Screening for Patent Quality: Examination, Fees, and the Courts*

Mark Schankerman[†] Florian Schuett[‡]

March 2017

Abstract

To study how governments can improve the quality of patent screening, we develop an integrative framework incorporating four main policy instruments: patent office examination, pre- and post-grant fees, and challenges in the courts. We show that examination and pre-grant fees are complementary, and that pre-grant fees screen more effectively than post-grant fees. Simulations of the model, calibrated on U.S. patent and litigation data, indicate that patenting is socially excessive and the patent office does not effectively weed out low-quality applications. We quantify the welfare effects of counterfactual policy reforms and show how they depend on the quality of the courts.

Keywords: innovation, patents, screening, litigation, courts, patent fees
JEL classification: D82, K41, L24, O31, O34, O38

1 Introduction

The patent system is one of the key instruments governments use to provide innovation incentives. However, there is growing concern among academic scholars and policy-makers that patent rights are becoming an impediment, rather than an incentive, to innovation. Critics claim that the proliferation of patents, and the fragmentation of ownership among firms, raise the transaction costs of doing R&D, and expose firms to hold-up through patent litigation (Heller and Eisenberg, 1998; Lemley and Shapiro, 2005; Bessen and Maskin, 2009). These dangers have been prominently voiced in public debates on patent policy (Federal Trade Commission, 2011; The Economist, 2015) as

*We are grateful to Klenio Barbosa, Özlem Bedre, Jay Pil Choi, Vincenzo Denicolò, Rafael Ferraz, Alberto Galasso, John Golden, Eirik Kristiansen, Patrick Legros, Mark Lemley, Matt Mitchell, Pierre Régibeau, Dave Rietzke, Monika Schnitzer, David Schwartz, and participants at the MaCCI Workshop in Bad Homburg, TILEC Conference in Amsterdam, IO Workshop in Lancaster, Munich Summer Institute, EARIE Conference in Lisbon, and seminars at universities in Tilburg, Bolzano, Florence, the Norwegian School of Economics, Telecom ParisTech, and Duke. We especially want to recognize William Matcham for outstanding research assistance. The simulation analysis in Section 5 was done with his collaboration. All errors are our own.

[†]London School of Economics and Centre for Economic Policy Research, E-mail: m.schankerman@lse.ac.uk

[‡]TILEC & CentER, Tilburg University. E-mail: f.schuett@uvt.nl

well as recent Supreme Court decisions in the United States (eBay Inc. v. MercExchange, 547 U.S. 338, 2006), and have resulted in the Leahy-Smith America Invents Act of 2011, the most significant statutory change to the U.S. patent system in half a century.

Critics claim that these problems arise in large part from ineffective patent office screening, granting patents to obvious inventions that do not represent a substantial inventive step, especially but not only in new areas such as business methods and software. This view is widely held and buttressed by anecdotes of egregious cases (Jaffe and Lerner, 2004), but there is very little direct evidence on the extent or the sources of the patent-quality problem.[1] Whatever the source of the problem, in the presence of costly and probabilistic review by courts, weak patents – with low probability of being upheld as valid – may end up being strong (Farrell and Shapiro, 2008). This can create greater opportunity for rent-extraction by owners of weak patents.

There is widespread criticism about low-quality patents and the need to make screening more effective. But how best to achieve this goal? The economic theory literature has typically focused more on the design of patent rights, such as optimal length and breadth (Scotchmer, 1999; Cornelli and Schankerman, 1999; Hopenhayn and Mitchell, 2000), but recently economists have begun to study the impact of patent screening. Schuett (2013a, 2013b) studies how patent examination intensity, examiner incentives, and application fees affect *ex ante* choice of research projects by inventors, while Kou, Rey and Wang (2013) study the impact of the non-obviousness threshold for patenting on project selection.[2]

Legal scholars have written extensively on ways to improve patent screening, including the use of external peer review (Noveck, 2006), the appropriate standard and application of the non-obviousness criterion (Eisenberg, 2004; Dreyfuss, 2008), and weakening the presumption of validity and evidentiary standards for invalidation in the courts (Lichtman and Lemley, 2007). Some scholars have argued that it may be rational for the patent office not to screen too rigorously, and that the remaining burden of screening should be picked up by the courts – the so-called 'rational ignorance' argument (Lemley, 2001). Taken to its extreme, this logic might suggest a pure registration system in which the patent office makes no substantive examination effort at all. How-

[1] Attempts at assessing the extent of the problem include Paradise, Andrews and Holbrook (2005), who examine claims in 74 human gene patents and find 38 percent to be problematic, and Henkel and Zischka (2015), who use invalidation rates in the German federal patent court and survey evidence to conclude that 75 percent or more of German patents are fully or partially invalid. Two recent papers study the potential sources of the problem. Frakes and Wasserman (forthcoming) show that examination intensity affects quality; reductions in the time U.S. patent examiners spend reviewing applications result in higher grant rates and patents of lower quality. Frakes and Wasserman (2015) provide evidence that the fact that examiners cannot issue final rejections (applicants can amend and refile repeatedly) leads them to grant more low-quality patents in periods of tight budget constraints.

[2] Other related studies include Caillaud and Duchene (2011) and Atal and Bar (2014).

ever, the ideas developed by legal scholars have not been subjected to formal economic analysis, embedding them in an equilibrium framework in which policy instruments affect optimal strategies of inventors and competitors.

This paper studies how policy-makers can most effectively use the instruments at their disposal to improve patent screening. We focus on four key policy instruments: the intensity of patent office examination, pre-grant (application) fees paid before patent examination, post-grant (issuance/renewal) fees paid by inventions that have passed examination, and review by the courts for patents challenged by a competitor. For most of the analysis, we assume that courts invalidate patents on obvious inventions with certainty; we relax this later and study how results change with the characterization of courts. To our knowledge, we are the first to develop an integrative framework incorporating all four policy instruments, and to study how they interact.

One of our contributions is that our framework allows us to identify the severity of the patent-quality problem. The reason for the dearth of evidence noted above is that it is hard to know how many low-quality applications and patents there are. First, the quality of applications is unobservable and can vary from year to year. Thus, changes in the grant rate do not directly tell us how good a job the patent office is doing. Second, because litigated patents are highly selected, litigation outcomes may not be reflective of the general population. In particular, invalidity rates do not directly tell us what share of all patents is of low quality. By putting structure on the problem and exploiting an equilibrium framework, our approach allows us to get around both of these issues.

We develop a model in which an inventor faces a competitor. The inventor is endowed with an idea for an invention which can be either *obvious* ('low type') or *non-obvious* ('high type'). The invention type is private information to the inventor. An obvious invention is profitable to develop in the absence of a patent, whereas a non-obvious one requires patent protection to be profitable. Since patent protection increases the profit for both types, however, owners of obvious inventions also have a private incentive to seek a patent. There is a net social cost (benefit) of granting patents for obvious (non-obvious) inventions, so effective screening is important for welfare. The inventor chooses whether to pay a pre-grant fee and, if subsequently approved (screening by the patent office is imperfect), whether to pay a post-grant fee to activate the patent. If the patent is activated, the inventor may choose to license the invention to the competitor, and the competitor chooses whether to challenge the validity of the patent in court. The baseline model has a perfect court that always invalidates an obvious patent and upholds a non-obvious one. In an appendix and in simulations, we also analyze a second version that allows both the patent office and the court to make two-sided errors – granting/upholding an obvious patent and rejecting/invalidating a non-obvious one. Formally, our model is a signaling game in which each decision by the inventor can reveal information about the invention type, and the competitor Bayesian updates.[3]

The key results are as follows. First, provided challenges are credible, the equilibrium is in mixed strategies, with low types either randomizing between offering low and high license fees, or between applying and not applying for patents. The decision they randomize over depends on the level of pre- and post-grant fees, as well as on the examination intensity and the cost of going to court. This highlights the fact that the policy instruments interact in shaping the equilibrium.

Second, we show that if the patent office makes no examination effort (a pure registration system), or if the pre-grant fee is zero and examination is imperfect, complete screening (where no low types obtain patents) cannot be achieved. This is important because it emphasizes that fees cannot completely screen in a pure registration system, and that pre-grant fees and examination are complements, not substitutes. Complete screening can be achieved by a combination of a pre-grant fee and an examination that is sufficiently rigorous. We also show that, despite our assumption that courts are mistake-free, they cannot eliminate all bad patents that are issued. This is because in equilibrium not all low-type patents are challenged by the competitor. This result raises serious doubts about over-reliance on the court system to weed out obvious patents.

Third, we study the optimal structure of fees and show that a social planner would always *frontload* fees, i.e., rely on pre-grant rather than post-grant fees. The intuition for frontloading of fees is that the low type prefers post-grant fees to pre-grant fees more strongly than the high type because the low type has a smaller chance of passing examination. This result calls into question the current structure of fees at the major patent offices around the world. Patent offices often backload a substantial part of their fees through post-grant charges for issuance and renewal.[4] However, there may be reasons outside our model for backloading fees. As Scotchmer (1999) and Cornelli and Schankerman (1999) showed, renewal fees can be used to ensure that more valuable inventions receive longer protection, which can increase welfare. In addition, capital market imperfections, especially for young, small firms, may be another argu-

[3]To our knowledge, Meurer (1989) was the first to analyze a signalling game of patent licensing negotiation in the shadow of litigation, but his focus was not on patent screening. Our model incorporates a licensing subgame but in a broader framework with multiple policy instruments. For other examples of patent litigation models, but without private information, see Choi (1998) and Crampes and Langinier (2002).

[4]At the USPTO, for example, a typical patent application (three independent claims or less) involves a total of $1,740 in pre-grant fees. If the application is granted and the patent is renewed to full term, total post-grant fees amount to $13,560 (not discounted). Of course, the applicant also incurs the legal cost of preparing the patent application, which typically is in the range of 10-20 thousand dollars.

ment against frontloading. Our analysis shows that any such benefits of backloading fees must be traded off against the costs of impairing the effectiveness of screening.

Fourth, we show that the private incentives to challenge a patent can be either too high or too low relative to the socially optimal level. This is noteworthy because the conventional wisdom suggests that the private incentives to challenge are inadequate due to the public-good nature of challenges (Farrell and Merges, 2004; Farrell and Shapiro, 2008). While this point is valid, it is also incomplete. We show that there are other factors that can either reinforce or counteract this effect. Private and social incentives generally diverge both on the cost and benefit sides. First, a challenger takes his own, but not the patentee's, litigation cost into account. Second, the private gains from a successful challenge, given by the challenger's incremental profit from invalidation, can be either larger or smaller than the social gains, given by the deadweight loss that is eliminated.

Finally, we simulate the model, calibrated on U.S. patent and litigation data. The simulations have two purposes: first, to characterize the severity of the current problem with patent quality and screening, and second, to study the welfare impact of various counterfactual policy reforms. We simulate the model both with a perfect court and an imperfect court. In both settings, we find that at least 75 percent of patent applications are made on inventions that would be developed even without patent rights, and that the patent office is relatively ineffective at screening them out, as it does so with a probability of only about 30 percent. This implies that between 65 and 81 percent of granted patents are invalid, depending on the specification of the courts. These findings highlight the current crisis in patent screening, and the need to develop effective policies to address it. We use the simulations to quantify the effects of policy reforms including frontloading fees, intensifying patent office examination, introducing a cap on litigation costs, and replacing examination with a pure registration system. The simulations indicate that the first three reforms significantly increase welfare, while a registration system reduces it. These results suggest that an effective strategy for improving patent screening may require both patent office and tort reforms.

Throughout the paper, we take the existence of a patent system as given. It is not *a priori* clear whether a patent system is the optimal mechanism in the environment we consider. However, since abolishing the patent system is not on the table in the foreseeable future, we believe that exploring how to improve the functioning of the existing system is a worthwhile endeavor in its own right. One of the advantages of the game-theoretic approach we adopt, as compared to the mechanism design alternative, is that it allows us to use the model to evaluate the performance of the current patent screening institutions and to assess the welfare effects from counterfactual policy reforms.[5]

The paper is organized as follows. Section 2 presents a simple model of patent screening without courts. Section 3 introduces court challenges, derives the equilibrium, and presents comparative statics results. Section 4 studies the effectiveness of screening and develops the welfare analysis. Section 5 presents simulations that illustrate welfare gains from policy reforms both for the model with perfect courts and imperfect courts. Section 6 discusses some of the key assumptions in the model and how the results extend to alternative settings, including heterogeneous value of inventions. We conclude with a brief summary and directions for further research.

2 A Simple Model of Patent Screening

There is a unit mass of inventors. Each inventor is endowed with an idea $\theta \in \{L, H\}$ for an invention. Developing an idea into an invention requires an R&D investment κ_θ. Ideas with $\theta = H$ correspond to inventions with high R&D cost κ_H and occur with probability λ, while ideas with $\theta = L$ correspond to inventions with low R&D cost κ_L and occur with probability $1 - \lambda$. The idea θ (and thus the R&D cost, κ_θ) is the inventor's private information. Once an invention has been developed, the inventor can apply for a patent.[6]

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 $\pi \geq 0$ and total welfare is $w \geq \pi$ (both are gross of R&D costs). Thus, π is a measure of the profits the inventor can appropriate without a patent.[7] With a patent, the inventor earns $\pi + \Delta$, and welfare is $w - D$, where $\Delta > 0$ denotes the patent premium and $D \geq 0$ is the deadweight loss from monopoly pricing.[8] This captures in the simplest possible way the classic tradeoff whereby patents increase the incentives to innovate but at the same time cause deadweight loss. We assume

[5] It is also worth noting that a patent system with the patentability requirement which we will analyze places lower informational requirements on the government than an alternative incentive mechanism such as a prize system. A prize system requires the government to know the gap between R&D costs and profits in the absence of patent protection, whereas the patent system we model only requires knowledge of the *sign* of the gap. For more discussion of our patentability requirement and how it relates to the existing statutory and judicial standards, see Section 2.

[6] We assume that patent applications can only be submitted on inventions, not ideas. While the legal treatment is more complicated, the basic patent ineligibility of abstract ideas was recently affirmed by the U.S. Supreme Court in the Bilski case (Bilski v. Kappos, 561 U.S. 593, 2010).

[7] We note two points. First, for the inventor to benefit from the invention even absent patent protection ($\pi > 0$), competition must not be too fierce, or there must be an alternative appropriation mechanism such as lead time. Second, in Section 6.1 we relax the assumption that all inventions have the same value.

[8] In principle, D can also capture other costs associated with patents such as patent congestion (thickets) and royalty stacking that can raise transaction costs for inventors to license inputs necessary to conduct R&D (Heller and Eisenberg, 1998; Shapiro, 2001). We do not explicitly model those costs, however.

that $w - D \geq \pi + \Delta$. This assumption says that the social returns exceed the private returns to R&D, which is consistent with the evidence in Bloom, Schankerman and Van Reenen (2013).

An important modeling choice concerns the requirements for an invention to be patentable. In what follows, we impose the patentability requirement $\kappa > \pi$, i.e., R&D costs must exceed the profits the inventor can appropriate without a patent. This corresponds to the notion that patents should be given only to those inventions that require the patent incentive to be developed, and not to those that society would have benefited from even absent a patent. This perspective is in line with the rationale courts and legal scholars typically give for the non-obviousness requirement in patent law (Eisenberg, 2004). For example, in the landmark case of Graham v. John Deere (383 U.S. 1 (1965)) the U.S. Supreme Court stated: "The inherent problem was to develop some means of weeding out those inventions which would not be disclosed or devised but for the inducement of a patent." Of course, while this theoretical patentability criterion makes economic sense, the courts have struggled with the practical aspects of how to implement it. The various judicial standards for non-obviousness, novelty etc. reflect an attempt by the courts to do this, the basic presumption being that if an invention is obvious to those skilled in the relevant scientific arts then it is probably cheap to develop, and one does not need a patent to induce it.

In our setup, where ideas arrive exogenously and are scarce, this is also the patentability requirement that a social planner would choose. To see this, consider the planner's decision about which types of inventors should be given patents. The planner is worried about deadweight loss, but also realizes that some socially valuable inventions will not be developed without the promise of a patent. Ideas with $\pi \geq \kappa$, however, are developed even without a patent, so giving them a patent can never be optimal. A second concern for the planner would be that giving patents to some ideas with $\pi < \kappa \leq \pi + \Delta$ leads to development of these ideas even though they are not socially valuable, i.e., $\kappa > w - D$. Our assumption that $w - D \geq \pi + \Delta$ rules this out, however, as it implies that ideas that are not socially valuable are not privately valuable either.

To make the problem interesting, we assume that $\kappa_H = k > \pi \geq \kappa_L = 0$. Setting $\kappa_L = 0$ is without loss of generality as type-L inventions are developed regardless of whether or not a patent is expected to be obtained.[9] In order to ensure that patent protection can provide sufficient incentive for type-H inventors to develop their ideas, we assume $\Delta \geq k - \pi$. Note that this implies $w - D \geq k$, so that investment by type-H inventors is socially valuable even if it comes at the expense of deadweight loss.

