes.

5. **TFP growth per invention**: Let ΔTFP denote TFP growth (*i.e.* the expected cost reduction) generated by patent applicants. Then TFP growth per invention is

$$TFPI = \frac{\Delta TFP}{A^L + A^H},$$

where

$$\Delta TFP = b\left[\int_{\hat{s}}^{1} sG_s(b\pi(s))dF(s) + \int_{s^*}^{1} s(G_s(\Pi(s)) - G_s(b\pi(s)))dF(s)\right].$$

Empirical target: TFP growth per invention is constructed for each 6-digit NAICS manufacturing industry, and then aggregated based on the number of patent grants. We use the average TFP growth over the period 1987–2007 for each sector. The TFP measure is the multifactor productivity index based on capital, production worker hours, non-production worker hours, energy and non-energy materials, constructed by the U.S. Census Bureau (NBER-CES Manufacturing Industry Database). For the number of inventions, we use the same measure as described under item 4 above.

The second-stage estimation of (γ_0, γ_1) requires two additional inputs: the patent office cost per application, which we construct from USPTO reports on labour and other costs for patent operations; and the elasticity of the examination cost function, which we infer using information on grant rates for examiners at different seniority levels from Frakes and Wasserman (2017).

4.3. *Results*

4.3.1. Baseline estimates. Table 2 presents baseline parameter estimates (Panel A) and three robustness checks (Panel B). Turning first to Panel A, the estimated parameters (β_0, β_1) for $F(s)$ show substantial variation in the size of inventions. The ratio of the 75th to the 25th percentile is 8.7.[40] The estimates of (k_0, k_1) for $G_s(\kappa)$ imply that the mean development cost is increasing in the invention size s. Finally, the estimates of (γ_0, γ_1) imply that the examination cost function is sharply convex.

The last four entries in the first row of Panel A provide metrics on key dimensions of interest:[41]

- *SHA*: share of high types among applications;
- *SHP*: share of high types among granted patents;
- SP_{cc}: share of patents for which challenges are credible;
- $1-\bar{y}$: share of low types above the challenge-credibility threshold that charge challenge-preempting license fees.

The estimated SHA shows that 39.9% of all patent *applications* are high-type inventions, which would not be developed were it not for patent rights. Thus 60.1% are low-type inventions which, from an economic point of view, should not be granted patents. Second, the patent office screens out about half of these low-type applications, $e = 47.9\%$. Taken together, these two findings imply that the share of high-type inventions among patent *grants* is SHP $= 56.0\%$. This estimate suggests

40. It is worth noting that the estimated parameters of our model imply that 45% of patents have value below the median estimated for the U.S. by Bessen (2008).

41. The formal definitions are provided in Appendix B.

TABLE 2
Baseline parameter estimates and robustness

Panel A. Baseline: estimates and welfare decomposition

	e %	β_0	β_1 $\times 10^{-6}$	k_0 $\times 10^3$	k_1 $\times 10^{10}$	γ_0	γ_1	SHA %	SHP %	SP_{cc} %	$1-\bar{y}$ %
Baseline	47.9	1.02	1.14	254.6	2.33	4.05	5.19	39.9	56.0	10.4	66.9

W $\times 10^6$	I_H $\times 10^6$	DWL $\times 10^6$	DA $\times 10^6$	LC $\times 10^6$	Γ $\times 10^6$	$W-I_L$ $\times 10^6$
2.680	0.221	0.187	0.004	0.032	0.002	0.004

Panel B. Robustness tests

	e %	β_0	β_1 $\times 10^{-6}$	k_0 $\times 10^3$	k_1 $\times 10^{10}$	γ_0	γ_1	SHA %	SHP %	SP_{cc} %	$1-\bar{y}$ %
$LR=2.38\%$	46.1	1.12	2.49	315.3	2.12	3.8	5.41	37.5	52.7	14.5	67.0
$1-b=0.279$, $r=0.107$	49.3	0.99	0.97	200.4	2.45	4.26	5.03	41.5	58.3	10.1	65.7
Fréchet $F(s)$	48.0	1.01	1.45	269.7	2.29	4.06	5.18	40.0	56.1	10.3	66.5

*In the log-logistic distribution, β_0 and β_1 represent the shape and scale parameters, respectively. For the Fréchet distribution, we list its parameters in the same order.

that the patent-quality problem is real – almost half of all patents are issued on low-type inventions where the patent rights are not needed to elicit their development, and thus impose unnecessary social costs.

The vast majority of granted patents are not at risk of litigation—they either become obsolete, or the size of the underlying invention (s) is insufficient to warrant a challenge. Only about 10% satisfy the challenge credibility constraint, as shown by SP$_{cc}$. The actual litigation rate is much lower, at 1.71%. This is both because competitors do not always challenge patents that charge the high license fee (due to the mixed-strategy nature of the equilibrium), and because some type-L patents avoid litigation by charging the low license fee. In particular, the parameter $1-\bar{y}$ shows that about 67% of type-L patentees above the challenge-credibility threshold preempt the competitor from challenging (and winning) by charging a license fee equal to her litigation cost. Many critics, and antitrust authorities, claim that this is the strategy adopted by so-called "patent trolls" (*e.g.* Federal Trade Commission, 2011). The key point to note is that, in our model, the frequency of such trolling behavior in equilibrium is endogenous to the design of the patent system. As we show in the counterfactual analysis below, some policy reforms that *improve* welfare have the side effect of *increasing* the level of trolling. This highlights the important point that policy should not focus on reducing trolling by itself.

The second row in Panel A provides various welfare-related measures:[42]

- W: welfare gains from innovation, $W=I^L+I^H-DWL+DA-LC-\Gamma$;
- I^θ: gross welfare gains from innovation by type $\theta=L,H$, net of R&D costs (but without netting out deadweight loss, examination costs, or litigation costs);
- DWL: gross deadweight loss (before litigation, but after examination);
- DA: deadweight loss avoided thanks to challenges (due to invalidation and, if negative fixed fees are prohibited, lower royalties charged by low types);
- LC: litigation costs;
- Γ: examination costs;

42. The formal definitions are provided in Appendix B.

Two points are worth noting. First, gross welfare generated by type-H inventions, I^H, accounts for only 8% of total gross welfare ($I^H + I^L$, not shown); the rest is due to type-L inventions. Thus, high types account for a significantly lower share of gross welfare than their share of patent applications. This is because by definition, for any given invention size s, the expected R&D cost of a high-type invention exceeds that of a low-type invention (*i.e.*, $E(\kappa|s,\kappa>\pi(s))>E(\kappa|s,\kappa\leq\pi(s))$). Second, the deadweight loss avoided by challenges, DA, is much smaller than the associated litigation cost, LC. Qualitatively, this conclusion is a reflection of the result in Proposition 4, which derives a condition for the net benefits from challenges to be positive.[43] With the model we use for the quantitative analysis, this condition does not hold.

The last column provides an estimate of the welfare generated by the patent system, which is equal to the welfare gains from innovation, W, net of the gross welfare due to low-type inventions, I^L. Since low-type inventions would be developed even without patent protection, $W - I_L$ is the measure of the social value of the patent system in our model. Our estimate is positive, indicating that the current patent system *is* welfare-improving. In the counterfactual analysis below, we discuss how various reforms affect the size of these gains (or, potentially, losses) from the patent system.

4.3.2. Robustness checks. Panel B presents three robustness checks. The first uses a different measure of the litigation rate—the number of patent suits per application—which includes multiple suits brought against the same patents. Our baseline measure is the percentage of patents involved in a suit, and thus excludes multiple suits. The second uses an alternative calibration of obsolescence and depreciation, based on Pakes (1986), whose estimates imply a lower rate of obsolescence than those of Lanjouw (1998) used in our baseline. The last exercise checks for robustness to functional form—we replace the log-logistic distribution for $F(s)$ with the Fréchet distribution, which also has fat-tailed characteristics.

All but one of the key parameters of the model, which we report in that table, are robust to these variations. The basic messages of the baseline model do not change.

4.4. *External validation exercises*

We conduct five different external validation checks on the analysis by comparing implications of our calibrated model against external information that played no role in its estimation. As we show below, these strongly validate our baseline calibration results.