Under these assumptions, type-H inventors should be given patents while type-L

[9] Note that for the simulations in Section 5 we do *not* impose $\kappa_L = 0$. The simulations actually generate an estimate of κ for both type-L and H inventions.

inventors should not.[10] The problem, of course, is that the type θ of an invention is privately observed by the inventor. Since type L also benefits from patent protection, society must put in place a screening mechanism. We start by considering the patent office; in Section 3 we introduce the courts as a second 'line of defense.'

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. (We relax the assumption that there are no erroneous rejections in Appendix B.) Inventions that pass the examination must pay a *post-grant fee* $\phi_P \geq 0$ in order to be issued a patent.[11] 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.[12] If the inventor does not apply, does not pass the examination, or does not pay ϕ_P, the invention falls into the public domain.

Suppose for the moment that challenging the validity of a patent in court is not possible. Then an inventor of type H invests in R&D, applies for a patent, and activates it if and only if

$$\Delta - (k - \pi) - \phi_P - \phi_A \geq 0. \tag{1}$$

An inventor of type L always invests as $\kappa_L = 0 < \pi$. He applies for a patent and activates (conditional on passing the examination) if and only if

$$(1 - e)(\Delta - \phi_P) - \phi_A \geq 0. \tag{2}$$

Although derived from a highly stylized model, this pair of inequalities leads to two key observations which, as we will show, apply much more generally:

Observation 1. In the absence of sufficiently rigorous patent examination, it is impossible to deter type L without also deterring type H.

Observation 2. Pre-grant fees screen more effectively than post-grant fees.

[10] Though it may seem counterintuitive to reward ideas with high R&D costs, we show in Section 6.2 that such a patentability requirement can make sense even in an environment where inventors choose which ideas to pursue.

[11] As we will show, our results suggest that negative post-grant fees might be optimal. In principle this could be implemented by imposing a penalty on failed patent applications or a reward to successful applicants. Of course, this might introduce excessive incentives for the patent office to reject applications.

[12] 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.

Observation 1 is related to the fact that holding a patent is worth the same to both types (namely, Δ), but type H needs to cover an additional $k - \pi$ to make (investing and) applying worthwhile. This implies that if $e = 0$, type L's payoff from applying strictly exceeds type H's for any (ϕ_A, ϕ_P), so setting fees high enough to deter type L will also deter type H. Achieving *full screening* – i.e., deterring applications from type L without discouraging investment by type H – requires

$$(1 - e)\Delta \leq \Delta - (k - \pi), \tag{3}$$

i.e., a minimum examination intensity $\bar{e} \equiv (k - \pi)/\Delta > 0$. If $e < \bar{e}$, there is no combination of fees that can achieve full screening.

In Observation 2, pre-grant fees screen more effectively in the sense that type L prefers fees to be backloaded more strongly than type H. 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: ϕ_A and ϕ_P are perfect substitutes for him. By contrast, type L strictly prefers post-grant fees to pre-grant fees. Again keeping the sum of fees constant, type L's expected total fee payment is $\phi_A + (1 - e)\phi_P = \phi_A + (1 - e)(\Phi - \phi_A)$, which is increasing in ϕ_A. The reason why type L prefers fees to be backloaded is that pre-grant fees must be paid whether or not he passes examination, whereas post-grant fees are only paid conditional on passing examination.

One implication of Observation 2 is that we cannot rely exclusively on post-grant fees if we want to achieve full screening. If $\phi_A = 0$, then (unless $e = 1$) the only way to deter type L is to set $\phi_P \geq \Delta$, but then type H's payoff from investing in R&D is negative. Hence, full screening requires strictly positive pre-grant fees, $\phi_A > 0$.

3 Introducing Court Challenges

We now extend the model from Section 2 by introducing 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 whose validity is challenged (in practice, usually by a competitor or other potential licensee). We thus consider an explicit licensing game between the inventor and a competitor active in the same industry. Second, in assessing validity the courts make fewer mistakes than the patent office.[13] To begin, we adopt the simplifying assumption that the courts do not make any mistakes: they always uphold valid patents (type H) and revoke invalid ones (type L). In Section 5.2 and Appendix B, we examine a more general screening technology in which both the patent office and courts can make type I and type II errors.

[13] Court cases have more time to gather evidence and hear arguments than the typical time patent examiners 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 better at screening.

Suppose the invention is a cost-reducing technology and each inventor has a single competitor. Once developed the invention can be freely used by the competitor unless it is protected by a patent. Thus, in the absence of a patent, both the inventor and competitor benefit from the cost reduction and obtain profit π each.[14] Denoting consumer surplus by $S \geq 0$, total welfare is $w = 2\pi + S$. An inventor holding a patent can enter into a license agreement with his competitor. We assume that the inventor makes a take-it-or-leave-it offer to the competitor to license the invention at a royalty rate that generates license revenue F (for brevity, we hereafter refer to F as the license fee). The competitor decides whether to accept or reject; if she rejects, she further decides whether to challenge the patent in court.

If the firms fail to agree on a license contract and the courts do not revoke the patent, only the inventor can use the new cost-reducing technology while the competitor has to use the backstop technology.[15] Under this asymmetric competition, the inventor earns $\pi + \Delta_I$ and the competitor $\pi - \Delta_C$, where $\Delta_I > 0$ and $\Delta_C > 0$. That is, the patent benefits the inventor and hurts the competitor. The competitor's profits in this case $(\pi - \Delta_C)$ constitute her outside option in the licensing negotiations.

If the firms reach agreement, the competitor pays the license fee to the inventor and both firms can use the invention in production. With a license agreement in place, 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 $\pi+m+F$ and the competitor $\pi-F$, where $m \geq 0$ is the extra profit due to market power.[16] Thus, total industry profit becomes $2\pi + m$. We assume that $m + \Delta_C \geq \Delta_I$, which will ensure that the inventor prefers to license his invention rather than exclude the competitor and compete with asymmetric costs. Accordingly, in what follows we let $\Delta \equiv m + \Delta_C$.[17]

If the competitor rejects the license contract, she has the option of challenging the patent in court at litigation cost l_C to herself and l_I to the inventor. We assume $l_C < \Delta_C$, since otherwise the competitor never has any incentive to challenge. The court then determines the validity of the patent. We assume that during litigation the court learns the invention's true type θ. If $\theta = L$, the court invalidates the patent and both firms can freely use the invention. If $\theta = H$, the court upholds the patent and the

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

[15]We are assuming that infringement is not at issue. If the competitor infringed the patent without challenging its validity, the inventor could sue for infringement and would be sure to prevail.

[16]The parameter m can be interpreted as a measure of how lenient or restrictive antitrust policy is towards license agreements. Because the inventor has all the bargaining power, it does not matter how the extra profit m is distributed between inventor and competitor.

[17]Letting $S - \Delta_S$ denote consumer surplus after innovation when there is a patent, we have $D \equiv \Delta_S - m$. The assumption that deadweight loss is positive amounts to $\Delta_S \geq m$.

inventor can offer a new license contract to the competitor.

This is a signaling game where the inventor's decisions to apply for and activate a patent, as well as which license fee to propose, potentially convey information about his type to the competitor. We now analyze the Perfect Bayesian Equilibrium of this game.

3.1 Equilibrium

Let $\tilde{\lambda}(F)$ denote the competitor's belief that she is facing a type-H inventor when offered a license at fee F. Suppose type H invests and applies, and that if he is successful in obtaining a patent, he activates it and charges a license fee $F^H = \Delta_C$ (we later show under what conditions this is an equilibrium). Let α denote the probability that a type-L inventor applies for a patent. Similarly, let ρ denote the probability that a type-L inventor activates the patent in case he passes examination, and y denote the probability that he offers a license at fee F^H. The competitor's belief that an activated patent is valid when offered a license at fee F^H is

$$\hat{\lambda} \equiv \tilde{\lambda}(F^H) = \frac{\lambda}{\lambda + (1-\lambda)(1-e)\alpha\rho y}. \tag{4}$$

The competitor prefers challenging to not challenging if and only if $(1-\hat{\lambda})\Delta_C \geq l_C$. The left-hand side is the expected benefit from a challenge, given by the probability that the court invalidates the patent times the increase in the competitor's profits if the patent is invalidated; the right-hand side is the cost of litigation. The lower bound on $\hat{\lambda}$ occurs if the type-L inventor pools with the type-H inventor, i.e., if $\alpha = \rho = y = 1$, and is given by

$$\underline{\lambda} \equiv \frac{\lambda}{\lambda + (1-\lambda)(1-e)}.$$

Because this is the lowest value that $\hat{\lambda}$ can take, we will say that a validity challenge is *credible* if and only if

$$(1-\underline{\lambda})\,\Delta_C \geq l_C. \tag{5}$$

When challenges are not credible, the competitor never challenges and accepts any license fee $F \leq \Delta_C$. Hence, both types of inventor propose $F = \Delta_C$, and holding a patent is worth the same to both of them, namely $\Delta = m + \Delta_C$. The outcome is thus exactly the same as in the simple model of Section 2.

When challenges are credible, holding a patent is generally not worth the same to both types of inventors. Because type L is less likely to survive a challenge than type H, the patent is worth less to type L for any given probability of being challenged. One might thus expect post-grant fees to be effective at screening out type-L inventors: it should suffice to set ϕ_P at (or slightly above) the expected value of a patent to type

L. Since the decision to challenge is endogenous, however, this argument does not work. To see this, suppose only type-H inventors pay the post-grant fee to activate their patents (a separating equilibrium). Then the competitor correctly infers that all activated patents are valid and never challenges. But then, the patent is again worth the same to both types of inventors, and the low type would want to activate as well. As this argument suggests, and Proposition 1 establishes formally, the equilibrium generally involves mixed strategies.

Proposition 1. *If $\phi_P > \Delta - \phi_A/(1-e)$, the type-L inventor does not apply and the competitor does not challenge. If challenges are credible and $\phi_P \leq \Delta - \phi_A/(1-e)$, there is a semi-separating equilibrium in which:*

(i) the type-H inventor invests, applies, activates, and proposes $F^H = \Delta_C$

(ii) the type-L inventor always randomizes over either

(a) the decision whether to apply or not (probabilities α and $1-\alpha$), or

(b) the license fee to propose, $F^L \in \{\Delta_C, l_C\}$ (probabilities y and $1-y$),

such that $(1-\hat{\lambda})\Delta_C = l_C$; he always activates ($\rho = 1$)

(iii) the competitor randomizes over the decision whether to challenge or not (probabilities x and $1-x$) if offered $F = \Delta_C$ and never challenges if offered $F = l_C$, with

$$x = \begin{cases} \tilde{x} \equiv \dfrac{\Delta_C - l_C}{\Delta + l_I} & \text{for } \phi_P < l_C + m - \dfrac{\phi_A}{1-e} \\[2ex] \hat{x} \equiv \dfrac{\Delta - \phi_P - \phi_A/(1-e)}{\Delta + l_I} & \text{for } l_C + m - \dfrac{\phi_A}{1-e} \leq \phi_P \leq \Delta - \dfrac{\phi_A}{1-e} \end{cases}$$

This semi-separating equilibrium requires that the high type make nonnegative profit, $\Delta - (k - \pi) - xl_I - \phi_A - \phi_P \geq 0$.

Proof. See Appendix A. □

In the equilibrium described in Proposition 1, the type-L inventor applies with probability α and charges a high license fee Δ_C, mimicking the type-H inventor, with probability y.[18] With probability $1-y$, type L charges a low license fee l_C, which

[18] The equilibrium is not unique. There exists a continuum of semi-separating equilibria with $\Delta_C - l_I \leq F^H \leq \Delta_C$ and (depending on parameters) also pooling equilibria in which both types charge $F^* \leq l_C + \underline{\lambda}\Delta_C$ and the competitor refrains from challenging. However, the semi-separating equilibrium with $F^H = \Delta_C$ is the only one that survives application of the D1 criterion. Details are available from the authors upon request.

prevents the competitor from challenging.[19] The competitor challenges with probability x whenever the inventor proposes a license fee of Δ_C.

When the rate of challenges is x, the type-H inventor's payoff from investing, applying, activating and proposing F^H is $\Pi^H \equiv \Delta - (k - \pi) - xl_I - \phi_P - \phi_A$: if a challenge occurs, the type-H inventor is sure to win but nevertheless bears the litigation cost l_I. Because the rate of challenges depends endogenously on the patent office fees, the type-H inventor's payoff is affected by ϕ_A and ϕ_P both directly and indirectly through their effect on x.

Let $\tilde{\alpha}$ (resp. $\tilde{y}$) denote the value of α (resp. y) that solves $(1 - \hat{\lambda})\Delta_C = l_C$ when $y = 1$ (resp. $\alpha = 1$) and $\rho = 1$. From (4) it follows that $\tilde{\alpha} = \tilde{y}$. Figure 1 depicts how the equilibrium that arises when challenges are credible depends on ϕ_A and ϕ_P. In region 1, where $\phi_P < l_C + m - \phi_A/(1 - e)$, fees are sufficiently low for the type-L inventor to always find it worthwhile to apply and activate ($\alpha = \rho = 1$), while randomizing over the license fee to offer with $y = \tilde{y}$. The rate of challenges is given by $\tilde{x}$. Moving toward the north-east into region 2, type L can no longer break even by proposing the low fee $F = l_C$; he now randomizes over the application decision, applying with probability $\alpha = \tilde{\alpha}$ while activating and offering the high license fee F^H with certainty ($\rho = y = 1$). The rate of challenges is given by $\hat{x}$. As fees increase further and we reach region 3, the type-L inventor no longer applies ($\alpha = 0$) and the rate of challenges drops to zero ($x = 0$). The figure also shows the condition under which type H finds it profitable to invest, as represented by the $\Pi^H = 0$ locus.[20]

3.2 Comparative statics

Patent office fees and examination intensity affect the equilibrium application rate of type-L inventors, the rate of challenges, and the license fees proposed. The next proposition characterizes these comparative statics.

Proposition 2. *Suppose the type-H inventor invests, i.e., $\Pi^H \geq 0$, and challenges are credible. Then:*

(i) An increase in ϕ_A or ϕ_P weakly decreases applications by type L (α), weakly decreases the rate of challenges (x), and weakly increases the probability that type

[19] This outcome, where the owner of the low quality patent charges a royalty that preempts a challenge, could be interpreted as behavior by a patent assertion entity. For recent analysis of this issue, see FTC (2016).

[20] Note that for certain parameters, this locus may be upward sloping, as shown in the figure. This implies that the type-H inventor's expected profit is *increasing* in ϕ_A over some range. The reason this can happen is that an increase in fees raises the perceived quality of patents, to the benefit of patent holders. While this argument is made in a reduced-form way in Atal and Bar (2014), here we endogenize the source of the benefits from higher perceived patent quality, namely, the fact that higher fees lead to fewer challenges and thus lower litigation costs. Under some conditions, this indirect effect of higher fees can dominate their direct effect on the high type's profit.

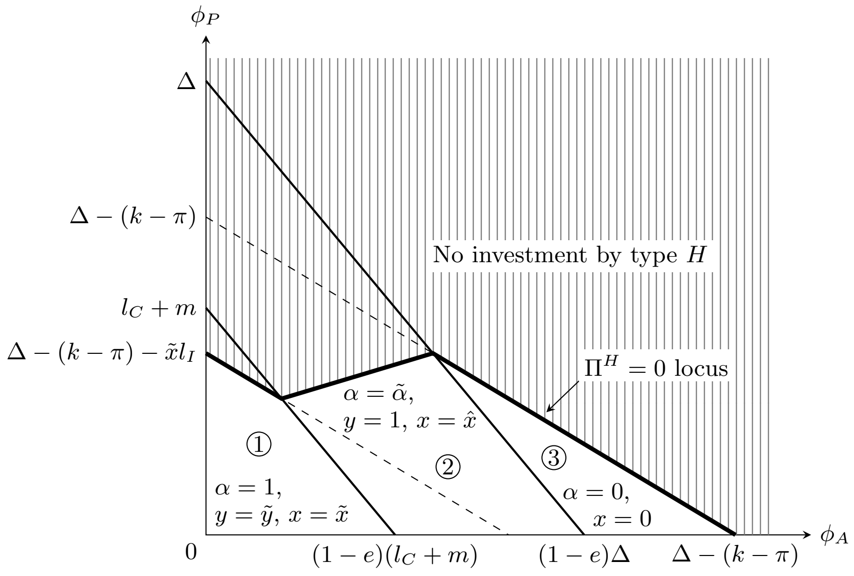

Figure 1: The equilibrium as a function of ϕ_A and ϕ_P when $\bar{e} < e < 1$

L charges the high license fee (y).

(ii) *An increase in e has ambiguous effects on applications by type L (α), weakly decreases the rate of challenges (x), and weakly increases the probability that type L charges the high license fee (y).*

Proof. See Appendix A. □

Proposition 2 shows that an increase in fees unambiguously decreases bad applications and challenges (in a weak sense). It also leads type L to charge the high license fee more often. As fees increase and we move from region 1 to 2, type-L inventors switch from randomizing over the license fee to randomizing over the application decision, and charge the high license fee with certainty. Perhaps more surprisingly, an increase in the examination intensity has ambiguous effects on applications by type-L inventors. While increasing the examination intensity directly discourages type-L inventors from applying, over some range, the application rate of type-L inventors actually increases with e. The intuition is that more rigorous examination makes it more likely that a granted patent is valid, other things equal. That is, higher e raises the competitor's posterior belief $\hat{\lambda}$. But in equilibrium, the competitor must be indifferent between challenging and not, which requires that $\hat{\lambda}$ be held constant. Therefore, in region 2, type L responds to an increase in e by adjusting the probability of applying (α) upward. Finally, an increase in e lowers the rate of challenges: as type L's payoff from applying declines, the rate of challenges needed to make him indifferent between applying and

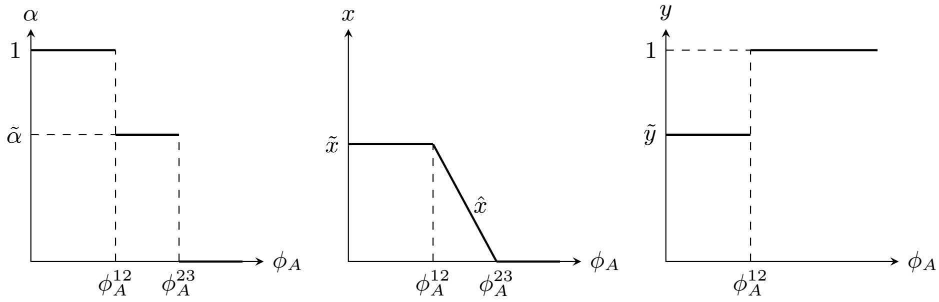

Figure 2: α, x, and y as a function of ϕ_A for $\bar{e} < e < 1$ and $\phi_P < \Delta - (k - \pi)/e$

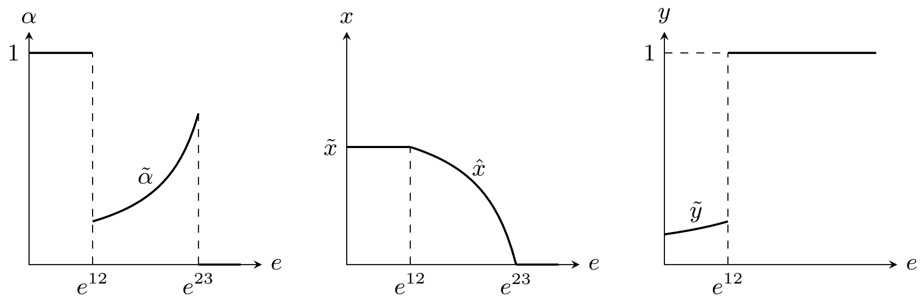

Figure 3: α, x, and y as a function of e for $\phi_A + \phi_P < l_C + m$ and $\phi_A > 0$

not is reduced.[21]

Figures 2 and 3 illustrate these results.[22] Figure 2 shows how the equilibrium values of α, x, and y vary with ϕ_A, holding ϕ_P and e constant. It is drawn for parameter values such that $\bar{e} < e < 1$ and $\phi_P < \Delta - (k - \pi)/e$, which ensures that for $\phi_A = 0$ the equilibrium is in region 1, while for ϕ_A sufficiently large the equilibrium is in region 3. The cutoffs ϕ_A^{12} and ϕ_A^{23} are defined as the threshold values of ϕ_A above which the equilibrium moves from region 1 to region 2 and from region 2 to region 3, respectively. Figure 3 shows how α, x and y vary with e, holding constant ϕ_A and ϕ_P. It is drawn for parameter values such that $\phi_A + \phi_P < l_C + m$ and $\phi_A > 0$, which ensures that for

[21] More stringent patent examination can reduce the incidence of litigation, but does not necessarily do so. Although a sufficiently large increase in the examination intensity (to the point where the challenge credibility constraint (5) no longer holds) always leads to a drop in litigation, the same is not true for marginal increases in e. To see this, note that the rate of challenges x in the model does *not* coincide with the observed rate of litigation. Letting LR denote the rate at which patents are litigated, in region 1 for example we have

$$LR = \frac{x(\lambda + (1-\lambda)(1-e)y)}{\lambda + (1-\lambda)(1-e)}.$$

That is, the litigation rate is equal to x multiplied by the percentage of patentees charging high royalties. Although raising e weakly reduces the rate of challenges, it may increase the litigation rate. This is the case in region 1, where x is constant but y increases with e; see Figure 3.

[22] Both figures are drawn under the implicit assumption that $\Pi^H \geq 0$ and that e is low enough for the challenge credibility constraint (5) to be satisfied.

$e = 0$ the equilibrium is in region 1, while for e sufficiently close to 1 the equilibrium is in region 3. The cutoffs e^{12} and e^{23} are defined similarly as above.

4 Screening and Welfare

In this section we revisit the conditions for full screening. We show that when court challenges are possible, full screening again requires a minimum examination intensity, and that its level is the same as in the absence of courts. We then set up the welfare function and study the optimal fee structure and the private and social incentives for court challenges.

4.1 Conditions for full screening

The following proposition examines when and how it is possible to deter type-L inventors without discouraging type-H inventors.

Proposition 3. *There exists a combination of fees (ϕ_A, ϕ_P) inducing full screening if and only if $e \geq \bar{e}$. For a given $e \geq \bar{e}$, any combination of fees satisfying*

$$(1-e)(\Delta - \phi_P) \leq \phi_A \leq \Delta - (k - \pi) - \phi_P \tag{6}$$

achieves full screening.

Proof. By Proposition 1, deterrence of type L ($\alpha = 0$) requires $\phi_A \geq (1-e)(\Delta - \phi_P)$. Investment by type H when type L is deterred (and thus, in equilibrium, $x = 0$) requires $\Delta - (k - \pi) - \phi_P \geq \phi_A$. A pre-grant fee ϕ_A satisfying both inequalities exists if and only if $\Delta - (k - \pi) - \phi_P \geq (1-e)(\Delta - \phi_P)$, or

$$e \geq \frac{k - \pi}{\Delta - \phi_P}. \tag{7}$$

Since the right-hand side increases with ϕ_P, the minimum level of e required to achieve full screening is obtained by evaluating (7) at $\phi_P = 0$, yielding $e \geq (k - \pi)/\Delta = \bar{e}$. □

Proposition 3 extends Observation 1 from the simple model of Section 2 by showing that full screening of inventors requires a minimum examination intensity, $\bar{e} > 0$. This happens despite the fact that we have assumed that courts are perfect at discriminating between valid and invalid patents. The presence of courts does not even change the level of rigor required to achieve full screening, which remains $\bar{e}$. This is because, although the courts are perfect, challenges must be initiated by the competitor, who is not. The competitor is Bayesian and updates her beliefs based on the inventor's equilibrium strategy. If there were an equilibrium in which only type H applies, the competitor would rationally expect any applicant to be of type H, and therefore refrain from challenging. But then, in the absence of patent examination, type L would also find it worthwhile to apply, hence such an outcome cannot be an equilibrium. Instead the equilibrium will be in mixed strategies, implying that at least some type-L inventors apply.

This result underlines the importance of the patent office. A key 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 type-L inventor's equilibrium strategy. This is the fundamental drawback of a pure registration system, relying entirely on the courts for screening.[23]

Proposition 3 also characterizes the fee structure that is necessary to achieve full screening. Because $\phi_P \leq \Delta - (k - \pi) < \Delta$ is needed in order for type H to invest, (6) implies that for any $e < 1$, the pre-grant fee ϕ_A must be strictly positive if type L is to be deterred. The intuition is that, if applications are costless ($\phi_A = 0$) and examination is less than perfect ($e < 1$), type-L inventors have nothing to lose from applying. Because of the mixed-strategy nature of the equilibrium, at least a fraction of type-L inventors will activate the patent for any $\phi_P < \Delta$. Completely deterring type L through the post-grant fee is possible only by setting $\phi_P \geq \Delta$, which also deters type H. In addition, satisfying (6) requires that ϕ_P be set sufficiently low, which is a variant of Observation 2 from the simple model of Section 2.

A further result of our analysis is that, despite the fact the courts are mistake-free, they cannot eliminate all bad patents that are issued. Eliminating all bad patents would require that $x = 1$ whenever $\alpha > 0$, i.e., all issued patents would need to be challenged. Alternatively, type-L inventors would have to reveal themselves so that they could be targeted by challenges. But neither of these is an equilibrium outcome. There is no equilibrium with $\alpha > 0$ and $x = 1$. There is also no equilibrium in which type-L inventors reveal themselves and then get challenged. Although for $\phi_P < l_C + m - \phi_A/(1 - e)$, type L sometimes reveals himself by offering $F = l_C$, the competitor optimally responds to this by *not challenging.*

[23]We implicitly assume that the patent office can commit to examining all patent applications, even if in equilibrium no type-L inventor applies (region 3). This assumption can be justified in two ways. First, patent examiner compensation contains few performance-dependent elements, so that there is little incentive for them to adjust their examination intensity according to their beliefs about the proportion of good and bad applications. Second, in a richer setting where inventions differ in their commercial value, there will always be high-value inventors who apply even though their inventions do not satisfy the patentability criteria; the patent office's examination intensity merely determines the threshold on value below which no type-L applications are submitted. (We develop this extension in Section 6.1.) 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. Provided the competitor observes the value, as we assume, our results on private challenges are unaffected.

4.2 Welfare

We now derive expected welfare as a function of the equilibrium variables α, x, and y. Assume that the cost of examining an application with intensity e is $\gamma(e)$. Denoting expected welfare by W, we have

$$\begin{aligned} W(\alpha, x, y) = 2\pi + S + \lambda(-D - x(l_C + l_I) - k - \gamma(e)) \\ + (1-\lambda)\alpha\big((1-e)\big[(y(1-x) + 1 - y)(-D) - xy(l_C + l_I)\big] - \gamma(e)\big). \end{aligned} \quad (8)$$

With probability λ, the invention is non-obvious (type H), in which case the inventor always applies and society incurs the deadweight loss D, the cost of investment k and the cost of examination $\gamma(e)$ with certainty, while it incurs the cost of challenges (l_C+l_I) with probability x. With probability $1-\lambda$, the invention is obvious (type L), in which case the inventor applies with probability α. Conditional on application, society incurs the deadweight loss with probability $(1-e)(y(1-x)+1-y)$, the cost of challenges with probability $(1-e)xy$, and the cost of examination with certainty. Since profits and consumer surplus prior to invention are normalized to zero, W in fact represents the expected welfare gains from innovation.

Optimal fee structure Let $\underline{e} \equiv [k - \pi - \Delta(\Delta_C - l_C)/(\Delta + l_I)]/(l_C + m)$. If $e < \underline{e}$ only region 1 is attainable, and in region 1 welfare is unaffected by ϕ_A and ϕ_P because $\alpha = 1$, $x = \tilde{x}$, and $y = \tilde{y}$, none of which depend on fees. If $e \geq \bar{e}$, region 3 is attainable, and any combination of fees within region 3 maximizes welfare. This is immediately clear from inspection of (8): holding e constant, welfare is maximum for $\alpha = 0$ and $x = 0$.

To make the problem interesting, assume in what follows that $\underline{e} < e < \bar{e}$ (that is, $(1-e)\Delta > \Delta - (k-\pi)$ and $\Delta - (k-\pi) - \tilde{x}l_I > (1-e)(l_C+m)$), so that both region 1 and region 2 are attainable but region 3 is not (i.e., full screening cannot be achieved). Geometrically, region 3 in Figure 1 disappears as the $\phi_P = \Delta - \phi_A/(1-e)$ line now shifts above the $\Pi^H = 0$ locus.

The social planner's problem is to choose ϕ_A and ϕ_P to maximize welfare subject to type H investing, taking e as given:

$$\max_{\phi_A, \phi_P} W(\alpha, x, y) \quad \text{subject to } \Pi^H \geq 0.$$

Although welfare does not depend directly on ϕ_A and ϕ_P, it depends on them indirectly through their effect on the equilibrium values of α, x, and y. The following proposition characterizes the welfare-maximizing mix of fees.

Proposition 4. *Suppose $\underline{e} < e < \bar{e}$. Welfare maximization always entails fees such that $\phi_A \geq (1-e)(l_C + m - \phi_P)$ (region 2). If $\Delta_C > (l_C/(l_C + l_I))D$, welfare is* *maximized by setting* $\phi_P = 0$ *and* ϕ_A *such that* $\Pi^H = 0$*, thus minimizing challenges. If* $\Delta_C < (l_C/(l_C + l_I))D$*, welfare is maximized by setting* $\phi_A = (1-e)(l_C + m - \phi_P)$ *and* $0 \leq \phi_P \leq \min\{l_C + m, [\Delta - (k - \pi) - \tilde{x}l_I - (1-e)(l_C + m)]/e\}$*, thus maximizing challenges.*

Proof. See Appendix A. □

Proposition 4 establishes two important results. First, it is always optimal to raise fees high enough to push the equilibrium into region 2. To understand this result, note that the welfare loss from type-L inventors in (8) can be written as

$$\alpha[(1-e)D + \gamma(e)] - (1-e)\alpha y x(D - (l_C + l_I)).$$

The first term represents the social cost that low types applying for patents would cause if there were no court challenges. The second term represents the deadweight loss that is eliminated by court challenges, net of litigation costs. Recall that in region 2 type L randomizes over the application decision ($\alpha = \tilde{\alpha}$, $y = 1$), whereas in region 1 he randomizes over the license fee to propose ($\alpha = 1$, $y = \tilde{y}$). Since $\tilde{\alpha} = \tilde{y}$, it follows that αy is the same in regions 1 and 2. This reflects the fact that in both regions αy is such that the competitor is indifferent between challenging and not. Thus holding x constant the second term in the expression above is also the same in both regions. Raising fees thus affects welfare only through the first term and through the rate of challenges x. Recall that x equals $\tilde{x}$ in region 1 (which is constant in fees) and $\hat{x}$ in region 2 (decreasing in fees). The proof shows that $\hat{x}$ tends to $\tilde{x}$ as fees approach the border between regions 1 and 2 (with equality for $\phi_A = (1-e)(l_C + m - \phi_P)$). Thus, we can achieve the same rate of challenges in region 2 as in region 1. Because the first term is increasing in α, and α is lower in region 2, the welfare loss is also lower in region 2.

Second, it is optimal to set fees to either maximize or minimize challenges, depending on parameters. The welfare effect of challenges is *a priori* ambiguous. Differentiating W with respect to x, we have

$$\frac{\partial W}{\partial x} = -\lambda(l_C + l_I) + (1-\lambda)(1-e)\alpha y(D - (l_C + l_I)). \tag{9}$$

On the one hand, challenges help society get rid of invalid patents, which raises welfare provided deadweight loss exceeds litigation costs, $D > l_C + l_I$. On the other hand, challenges also create wasteful litigation of valid patents. Rewriting (9) using the definition of $\hat{\lambda}$ yields

$$\frac{\partial W}{\partial x} = (\lambda + (1-\lambda)(1-e)\alpha y)\left[(1-\hat{\lambda})D - (l_C + l_I)\right]. \tag{10}$$

The term in square brackets represents the social incentive to challenge a patent whose probability of being valid is $\hat{\lambda}$. It equals the posterior probability of the inventor being of type L (so that the courts will invalidate) times the deadweight loss that is saved if the patent is invalidated, minus the social cost of a challenge.

Now consider the private incentive to challenge. In equilibrium (be it in region 1 or region 2), we have $(1-\hat{\lambda})\Delta_C = l_C$: the competitor must be indifferent between challenging and not when the inventor charges the high license fee $F = \Delta_C$. Evaluating (10) at $1-\hat{\lambda} = l_C/\Delta_C$, we see that welfare is decreasing in x if and only if

$$\Delta_C \geq D\left(\frac{l_C}{l_C+l_I}\right). \tag{11}$$

The social and private incentives to challenge generally diverge: if (11) holds (with strict inequality), the equilibrium rate of challenges is socially excessive; if (11) does not hold, the rate of challenges is socially insufficient. In either case, the private challenge decision is inefficient.[24]

When equilibrium challenges are socially excessive, a social planner wants to minimize challenges. Proposition 4 shows that to achieve this, the planner should *frontload fees*, setting post-grant fees to zero while relying exclusively on pre-grant fees, and set pre-grant fees at the *highest level compatible with investment by type H*, so that the constraint $\Pi^H \geq 0$ is binding. The reason for this is that the rate of challenges goes hand in hand with type L's payoff: fewer challenges are necessary to make type L indifferent between applying and not when his payoff from applying is small. Minimizing challenges subject to investment by type H requires minimizing type L's payoff while holding type H's payoff constant. Because type L prefers fees to be backloaded more strongly than type H, this is accomplished through a mix of fees that emphasizes pre-grant over post-grant fees.