4.4.1. Share of high-type inventions. Recall that high-type inventions are those that would not be developed without the patent incentive. Our baseline parameters put the share of high types among patent applicants at 39.9%. To our knowledge, the only external evidence on this is a study by Mansfield (1986), based on a survey of 100 manufacturing firms. Mansfield reports the percentage of *innovations* that firms in twelve different sectors claim would not have been developed without patent protection. The unweighted average is about 14%, and the weighted average across sectors (using patent applications in 2011 as weights) is 24%. Since innovations can be protected by other means outside our model (*e.g.*, secrecy), we translate Mansfield's figures into the share of high types in *patent applications* by dividing them by the aggregate patent propensity, which we compute to be 48% (see Supplementary Appendix G). This yields a share of high types in patent applications of between 28% and 50%. The share computed with our estimated parameters falls within that range.

43. Note that $\int_{s^{cc}}^{1} \chi(s)dF(s) = DA - LC$.

4.4.2. Elasticity of patent applications to fees. The second check is to compare the implied elasticity of patent applications with respect to pre-grant fees, based on our baseline calibration, to econometric estimates in the literature. We estimate this elasticity by making a small perturbation in pre-grant fees and recalculating the equilibrium, which yields −0.11. Thus the "demand for patents" is highly inelastic. de Rassenfosse and van Pottelsberghe (2012) use panel data from the U.S., Japanese and European patent offices to estimate this long-run elasticity. They also find the demand to be inelastic, and their elasticities are in the range −0.15 to −0.49 (depending on econometric specification).

4.4.3. Elasticity of patent grants to R&D. Third, we calculate the impact of changes in R&D on the number of patent grants in equilibrium and compare the implied elasticity from our baseline calibration to econometric estimates from the literature. Of course, the level of R&D in the model is endogenous, depending among other things on the parameters of the distribution of invention size, $F(s)$, and the distribution of costs, $G_s(\kappa)$. Both of these distributions are taken as exogenous in the model. To compute the implied elasticity, we introduce an exogenous shift in the distribution of invention sizes by marginally changing the scale parameter β_1. We then recalculate the equilibrium of the model, allowing for optimal adjustment of the level of R&D and the associated equilibrium change in the number of patents. This yields an estimate of the elasticity of patent grants with respect to (induced) R&D of 0.30. There is a large literature on estimating patent production functions at the firm level. A classic paper by Hall *et al.* (1986) estimates the long-run elasticity of granted patents to R&D in the range of 0.29–0.66. In a more recent paper that incorporates technology and product market spillovers as well as own R&D, Bloom *et al.* (2013, Table IV) estimate it at between 0.22 and 0.50. Again, our calibration-based estimate of the elasticity is within this range.

4.4.4. Cost saving from a registration system. Using our calibrated parameters, we can estimate the cost of processing a patent application under a pure registration system, where there is no examination ($e=0$). This yields $\gamma(0)=$ \$1,425. Given the USPTO patent operations budget per application of \$4,117, this implies that moving to a registration system would result in a cost saving of about 65%. To validate this estimate, we compute the cost savings using USPTO budgetary data for patent operations (*i.e.*, excluding trademarks). Specifically, we assume that under a registration system, the cost of patent examiner salaries would be avoided. We express this cost saving as a percentage of overall personnel costs, and assume that this percentage saving applies equally to all other (direct and allocated) costs. Translating this into the percentage saving in the overall patents operations budget, we get 59% (examiner-weighted average for 2005, 2010, and 2015). As an additional check, we compute how much pre-grant fees would decline if search and examination fees were set to zero (maintaining the filing fee, which would still be needed in a registration system). Under the assumption that fees are proportional to screening costs, this approach gives us an implied cost saving of about 75% (patent-weighted average for 2005, 2010, and 2015). Our simulated savings of 65% is within the range of these external estimates.

4.4.5. Ratio of licensing revenue to R&D. We use the calibrated parameters to compute the equilibrium level of licensing revenue and R&D costs in our model.[44] This yields a ratio of licensing revenue to R&D of 35.4%. For external validation, we use Robbins (2009), which is the most comprehensive analysis of U.S. corporate income from licensing intangible property. Her

44. See Supplementary Appendix F for the expressions.

analysis covers four types of intangibles, but we use her estimates for industrial property (covered by patents and trade secrets). Robbins estimates total licensing revenue for U.S. manufacturing corporations in 2002 at \$59.5 billion (current dollars). As our model corresponds to licensing to non-affiliated firms, we adjust by the percentage of licensing receipts accounted for by unaffiliated firms (32.4%, according to data on foreign transactions by U.S. firms from the Bureau of Economic Analysis). We then divide by manufacturing business enterprise R&D (\$101.3 billion in 2002, according to National Science Foundation data), scaled down by the aggregate patent propensity in manufacturing (since R&D also generates unpatented innovations). This yields the ratio of total licensing revenue to R&D of 39.3%, which is very close to the estimate based on our model.

5. ANALYSIS OF POLICY REFORMS

In this section, we analyse the impact of various policy reforms, including policies introduced recently by the U.S. Congress and the Supreme Court, as well as proposals by legal scholars active in the policy debates. We also analyse the impact of one of our theoretical results (frontloading fees) and the implementation of an optimal patent policy. We focus on how much, and the channels through which, each reform affects welfare, and highlight some indirect consequences that our integrated framework reveals. Table 3 presents the key results for reforms to patent fees and examination intensity. Table 4 covers reforms to the judicial rules governing patent litigation and licensing.

5.1. *Fee structure and examination intensity*

5.1.1. Frontloading fees. The first reform we analyze is frontloading all fees—shifting issuance and renewal fees from the post-grant to the pre-grant stage. This policy has become more relevant since the America Invents Act (AIA) of 2011, which gave the USPTO wide-ranging discretion over the level and structure of patent fees and the right to reinvest them to improve screening. We implement frontloading by setting post-grant fees to zero and by raising pre-grant fees by $b\phi_P$, which is the expected amount of post-grant fees an inventor would have paid, given obsolescence. Our theoretical analysis showed that frontloading is welfare-increasing (Proposition 7); the simulation will show by how much. At the same time, frontloading increases fee revenue, since low-type applicants pay the enhanced pre-grant fees, even if they are later rejected.[45] We simulate frontloading in two scenarios: one in which the extra fee revenue is returned to the government, and a second in which the revenue is reinvested by the patent office. This second version is a revenue-neutral increase in examination intensity.[46]

The results show that frontloading *by itself* slightly increases the share of high-type patents (as the theory showed, this works through discouraging low-type applicants, whose number declines by 0.6% (not reported)), but it has almost no measurable effect on welfare (because frontloading discourages only those low-type inventors whose inventions are not very valuable). This highlights the importance of the simulations as a way to check whether theoretical features actually have any quantitative significance, and thus relevance to policy analysis.

In contrast, we find that frontloading *with reinvestment* has a number of positive effects. It funds an increase in examination intensity from 47.9% to 56.0%, resulting in a welfare increase of

45. It has a countervailing effect by discouraging marginal (low-value) applications by low-type inventors. But the simulations show that the net effect is to raise fee revenue.

46. We solve for the value of e such that $FR(e)-\Gamma(e)=FR_0-\Gamma_0$, where FR_0 and Γ_0 are fee revenue and (total) examination cost in the baseline, while $FR(e)$ and $\Gamma(e)$ are fee revenue and examination cost when fees are frontloaded and the examination intensity is e.

TABLE 3
Policy reforms: fee structure and examination intensity

	e %	ϕ_A $\times 10^3$	SHP %	$1-\bar{y}$ %	ΔW %	ΔI_H %	ΔDWL %	ΔDA %	ΔLC %	$\Delta\Gamma$ %
Baseline	47.9	16.3	56.0	66.9	–	–	–	–	–	–
Frontload fees	47.9	22.5	56.2	66.9	0.0	0.0	0.0	0.0	0.0	−0.4
Frontload and reinvest	56.0	22.5	60.5	65.5	0.9	0.0	−12.6	−3.1	−3.2	41.6
Registration system	0.0	18.0	40.2	72.1	−5.3	0.2	74.7	11.6	11.8	−64.6
Optimal patent policy	82.6	132.4	83.8	56.5	3.0	−6.4	−56.8	−19.7	−19.9	796.3
Double PO budget/app	62.0	24.8	64.0	64.3	1.5	−0.1	−22.0	−5.8	−5.9	94.8
Triple PO budget/app	67.7	29.0	67.8	62.8	2.1	−0.3	−30.9	−8.8	−8.9	183.2

Notes: The column headed ϕ_A refers to the sum of pre-grant fees and the legal costs of drafting a patent application.