Conversely, when challenges are socially insufficient, the planner wants to maximize challenges. To this end, the planner should set fees *as low as possible within region 2*. Any mix of fees located on the boundary between regions 1 and 2 (where $\phi_A = (1-e)(l_C+m-\phi_P)$) and for which $\Pi^H \geq 0$ (which requires $\phi_P \leq [\Delta-(k-\pi)-\tilde{x}l_I-(1-e)(l_C+m)]/e$) leads to $x=\tilde{x}$, which is the maximum rate of challenges that can be sustained. The optimal mix of fees is no longer uniquely determined. In general, the constraint $\Pi^H \geq 0$ does not bind at the optimum. Notice, however, that $\phi_P = 0$

[24]Two points should be noted. First, there is another reason outside our model that leads to insufficient incentives to challenge patents: when there are multiple licensees, a patent challenge has the nature of a public good (Farrell and Merges, 2004; Farrell and Shapiro, 2008). Second, for completeness, we want to point out that there is also a second inefficiency associated with the competitor's challenge decision which cuts in the opposite direction: when $F = l_C$, the inventor is revealed as being of type L, so a challenge is socially desirable but does not occur. The planner cannot address this second inefficiency through patent office fees, however, because the fees have no effect on the rate of challenges when $F = l_C$ (it is always zero).

continues to be in the solution set. Thus, insufficient incentives to challenge do not provide an argument against frontloading of fees.[25]

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 simulations in Section 5), we have $\Delta_C > D$, i.e., the competitive disadvantage that the patent inflicts on the competitor is greater than the deadweight loss it causes. This alone is enough to imply that eq. (11) holds, and thus that challenges are socially excessive. 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 it imposes on the inventor (and society) by challenging. If litigation costs for both parties are identical ($l_C = l_I$), for example, (11) only requires $\Delta_C > D/2$.

That challenges are undesirable does not mean that the *threat* of challenges is undesirable. The presence of courts induces randomization by the type-L inventor and thus leads to lower license fees (region 1) or fewer bad applications (region 2). Our results do suggest, however, that the value of courts stems not so much from invalidating bad patents – when incentives to challenge are excessive, the expected reduction in deadweight loss is outweighed by litigation costs in equilibrium – but from deterring bad applications and keeping a lid on royalties charged by the low type.

Fee structure and examination intensity when the patent office is budget constrained Proposition 4 calls for an increase in patent office fees at least to the point where $\phi_P + \phi_A/(1-e) = l_C + m$ (region 2). Given that litigation costs can easily run a million dollars or more (AIPLA, 2011), such a steep hike may not be politically feasible. In a richer model, it might also adversely affect innovation incentives. But if patent policy is limited to operating within region 1, where patent office fees have no effect on any of the equilibrium variables, is there still an argument for frontloading? We now argue that there is, because frontloading raises fee revenue.

Suppose the patent office operates in region 1 and is constrained in the examination intensity it can set by the revenue it collects from fees, and that this constraint is binding. As a result the patent office operates at some e below the socially optimal level, so that an increase in e would raise welfare. Now consider a shift in the fee structure that leaves the sum of fees constant but shifts fees from post-grant to pre-grant. Because $\phi_A+\phi_P$ is constant and x remains at $\tilde{x}$, the profits of the type-H inventor are unaffected. The change in the fee structure, however, leads to more revenue being collected from type-L inventors who are refused patent protection and thus do not pay

[25] Among all solutions, $\phi_P = 0$ is the one that maximizes the high type's payoff.

the post-grant fee but do pay the pre-grant fee. This relaxes the patent office's budget constraint and allows it to implement a welfare-improving tightening of examination. We pursue this idea further in Section 5, where we consider a policy experiment along these lines.

5 Simulation Analysis

5.1 Baseline simulations: perfect courts

The model generates predictions on the grant rate, litigation rate and patent validation rate in court as functions of underlying parameters. In this section we use the observed values of these outcomes to calibrate the model's parameters, and then simulate the model to quantify the welfare effects of different policy reforms. We first summarize the simulation procedure and present the parameters from the baseline model (for more details, see Appendix C). We then describe the policy experiments and present the results. The observed patent fees and litigation costs are such that we are in region 1 (Figure 1, Section 3), hence $\alpha = 1$. For the simulations we set litigation cost to be the same for the inventor and competitor, $l_I = l_C = l$.

The simulations exploit equilibrium relationships from the theoretical model, together with Cournot competition between the incumbent innovator and a single competitor. It is a noteworthy feature of our model that, when assuming Cournot competition, the grant rate, litigation rate and validation rate uniquely pin down the parameters λ and e (without needing any other information). The main elements of the simulation structure are summarized below:

1. *Computed litigation rate:* To compute the litigation rate, we use the observed litigation rate for all patents and adjust it upward using the distribution of the value of patent rights estimated by Schankerman and Pakes (1986). The adjustment is made to account for the fact that only a subset of patents is worth litigating (i.e., they satisfy the challenge credibility constraint in the theoretical model).

2. *Equilibrium grant rate, GR:* The grant rate depends on the fraction of type-L patents, λ, and the examination intensity, e. This gives an equation $e = e(\lambda; GR)$.

3. *Equilibrium litigation rate, LR:* In the model this depends on λ, GR and validation rate VR. Thus we can compute λ from the implied relationship $\lambda = \lambda(LR, GR, VR)$. From Step 2, we thus get e. Using the observed examination cost per patent application at the USPTO, this allows us to back out the implied marginal cost of e, denoted by MCE.[26]

[26] We assume that examination costs are linear in examination intensity. We could easily allow for convex costs, given an assumption about the cost elasticity.

4. *Cournot model of competition:* An incumbent innovator faces a single competitor and a linear market demand. The incumbent can use his cost-reducing invention and also license it to the competitor. He either sets a high royalty rate, which generates income $F = \Delta_C$ or a low royalty rate with income $F = l_C$. In this model, the elasticity of demand in equilibrium, η, depends on the demand scale parameter a, initial unit cost c, and percentage cost reduction from the innovation, s. This implies $a = a(c; \eta, s)$.

5. *Equilibrium patent validation rate in court, VR:* In the model, the patent validation rate depends on (a, c, l). This gives an equation $c = c(a, l; VR)$.

6. *R&D equation:* We assume that expected profit from innovations is equal to observed R&D expenditures per patent, R, adjusted upward for a private rate of return. This implies a relationship $R = R(l, a, c; \lambda, s, \eta, VR, GR)$, where λ is obtained from Step 3. We thus have three (non-linear) equations – R&D equation, $a = a(c; \eta, s)$ from Step 4, and $c = c(a, l; VR)$ from Step 5. We solve these equations to obtain (unique) values of (a, c, l).

7. Finally, to compute development costs for type-L and H inventions, we use the inequalities that ensure that expected profit for the type-L inventor without patent protection is above κ_L, profit for the type-H inventor is below κ_H, and profit for the type-H inventor with the patent exceeds κ_H. This yields a set of feasible (κ_H, κ_L) pairs; for the simulations we use the mean of these values but the results are robust.

To summarize, the main inputs to the simulations are: patenting costs, including patent office fees ϕ_A and ϕ_P plus cost of drafting the application; demand elasticity, η; cost saving from innovation, s; patent validation rate, VR; grant rate, GR; patent litigation rate, LR; and R&D per patent, R. The main outputs of the simulations are the fraction of non-obvious inventions, λ; examination intensity, e; demand and cost parameters, (a, c); and development costs, (κ_H, κ_L). These parameters allow us to compute all components of the welfare function.

5.1.1 Simulation Results

Table 1 presents the baseline results and robustness checks. The two most striking results are that only 14 percent of patent applications are 'non-obvious' in the sense that they would not be developed in the absence of patent protection ($\lambda = 0.14$), and the patent office appears relatively ineffective at screening out 'obvious' patent applications, as they do so with a probability $e = 0.29$. These parameters imply that 81 percent of granted patents cover inventions that are 'obvious' (computed by $(1-\lambda)(1-e)/[\lambda+(1-\lambda)(1-e)]$), which highlights the severity of the patent-quality problem.

Table 1. Simulation Results: Perfect Courts

Panel A. Baseline

λ	e	l ($\times 10^6$)	y	x	Ratio	κ_L ($\times 10^6$)	κ_H ($\times 10^6$)	MCE
0.14	0.29	1.26	0.15	0.45	0.11	1.52	4.43	126

Notes: Baseline parameters: $s = 0.025$, $\eta = 2$, $R = \$2.4$ ($\times 10^6$), $l_{\min} = \$350$ ($\times 10^3$).

The column labeled 'Ratio' denotes the simulated value of $\frac{\Delta}{\Delta+\pi_I(0)}$ where Δ is the gain to the inventor from having a patent and $\pi_I(0)$ is the inventor's profit from an unpatented invention.

Panel B. Robustness

Perturbation	λ	e	l ($\times 10^6$)	y	x	*Ratio*	κ_L ($\times 10^6$)	κ_H ($\times 10^6$)	MCE
$R = \$2$ ($\times 10^6$)	0.14	0.29	1.05	0.15	0.45	0.11	1.26	3.69	126
$R = \$3$ ($\times 10^6$)	0.14	0.29	1.57	0.15	0.45	0.11	1.89	5.56	126
$s = 0.01$	0.14	0.29	1.27	0.15	0.45	0.047	1.51	4.45	126
$s = 0.05$	0.14	0.29	1.25	0.15	0.45	0.19	1.52	4.40	126
$\eta = 1$	0.14	0.29	1.27	0.15	0.45	0.04	1.51	4.46	126
$\eta = 3$	0.14	0.29	1.26	0.15	0.45	0.17	1.52	4.41	126

The simulations also generate three predictions that can be validated against external and independent sources of information. First, the simulations imply a litigation cost of \$1.26 million (2015 dollars) for each party, which is consistent with the survey evidence reported by the American Intellectual Property Law Association (AIPLA, 2011).[27] Second, the simulated parameters imply that patent rights account for 11 percent of the total returns to an invention (column 'Ratio'). This is very similar to the estimates generated by patent renewal models, which are typically in the range of 5-15 percent depending on the technology field (Schankerman, 1998). Third, Frakes and Wasserman (forthcoming) estimate the impact of changes in time spent by examiners per application on the probability that the patent is granted. Their estimates imply an elasticity of the grant rate with respect to time spent from about minus one third to minus one half. Our model allows us to compute a similar elasticity with respect to

[27]Their survey reports the typical patent litigation costs for different value at stake: less than one million, 1-25 and greater than 25. We take a weighted average with weight of 90% on the first, reflecting the fact that more than 90% of cases are settled before the trial begins (Lanjouw and Schankerman, 2004). This gives $l = \$1.5$ million.

examination intensity (using the fact that the grant rate is $GR = \lambda + (1 - \lambda)(1 - e)$). Our baseline estimates of λ and e imply an elasticity of -0.33.

The implied investment costs for type-L and H inventions are \$1.52 and \$4.43 million, respectively. The cost per patent application of increasing examination intensity by one percentage point, denoted by MCE – e.g., the cost of moving from $e = 0.29$ to $e = 0.30$ – is \$126. The probability that a type-L inventor mimics a type-H inventor by charging the high royalty rate is $y = 0.15$. This finding is noteworthy because it means that 85 percent of obvious inventions are never challenged because the low royalty rate is set at the competitor's litigation cost in order to preempt challenges. For those type-L inventors who charge the high royalty, the probability that a competitor challenges is $x = 0.45$.

The simulation results are very robust to changes in the assumed parameters. In each row of Panel B we vary the one parameter indicated, relative to the baseline specification. Reducing the assumed R&D cost per patent to \$2 million, or increasing it to \$3 million, affects only the implied litigation cost l and development costs (κ_L, κ_H). The baseline cost reduction for the invention is 2.5 percent (calibrated at roughly the economy-wide rate of total factor productivity growth). The main effect of changing it to 1 percent or 5 percent is to alter the simulated demand and cost parameters (not reported) and thus the gains from the patent relative to the total returns to R&D: this ratio varies from 5 to 19 percent. The results are similar when we vary the assumed elasticity of demand, down to $\eta = 1$ or up to $\eta = 3$.

5.1.2 Policy Experiments

In this section we analyze the change in the welfare gains from innovation generated by counterfactual policy reforms. The pre-reform welfare gain from innovation is defined as welfare in the baseline patent regime (pre-reform) minus welfare without innovation. The post-reform welfare gain from innovation is defined analogously. For brevity, we will refer to these changes in welfare gains from innovation as 'welfare changes'.

Table 2 summarizes the simulation results for five policy experiments. First, we frontload post-grant fees in a revenue neutral way. Frontloading raises additional revenue, because all type-L applicants pay the pre-grant fees, whereas only successful type-L applicants pay the post-grant fees. We use to this revenue to increase the examination intensity.[28] This frontloading reform generates enough extra revenue to double the examination intensity. The direct effect of the increase in examination intensity

[28] In these simulations, the parameters are such that we remain in region 1 (Figure 1 in Section 3), so that the application rate of type-L inventions does not change ($\alpha = 1$). Raising patent fees so high that region 2 would become relevant is not politically feasible, and would require us to incorporate the potential effects of these fees on the supply of innovation.

is that the patent office detects and denies patent rights to more type-L patents that apply, which increases welfare. There is also an indirect effect. The increase in examination intensity raises the probability that a type-L inventor charges the high royalty (and thus average royalties in the economy) because there is a stronger presumption by the competitor that a high royalty signals a type-H invention. This indirect effect reduces welfare, but it is dominated by the direct effect, so that the reform *increases* welfare by 1.90 percent, with consumers gaining at the expense of firms (inventor and competitor).

The second experiment is to cap litigation costs at half the level given by the baseline simulation. This reform has a *very large* and positive impact on welfare, 4.92 percent, and both inventors and consumers gain. The reason is that litigation costs are high and reducing them saves inventors money when challenged. At the same time, type-L inventors are less likely to charge the high royalty rate (i.e., y declines from 0.15 to 0.06) because challenges are more attractive given the lower litigation costs. This experiment highlights the potential importance of tort reform in the patent arena.[29]

In the third experiment we 'turn off' court challenges entirely (e.g., as would occur if litigation costs were high enough to violate the challenge credibility constraint). This reform has a surprisingly small (negative) effect on welfare, which reflects the savings on the very substantial litigation costs when challenges are feasible. While inventors benefit – especially those with obvious inventions because they can now charge the high royalty without fear of being challenged – consumers lose heavily from the consequently higher output prices. The presence of courts as a second screening device appears much less significant for welfare than one might have thought, even with perfect courts as here – essentially because litigation costs are so high.

The fourth experiment replaces the patent system with a pure registration system in which there is no patent examination but retains the option of challenging the registered right in court. In this case, we set both pre- and post-grant fees to zero, but the cost of preparing the application for registration is maintained (as this would be needed as the basis for any court challenge). This reform *reduces* welfare substantially, by 3.1 percent, coming at the expense of consumers while benefitting firms.

Lastly, we increase the examination intensity from the baseline value to 0.5 and 0.7, and cover the incremental cost by increasing pre-grant fees. This increases welfare, with consumers gaining at the expense of firms. Of course, we have assumed a constant marginal cost of examination, but if costs are convex the welfare gains would be smaller.

[29]For recent empirical evidence of the impact of tort reform (limited liability) on innovation in the healthcare sector, see Galasso and Luo (2016).

Table 2. Policy Experiments with Perfect Courts

Experiment	e	y	x	$\%\mathbb{E}\Delta W$	$\%\mathbb{E}\Delta\pi$	$\%\ \mathbb{E}\Delta CS$
Status quo	0.29	0.15	0.45	-	-	-
Frontloading	0.56	0.24	0.45	1.90	-5.88	5.72
Halving litigation costs	0.29	0.06	0.69	4.92	0.78	8.24
Shutting down courts	0.29	1.0	0	-0.74	42.0	-22.3
Registration system	0	0.14	0.45	-3.10	4.76	-6.50
More stringent examination I	0.50	0.21	0.45	1.49	-4.42	4.49
More stringent examination II	0.70	0.36	0.45	2.91	-8.63	8.76

Note: Changes in welfare, profit and consumer surplus are measured net of the no-invention levels.

In the baseline model, we assume that perfect courts always uphold type-H inventions and completely invalidate type-L inventions, i.e., all claims in the type-L patent are revoked. However, in practice the courts typically invalidate only a fraction of claims. To account for this, we formulated a modified version of the baseline model with perfects courts to allow for partial invalidation.[30] In this model we treat patent claims as independent of each other, so that if a fraction f of claims is invalidated, the cost reduction equal to fs is now available to the competitor and the royalty that the incumbent can extract is only on the remaining proprietary cost reduction of $(1-f)s$.[31]

We simulated this alternative model to check robustness of the findings in Table 1 and 2 (results not reported for brevity). Allowing for partial invalidation changes the simulated parameters in three main ways. First, the fraction of non-obvious inventions rises from $\lambda = 0.14$ to $\lambda = 0.28$, and the effectiveness of patent office screening increases from $e = 0.29$ to $e = 0.35$. Second, the implied cost of litigation declines from \$1.26 to \$0.79 million. Third, the probability that a type-L inventor mimics the type-H inventor by charging the high royalty rate increases sharply from $y = 0.15$ to $y = 0.40$. In the policy experiments, we again find that frontloading fees, capping litigation costs and raising examination intensity all increase welfare, and introducing a registration system reduces it. However, with partial invalidation, shutting off court challenges actually increases welfare (rather than modestly reducing it as with full invalidation). The reason is that we are now making court screening imperfect and, with substantial legal costs, they are no longer worthwhile.