0.9%. The increase in welfare is driven by better detection (due to higher e) and better deterrence (due to higher e and ϕ_A), which combine to reduce deadweight loss before challenges (DWL) by 12.6%. Higher e also leads to more high-type innovation (by pushing up the challenge-credibility threshold and thereby reducing litigation exposure), but the magnitude is small. Although the reform reduces the deadweight loss avoided by challenges (DA), this effect is more than offset by a reduction in litigation costs (LC). These various welfare gains easily compensate for the substantial increase in examination costs, which go up by 41.6%.

5.1.2. Registration system. The second, more radical, reform we study is to replace the patent system with a simple registration system. While inventors would still have to register and pay a filing fee, there would be no substantive screening for patentability ($e=0$). This would also save the resource costs of screening. We therefore accompany the reduction in e by a reduction in fees that makes the reform revenue-neutral.[47]

In a widely known paper, Lemley (2001) argues that it may not make sense to spend resources on patent examination since very few patents are ever important enough to be litigated. He contends that it might be more efficient for the patent office to be "rationally ignorant" and to shift more of the burden of screening to the courts. Kieff (2003) actually proposes adoption of a registration system. One of the key points from our theoretical analysis is that screening is imperfect even if the courts are perfect, due to selection into litigation and particularly the fact that low-type inventors can preempt validity challenges by strategically charging a low license fee. But it is possible that the cost saving from a registration system outweighs the welfare costs of relying on *ex post* screening. Even though Proposition 8 showed that, when fees cover costs, it is optimal to *marginally* increase e above zero, this does not mean that the current system, with an estimated $e=47.9\%$, outperforms a registration system.

The table shows that introducing a registration system sharply reduces welfare, by 5.3%, despite a steep decline in patent office costs Γ and a modest increase in high-type innovation I^H. With no weeding out of, or deterrence of, type-L applications, deadweight loss before challenges jumps by 74.7%. And although the deadweight loss avoided by challenges increases significantly, litigation costs increase by even more, as the reduction in e lowers the challenge-credibility threshold. Also note that a registration system increases the level of trolling—72% of type-L patents now shield themselves from litigation by charging the challenge-preempting license fee. This illustrates why, in the presence of endogenous challenges, even mistake-free courts cannot substitute for patent office screening.

47. Note that, even though the value of ϕ_A given in Table 3 is higher than the one in the baseline, we also set $\phi_P=0$, so that the sum of fees decreases.

Finally, the sharp reduction in welfare implies that the social value of the registration regime (*i.e.*, a patent system with $e=0$ and lower fees) is *negative*. In other words, it is worse to have a registration system that gives temporary monopoly patent rights to all applicants than it is to have *no* patent rights. This highlights the critical point that a sufficiently poorly designed patent system can be worse than none at all.

5.1.3. Optimal patent policy and approximations to it. The third reform is to implement the patent policy that maximizes welfare. To do this, we compute via numerical optimization the welfare-maximizing combination of examination intensity and pre-grant fees, (e^*, ϕ_A^*) (from Proposition 7, we know frontloading fees is part of the optimal policy). We do not restrict attention to policies that are revenue-neutral or otherwise impose a budget constraint.

Moving to the optimal patent policy substantially increases welfare, by 3%. The optimal examination intensity is 82.6%, as compared to 47.9% in the current system. At the same time, however, optimal fees—including the legal costs of drafting a patent—are about \$130,000 per application, as compared to \$27,073 in the baseline (pre-grant fees including drafting costs plus post-grant fees). Patent office fees alone rise dramatically, from \$12,073 to \$117,400.[48] More intense screening and higher fees discourage low-type applicants—the share of high types among applicants rises from 40% to 47% (not shown)—and the increased scrutiny weeds out more low types among inventors that continue to apply, with the share of high types among granted patents reaching 84%. As a result, gross deadweight loss is reduced by about 57%, and trolling declines sharply. The optimal policy also pushes up the challenge-credibility threshold, thereby reducing litigation costs by 20%. Finally, although the examination cost *per application* is multiplied by a factor of 20, the total cost of examination increases by a comparatively smaller 796%, as applications plummet.

Critics of proposals to increase patent office fees point to their likely deleterious effects on innovation. Indeed, the optimal policy leads high-type applications to fall by 46% (not shown). The gross welfare from high-type innovation, I^H, falls by a more modest, but still considerable, 6.4% (the discouragement effect of higher fees is concentrated among low-value inventions). Compared to the current policy, the optimal policy thus trades a share of the welfare gains from high-type innovation for a large reduction in deadweight loss and litigation costs. This result underscores the tradeoff between screening and innovation that patent-policy design must confront.

In addition to the optimal policy, we consider two scenarios in which the patent office doubles or triples the expenditure on screening per application, and raises fees to cover the extra costs. The purpose of these counterfactuals is to allow us to assess how close, in terms of welfare gains, these more modest budgetary increases would get to the optimal patent policy.

The two rows below the optimal policy in Table 3 show that much of the welfare gains from an optimal policy can be achieved by revenue-neutral policies that require only modest increases in pre-grant fees. Doubling the budget per application (*e.g.* the time allotted to patent examiners) raises welfare by 1.5%, which is 50% of the welfare gain from the optimal system. Tripling the budget increases welfare by 2.1%, or 70% of the gain from the optimal system.[49] Finally, as with the optimal policy, the increase in total patent office costs associated with these reforms is smaller

48. Two qualifications in interpreting these results should be noted. First, our calibration of the examination cost function is based on a local measure of its elasticity. We are using it to extrapolate to the much higher optimal level of examination intensity. Second, with other parameter configurations than the one in our baseline, the optimal fee is sometimes even higher.

49. We should point out that throwing money at the problem is not the only way to improve detection. Another approach is to design better incentives. Schuett (2013b) theoretically studies the design of incentives for patent examiners in a setting where examination is a moral-hazard problem followed by an adverse-selection problem. He shows that

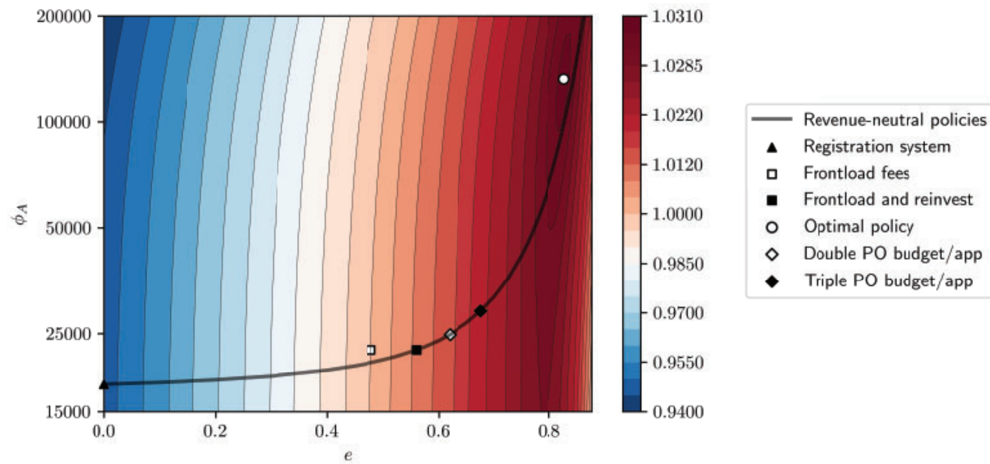

FIGURE 2

Map of iso-welfare curves with counterfactual patent policies.

than the increase in the per-application cost because applications decline, with Γ going up by 95% and 183%, respectively.

To get a broader view as to the welfare associated with different patent policies, Figure 2 presents a "heat map" of iso-welfare curves as a function of examination intensity (horizontal axis) and pre-grant fees including patent drafting costs (vertical axis, logarithmic scale), with post-grant fees set to zero. The color bar on the right hand side of the figure gives the normalized welfare level relative to the current system. We depict the optimal patent policy along with the other reforms discussed above, and plot the locus of revenue-neutral policy options. (Current patent policy cannot be depicted because it involves $\phi_P > 0$, but its welfare level corresponds roughly to the one for frontloading without reinvestment.) This can be helpful in thinking about what may be politically feasible policy choices (in terms of fees, for example) and understanding the welfare cost of not being more ambitious.

The figure also shows that fees have little impact on welfare unless e is very high, with iso-welfare curves being almost vertical up to a fee of about \$50,000. For a given level of e, the optimal fee is found at the point of tangency between a vertical at e and the iso-welfare curves. This conditionally optimal fee increases with e. The intuition is that, as examination is intensified, fees become a more effective deterrent against low types (recall that the application cutoff for low types is determined by $\phi_A/[b(1-e)]$). Moreover, the examination costs per application increase, so it also becomes more important to keep the number of applications in check. The figure thus highlights the complementarity between the two patent-policy instruments.