[30] Details are available on request. We are grateful to John Golden for pointing this out and suggesting a robustness check with this more realistic specification of the courts. Using information on cases at the Court of Appeals for the Federal Circuit (Galasso and Schankerman, 2015), the fraction of claims revoked in cases of patent invalidation is about 75 percent, and we use this figure to calibrate this model.

[31] Independence of claims seems a natural assumption for this analysis. A firm has no incentive to include claims that are substitutes for each other. At the other extreme, if all claims are perfect complements, the original model applies since loss of any one claim is equivalent to the loss of the entire patent.

5.2 Simulations: imperfect courts

5.2.1 A more general screening technology

Many observers argue that patents are "probabilistic" in nature, with court decisions subject to a considerable degree of randomness (Lemley and Shapiro, 2005). We have assumed that courts do not make mistakes in assessing validity. In Appendix B we analyze a more general screening technology where both the patent office and courts sometimes make mistakes. This technology encompasses as special cases both the basic model (with perfect courts) and completely random courts (where the probability a patent is upheld is independent of the invention type, θ). The appendix shows that the basic insights from the baseline model are robust to this generalization, provided that court screening is sufficiently accurate for the competitor to challenge when certain of facing a type-L inventor and not to challenge when certain of facing a type-H inventor.

For the simulation analysis with imperfect courts that follows, we will use a special case of the more general screening technology. We allow the court to screen out invalid patents with a probability that lies between one (perfect court) and the probability in the patent office. In addition, we allow the court and patent office to apply different rules in the event the evidence is weak. We assume that the patent office only rejects if it finds strong evidence that the invention is type L. By contrast, the courts apply a "presumption of validity" that depends on the updated posterior on the invention type, which in turn depends on the intensity of patent office examination. Other things equal, when patent office screening is more intensive, the court applies a stronger presumption that the patent is valid, in the absence of strong evidence against it.

Specifically, with probabilities $e_1 = e$ and $e_2 = \beta + (1-\beta)e$, respectively, the patent office and court each receive an independent draw of a signal that reveals the true invention type with certainty, where $\beta \in [0, 1]$. The case $\beta = 1$ corresponds to the baseline model of perfect courts. In the absence of a signal revealing the invention to be of type L, the patent office always accepts the patent, whereas the court upholds it with probability $\hat{\lambda}$ and invalidates it with probability $1 - \hat{\lambda}$, where $\hat{\lambda}$ is the posterior probability that the invention is of type H.[32] Comparing the simulation results for perfect and imperfect courts will allow us to check how our assessment of patent screening institutions, and of the welfare effects of different policy reforms, varies with the characterization of court behavior.

5.2.2 Simulation results

Table 3 presents the simulated parameters for the model with imperfect courts. We present results for two different scenarios: low quality courts ($\beta = 1/3$) and high quality

[32]For more details, see Appendix C.5.

courts ($\beta = 2/3$).

Table 3. Simulated Parameters with Imperfect Courts

Panel A. High Quality Courts

λ	e	l ($\times 10^6$)	y	x	Ratio	κ_L ($\times 10^6$)	κ_H ($\times 10^6$)	MCE
0.17	0.30	1.12	0.20	0.39	0.11	1.42	4.32	121

Panel B. Low Quality Courts

λ	e	l ($\times 10^6$)	y	x	Ratio	κ_L ($\times 10^6$)	κ_H ($\times 10^6$)	MCE
0.26	0.34	1.02	0.35	0.32	0.11	1.25	3.83	109

With high quality but imperfect courts, the simulated parameters are very similar to the case of perfect courts (compare Tables 1 and Panel A, Table 3). The imperfect court model gives a slightly higher proportion of non-obvious patent applications (λ) and effectiveness of patent office screening (e). In addition, the probability that a type-L inventor charges the high royalty (y) is somewhat higher (implying that there is a smaller probability he preempts a challenge). This is because the imperfect court may actually uphold the type-L patent, making it more attractive to take the risk of charging the high royalty. The other parameters are very similar to the earlier results. Panel B in Table 3 shows that the results are generally robust even with a low-quality court. The main difference with the perfect court results is that the proportion of non-obvious inventions nearly doubles, from 0.14 with perfect courts to 0.26 and, not surprisingly, the probability that a type-L inventor charges the high royalty rate increases sharply from 0.15 to 0.35. The simulated values of λ and e imply that between 65 and 77 percent of granted patents are invalid in the specifications with low-quality and high-quality (imperfect) courts, respectively.

Table 4 presents the simulation results for the policy experiments with both low and high quality (imperfect) courts. Comparing these results to the model with perfect courts (Table 2), we observe the same overall pattern emerging. Frontloading fees, capping litigation costs and increasing examination intensity all generate welfare gains, while moving to a registration system reduces welfare. The only difference is that, while shutting down the courts modestly reduced welfare in the case of perfect courts, the policy actually improves welfare when courts are imperfect, especially when they are low quality. It is also worth noting that frontloading fees (using the extra revenue to fund higher examination intensity) generates significantly larger welfare gains when courts are imperfect (compare Tables 2 and 4).

Table 4. Policy Experiments with Imperfect Courts

Panel A. High Quality Courts

Experiment	e	y	x	$\%\mathbb{E}\Delta W$	$\%\mathbb{E}\Delta\pi$	$\%\ \mathbb{E}\Delta CS$
Status quo	0.30	0.20	0.39	-	-	-
Frontloading	0.58	0.33	0.39	2.82	-9.14	8.23
Halving litigation costs	0.30	0.08	0.63	5.41	3.13	6.52
Shutting down courts	0.30	1.0	0	1.94	43.66	-17.36
Registration system	0	0.14	0.39	-3.06	10.93	-8.93
More stringent examination I	0.50	0.28	0.39	2.00	-6.26	5.84
More stringent examination II	0.70	0.35	0.39	3.01	-9.43	8.80

Panel B. Low Quality Courts

Experiment	e	y	x	$\%\mathbb{E}\Delta W$	$\%\mathbb{E}\Delta\pi$	$\%\ \mathbb{E}\Delta CS$
Status quo	0.34	0.35	0.32	-	-	-
Frontloading	0.65	0.66	0.32	3.89	-13.82	10.88
Halving litigation costs	0.34	0.13	0.55	5.45	7.55	4.67
Shutting down courts	0.34	1.0	0	6.61	52.23	-11.64
Registration system	0	0.23	0.32	-4.22	16.38	-11.80
More stringent examination I	0.50	0.46	0.32	2.04	-7.02	5.70
More stringent examination II	0.70	0.77	0.32	4.54	-15.64	12.69

6 Extensions

6.1 Heterogeneous invention values

The baseline model assumes that inventions have identical private (and social) value, but differ in their development cost. The key distinction is between the high type, which needs patent protection to cover this cost, and the low type, which does not. In this section we study a modified version of the model where inventions differ in both value and development cost. The distinction between high and low types is still based on whether patent rights are needed to induce development, but here this varies with both cost and value. In this version, as in the baseline, the supply of ideas for inventions is exogenous.

We show that two key ideas from the baseline model also hold in this context. First, we cannot screen out low types of a given value without using *both* examination *and* pre-grant fees (i.e., these two instruments are complements). Second, holding the incentive to invest for high types fixed, frontloading patent fees reduces the set of low types that apply.

Suppose ideas differ along two dimensions: value, indexed by $v \in [\underline{v}, \overline{v}]$, and development cost $\kappa \in [\underline{\kappa}, \overline{\kappa}]$. In the absence of patent protection an idea of value v is associated with profit $\pi(v)$ and welfare $w(v)$. With a patent the inventor earns an additional $\Delta(v) = m(v) + \Delta_C(v)$ while welfare is reduced by $D(v)$. Ideas (v, κ) are drawn from some continuous distribution. Both the inventor and the competitor observe v, but only the inventor observes κ. Translated to this environment, our patentability requirement becomes $\kappa > \pi(v)$. To remain consistent with our previous terminology, we refer to types with $\kappa > \pi(v)$ as high types and to those with $\kappa \leq \pi(v)$ as low types. For simplicity, assume that $l_I = 0$ and $l_C = l(v) < \Delta_C(v)$ for all v. Furthermore, assume that all functions of v are continuous and increasing.

The following proposition characterizes the equilibrium.

Proposition 5. *There exists a cutoff v^*, defined by*

$$\Delta(v^*) = \phi_A + \phi_P, \tag{12}$$

such that high types never invest for $v < v^*$. For $v \geq v^*$, they invest, apply, activate and propose $F = \Delta_C(v)$ if $\kappa \leq \pi(v) + \Delta(v) - \phi_A - \phi_P$ and do not invest otherwise. Similarly, there exist two cutoffs $\hat{v}$ and $\tilde{v}$, defined by

$$\Delta(\hat{v}) = \phi_P + \frac{\phi_A}{1-e} \tag{13}$$

$$l(\tilde{v}) + m(\tilde{v}) = \phi_P + \frac{\phi_A}{1-e}, \tag{14}$$

such that low types behave as follows:

(i) for $v < \hat{v}$: $\alpha = 0$ and $y = 1$

(ii) for $\hat{v} \leq v < \tilde{v}$: (a) if challenges are credible, $\alpha = \tilde{\alpha}(v)$ and $y = 1$, (b) if challenges are not credible, $\alpha = y = 1$

(iii) for $\tilde{v} \leq v$: (a) if challenges are credible, $\alpha = 1$ and $y = \tilde{y}(v)$, (b) if challenges are not credible, $\alpha = y = 1$

where $\tilde{\alpha}(v)$ and $\tilde{y}(v)$ are such that the competitor is indifferent between challenging and not.

This is a straightforward extension of the analysis from the baseline model.[33] The proposition says that high types only invest and apply if their ideas are sufficiently

[33]The intuition is as follows. For high types, v^* is the threshold value at which the gains from patenting exceed the costs, so – provided the development cost is low enough – they apply for and activate their patent and, as in the baseline model, charge the high royalty. For low types, $\hat{v}$ is the corresponding threshold – they earn at most $\Delta(v)$ from a patent and pay the post-grant fee, with probability $(1-e)$, plus pay the application fee. Thus for $v < \hat{v}$ they do not apply. If challenges are not credible, then for any $v > \hat{v}$, low types always apply and charge the high royalty since they are not at risk of litigation. If challenges are credible, low types randomize. If v is high enough for the low royalty $l(v) + m(v)$ (the license fee that preempts challenges plus the extra rent from the incremental market power) to cover fees, $v > \tilde{v}$, they always apply and randomize over the royalty rate. If the low royalty does not generate sufficient revenue ($\hat{v} < v < \tilde{v}$), they randomize over application and charge the high royalty.

valuable. Low types always invest, and apply for patents at least some of the time if their ideas are sufficiently valuable. The rate at which they apply, α, depends on v and on whether challenges are credible given v.[34]

The application behavior is illustrated in Figure 4 for the case where challenges are credible for all $v \geq \hat{v}$. The figure depicts the space of ideas (v, κ), whose support is represented by the dashed box. The grey shaded area shows the set of high types that invest and apply, i.e., those for which $\pi(v) < \kappa \leq \pi(v) + \Delta(v) - \phi_P - \phi_A$. The figure also shows how different sets of low types ($\kappa \leq \pi(v)$) differ in their application behavior: those with $v < \hat{v}$ never apply, those with $\hat{v} \leq v < \tilde{v}$ randomize over the application decision, and those with $v \geq \tilde{v}$ apply with certainty.[35] The figure shows how inventor behavior varies with value, holding patent fees constant. One could easily add another dimension to the figure to show how the equilibrium varies with patent fees. It should be clear that changes in fees would have differential effects on high- and low-value inventors.

Two remarks emerge from this analysis. First, $\hat{v}$, and thus the set of low types that apply at least some of the time, does not depend on whether court challenges are credible (or, equivalently, whether they are available). In other words, court challenges have no effect on the *extensive margin* of bad applications (i.e., the interval of values over which the low types apply); they affect only the intensive margin (i.e., the probability that they apply). Like in the baseline model, the intuition is that the threat of challenges cannot completely deter a low type from applying, since this would lead his (Bayesian) competitor to refrain from challenging.

Second, any change in patent office fees will affect R&D investment decisions. The level of innovation investment (by both high and low types) is fully characterized by v^*; clearly, higher fees raise v^* and reduce investment. (Note that a fee increase leads the set of types that invest to shrink in both dimensions of the idea space.)

We then have the following results.

Proposition 6. *The cutoff on v above which high types apply is always weakly below the corresponding cutoff for low types: $v^* \leq \hat{v}$, with equality if either $e = 0$, or $\phi_A = 0$*

[34] Formally, let $G_v(\kappa)$ denote the cumulative distribution function of κ conditional on v. Challenges are credible for some $v \geq v^*$ if and only if $(1 - \underline{\lambda}(v))\Delta_C(v) \geq l(v)$, where

$$\underline{\lambda}(v) = \frac{G_v(\pi(v) + \Delta(v) - \phi_A - \phi_P) - G_v(\pi(v))}{G_v(\pi(v) + \Delta(v) - \phi_A - \phi_P) - eG_v(\pi(v))}.$$

Making the competitor indifferent requires $(1 - \hat{\lambda}(v))\Delta_C(v) = l(v)$, where

$$\hat{\lambda}(v) = \frac{G_v(\pi(v) + \Delta(v) - \phi_A - \phi_P) - G_v(\pi(v))}{G_v(\pi(v) + \Delta(v) - \phi_A - \phi_P) - G_v(\pi(v)) + (1-e)\alpha y G_v(\pi(v))}.$$

[35] The assumption $l(v) < \Delta_C(v)$ ensures that $\hat{v} < \tilde{v}$.

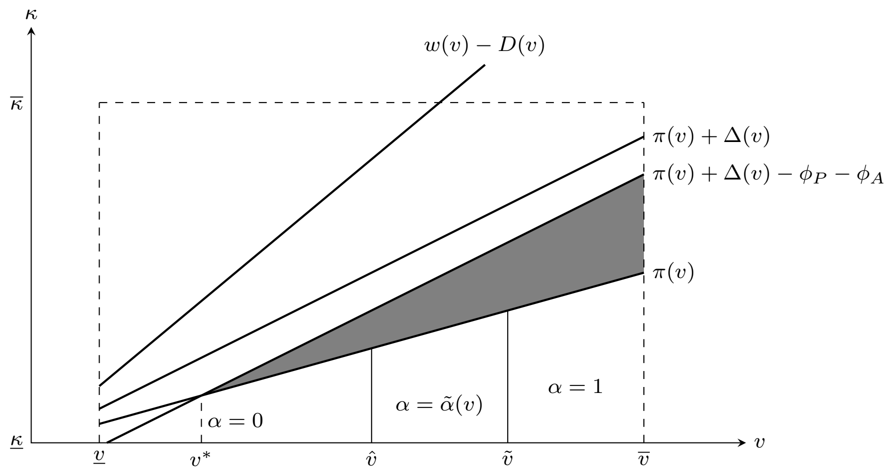

Figure 4: The equilibrium with continuous types

and $e < 1$.

Proof. Because $\Delta(v)$ is an increasing function and the right-hand side of (13) weakly exceeds that of (12), we obtain $v^* \leq \hat{v}$. If $e = 0$, or if $\phi_A = 0$ and $e < 1$, the right-hand sides of the two expressions coincide and so $v^* = \hat{v}$. □

The first result in Proposition 6 says that there cannot be a value v such that low types apply but high types do not. The reason is that the high type with $\kappa = \pi(v) + \varepsilon$ (invests and) applies for a patent if and only if $\Delta(v) \geq \phi_P + \phi_A + \varepsilon$. Letting $\varepsilon \to 0$, this is also the condition for low types to ever find applying worthwhile. The second result in Proposition 6 is the counterpart of Observation 1 from the simple model of Section 2. It says that, in the absence of either patent examination or pre-grant fees, we cannot deter low types with ideas of value v without also deterring high types with ideas of the same value. This is related to the fact that, as argued above, court challenges do not affect the extensive margin of bad applications; hence, the set of low types that apply at least some of the time when there is no examination is the same as when there is neither examination nor court review. Together with $v^* \leq \hat{v}$, this implies that for $e = 0$ the two cutoffs must coincide.

Proposition 7. *Fix* v^* *and* e. *Then applications from low types are minimized by frontloading fees, i.e., setting* $\phi_P = 0$ *and* $\phi_A = \Delta(v^*)$.

Proof. See Appendix A. □

Proposition 7 is the counterpart of Observation 2 from Section 2. It says that, if we fix the level of innovation in the economy at v^* (and hence the sum of patent office fees at $\phi_A + \phi_P = \Delta(v^*)$) and the examination intensity at some e, frontloading fees reduces the number of applications filed by low types. This is because the cutoffs $\hat{v}$ and $\tilde{v}$ are increasing in ϕ_A when holding $\phi_A + \phi_P$ fixed – for the same reason as in the baseline model, namely, that low types prefer post-grant fees to pre-grant fees whereas high types are indifferent.[36]

6.2 Endogenous inventions

One possible objection to our patentability requirement is that it appears to encourage high-cost inventions. This concern does not apply to our model in which the population of obvious (low cost) and non-obvious (high cost) inventions is given exogenously. The objection more naturally arises in a different environment where the mix of inventions is endogenous. However, even in such an environment, our patentability requirement can be justified.

To see this, consider the two-dimensional model from the previous subsection but suppose that, instead of continuous support, we have $v \in \{v_L, v_H\}$ and $\kappa \in \{\kappa_L, \kappa_H\}$. That is, as in the baseline model there are two levels of development cost, but now there are also two levels of value, with associated profits $\pi(v_L)$ and $\pi(v_H) > \pi(v_L)$.

Suppose the inventor can choose between different ideas but when deciding whether to invest only observes κ and not v. Assume, however, that κ is a signal of value: that is, letting $p_\kappa \equiv \Pr(v = v_H|\kappa)$, $p_{\kappa_H} > p_{\kappa_L}$. Then expected welfare can be written $E(w(v)|\kappa) = p_\kappa w(v_H) + (1 - p_\kappa)w(v_L)$, where $w(v)$ denotes welfare when the invention has value v. If $E(w(v)|\kappa_H) - \kappa_H > E(w(v)|\kappa_L) - \kappa_L$, or equivalently

$$(p_{\kappa_H} - p_{\kappa_L})(w(v_H) - w(v_L)) > \kappa_H - \kappa_L,$$

the planner wants to encourage the inventor to go for ideas with $\kappa = \kappa_H$, even though they have a high development cost. That is, the planner would like to promote "ambitious" research if it generates a sufficiently high payoff in a probabilistic sense.

7 Concluding remarks

This paper develops a framework to examine how governments can improve the quality of patent screening, incorporating four policy instruments: patent office examination, pre- and post-grant fees, and challenges in the courts. The analysis yields three key theoretical results. First, if the patent office makes no examination effort (pure registration system), or if the pre-grant fee is zero and examination is imperfect, complete screening (only high types obtaining patents) cannot be achieved. Thus pre-grant fees

[36] As shown in the proof of Proposition 7, frontloading fees leaves $\tilde{\alpha}(v)$ unaffected. Hence, if $\hat{v}$ increases to $\hat{v}'$ and $\tilde{v}$ to $\tilde{v}'$, applications by low types decrease for $v \in [\hat{v}, \hat{v}']$ (from $\tilde{\alpha}(v)$ to 0) and $v \in [\tilde{v}, \tilde{v}']$ (from 1 to $\tilde{\alpha}(v)$) while remaining unchanged for all other values of v.

and examination are complements, not substitutes as one might think. Second, even if courts are mistake-free, they cannot eliminate all bad patents that are issued because in equilibrium not all such patents are challenged. This result raises serious doubts about over-reliance on the court system to weed out obvious patents. Third, when examination is not sufficiently rigorous to achieve complete screening, we show that a social planner would always *frontload* fees.

We simulate the model, calibrated on U.S. patent and litigation data, to identify key parameters and study the welfare effects of different policy reforms. The simulations indicate that 75 percent or more of patent applications are low quality in the sense that they are made on inventions that would be developed even without patent rights. At the same time, the patent office only screens out about 30 percent of these low-quality applications. Our simulations imply that between 65 and 81 percent of granted patents are invalid, depending on the specification of the courts. These findings highlight the crisis in patent screening and the need to develop effective policies to address it. The simulations show that intensifying patent office examination, frontloading patent fees, and capping litigation costs all generate welfare gains, while replacing examination with a pure registration system reduces welfare.

We believe the framework developed in this paper can be used to study the welfare impacts of tort reform and patent litigation insurance, both of which have featured prominently in the public debate. Our model can potentially provide a useful analytical platform for studying other patent reforms as well. In principle, the simulations of the model, or suitable extensions of it, could also be calibrated for other patent systems. This might allow one to analyze the welfare effects of more ambitious, international patent reforms, such as harmonized screening.

References

[1] Allison, John R., Mark A. Lemley and David L. Schwartz (2014), "Understanding the Realities of Modern Patent Litigation," *Texas Law Review* (92): 1769-1801

[2] American Intellectual Property Law Association (2011), *Report of the Economic Survey 2011* (Washington, D.C.: AIPLA)

[3] Atal, Vidya and Talia Bar (2014), "Patent Quality and a Two-Tiered Patent System," *Journal of Industrial Economics*, 62, 503-540

[4] Bessen, James and Eric Maskin (2009), "Sequential innovation, patents, and imitation," *RAND Journal of Economics*, 40, 611-635

[5] Bloom, Nick, Mark Schankerman, and John Van Reenen (2013), "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica,* 81: 1347-1393

[6] Caillaud, Bernard and Anne Duchene (2011), "Patent Office in Innovation Policy: Nobody's Perfect," *International Journal of Industrial Organization*, 29: 242-252

[7] Carley, Michael, Deepak Hegde, and Alan Marco (2015), "What is the Probability of Receiving a U.S. Patent?" *Yale Journal of Law & Technology*, 17(1): 203-223

[8] Choi, Jay Pil (1998), "Patent Litigation as an Information-Transmission Mechanism," *American Economic Review*, 88(5): 1249-1263

[9] Cornelli, Francesca and Mark Schankerman (1999), "Patent Renewals and R&D Incentives," *RAND Journal of Economics*, 30: 197-213

[10] Crampes, Claude and Corinne Langinier (2002), "Litigation and Settlement in Patent Infringement Cases," *RAND Journal of Economics*, 33(2): 258-274

[11] Dreyfuss, Rochelle Cooper (2008), "Non-obviousness: A Comment on Three Learned Papers," *Lewis and Clark Law Review*, 12: 431-441

[12] Eisenberg, Rebecca (2004), "Obvious to Whome – Evaluating Inventions from the Perspective of PHOSITA," *Berkeley Technology Law Journal*, 19(3): 885-906

[13] Farrell, Joseph and Robert P. Merges (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(3): 943-970

[14] Farrell, Joseph and Carl Shapiro (2008), "How Strong are Weak Patents?," *American Economic Review*, 98: 1347-1369

[15] Fauli-Oller, Ramon and Joel Sandonis (2002), "Welfare Reducing Licensing," *Games and Economic Behavior*, 41: 192-205

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

[17] Federal Trade Commission (2016), *Patent Assertion Entity Activity: An FTC Study* (Washington D.C.: Government Printing Office)

[18] Frakes, Michael D. and Melissa F. Wasserman (2015), "Does the U.S. Patent and Trademark Office Grant Too Many Bad Patents?: Evidence from a Quasi-Experiment," *Stanford Law Review*, 67: 613-676

[19] Frakes, Michael D. and Melissa F. Wasserman (forthcoming), "Is the Time Allocated to Review Patent Applications Inducing Examiners to Grant Invalid Patents?: Evidence from Micro-Level Application Data," *Review of Economics and Statistics*

[20] Galasso, Alberto and Hong Luo (2016), "Tort Reform and Innovation," NBER Working Paper 22712

[21] Galasso, Alberto and Mark Schankerman (2015), "Patents and Cumulative Innovation: Causal Evidence from the Courts," *Quarterly Journal of Economics*, 130: 317-369

[22] Heller, Mark and Rebecca Eisenberg, "Can Patents Deter Innovation? The Anticommons in Biomedical Research," *Science*, 280 (1998), 698-701

[23] Henkel, Joachim and Hans Zischka (2015), "Why most Patents are Invalid – Extent, Reasons, and Potential Remedies of Patent Invalidity," mimeo, Technical University of Munich

[24] Hopenhayn, Hugo and Matthew Mitchell (2000), "Innovation Variety and Patent Breadth," *RAND Journal of Economics*, 32(1): 152-166

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

[26] Kou, Zonglai, Patrick Rey and Tong Wang (2013), "Non-Obviousness and Screening," *Journal of Industrial Economics,* LXI(3): 700-732

[27] Lanjouw, Jean and Mark Schankerman (2004), "Protecting Intellectual Property Rights: Are Small Firms Handicapped?" *Journal of Law and Economics*, 47: 45-74

[28] Lemley, Mark A. (2001), "Rational Ignorance at the Patent Office," *Northwestern University Law Review*, 95(4): 1495-1532

[29] Lemley, Mark A. and Carl Shapiro (2005), "Probabilistic Patents," *Journal of Economic Perspectives*, 19(2): 75-98

[30] Lichtman, Doug and Mark A. Lemley (2007), "Rethinking Patent Law's Presumption of Validity," *Stanford Law Review*, 60(1): 42-72

[31] Meurer, Michael J. (1989), "The Settlement of Patent Litigation," *RAND Journal of Economics,* 20(1): 77-91

[32] Noveck, Beth (2006), "Peer to Patent: Collective Intelligence, Open Review, and Patent Reform," *Harvard Journal of Law and Technology*, 20(1): 123 - 162

[33] Paradise, Jordan, Lori Andrews and Timothy Holbrook (2005). "Patents on Human Genes: an Analysis of Scope and Claims," *Science*, 307(5715): 1566-1567

[34] Schankerman, Mark (1998), "How valuable is patent protection? Estimates by technology field," *RAND Journal of Economics*, 29: 77-107

[35] Schankerman, Mark and Ariel Pakes (1986), "Estimates of the Value of Patent Rights in European Countries during the Post-1950 Period," *Economic Journal,* 96: 1052-1076

[36] Scotchmer, Suzanne (1999), "On the Optimality of the Patent Renewal System," *RAND Journal of Economics*, 30: 181-196

[37] Schuett, Florian (2013a), "Patent Quality and Incentives at the Patent Office," *RAND Journal of Economics,* 44: 313-336

[38] Schuett, Florian (2013b), "Inventors and Imposters: An Analysis of Patent Examination with Self-Selection of Firms into R&D," *Journal of Industrial Economics,* LXI(3): 660-699

[39] Shapiro, Carl (2001), "Navigating the Patent Thicket: Cross-Licenses, Patent Pools, and Standard-Setting," in: Jaffe, A., J. Lerner, and S. Stern (eds.), *Innovation Policy and the Economy*: 119-150 (Cambridge, MA: MIT Press)

[40] The Economist (2015), "Time to Fix Patents," August 8th-14th, p. 9

APPENDICES – FOR ONLINE PUBLICATION

Appendix A Proofs

Proposition 1. *If $\phi_P > \Delta - \phi_A/(1-e)$, the type-L inventor does not apply and the competitor does not challenge. If challenges are credible and $\phi_P \leq \Delta - \phi_A/(1-e)$, there is a semi-separating equilibrium in which:*

(i) the type-H inventor invests, applies, activates, and proposes $F^H = \Delta_C$

(ii) the type-L inventor always randomizes over either

(a) the decision whether to apply or not (probabilities α and $1-\alpha$), or

(b) the license fee to propose, $F^L \in \{\Delta_C, l_C\}$ (probabilities y and $1-y$),

such that $(1-\hat{\lambda})\Delta_C = l_C$; he always activates ($\rho = 1$)

(iii) the competitor randomizes over the decision whether to challenge or not (probabilities x and $1-x$) if offered $F = \Delta_C$ and never challenges if offered $F = l_C$, with

$$x = \begin{cases} \tilde{x} \equiv \dfrac{\Delta_C - l_C}{\Delta + l_I} & \text{for } \phi_P < l_C + m - \dfrac{\phi_A}{1-e} \\[2ex] \hat{x} \equiv \dfrac{\Delta - \phi_P - \phi_A/(1-e)}{\Delta + l_I} & \text{for } l_C + m - \dfrac{\phi_A}{1-e} \leq \phi_P \leq \Delta - \dfrac{\phi_A}{1-e} \end{cases}$$

This semi-separating equilibrium requires that the high type make nonnegative profit, $\Delta - (k - \pi) - x l_I - \phi_A - \phi_P \geq 0$.

Proof. We first derive the equilibrium behavior of type-L inventor and competitor assuming that the type-H inventor invests, applies, activates, and proposes $F^H = \Delta_C$ in Lemma 1. We then show under what conditions type H indeed finds this behavior optimal in Lemma 2.

Lemma 1. *Suppose* (5) *holds and the type-H inventor invests, applies, activates, and proposes F^H. Then, the equilibrium behavior of the type-L inventor and the competitor can be characterized as follows:*

*(i) If $\phi_P \geq \Delta - \phi_A/(1-e)$ [**Region 3**], the type-L inventor does not apply ($\alpha = 0$), and the competitor does not challenge ($x = 0$). If granted a patent the type-L inventor would always propose F^H ($y = 1$).*

(ii) If $\phi_P < \Delta - \phi_A/(1-e)$, the type-L inventor applies with strictly positive probability ($\alpha > 0$) and randomizes such that

$$\left(1 - \frac{\lambda}{\lambda + (1-\lambda)(1-e)\alpha\rho y}\right)\Delta_C = l_C. \tag{15}$$

The competitor always accepts when offered $F = l_C$ ($x = 0$); she always challenges the patent when offered $F \notin \{l_C, F^H\}$ ($x = 1$). When offered $F = F^H$ she randomizes between accepting and challenging. Specifically:

*(a) for $\phi_P < l_C + m - \phi_A/(1-e)$ **[Region 1]**, the type-L inventor chooses $\alpha = \rho = 1$ and $y \in (0,1)$ solving (15). The competitor challenges with probability x such that*

$$(1-x)\Delta - x l_I = l_C + m; \tag{16}$$

*(b) for $\phi_A > 0$ and $l_C + m - \phi_A/(1-e) < \phi_P < \Delta - \phi_A/(1-e)$ **[Region 2]**, the type-L inventor chooses $\rho = y = 1$ and $\alpha \in (0,1)$ solving (15). The competitor challenges with probability x such that*

$$(1-e)\left[(1-x)\Delta - x l_I - \phi_P\right] = \phi_A; \tag{17}$$

(c) for $\phi_A = 0$ and $l_C + m < \phi_P < \Delta$, the type-L inventor chooses $y = 1$ and $(\alpha, \rho) \in [0,1]^2$ solving (15). The competitor challenges with probability x such that

$$(1-x)\Delta - x l_I = \phi_P. \tag{18}$$

Proof of Lemma 1. For $\phi_P \geq \Delta - \phi_A/(1-e)$, type L's payoff from applying is weakly less than his payoff from not applying even if there are no challenges, so $(\alpha = 0, x = 0)$ is an equilibrium, establishing (i). (This equilibrium is unique for $\phi_P > \Delta - \phi_A/(1-e)$.) In what follows, we first show that for $\phi_P < \Delta - \phi_A/(1-e)$, there is no equilibrium in which $\alpha = 0$, $\rho = 0$, or $y = 0$, that there is none in which $\alpha = \rho = y = 1$, and that there is also none in which $x = 0$ or $x = 1$, implying that the equilibrium must be in mixed strategies. We then prove the more specific claims made in (ii.a)-(ii.c).

Suppose first there were an equilibrium with $\alpha = 0$, $\rho = 0$, or $y = 0$ when $\phi_P < \Delta - \phi_A/(1-e)$. Then the competitor's belief when observing $F^H = \Delta_C$ would be $\hat{\lambda} = 1$, and hence she would not challenge. But this means that the type-L inventor could secure a strictly positive expected payoff of $(1-e)(\Delta - \phi_P) - \phi_A > 0$ by applying, activating and offering F^H, contradicting the optimality of $\alpha = 0$ or $\rho = 0$. Moreover, by offering $F = \Delta_C$ type L obtains a higher payoff than by offering $F = l_C$ (because $\Delta_C > l_C$), contradicting the optimality of $y = 0$. Hence, under the assumed condition on ϕ_P, in any equilibrium we must have $\alpha > 0$, $\rho > 0$, and $y > 0$. Next, suppose there were an equilibrium with $\alpha = \rho = y = 1$. Then $\hat{\lambda} = \underline{\lambda}(e)$, and by (5), the challenger would always challenge. But then type L would be better off deviating in some dimension. (If $\phi_P > 0$, type L would be better off not activating or not applying. If $\phi_P = 0 < l_C + m$, type L would be better off offering $0 < F \leq l_C$.)

Now consider the competitor's decision to challenge. If there were an equilibrium in which she never challenges ($x = 0$), then all type-L inventors would apply, activate and offer F^H ($\alpha = \rho = y = 1$). But in that case, (5) implies that the competitor would strictly prefer to challenge. If there were an equilibrium in which she always challenges ($x = 1$), then type L would prefer not to apply, in which case the competitor would be better off not challenging.

Hence, the equilibrium must be in mixed strategies. For the competitor to be indifferent between challenging and not, his beliefs about the type of inventor he faces must be such that the payoff from challenging is the same as the payoff from not challenging, i.e., $(1-\hat{\lambda})\Delta_C = l_C$. Using the definition of $\hat{\lambda}$ yields (15). The conditions for the type-L inventor to be indifferent depend on parameters and are specified below.

Claim (ii.a) [$\phi_P < l_C + m - \phi_A/(1-e)$]: Because the competitor always accepts the offer $F = l_C$, the type-L inventor can guarantee himself a payoff of $l_C + m$ following activation. Hence, if $\phi_P < l_C + m$, type L strictly prefers activating to not activating, implying $\rho = 1$. By the same argument, $\phi_P < l_C + m - \phi_A/(1-e)$ implies that type L strictly prefers applying to not applying, so $\alpha = 1$. The only randomization variable that remains is y. For the type-L inventor to be indifferent between offering $F = l_C$ and $F = \Delta_C$, both must procure him the same payoff. This requires $\pi + l_C + m = \pi + (1-x)\Delta - xl_I$, or (16).