5.2. *Judicial rules governing litigation and licensing*

5.2.1. Patent Trial and Appeal Board (PTAB). In landmark legislation—the America Invents Act (2011)—the U.S. Congress expanded the scope of post-grant opposition procedures

intrinsic motivation plays a key role, and that monetary incentives can be made more effective by conditioning on the outcome of an internal review of examiners' decisions. In ongoing work, Matcham, Schankerman, and Drabik (2020) develop a structural model of the bargaining game between applicants and examiners in order to estimate the impact of intrinsic and extrinsic rewards on patent screening outcomes.

TABLE 4
Policy reforms: judicial rules governing litigation and licensing

	q^H %	q^L %	LR %	$1-\bar{y}$ %	ΔW %	ΔI_H %	ΔDWL %	ΔDA %	ΔLC %
Baseline	100.0	0.0	1.7	66.9	–	–	–	–	–
PTAB: preponderance of evidence	89.9	16.3	4.2	77.8	0.8	−0.3	−0.1	−36.8	−68.2
PTAB: clear and convincing evidence	97.8	32.7	3.4	74.5	0.8	0.0	0.0	−68.4	−73.8
Prohibiting negative fixed fees	100.0	0.0	2.1	66.9	3.6	−0.1	0.0	2550.4	24.8
Loser pays	100.0	0.0	3.0	59.0	−0.4	0.2	0.1	78.8	45.3

and established the Patent Trial and Appeal Board, an administrative law body within the USPTO.[50] This allows third parties to challenge the validity of granted patents at much lower cost than doing so in court. Estimates of the cost of a PTAB opposition are about \$350,000, as compared to litigation costs which (even for modestly valued patents) run into the millions (Supplementary Appendix G).[51] As discussed in Section 3.2.3, a reduction in litigation cost has an ambiguous effect on welfare: it saves costs per litigation, but it can also increase the rate of challenges, and thus the amount of litigation.

Because the PTAB spends less time and resources on reviewing patents, it is arguably more prone to adjudication errors than courts—upholding low-type patents, or revoking high-type patents.[52] Moreover, critics have alleged that the PTAB is too lax in revoking patents that are challenged. The main reason for this criticism is that the PTAB uses a weaker evidentary standard than the one universally used by federal district and appellate courts in patent cases: "preponderance of the evidence" rather than "clear and convincing evidence." There have been proposals to align standards by having the PTAB move to the more restrictive standard.

In this counterfactual, we model the PTAB review as lowering litigation costs from $l(s) = l_0 + l_1\Delta^H(s)$ to \$350,000 (independent of s), and introducing imperfect adjudication with a "preponderance of the evidence" standard. For this purpose, we generalize the model to allow for adjudication errors by specifying that the courts uphold high types with probability q^H and low types with probability $q^L < q^H$. This generalization is described in Section 3.2.4 and Supplementary Appendix C.

To microfound the probabilities q^H and q^L, in Supplementary Appendix H, we develop a model in which a Bayesian adjudicator (PTAB) receives signals that are correlated with the inventor's true type, and whose precision depends on the adjudicator's effort, which we set at the level of the examination intensity estimated in the baseline. Since PTAB review involves a panel of senior examiners and takes more time and evidence than the original examination, the resulting accuracy should be a lower bound. We set the adjudicator's prior at the share of high types among patentees estimated in the baseline, and assume that they uphold the patent if

50. PTAB trials include inter partes, post-grant, and covered business method patent reviews. These differ in various important aspects, such as the allowable statutory grounds to challenge and the scope for evidence discovery. PTAB reviews are held before a panel of three senior examiners. Prior to the AIA, internal challenges were adjudicated by one examiner via the inter partes re-examination procedure. The procedure was perceived as unattractive and thus rarely used (Graham *et al.*, 2002).

51. Initiating a PTAB challenge does not prevent a party from also bringing a patent challenge in court. In our model, we do not incorporate both avenues, so our counterfactual focuses on the reduction in litigation costs and the change in accuracy of adjudication.

52. PTAB proceedings are much shorter than typical federal court trials, being completed within 12 months, and they are much more limited in terms of the inputs provided by the contesting parties (*e.g.* very limited discovery of new evidence, typically only written submission rather than oral testimony, etc.). We believe these features make it likely that they have lower accuracy than courts (at least, compared to bench trials).

and only if their posterior that the patent is of high type exceeds a threshold determined by the evidentiary standard—50%, in the case of the "preponderance of the evidence" standard.

We find that the PTAB increases welfare by 0.75%. This is despite the fact that, counterintuitively, the deadweight loss avoided through patent challenges, *DA*, is cut by 37%, and only a quarter of it comes from eliminating low types (not reported). One might have expected that the introduction of PTAB, by lowering the challenge-credibility threshold, would *raise DA*. However, low types react by ramping up the use of challenge-preempting license contracts, thereby escaping scrutiny: trolling sharply *increases* from 67% to 78%. Moreover, the prevalence of trolling rises with invention value. Our analysis shows that, in the presence of PTAB, the most valuable 15% of low-type patents preempt challenges 100% of the time. This once again underlines the importance of modelling challenges as endogenous.

The explanation for the positive welfare effect lies in the PTAB's effect on litigation costs, which decline by 68%. This precipitous fall in litigation costs makes the PTAB welfare-enhancing, despite the fact that it eliminates only 17% as much of the deadweight loss from low types relative to perfect courts in the baseline. Finally, we find that the PTAB does have a negative effect on high-type innovation (I_H), which declines by 0.3%. Contrary to critics' fears, however, this effect is modest, suggesting that concerns the PTAB lowers the incentives for R&D investment may be overblown.

5.2.2. Strengthening the evidentiary standard for PTAB. In this counterfactual, we evaluate the impact of shifting from the current evidentiary standard used by the PTAB to the more restrictive "clear and convincing evidence" standard. This would make it harder to revoke patents—both type *L* and *H* patents—through the PTAB. For the simulations, we need to translate this evidentiary standard into a cutoff on the posterior above which the PTAB upholds the patent. In line with legal scholarship (McCauliff, 1982), we assume that, under the "clear and convincing evidence" standard, the PTAB must assign a posterior of at least 75% to the inventor being of low type in order to revoke the patent.

Relative to the baseline, the PTAB with the stronger evidentiary standard raises welfare by 0.79%. However, note that this is only a small gain relative to the PTAB under the existing standard. On the one hand, a tougher standard for revoking patents makes it less likely that the PTAB invalidates a type-*H* patent (which is a primary criticism of the PTAB), which raises high-type innovation slightly as compared to the current standard. On the other hand, it means that fewer type-*L* patents are invalidated, and less deadweight loss is avoided. This is evident by looking at the (calibrated) probabilities that type-*H* and type-*L* patentees win a challenge in the PTAB—given by q_H and q_L. With the current standard, 89.9% of high types and 16.3% of low types are upheld, but that rises to 97.8% and 32.7%, respectively, under the tougher standard.

5.2.3. Prohibiting negative fixed fees in license contracts. In a 2013 decision, the U.S. Supreme Court ruled that "pay-for-delay" settlements—according to which drug companies pay to resolve a patent dispute and delay entry by a generic firm—may be illegal under the antitrust laws (*FTC v. Actavis, Inc.*, 570 US 136 (2013)).[53] The FTC asserted that such agreements violate antitrust laws by restricting entry and raising prices (in effect, allowing the potential entrant to share in the monopoly profits from the patent). The court recognized that such arrangements pose a trade-off: they may prevent costly litigation, but they can also be used to prevent challenges of patents whose validity is questionable and thereby raise prices (or royalties). While the court did

53. For discussion of this issue before the Actavis decision, see Hovenkamp *et al.* (2003) and Shapiro (2003).

not prohibit licensing agreements with negative fixed payments, it expressed concern when such payments are not proportionate to the expected litigation costs that might be saved. Motivated by the Actavis decision, this counterfactual evaluates the welfare effect of prohibiting negative fixed fees.

This judicial reform has a large, positive impact on welfare—a 3.6% gain. The main source of the gain is the avoided deadweight loss, which increases by 2550%. When negative fixed fees are prohibited, the low-type patentee can no longer charge the competitor the high royalty rate and compensate her for refraining from a challenge through the fixed fee. To preempt challenges, the low type must instead charge a lower royalty, which reduces prices and deadweight loss. At the same time, litigation costs rise by about 25%. This tradeoff between lower prices and more litigation exactly mirrors the debate in the Actavis case, but our simulations allow us to assess the net effect. Unlike in all the other counterfactuals we consider, the net benefit from challenges—or more precisely, from challenge credibility—becomes positive, as DA now exceeds LC. As a result, both welfare and the social value of the patent system increase substantially.