Claim (ii.b) [$\phi_A > 0$ and $l_C + m - \phi_A/(1-e) < \phi_P < \Delta - \phi_A/(1-e)$]: Because $l_C + m - \phi_A/(1-e) < \phi_P$, the type-$L$ inventor cannot break even offering $F = l_C$; hence, $y = 1$. To see that type L will necessarily randomize over the application decision rather than the activation decision, suppose to the contrary $\rho < 1$. This would require $(1-x)\Delta - xl_I = \phi_P$. But in that case, type L's payoff from applying would be zero, and given $\phi_A > 0$ he would prefer not to apply. Hence, $\rho = 1$. The only randomization variable that remains is α. For type L to be indifferent between applying and not, it must be that $\pi = \pi + (1-e)[(1-x)\Delta - xl_I - \phi_P] - \phi_A$, or (17).

Claim (ii.c) [$\phi_A = 0$ and $l_C + m < \phi_P < \Delta$]: Because $\phi_P > l_C + m$, the inventor will never offer $F = l_C$, as he would be sure to make a loss then. Hence, $y = 1$. In any equilibrium in which the type-L inventor is indifferent between activating and not, which requires $\pi = \pi + (1-e)[(1-x)\Delta - xl_I - \phi_P]$, or (18), his payoff from applying for a patent will be zero. Because $\phi_A = 0$, he will also be indifferent between applying and not. Hence, any combination of α and ρ which (given $y = 1$) solves (15) constitutes an equilibrium. $\square$

Remark: For the case $\phi_A = (1-e)(\Delta - \phi_P)$, type L is indifferent between applying and not. Thus in principle there is also an equilibrium in which type L applies with strictly positive probability $\alpha > 0$ satisfying (15) and the competitor never challenges ($x = 0$). We ignore this equilibrium as it seems uninteresting and concerns a knife-edge case.

For the purposes of the next lemma, let

$$\underline{\phi} \equiv \left(\frac{1-e}{e}\right)\left[k - \pi - \frac{(\Delta_C - l_C)\Delta}{\Delta + l_I}\right]$$
$$\overline{\phi} \equiv \left(\frac{1-e}{e}\right)[k - \pi].$$

Lemma 2. *Suppose* (5) *holds. The type H inventor invests, applies, activates, and proposes $F^H = \Delta_C$ if $\phi_P \le h(\phi_A)$, where*

$$h(\phi_A) = \begin{cases} \Delta - (k-\pi) - \dfrac{(\Delta_C - l_C)l_I}{\Delta + l_I} - \phi_A & \text{for } \phi_A \le \underline{\phi} \\ \Delta - \left(1 + \dfrac{l_I}{\Delta}\right)(k-\pi) - \left(1 - \dfrac{e l_I}{(1-e)\Delta}\right)\phi_A & \text{for } \underline{\phi} < \phi_A \le \overline{\phi} \\ \Delta - (k-\pi) - \phi_A & \text{for } \phi_A > \overline{\phi}, \end{cases} \tag{19}$$

and does not invest otherwise.

Proof of Lemma 2. The type-H inventor's payoff from investing, applying, activating and proposing $F^H = \Delta_C$ is

$$\Pi^H = \Delta - (k-\pi) - x l_I - \phi_P - \phi_A.$$

Consider first the case where $\phi_P < l_C + m - \phi_A/(1-e)$, so that $x = \tilde{x}$. Then $\Pi^H \ge 0$ if and only if

$$\phi_P \le \Delta - (k-\pi) - \frac{(\Delta_C - l_C)l_I}{\Delta + l_I} - \phi_A. \tag{20}$$

Next, consider the case where $\phi_P > \Delta - \phi_A/(1-e)$, so that $x = 0$. Then, $\Pi^H \ge 0$ if and only if

$$\phi_P \le \Delta - (k-\pi) - \phi_A. \tag{21}$$

Finally, consider the intermediate case where $l_C + m - \phi_A/(1-e) < \phi_P < \Delta - \phi_A/(1-e)$, so that $x = \hat{x}$. Then $\Pi^H \ge 0$ if and only if

$$\phi_P \le \Delta - (k-\pi) - \frac{l_I(\Delta - \phi_P - \phi_A/(1-e))}{\Delta + l_I} - \phi_A$$

$$\Leftrightarrow \quad \phi_P \leq \Delta - \left(1 + \frac{l_I}{\Delta}\right)(k - \pi) - \left(1 - \frac{e l_I}{(1-e)\Delta}\right)\phi_A. \qquad (22)$$

The threshold $\underline{\phi}$ is obtained by equalizing the right-hand sides of $\phi_P < l_C + m - \phi_A/(1-e)$ and (20):

$$l_C + m - \phi_A/(1-e) = \Delta - (k - \pi) - \frac{(\Delta_C - l_C)l_I}{\Delta + l_I} - \phi_A.$$

Solving for ϕ_A yields $\underline{\phi}$. The threshold $\overline{\phi}$ is obtained by equalizing the right-hand sides of $\phi_P < \Delta - \phi_A/(1-e)$ and (21):

$$\Delta - \phi_A/(1-e) = \Delta - (k - \pi) - \phi_A.$$

Solving for ϕ_A yields $\overline{\phi}$.

What remains to be shown is that type H finds it optimal to propose $F^H = \Delta_C$, knowing that this results in a challenge with probability x. His best deviation is to $F = l_C$, which is the highest license fee that would avoid a challenge; all other license fee offers are either rejected or lead to a challenge with probability one. To see that deviating to $F = l_C$ is unprofitable for type H, note that in any equilibrium with $x > 0$, type L is either indifferent between Δ_C and l_C or strictly prefers Δ_C. Because type H knows that he will win in court while type L knows he will lose, it follows that type H must strictly prefer Δ_C over l_C. □

This completes the proof. □

Proposition 2. *Suppose the type-H inventor invests, i.e., $\Pi^H \geq 0$, and challenges are credible. Then:*

(i) An increase in ϕ_A or ϕ_P weakly decreases applications by type L (α), weakly decreases the rate of challenges (x), and weakly increases the probability that type L charges the high license fee (y).

(ii) An increase in e has ambiguous effects on applications by type L (α), weakly decreases the rate of challenges (x), and weakly increases the probability that type L charges the high license fee (y).

Proof. Claim (i): Within each of the three regions shown in Figure 1, α and y are constant with respect to ϕ_A and ϕ_P ($\tilde{\alpha}$ and $\tilde{y}$ do not depend on the fees). An increase in ϕ_A or ϕ_P can move the equilibrium from region 1 to region 2, in which case α decreases from $\alpha = 1$ to $\alpha = \tilde{\alpha} < 1$ and y increases from $y = \tilde{y}$ to $y = 1$, or from region 2 to region 3, in which case α decreases from $\alpha = \tilde{\alpha}$ to $\alpha = 0$ and y remains constant. The rate of challenges x is constant with respect to ϕ_A and ϕ_P within region 1 ($\tilde{x}$ does not depend on the fees). Within region 2, x is decreasing in ϕ_A and ϕ_P (differentiating $\hat{x}(\phi_A, \phi_P)$ yields $\partial\hat{x}/\partial\phi_A = -1/[(1-e)(\Delta + l_I)] \leq 0$ and $\partial\hat{x}/\partial\phi_P = -1/(\Delta + l_I) < 0$). What remains to be shown is that $\tilde{x} \geq \hat{x}(\phi_A, \phi_P)$ for any combination of fees in region 2, i.e., (ϕ_A, ϕ_P) such that $l_C + m - \phi_A/(1-e) \leq \phi_P \leq \Delta - \phi_A/(1-e)$. Since $\hat{x}$ is decreasing in ϕ_A and ϕ_P, its maximum is attained for some combination of fees such that $l_C + m - \phi_A/(1-e) = \phi_P$. For any (ϕ_A, ϕ_P) satisfying this equality we have

$$\hat{x}(\phi_A, \phi_P)|_{l_C+m-\phi_A/(1-e)=\phi_P} = \frac{\Delta_C - l_C}{\Delta + l_I} = \tilde{x}.$$

Hence, $\hat{x}(\phi_A, \phi_P) \leq \tilde{x}$ for $l_C + m - \phi_A/(1-e) \leq \phi_P$.

Claim (ii): Within region 1, α and x are constant with respect to e ($\alpha = 1$ and $x = \hat{x}$, neither of which depends on e), while y increases with e: differentiating $\tilde{y}$ with respect to e yields

$$\frac{\partial\tilde{y}}{\partial e} = \frac{\lambda l_C}{(1-\lambda)(\Delta_C - l_C)(1-e)^2} > 0.$$

An increase in e can move the equilibrium from region 1 to region 2, in which case α decreases from $\alpha = 1$ to $\alpha = \tilde{\alpha} < 1$, x decreases from $x = \tilde{x}$ to $x = \hat{x} \leq \tilde{x}$ (the inequality having been established in the proof of Claim (i)), and y increases from $y = \tilde{y}$ to $y = 1$. Within region 2, α is increasing and x is decreasing in e: differentiating $\tilde{\alpha}$ and $\hat{x}$ with respect to e yields

$$\frac{\partial\tilde{\alpha}}{\partial e} = \frac{\lambda l_C}{(1-\lambda)(\Delta_C - l_C)(1-e)^2} > 0,$$
$$\frac{\partial\hat{x}}{\partial e} = -\frac{\phi_A}{(1-e)^2(\Delta + l_I)} \leq 0;$$

y is constant in e ($y = 1$). An increase in e can move the equilibrium from region 2 to region 3, in which case α decreases from $\alpha = \tilde{\alpha}$ to $\alpha = 0$ and x from $x = \hat{x}$ to $x = 0$, while y remains constant at $y = 1$. □

Proposition 4. *Suppose $\underline{e} < e < \bar{e}$. Welfare maximization always entails fees such that $\phi_A \geq (1-e)(l_C + m - \phi_P)$ (region 2). If $\Delta_C > (l_C/(l_C + l_I))D$, welfare is maximized by setting $\phi_P = 0$ and ϕ_A such that $\Pi^H = 0$, thus minimizing challenges. If $\Delta_C < (l_C/(l_C + l_I))D$, welfare is maximized by setting $\phi_A = (1-e)(l_C + m - \phi_P)$ and $0 \leq \phi_P \leq \min\{l_C + m, [\Delta - (k - \pi) - \tilde{x}l_I - (1-e)(l_C + m)]/e\}$, thus maximizing challenges.*

Proof. Notice that within region 1, welfare does not depend on either ϕ_A or ϕ_P, as $\alpha = 1$, $x = \tilde{x}$, and $y = \tilde{y}$ are all constant in ϕ_A and ϕ_P. Similarly, within region 2, welfare depends on ϕ_A and ϕ_P only through $x = \hat{x}(\phi_A, \phi_P)$ (which we make explicit by including the fees as arguments) and not through $\alpha = \tilde{\alpha}$ or $y = 1$. Thus the welfare maximization problem is reduced to a choice between $W_1^* \equiv W(1, \tilde{x}, \tilde{y})$ and

$W_2^* \equiv \max_{\phi_A \geq 0, \phi_P \geq 0} W(\tilde{\alpha}, \hat{x}(\phi_A, \phi_P), 1)$ subject to $l_C + m - \phi_A/(1-e) \leq \phi_P \leq h(\phi_A)$, where $h(\phi_A)$ was defined in the proof of Lemma 2 (recall that $\phi_P \leq h(\phi_A) \Leftrightarrow \Pi^H \geq 0$). Notice also that αy is the same in regions 1 and 2, as $\alpha y = \tilde{y}$ in region 1, $\alpha y = \tilde{\alpha}$ in region 2, and $\tilde{\alpha} = \tilde{y}$.

The proof proceeds as follows. We first show that, keeping αy fixed, welfare is decreasing in α. Together with the fact that $\tilde{x} \geq \hat{x}(\phi_A, \phi_P)$ for any (ϕ_A, ϕ_P) in region 2, with equality for $\phi_A = (1-e)(l_C + m - \phi_P)$ (see the proof of Proposition 2), this implies that $W_2^* > W_1^*$, so that the welfare-maximizing combination of fees lies in region 2. Second, we show that fixing αy (and thus $\hat{\lambda}$) at its equilibrium value in regions 1 and 2, welfare is decreasing in x if and only if $\Delta_C > (l_C/(l_C + l_I))D$. Third, we show that $h'(\phi_A) < 0$ in region 2 for $\underline{e} < e < \bar{e}$, which is a useful property for what follows. We then maximize welfare over all combinations of fees lying within region 2.

Claim 1: For any $(\alpha, y) \in (0,1]^2$ and $(\alpha', y') \in (0,1]^2$ such that $\alpha y = \alpha' y'$, $W(\alpha, x, y) \geq W(\alpha', x, y')$ if and only if $\alpha \leq \alpha'$. Rewriting W we have

$$\begin{aligned} W(\alpha, x, y) = 2\pi + S - \lambda(D + x(l_C + l_I) + k + \gamma(e)) \\ + (1-\lambda)\big((1-e)\alpha y x(D - (l_C + l_I)) - \alpha[(1-e)D + \gamma(e)]\big). \end{aligned} \tag{23}$$

For $\alpha y = \alpha' y'$, we have $W(\alpha, x, y) - W(\alpha', x, y') = (1-\lambda)(\alpha' - \alpha)[(1-e)D + \gamma(e)] \geq 0$ if and only if $\alpha \leq \alpha'$.

Claim 2: $(\partial/\partial x)W(\alpha, x, y)|_{\alpha y = \tilde{\alpha}} < 0$ if and only if $\Delta_C > (l_C/(l_C + l_I))D$. We have

$$\begin{aligned} \frac{\partial W}{\partial x} &= -\lambda(l_C + l_I) + (1-\lambda)(1-e)\alpha y[D - (l_C + l_I)] \\ &= (1-\lambda)(1-e)\alpha y D - \big[\lambda + (1-\lambda)(1-e)\alpha y\big](l_C + l_I). \end{aligned}$$

Using the definition of $\hat{\lambda}$ in (4), we can rewrite this as

$$\frac{\partial W}{\partial x} = \big[\lambda + (1-\lambda)(1-e)\alpha y\big]\Big((1-\hat{\lambda})D - (l_C + l_I)\Big).$$

By Proposition 1, in the equilibrium in regions 1 and 2 we have $1 - \hat{\lambda} = l_C/\Delta_C$. Hence,

$$\frac{\partial W}{\partial x} < 0 \quad \Leftrightarrow \quad \frac{l_C}{\Delta_C}D - (l_C + l_I) < 0 \quad \Leftrightarrow \quad \Delta_C > \left(\frac{l_C}{l_C + l_I}\right)D.$$

Claim 3: $h'(\phi_A) < 0$ for $l_C + m - \phi_A/(1-e) \leq \phi_P \leq \Delta - \phi_A/(1-e)$ and $\underline{e} < e < \bar{e}$. From (19) we have that, in region 2,

$$h'(\phi_A) = -\frac{(1-e)\Delta - e l_I}{\Delta}, \tag{24}$$

which is negative if and only if $(1-e)\Delta > e l_I$. To establish that this condition holds, we first show that $e > \underline{e}$ implies that $h(\underline{\phi}) > 0$. By definition of $\underline{\phi}$,

$$h(\underline{\phi}) = \Delta - \tilde{x} l_I - (k - \pi) - \underline{\phi} \tag{25}$$

$$= l_C + m - \frac{\underline{\phi}}{1-e}. \tag{26}$$

Replacing $\underline{\phi}$ in (26) we have that

$$h(\underline{\phi}) = l_C + m - \frac{1}{e}\left[l_C + m - \Delta + \tilde{x}l_I + k - \pi\right] > 0$$
$$\Leftrightarrow \quad (1-e)(l_C + m) < \Delta - \tilde{x}l_I - (k - \pi)$$
$$\Leftrightarrow \quad e > \underline{e}.$$

Next, replacing $\underline{\phi}$ in (25) yields

$$\begin{aligned} h(\underline{\phi}) &= \Delta - (k-\pi) - \tilde{x}l_I - \frac{1-e}{e}[k - \pi - \tilde{x}\Delta] \\ &= \Delta - \frac{k-\pi}{e} + \frac{\tilde{x}}{e}[(1-e)\Delta - el_I]. \end{aligned}$$

Since $\Delta - (k-\pi)/e < 0$ by the assumption that $e < \bar{e}$ and we have just established that $h(\underline{\phi}) > 0$ when $e > \underline{e}$, it must be that $(1-e)\Delta > el_I$, and hence that $h'(\phi_A) < 0$ when $\underline{e} < e < \bar{e}$.