One qualification with respect to these results is that our model may exaggerate the incidence of licensing, and thus the effectiveness of policies targeting license contracts. In our model, almost all patents that do not become obsolete end up being licensed. This is unlikely to be the case in practice. However, our external validation exercise on the ratio of licensing revenue to R&D spending suggests that the model accurately captures the *importance* of licensing, if not its *incidence*.[54] The welfare impact of policies that target licensing arguably depends more strongly on the importance than on the incidence of licensing.

5.2.4. Changing the allocation of legal costs. The default rule for patent cases in the U.S. is that each party bears their own legal costs. In two important patent cases in 2014, the U.S. Supreme Court relaxed the conditions for the shifting of legal costs to the losing party (Highmark Inc. v. Allcare Health Management Systems, Inc., 134 S.C. 1744; Octane Fitness, Inc. v. ICON Health & Fitness, Inc., 134 S.C. 1756). Some commentators welcomed this development as they expected it to discourage so-called "patent trolls." However, the law and economics literature shows that a loser-pays rule generally increases litigation (Spier, 2007).

Our last counterfactual simulates the welfare effects of cost shifting. To do this, we augment the model to allow for a fraction ζ of the winner's legal costs to be allocated to the loser, as described in Section 3.2.5 and Supplementary Appendix E. We find that cost shifting of any degree $\zeta > 0$ reduces welfare, compared to the baseline with $\zeta = 0$. Table 4 reports the results for full cost shifting, $\zeta = 1$. Such a rule is found to decrease welfare by 0.4%. A loser-pays rule benefits high types and thus leads to an increase in I_H. By lowering the challenge-credibility threshold and raising the rate of challenges, it also leads to substantially more litigation, however: the litigation rate almost doubles, to about 3%, and litigation costs rise by 45%. As a result, although a loser-pays rule hurts low types and benefits high types, its overall welfare effect is negative. Finally, it is worth noting that the magnitude of the welfare decrease is such that the social value of the patent system drops below zero (not reported in the table).

6. CONCLUSION

This article develops a framework to examine how governments can improve the quality of patent screening, focusing on patent office examination, pre- and post-grant fees, and challenges in the

54. A major reason why valuable patents are not licensed in practice is the presence of transaction costs. Since transaction costs disproportionately affect the licensing of low-value patents, it is plausible that the model better reflects the importance than the incidence of licensing.

courts. The theoretical analysis yields three key results. First, even if courts are mistake-free, they cannot eliminate all bad patents that are issued because not all such patents are challenged. This result raises serious doubts about over-reliance on the court system to weed out bad patents. Second, it is welfare improving to frontload fees (especially if the additional revenue raised is reinvested to intensify patent examination). Third, there is no theoretical presumption that the level of patent challenges is too low. Private incentives to challenge can be either smaller or larger than the social incentives.

We calibrate the model and estimate the key structural parameters, using U.S. patent, litigation, and other data, and study the welfare effects of policy reforms. The quantitative analysis confirms that patent screening is very imperfect—nearly half of granted patents are on inventions that would have been developed even without patent rights, imposing unnecessary social costs. The findings indicate that the problem is real and we need to develop effective policies to address it. Our quantitative analysis of policy reforms identifies those reforms that generate the largest welfare gains, and others that impose welfare losses.

We believe that the framework developed in this article can be used to study other patent reforms—including policies encouraging litigation insurance—and to evaluate how combinations of reforms affect welfare. Avenues for future research include analyzing how the presence of imperfect courts affects the appropriate patentability standard, and studying the optimal accuracy of courts as part of a broader patent policy. In principle, our framework could be adapted to and calibrated for other patent systems. This might allow one to analyze the welfare effects of more ambitious, international patent reforms, such as harmonized screening.

Data Availability Statement

The data and code underlying this research are available on Zenodo at https://doi.org/10.5281/zenodo.5555850.

Acknowledgments. We are very grateful to the editor, Dirk Krueger, and five anonymous referees for their detailed and very constructive comments and suggestions. We also thank Özlem Bedre, Marc Bourreau, Jay Pil Choi, Vincenzo Denicolò, Alberto Galasso, John Golden, Dan Klerman, Patrick Legros, Matt Mitchell, Geert van Moer, Pierre Régibeau, Monika Schnitzer, Tim Simcoe, David Schwartz, and Mariagrazia Squicciarini for helpful comments on an earlier version of this paper, circulated under the title "Screening for Patent Quality." We also benefitted from seminar participants at, among others, Helsinki, Tel Aviv, Duke, Max Planck Institute for Innovation and Competition, and ETH Zurich. We thank William Matcham, Mia Lu, and Haoran Pan for their excellent research assistance on parts of the quantitative analysis. Mark Schankerman thanks the Tilburg Law and Economics Center (TILEC) for its hospitality. Part of this research was undertaken while he was the Ronald Coase Visiting Professor at TILEC. Mark Schankerman gratefully acknowledges support from the European Research Council, under the Grant Agreement no. 340903, and the Sackler Fellowship at Tel Aviv University. All errors are our own.

Supplementary Data

Supplementary data are available at *Review of Economic Studies* online. And the replication packages are available at https://dx.doi.org/10.5281/zenodo.5555850.

APPENDIX

A. PROOFS

Proof of Proposition 1. By definition, high-type innovation requires $\Pi_{nc}(v) > \pi(v)$. For (i), observe that if $\phi_A = 0$ or $e = 0$, then for (2) to hold we must have $\phi_P \geq \Delta^H(v)$ and thus $\Pi_{nc}(v) \leq \pi(v)$. For (ii), note first that maximizing high-type innovation conditional on deterrence requires that (2) hold with equality; otherwise high-type innovation could be increased by lowering ϕ_A without jeopardizing deterrence. Fixing ϕ_A such that (2) holds with equality yields $\Pi_{nc}(v) = \pi(v) + e(\Delta^H(v) - \phi_P)$, which is increasing in e for any $\phi_P \leq \Delta^H(v)$ and decreasing in ϕ_P for any $e \geq 0$. For (iii), fix $\Pi_{nc}(v) = \Pi > \pi(v)$ so that $\phi_A = \Delta^H(v) - \phi_P - (\Pi - \pi(v))$. Then for (2) to hold we need $(1-e)(\Delta^H(v) - \phi_P) \leq \pi(v) + \Delta^H(v) - (\Pi - \pi(v))$, or $e \geq (\Pi - \pi(v))/(\Delta^H(v) - \phi_P)$, the right-hand side of which is increasing in ϕ_P. ‖

Proof of Proposition 2. Suppose, to the contrary, that there is a policy $(e,\phi_A,\phi_P)\notin S_{nc}$ for which there exists an equilibrium in which no low types apply but some high types do; the latter requires that $\phi_A+\phi_P<\Delta^H(v)$. Since beliefs on the equilibrium path are derived from equilibrium strategies, the competitor correctly infers that any patent that is issued must be of high type, and by sequential rationality, she does not challenge. But then, a low-type inventor can obtain $(1-e)(\Delta^H(v)-\phi_P)-\phi_A>0$ by applying, where the inequality follows from the fact that $\phi_A+\phi_P<\Delta^H(v)$ and $(e,\phi_A,\phi_P)\notin S_{nc}$, leading to a contradiction. ‖

Proof of Proposition 3. The result is a direct application of Proposition 9 in Supplementary Appendix C. The assumption that $\phi_A<(1-e)(\Delta^L(v)-\phi_P)$ implies $v>\tilde{v}$. Because $q^H=1$, we have $(1-q^H)\Delta_C(v)<l_C(v)$, and thus $v<v_1$ and $R^H(v)=\Delta_C(v)$. For $l_C(v)\leq\Delta_C(v)$, we then have $R^H(v)+m(v)\geq\Delta^L(v)$, and hence $v>\hat{v}$. The assumption that (4) holds for $R^H=\Delta_C(v)$ and $x=0$ implies $v>v^{nc}$. Thus, we are in case (f) of Proposition 9, with $v<v_1$, so that the unique equilibrium has $(\alpha,x,y)=(1,\tilde{x}(v),\tilde{y}(v))$. ‖

Proof of Proposition 4. The patent policy (e,ϕ_A,ϕ_P) satisfies the conditions of Proposition 3. Using the expression for $\tilde{y}(v)$ from Proposition 3, we have

$$\chi=(G_v(\Pi_{cc}(v))-G_v(\pi(v)))\tilde{x}(v)\left[\frac{l_C(v)[D(v)-l_C(v)-l_I(v)]}{\Delta_C(v)-l_C(v)}-(l_C(v)+l_I(v))\right]$$
$$=(G_v(\Pi_{cc}(v))-G_v(\pi(v)))\tilde{x}(v)\left(\frac{l_C(v)D(v)-(l_C(v)+l_I(v))\Delta_C(v)}{\Delta_C(v)-l_C(v)}\right).$$

Using the equilibrium value of $\tilde{x}(v)$ yields the expression for χ in the proposition. The result on the sign of χ is due to the fact that $\Pi_{cc}(v)>\pi(v)$ for $\phi_A<(1-e)(\Delta^L(v)-\phi_P)$ because

$$\phi_A+\phi_P\leq\frac{\phi_A}{1-e}+\phi_P<\Delta^L(v)=(1-\tilde{x})\Delta^H(v)-\tilde{x}l_I(v)\leq\Delta^H(v)-\tilde{x}l_I(v),$$

where the equality follows from low types' indifference between $R=l_C(v)$ and $R=\Delta_C(v)$ and the last inequality from the fact that $\tilde{x}(v)\geq0$ for $\Delta_C(v)\geq l_C(v)$. ‖

Proof of Proposition 5. Welfare in the absence of a patent system is $\int_{\underline{v}}^{\overline{v}}\int_{\underline{\kappa}}^{\pi(v)}(v-\kappa)dG_v(\kappa)dF(v)$. Consider a patent policy $(e,\phi_A,\phi_P)=(\varepsilon,(1-\varepsilon)\Delta^H(\overline{v}),0)$, where $\varepsilon\in[0,1)$. By construction, $\hat{v}=\overline{v}$ under such a policy, and because no low types apply, challenges are not credible for any v. Let $\tilde{W}(\varepsilon)$ denote the associated welfare, given by

$$\tilde{W}(\varepsilon)=\int_{\underline{v}}^{\overline{v}}\int_{\underline{\kappa}}^{\pi(v)}(v-\kappa)dG_v(\kappa)dF(v)+\int_{v^*(\varepsilon)}^{\overline{v}}\int_{\pi(v)}^{\tilde{\Pi}_{nc}(v)}(v-D(v)-\kappa-\gamma(\varepsilon))dG_v(\kappa)dF(v),$$

where $\tilde{\Pi}_{nc}(v)\equiv\pi(v)+\Delta^H(v)-(1-\varepsilon)\Delta^H(\overline{v})$ and $v^*(\varepsilon)$ is defined by $\Delta^H(v^*)=(1-\varepsilon)\Delta^H(\overline{v})$. Notice that $v^*(0)=\overline{v}$ and $\tilde{W}(0)=\int_{\underline{v}}^{\overline{v}}\int_{\underline{\kappa}}^{\pi(v)}(v-\kappa)dG_v(\kappa)dF(v)$. Thus, to establish the claim it suffices to show that $\tilde{W}(\varepsilon)>\tilde{W}(0)$ for some $\varepsilon\in(0,1)$.

By assumption, $\Delta^H(\overline{v})>\gamma(0)$. Since γ is strictly increasing and $\lim_{e\to1}\gamma(e)=\infty$, by continuity there exists $\bar{\varepsilon}>0$ defined by $(1-\bar{\varepsilon})\Delta^H(\overline{v})=\gamma(\bar{\varepsilon})$, such that $(1-\varepsilon)\Delta^H(\overline{v})>\gamma(\varepsilon)$ if and only if $\varepsilon<\bar{\varepsilon}$. We then have, for $v\in[v^*(\bar{\varepsilon}),\overline{v}]$ and $\kappa\in[\pi(v),\pi(v)+\Delta^H(v)-(1-\bar{\varepsilon})\Delta^H(\overline{v})]$,

$$v-D(v)-\kappa-\gamma(\bar{\varepsilon})\geq v-D(v)-[\pi(v)+\Delta^H(v)-(1-\bar{\varepsilon})\Delta^H(\overline{v})]-\gamma(\bar{\varepsilon})$$
$$=v-D(v)-\pi(v)-\Delta^H(v)\geq0,$$

where the first inequality follows from $\Delta^H(v)-(1-\bar{\varepsilon})\Delta^H(\overline{v})\geq0$ for all $v\geq v^*(\bar{\varepsilon})$ (note that it is strict for $v>v^*(\bar{\varepsilon})$ and $\kappa<\pi(v)+\Delta^H(v)-(1-\bar{\varepsilon})\Delta^H(\overline{v})$), the equality from the definition of $\bar{\varepsilon}$, and the second inequality from Assumption 1.

Because Δ^H is strictly increasing,

$$\frac{dv^*}{d\varepsilon}=-\frac{\Delta^H(\overline{v})}{d\Delta^H(v)/dv}<0.$$

The fact that $\bar{\varepsilon}>0$ then implies $v^*(\bar{\varepsilon})<\overline{v}$ and hence

$$\tilde{W}(\bar{\varepsilon})-\tilde{W}(0)=\int_{v^*(\bar{\varepsilon})}^{\overline{v}}\int_{\pi(v)}^{\pi(v)+\Delta^H(v)-(1-\bar{\varepsilon})\Delta^H(\overline{v})}(v-D(v)-\kappa-\gamma(\bar{\varepsilon}))dG_v(\kappa)dF(v)>0.$$

‖

Proof of Lemma 1. Differentiating $\underline{\lambda}(v)$, defined in (13), with respect to v yields

$$\frac{d\underline{\lambda}}{dv}=\frac{\left[\Pi_{cc}'g_v(\Pi_{cc})+\partial G_v(\Pi_{cc})/\partial v-\left(\pi'g_v(\pi)+\partial G_v(\pi)/\partial v\right)\right](G_v(\Pi_{cc})-eG_v(\pi))}{(G_v(\Pi_{cc})-eG_v(\pi))^2}$$
$$-\frac{(G_v(\Pi_{cc})-G_v(\pi))\left[\Pi_{cc}'g_v(\Pi_{cc})+\partial G_v(\Pi_{cc})/\partial v-e\left(\pi'g_v(\pi)+\partial G_v(\pi)/\partial v\right)\right]}{(G_v(\Pi_{cc})-eG_v(\pi))^2}\leq0,$$

$$\Leftrightarrow G_v(\Pi_{cc})(1-e)\big(\pi' g_v(\pi)+\partial G_v(\pi)/\partial v\big) \geq G_v(\pi)(1-e)\big(\Pi'_{cc} g_v(\Pi_{cc})+\partial G_v(\Pi_{cc})/\partial v\big),$$

which, after simplification, yields (15). ‖

Proof of Proposition 6. The result is a direct application of Lemma 18 and Proposition 9 in Supplementary Appendix C. Because $q^H=1$, Assumption 3 implies that $(1-q^H)\Delta_C(v)\geq l_C(v)$ for all $v\in[\underline{v},\overline{v}]$, and thus $v_1=\overline{v}$ and $R^H(v)=\Delta_C(v)$ for all v. By Lemma 18, we have $\hat{v}<v^{cc}$ and $\tilde{v}<v^{cc}$ for $\phi_A<(1-e)(\Delta_L(v^{cc})-\phi_P)$. Since $v^{cc}\leq v^{nc}$, we can thus ignore cases (d) and (e) of Proposition 9. Cases (a), (b), (c), and (f) with $v<v_1$, establish the result. ‖

Proof of Proposition 7. The assumption that $\phi_A+\phi_P\leq(1-e)\Delta^L(v^{cc})$ ensures that the policy $(e,\phi_A+\phi_P,0)$ (and thus also the policy (e,ϕ_A,ϕ_P)) satisfies the conditions of Proposition 6, and hence that welfare is given by W in (16). Letting $W=W(e,\phi_A,\phi_P)$ and $\phi=\phi_A+\phi_P$, what we need to show is that $W(e,\phi-\phi_P,\phi_P)$ is decreasing in ϕ_P for $\phi_P\in[0,\phi]$. Note that, for $\phi_A+\phi_P=\phi$, we have $\Pi_{nc}(v)=\pi(v)+\Delta^H(v)-\phi$ and $\Pi_{cc}(v)=\pi(v)+\Delta^H(v)-\tilde{x}(v)l_I(v)-\phi$, neither of which depend on ϕ_P (since $\tilde{x}(v)=(\Delta_C(v)-l_C(v))/(\Delta^H(v)+l_I(v))$). Furthermore, v^* and v^{cc} are also invariant to ϕ_P, while $\hat{v}$ is defined by

$$\Delta^H(\hat{v})=\frac{\phi-\phi_P}{1-e}+\phi_P.$$

Thus

$$\frac{dW(e,\phi-\phi_P,\phi_P)}{d\phi_P}=\frac{\partial\hat{v}}{\partial\phi_P}G_{\hat{v}}(\pi(\hat{v}))\big((1-e)D(\hat{v})+\gamma(e)\big)f(\hat{v})\leq 0,$$

where the inequality follows from the fact that, because Δ^H is strictly increasing,

$$\frac{\partial\hat{v}}{\partial\phi_P}=-\frac{e}{(1-e)(d\Delta^H(v)/dv)}\leq 0,$$

with strict inequality for $e>0$ and $\Delta^H(\underline{v})<\phi_A/(1-e)+\phi_P$, so that $\underline{v}<\hat{v}$. ‖

Proof of Proposition 8. What we need to show is that $\partial W/\partial e>0$ when evaluated at $(e,\phi_A,\phi_P)=(0,\phi,0)$. The assumption that $\phi\leq(1-e)\Delta^H(v^{cc})$ implies that $\partial W/\partial e$ is given by (17). By (19), the first and second terms in (17) (labelled *detection* and *deterrence*) are positive, and strictly so for at least the former. The assumption that $\Delta_C(v^{cc})\geq D(v^{cc})l_C(v^{cc})/(l_C(v^{cc})+l_I(v^{cc}))$ implies that $\chi(v^{cc})\leq 0$ by Proposition 4. If $\gamma'(0)=0$, a sufficient condition for the sign of $\partial W(0,\phi,0)/\partial e$ to be strictly positive therefore is

$$\frac{\partial v^{cc}}{\partial e}\int_{\Pi_{cc}(v^{cc}))}^{\Pi_{nc}(v^{cc}))}(v^{cc}-D(v^{cc})-\kappa-\gamma(e))dG_{v^{cc}}(\kappa)f(v^{cc})\geq 0$$

when evaluated at $(e,\phi_A,\phi_P)=(0,\phi,0)$. Since $\partial v^{cc}/\partial e\geq 0$ by (19), it suffices to show that the integrand is positive for all $\kappa\in[\Pi_{cc}(v^{cc}),\Pi_{nc}(v^{cc})]$, and hence that

$$v^{cc}-D(v^{cc})-\Pi_{nc}(v^{cc})-\gamma(0)=v^{cc}-D(v^{cc})-\pi(v^{cc})-\Delta^H(v^{cc})+\phi-\gamma(0)\geq 0,$$

which, by Assumption 1, is true if $\phi\geq\gamma(0)$. ‖

B. METRICS USED IN QUANTITATIVE ANALYSIS

This appendix contains the expressions defining the metrics reported in Tables 2–4. The first set of indicators is:

$$SHA=A^H/(A^L+A^H)\qquad SHP=A^H/((1-e)A^L+A^H)$$

$$SP_{cc}=\frac{bP_{cc}}{A^H+(1-e)A^L}\qquad \bar{y}=\frac{\int_{s^{cc}}^1\tilde{y}(s)dF(s)}{1-F(s^{cc})},$$

where P_{cc} denotes the number of patents above s^{cc}, given by

$$P_{cc}=\int_{s^{cc}}^1(G_s(\Pi_{cc}(s))-eG_s(b\pi(s)))dF(s).$$

The second set of indicators, representing the various welfare components, is:

$$I^L=\int_0^1\int_0^{b\pi(s)}(bv(s)-\kappa)dG_s(\kappa)dF(s)\qquad I^H=\int_{s^*}^1\int_{b\pi(s)}^{\Pi(s)}(bv(s)-\kappa)dG_s(\kappa)dF(s)$$

$$DWL=\int_{\hat{s}}^1 b(1-e)G_s(b\pi(s))D(s,\rho^H(s))dF(s)+\int_{s^*}^1 b(G_s(\Pi(s))-G_s(b\pi(s)))D(s,\rho^H(s))dF(s)$$

$$DA = \int_{s^{cc}}^{1} b(1-e)G_s(b\pi(s))\left[\tilde{x}(s)\tilde{y}(s)D(s,\rho^H(s)) + (1-\tilde{y}(s))[D(s,\rho^H(s)) - D(s,\rho^L(s))]\right]dF(s)$$

$$LC = \int_{s^{cc}}^{1} b(1-e)G_s(b\pi(s))\tilde{x}(s)\tilde{y}(s)2l(s)dF(s) + \int_{s^{cc}}^{1} b(G_s(\Pi(s)) - G_s(\pi(s)))\tilde{x}(s)2l(s)dF(s)$$

$$\Gamma = (A^L + A^H)\gamma(e).$$

These expressions are valid for the baseline, with perfect courts ($q^H = 1$ and $q^L = 0$). For the expressions with imperfect courts, see Supplementary Appendix F.

REFERENCES

ACEMOGLU, D., AKCIGIT, U., ALP, H. et al. (2018), "Innovation, Reallocation, and Growth", *American Economic Review*, **108**, 3450–3491.

AKCIGIT, U., CELIK, M. A. and GREENWOOD, J. (2016), "Buy, Keep, or Sell: Economic Growth and the Market for Ideas", *Econometrica*, **84**, 943–984.

AKCIGIT, U., HANLEY, D. and SERRANO-VELARDE, N. (2021), "Back to Basics: Basic Research Spillovers, Innovation Policy, and Growth", *Review of Economic Studies*, **88**, 1–43.

AKCIGIT, U., HANLEY, D. and STANTCHEVA, S. (2016), "Optimal Taxation and R&D Policies" (NBER Working Paper No. 22908).

ALLISON, J. R., LEMLEY, M. A. and SCHWARTZ, D. L. (2014), "Understanding the Realities of Modern Patent Litigation", *Texas Law Review*, **92**, 1769–1801.

American Intellectual Property Law Association (2005), *2005 Report of the Economic Survey* (Arlington, VA).

American Intellectual Property Law Association (2011), *2011 Report of the Economic Survey* (Arlington, VA).

American Intellectual Property Law Association (2017), *2017 Report of the Economic Survey* (Arlington, VA).

ARUNDEL, A. and KABLA, I. (1998), "What Percentage of Innovations are Patented? Empirical Estimates for European Firms", *Research Policy*, **27**, 127–141.

ATAL, V. and BAR, T. (2014), "Patent Quality and a Two-tiered Patent System", *Journal of Industrial Economics*, **62**, 503–540.

BAGNOLI, M. and BERGSTROM, T. (2005), "Log-concave Probability and Its Applications", *Economic Theory*, **26**, 445–469.

BECKER, R., GRAY, W. and MARVAKOV, J. (2016), "NBER-CES Manufacturing Industry Database: Technical Notes" (National Bureau of Economic Research).

BESSEN, J. (2008), "The Value of U.S. Patents by Owner and Patent Characteristics", *Research Policy*, **37**, 932–945.

BESSEN, J. and MASKIN, E. (2009), "Sequential Innovation, Patents, and Imitation", *RAND Journal of Economics*, **40**, 611–634.

BLOOM, N., SCHANKERMAN, M. and VAN REENEN, J. (2013), "Identifying Technology Spillovers and Product Market Rivalry", *Econometrica*, **81**, 1347–1393.

BOLDRIN, M. and LEVINE, D. K. (2008), *Against Intellectual Monopoly* (Cambridge, MA: Cambridge University Press).

CAILLAUD, B. and DUCHÊNE, A. (2011), "Patent Office in Innovation Policy: Nobody's Perfect", *International Journal of Industrial Organization*, **29**, 242–252.

CARLEY, M., HEGDE, D. and MARCO, A. (2015), "What is the Probability of Receiving a U.S. Patent?", *Yale Journal of Law and Technology*, **17**, 203–223.

CHOI, J. P. (2005), "Live and Let Live: A Tale of Weak Patents", *Journal of the European Economic Association*, **3**, 724–733.

COHEN, W. M., NELSON, R. R. and WALSH, J. P. (2000), "Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (Or Not)" (NBER Working Paper No. 7552).

CORNELLI, F. and SCHANKERMAN, M. (1999), "Patent Renewals and R&D Incentives", *RAND Journal of Economics*, **30**, 197–213.

CROUCH, D. (2008), "The Rising Size and Complexity of the Patent Document" (University of Missouri School of Law Legal Studies Research Paper No. 2008-04).

DE RASSENFOSSE, G. and VAN POTTELSBERGHE, B. (2012), "On the Price Elasticity of Demand for Patents", *Oxford Bulletin of Economics and Statistics*, **74**, 58–77.

DUFFY, J. F. (2007), "Inventing Invention: A Case Study of Legal Innovation", *Texas Law Review*, **86**, 1–72.

EISENBERG, R. S. (2004), "Obvious to Whom – Evaluating Inventions from the Perspective of PHOSITA", *Berkeley Technology Law Journal*, **19**, 885–906.

FARRELL, J. and MERGES, R. P. (2004), "Incentives to Challenge and Defend Patents: Why Litigation Won't Reliably Fix Patent Office Errors and Why Administrative Patent Review Might Help", *Berkeley Technology Law Journal*, **19**, 943–970.

FARRELL, J. and SHAPIRO, C. (2008), "How Strong Are Weak Patents?", *American Economic Review*, **98**, 1347–1369.

FAULI-OLLER, R. and SANDONIS, J. (2002), "Welfare Reducing Licensing", *Games and Economic Behavior*, **41**, 192–205.

FEDERAL TRADE COMMISSION (2003), To Promote Innovation: The Proper Balance Between Competition and Patent Law and Policy" Report, available at http://www.ftc.gov/os/2003/10/innovationrpt.pdf.
FEDERAL TRADE COMMISSION (2011), "The Evolving IP Marketplace: Aligning Patent Notice and Remedies with Competition. Report" available at https://www.ftc.gov/sites/default/files/documents/reports/evolving-ip-marketplace-aligning-patent-notice-and-remedies-competition-report-federal-trade/110307patentreport.pdf.
FRAKES, M. D. and WASSERMAN, M. F. (2017), "Is the Time Allocated to Review Patent Applications Inducing Examiners to Grant Invalid Patents? Evidence from Microlevel Application Data", *Review of Economics and Statistics*, **99**, 550–563.
GALASSO, A., MITCHELL, M. and VIRAG, G. (2016), "Market Outcomes and Dynamic Patent Buyouts", *International Journal of Industrial Organization*, **48**, 207–243.
GRAHAM, S. J., HALL, B. H., HARHOFF, D. and MOWERY, D. C. (2002), "Post-issue Patent 'Quality Control': A Comparative Study of US Patent Re-examinations and European Patent Oppositions" (NBER Working Paper no. 8807).
HALL, B. H., GRILICHES, Z. and HAUSMAN, J. A. (1986), "Patents and R and D: Is there a lag?", *International Economic Review*, **27**, 265–283.
HELLER, M. A. and EISENBERG, R. S. (1998), "Can Patents Deter Innovation? The Anticommons in Biomedical Research", *Science*, **280**, 698–701.
HOPENHAYN, H. A., LLOBET, G. and MITCHELL, M. F. (2006), "Rewarding Sequential Innovators: Prizes, Patents, and Buyouts", *Journal of Political Economy*, **114**, 1041–1068.
HOVENKAMP, H., JANIS, M. and LEMLEY, M. A. (2003), "Anticompetitive Settlement of Intellectual Property Disputes", *Minnesota Law Review*, **87**, 1719–1766.
HUNT, R. M. (2004), "Patentability, Industry Structure, and Innovation", *Journal of Industrial Economics*, **52**, 401–425.
JAFFE, A. B. and LERNER, J. (2004), *Innovation and Its Discontents: How Our Broken Patent System is Endangering Innovation and Progress, and What to Do About It* (Princeton, NJ: Princeton University Press).
KIEFF, F. S. (2003), "The Case for Registering Patents and the Law and Economics of present Patent-obtaining Rules", *Boston College Law Review*, **45**, 55–123.
KITCH, E. W. (1966), "Graham v. John Deere Co.: New Standards for Patents", *Supreme Court Review*, **1966**, 293–346.
KOU, Z., REY, P. and WANG, T. (2013), "Non-obviousness and Screening", *Journal of Industrial Economics*, **61**, 700–732.
LANJOUW, J. O. (1998), "Patent Protection in the Shadow of Infringement: Simulation Estimations of Patent Value", *Review of Economic Studies*, **65**, 671–710.
LANJOUW, J. O. and SCHANKERMAN, M. (2001), "Characteristics of Patent Litigation: A Window on Competition", *RAND Journal of Economics*, **32**, 129–151.
LEMLEY, M. A. (2001), "Rational Ignorance at the Patent Office", *Northwestern University Law Review*, **95**, 1497–1532.
LEMLEY, M. A. and SHAPIRO, C. (2005), "Probabilistic Patents", *Journal of Economic Perspectives*, **19**, 75–98.
LLOBET, G., PARRA, A. and SUAREZ, J. (2021), "On the Interaction Between Patent Screening and Its Enforcement" (Mimeo).
MANSFIELD, E. (1985), "How Rapidly Does New Industrial Technology Leak Out?", *Journal of Industrial Economics*, **34**, 217–223.
MANSFIELD, E. (1986), "Patents and Innovation: An Empirical Study", *Management Science*, **32**, 173–181.
MATCHAM, W., SCHANKERMAN, M. and DRABIK, J. (2020), "The Patent Bazaar: Incentives, Bargaining, and Screening in the Patent System" (Mimeo).
MCCAULIFF, C. M. (1982), "Burdens of Proof: Degrees of Belief, Quanta of Evidence, or Constitutional Guarantees?", *Vanderbilt Law Review*, **35**, 1293–1335.
MERGES, R. P. (1992), "Uncertainty and the Standard of Patentability", *High Technology Law Journal*, **7**, 1–70.
MEURER, M. J. (1989), "The Settlement of Patent Litigation",*RAND Journal of Economics*, **20**, 77–91.
MEURER, M. J. and STRANDBURG, K. J. (2008), "Patent Carrots and Sticks: A Model of Nonobviousness", *Lewis & Clark Law Review*, **12**, 547–578.
MITCHELL, M. and SCHUETT, F. (2020), "Is This Obvious?" (Mimeo).
National Academy of Sciences (2004), "A Patent System for the 21st Century." Available at http://www.nap.edu/catalog/10976.html.
O'DONOGHUE, T. (1998), "A Patentability Requirement for Sequential Innovation", *RAND Journal of Economics*, **29**, 654–679.
PAKES, A. (1986), "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks", *Econometrica*, **54**, 755–784.
REINGANUM, J. F. and WILDE, L. L. (1986), "Settlement, Litigation, and the Allocation of Litigation Costs", *RAND Journal of Economics*, **17**, 557–566.
ROBBINS, C. A. (2009), "Measuring Payments for the Supply and Use of Intellectual Property", in Reinsdorf, M. and Slaughter, M. J. (eds), *International Trade in Services and Intangibles in the Era of Globalization* (Chicago, IL: University of Chicago Press) 139–171.
SCHANKERMAN, M. and PAKES, A. (1986), "Estimates of the Value of Patent Rights in European Countries During the Post-1950 Period", *Economic Journal*, **96**, 1052–1076.
SCHUETT, F. (2013a), "Inventors and Impostors: An Analysis of Patent Examination with Self-selection of Firms into R&D", *Journal of Industrial Economics*, **61**, 660–699.
SCHUETT, F. (2013b), "Patent Quality and Incentives at the Patent Office", *RAND Journal of Economics*, **44**, 313–336.

SCOTCHMER, S. (1996), "Protecting Early Innovators: Should Second-generation Products be Patentable?", *RAND Journal of Economics*, **27**, 322–331.

SCOTCHMER, S. (1999), "On the Optimality of the Patent Renewal System", *RAND Journal of Economics*, **30**, 181–196.

SCOTCHMER, S. (2004), *Innovation and Incentives* (Cambridge, MA: MIT Press).

SCOTCHMER, S. and GREEN, J. (1990), "Novelty and Disclosure in Patent Law", *RAND Journal of Economics*, **21**, 131–146.

SHAPIRO, C. (2003), "Antitrust Limits to Patent Settlements", *RAND Journal of Economics*, **34**, 391–411.

SPIER, K. E. (2007), "Litigation", in Polinsky, A. M. and Shavell, S. (eds), *Handbook of Law and Economics*, Vol. 1 (North Holland: Elsevier) 259–342.

The ECONOMIST (2015), "Time to Fix Patents" August 8th-14th, p. 9.

WEYL, E. G. and TIROLE, J. (2012), "Market Power Screens Willingness-to-pay", *Quarterly Journal of Economics*, **127**, 1971–2003.