Having established these claims, we now solve for the welfare-maximizing fees. The planner's problem can be written as

$$\max_{\phi_A \geq 0, \phi_P \geq 0} W(\tilde{\alpha}, \hat{x}(\phi_A, \phi_P), 1)$$
$$\text{subject to } \begin{cases} \phi_A \geq (1-e)(l_C + m - \phi_P) & [\mu_1] \\ \phi_P \leq h(\phi_A) & [\mu_2] \end{cases}$$

where μ_1 and μ_2 are the multipliers associated with the constraints. The first-order conditions are

$$\frac{\partial W}{\partial x}\frac{\partial \hat{x}}{\partial \phi_A} + \mu_1 + \mu_2 h'(\phi_A) = 0 \tag{27}$$
$$\frac{\partial W}{\partial x}\frac{\partial \hat{x}}{\partial \phi_P} + \mu_1(1-e) - \mu_2 = 0. \tag{28}$$

We have to distinguish two cases: $\partial W/\partial x < 0$ and $\partial W/\partial x > 0$.

Case 1: $\partial W/\partial x < 0$. Combining $h'(\phi_A) < 0$ with the fact that $\partial W/\partial x < 0$ and $\partial \hat{x}/\partial \phi_A < 0$, (27) implies $\mu_2 > 0$, so the constraint $\phi_P \leq h(\phi_A)$ must be binding at the optimum. Using (27) to replace μ_1, (28) becomes

$$\begin{aligned} \frac{\partial W}{\partial x}\frac{\partial \hat{x}}{\partial \phi_P} - (1-e)\left(\frac{\partial W}{\partial x}\frac{\partial \hat{x}}{\partial \phi_A} + \mu_2 h'(\phi_A)\right) - \mu_2 = \frac{\partial W}{\partial x}&\left(\frac{\partial \hat{x}}{\partial \phi_P} - (1-e)\frac{\partial \hat{x}}{\partial \phi_A}\right) \\ &- \mu_2\left((1-e)h'(\phi_A) + 1\right). \end{aligned}$$

Noting that $\partial \hat{x}/\partial \phi_P = (1-e)\partial \hat{x}/\partial \phi_A$ and substituting for $h'(\phi_A)$ from (24), this expression simplifies to

$$-\mu_2\left[1 - \frac{(1-e)\Delta - el_I}{\Delta}\right] = -\mu_2\frac{e(\Delta + l_I)}{\Delta} < 0.$$

Hence, when evaluated at the binding constraint $\phi_P \leq h(\phi_A)$, the objective is decreasing in ϕ_P. It follows that the optimal solution is such that $\phi_P = 0$ and $h(\phi_A) = \phi_P = 0$ (which is equivalent to $\Pi^H = 0$) as claimed.

Case 2: $\partial W/\partial x > 0$. Combining $h'(\phi_A) < 0$ with the fact that $\partial W/\partial x > 0$ and $\partial \hat{x}/\partial \phi_A < 0$, (27) implies $\mu_1 > 0$, so the constraint $\phi_A \geq (1-e)(l_C + m - \phi_P)$ must be binding at the optimum. We have

$$W(\tilde{\alpha}, \hat{x}(\phi_A, \phi_P), 1)|_{\phi_A=(1-e)(l_C+m-\phi_P)} = W(\tilde{\alpha}, \tilde{x}, 1),$$

which is constant in (ϕ_A, ϕ_P). Hence, any combination of fees such that $\phi_A = (1-e)(l_C + m - \phi_P)$, $\phi_A \geq 0$, $\phi_P \geq 0$, and $\phi_P \leq h(\phi_A)$ is a solution. Combining the three inequalities and noting that $l_C + m - \phi_A/(1-e)$ and $h(\phi_A)$ intersect at $\underline{\phi}$ yields $0 \leq \phi_P \leq \min\{l_C + m, h(\underline{\phi})\} = \min\{l_C + m, [\Delta - (k-\pi) - \tilde{x}l_I - (1-e)(l_C+m)]/e\}$. □

Proposition 7. *Fix v^* and e. Then applications from low types are minimized by frontloading fees, i.e., setting $\phi_P = 0$ and $\phi_A = \Delta(v^*)$.*

Proof. Fixing v^*, we have $\phi_A + \phi_P = \Delta(v^*)$. Then $\hat{v}$ and $\tilde{v}$ solve

$$\Delta(\hat{v}) = \Delta(v^*) + \left(\frac{e}{1-e}\right)\phi_A$$
$$l(\tilde{v}) + m(\tilde{v}) = \Delta(v^*) + \left(\frac{e}{1-e}\right)\phi_A.$$

Because by assumption Δ, l, and m are increasing functions, $\hat{v}$ and $\tilde{v}$ increase with ϕ_A. Given the nonnegativity constraint on ϕ_P, $\hat{v}$ and $\tilde{v}$ are thus maximized for $\phi_A = \Delta(v^*)$ and $\phi_P = 0$.

To complete the proof, we show that $\tilde{\alpha}(v)$ depends only on the sum of pre- and post-grant fees and not on how this sum is apportioned. For the competitor to be indifferent, it must be that, for all v, $(1 - \hat{\lambda}(v))\Delta_C(v) = l(v)$. Using the expression for $\hat{\lambda}(v)$ in footnote 34, $\tilde{\alpha}(v)$ is given by the α that solves

$$\left[1 - \frac{G_v(\pi(v) + \Delta(v) - \phi_A - \phi_P) - G_v(\pi(v))}{G_v(\pi(v) + \Delta(v) - \phi_A - \phi_P) - G_v(\pi(v)) + (1-e)\alpha y G_v(\pi(v))}\right]\Delta_C(v) = l(v).$$

Thus, for $\phi_A + \phi_P = \Delta(v^*)$,

$$\tilde{\alpha}(v) = \left(\frac{G_v(\pi(v) + \Delta(v) - \Delta(v^*)) - G_v(\pi(v))}{(1-e)G_v(\pi(v))}\right)\left(\frac{l(v)}{\Delta_C(v) + l(v)}\right)$$

is invariant with respect to ϕ_A and ϕ_P. □

Appendix B A more general screening technology

In this appendix, we introduce a more general screening technology that allows for both type-I and type-II errors at both the level of the patent office and the courts, and we examine how our results extend to that case. Suppose the patent office and the courts review applications with intensity e_1 and e_2, respectively. The intensity of review equals the probability that they find out the inventor's true type. With probability $1 - e_i$, $i = 1, 2$, they find no strong evidence either way, in which case the patent office allows the application with probability q_1 while the courts uphold the patent with probability q_2. This leads to stage-i probabilities of acceptance and rejection for each type of inventor given in Table B.1, with $i = 1, 2$, stage 1 corresponding to patent office review and stage 2 corresponding to court review. Our basic model is a special case of this setup with $e_1 = e$, $q_1 = 1$, and $e_2 = 1$ (q_2 is indeterminate). Assume $(1 - e_2)(1 - q_2)\Delta_C < l_C < (e_2 + (1 - e_2)(1 - q_2))\Delta_C$; otherwise the competitor either would not want to challenge even when certain of facing a type-L inventor or would want to challenge even when certain of facing a type-H inventor.

Table B.1: Probability of acceptance and rejection at review stage i by type

	Acceptance	Rejection
Type H	$e_i + (1 - e_i)q_i$	$(1 - e_i)(1 - q_i)$
Type L	$(1 - e_i)q_i$	$e_i + (1 - e_i)(1 - q_i)$

B.1 No challenges

Consider first the case where challenges are not possible. Then the two types of inventors have the following expected payoff from investing and applying:

$$\Pi^H = (e_1 + (1 - e_1)q_1)(\pi + \Delta - \phi_P) + (1 - e_1)(1 - q_1)\pi - k - \phi_A$$
$$\Pi^L = (1 - e_1)q_1(\pi + \Delta - \phi_P) + (e_1 + (1 - e_1)(1 - q_1))\pi - \phi_A.$$

Fixing the sum of fees, $\phi_A + \phi_P$, both types now prefer post-grant to pre-grant fees. However, type L has a stronger preference for post-grant fees than type H. To see this, we can compute the slope of the iso-profit lines:

$$\frac{\partial \Pi^H / \partial \phi_A}{\partial \Pi^H / \partial \phi_P} = \frac{1}{e_1 + (1 - e_1)q_1}$$
$$\frac{\partial \Pi^L / \partial \phi_A}{\partial \Pi^L / \partial \phi_P} = \frac{1}{(1 - e_1)q_1}.$$

For any $e_1 > 0$, type L's iso-profit lines are steeper than type H's. To obtain a one dollar reduction in ϕ_A, the type-L inventor is willing to accept a $1/[(1 - e_1)q_1]$ dollar increase in ϕ_P, while the type-L inventor is willing to accept only a smaller increase of $1/[e_1 + (1 - e_1)q_1]$ dollars.

The type-L inventor invests if and only $\Pi^H \geq 0$. The type-L inventor always invests; he applies for a patent if $\Pi^L \geq \pi$. Thus the condition for full screening is

$\Pi^H \geq 0 \geq \Pi^L - \pi$, or

$$(e_1 + (1 - e_1)q_1)(\Delta - \phi_P) - (k - \pi) \geq \phi_A \geq (1 - e_1)q_1(\Delta - \phi_P).$$

A necessary condition is

$$e_1 \geq \frac{k - \pi}{\Delta}.$$

That is, the minimum examination intensity required to achieve full screening in the absence of court challenges is the same as in the basic model.

B.2 Challenges

Now consider the possibility of court challenges. Suppose type H invests and applies; moreover, suppose that if he is successful in obtaining a patent, he activates it and charges a license fee F^H. Type H's expected payoff then is

$$\Pi^H = (e_1 + (1 - e_1)q_1)\Big[x\big[(e_2 + (1 - e_2)q_2)(\pi + \Delta) + (1 - e_2)(1 - q_2)\pi - l_I\big] + (1 - x)(\pi + m + F^H) - \phi_P\Big] + (1 - e_1)(1 - q_1)\pi - k - \phi_A, \quad (29)$$

where x is the rate of challenges given F^H. The competitor's belief that an activated patent is valid when a license is offered at fee F^H is

$$\hat{\lambda} = \frac{\lambda(e_1 + (1 - e_1)q_1)}{\lambda(e_1 + (1 - e_1)q_1) + (1 - \lambda)(1 - e_1)q_1\alpha\rho y}. \quad (30)$$

The lower bound of $\hat{\lambda}$ is attained at $\alpha\rho y = 1$ and given by

$$\underline{\lambda} = \frac{\lambda(e_1 + (1 - e_1)q_1)}{\lambda(e_1 + (1 - e_1)q_1) + (1 - \lambda)(1 - e_1)q_1}.$$

Challenges are credible if and only if

$$\pi - \Delta_C \leq \underline{\lambda}\left[(e_2 + (1 - e_2)q_2)(\pi - \Delta_C) + (1 - e_2)(1 - q_2)\pi\right] + (1 - \underline{\lambda})\left[(1 - e_2)q_2(\pi - \Delta_C) + (e_2 + (1 - e_2)(1 - q_2))\pi\right] - l_C \quad (31)$$

or

$$[1 - (1 - e_2)q_2 - e_2\underline{\lambda}]\Delta_C \geq l_C. \quad (32)$$

When challenges are not credible, the outcome is the same as in the absence of challenges (see the previous subsection). The equilibrium when challenges are credible is characterized in the following proposition.

Proposition 8. *Consider the generalized screening technology specified in Table B.1. If challenges are credible and $\phi_P \leq \Delta - \phi_A/((1 - e_1)q_1)$, there is a semi-separating equilibrium in which:*

(i) the type-H inventor invests, applies, activates, and proposes $F^H = \Delta_C$*;*

(ii) the type-L inventor always randomizes over either

(a) the decision whether to apply or not (probabilities α *and* $1-\alpha$*), or*

(b) the license fee to propose, $F^L \in \{\Delta_C, l_C + (1-e_2)q_2\Delta_C\}$ *(probabilities* y *and* $1-y$*),*

such that $[\hat{\lambda}(1-e_2)(1-q_2) + (1-\hat{\lambda})(e_2 + (1-e_2)(1-q_2))]\Delta_C = l_C$*; he always activates (*$\rho = 1$*);*

(iii) the competitor randomizes over the decision whether to challenge or not (probabilities x *and* $1-x$*) if offered* $F = \Delta_C$ *and never challenges if offered* $F = l_C + (1-e_2)q_2\Delta_C$*, with*

$$x = \begin{cases} \tilde{x} & \text{for } \phi_P \leq l_C + m + (1-e_2)q_2\Delta_C - \dfrac{\phi_A}{(1-e_1)q_1} \\ \\ \hat{x} & \text{for } l_C + m + (1-e_2)q_2\Delta_C - \dfrac{\phi_A}{(1-e_1)q_1} < \phi_P \leq \Delta - \dfrac{\phi_A}{(1-e_1)q_1}; \end{cases}$$

where

$$\tilde{x} = \frac{(e_2 + (1-e_2)(1-q_2))\Delta_C - l_C}{(e_2 + (1-e_2)(1-q_2))\Delta + l_I}$$
$$\hat{x} = \frac{\Delta - \phi_P - \phi_A/[(1-e_1)q_1]}{(e_2 + (1-e_2)(1-q_2))\Delta + l_I},$$

provided $(e_1 + (1-e_1)q_1)[(1 - x(1-e_2)(1-q_2))\Delta - xl_I - \phi_P] - (k-\pi) - \phi_A \geq 0$*. If* $\phi_P > \Delta - \phi_A/((1-e_1)q_1)$*, the type-L inventor does not apply and the competitor does not challenge.*

Proof. The proof is similar to that of Proposition 1; we therefore focus on the key points here. Suppose type H invests, applies, activates and charges $F^H = \Delta_C$. Consider the behavior of type L in the case where challenges are credible, i.e., (32) holds, and fees are sufficiently low as to make applying and activating profitable (i.e., $\alpha = \rho = 1$). Let us look for an equilibrium in which type L randomizes over the license fee F_L as follows:

$$F_L = \begin{cases} \Delta_C & \text{with probability } y \\ \tilde{F} & \text{with probability } 1-y, \end{cases}$$

where $\tilde{F}$ is chosen such that the competitor does not find it worthwhile to challenge.

The competitor's beliefs on the equilibrium path are $\tilde{\lambda}(F^H) = \hat{\lambda}$ and $\tilde{\lambda}(\tilde{F}) = 0$. The out-of-equilibrium belief most likely to support the equilibrium is $\tilde{\lambda}(F) = 0$ for

$F \neq F^H, \tilde{F}$. For the competitor to refrain from challenging when observing $F \neq F^H$ despite assigning probability 1 to the patent being invalid, it must be that

$$(1-e_2)q_2(\pi - \Delta_C) + (e_2 + (1-e_2)(1-q_2))\pi - l_C \leq \pi - F.$$

The highest fee that satisfies this inequality is

$$\overline{F} = l_C + (1-e_2)q_2\Delta_C.$$

Let $s(F) \in [0,1]$ denote the competitor's probability of challenging the patent when observing a license fee offer F. Sequential rationality requires $s(F) = 0$ for $F \leq \overline{F}$ and $s(F) = 1$ for $F > \overline{F}$, $F \neq F^H$. Thus, $\tilde{F} = \overline{F}$. We also have $s(F^H) = x$, which depends on $\hat{\lambda}$. For the type-L inventor to be indifferent between offering F^H and $\overline{F}$, it must be that

$$\pi + m + \overline{F} = x[(1-e_2)q_2(\pi + \Delta) + (e_2 + (1-e_2)(1-q_2))\pi - l_I] \\ + (1-x)(\pi + m + F^H).$$

Replacing $\overline{F}$ and F^H yields

$$x = \frac{(e_2 + (1-e_2)(1-q_2))\Delta_C - l_C}{(e_2 + (1-e_2)(1-q_2))\Delta + l_I} = \tilde{x}.$$

For the competitor to be willing to randomize between challenging and not, it must be that (31) holds with equality when replacing $\underline{\lambda}$ by $\hat{\lambda}$, i.e.,

$$(1 - (1-e_2)q_2 - e_2\hat{\lambda})\Delta_C = l_C. \tag{33}$$

Finally, we need to check that the type-H inventor has no incentive to deviate. The best deviation would be to $F = \overline{F}$. But since x is chosen so as to make the type-L inventor indifferent between F^H and $\overline{F}$, and type H has a higher probability of winning in court than type L, type H must strictly prefer F^H to $\overline{F}$.

The above equilibrium was derived under the assumption that fees are sufficiently low for the type-L inventor to find it profitable to apply. Since the type-L inventor's payoff in this equilibrium is equal to his payoff when offering $\overline{F}$, this requires

$$\phi_P \leq l_C + m + (1-e_2)q_2\Delta_C - \phi_A/[(1-e_1)q_1]. \tag{34}$$

Suppose instead (34) does not hold. Then, the above strategy profile cannot be part of an equilibrium. However, if the type-L inventor's payoff from applying is positive when there are no challenges, i.e.,

$$\phi_P \leq \Delta - \phi_A/[(1-e_1)q_1], \tag{35}$$

then type L will randomize over the decision to apply, and do so such that (33) holds.

The competitor chooses x to make the type-L inventor indifferent between applying and not:

$$\pi = (1-e_1)q_1\big[x[(1-e_2)q_2(\pi+\Delta) + (e_2 + (1-e_2)(1-q_2))\pi - l_I] \\ + (1-x)(\pi + m + \Delta_C)\big] + (e_1 + (1-e_1)(1-q_1))\pi - \phi_A.$$

Solving for x yields

$$x = \frac{\Delta - \phi_P - \phi_A/[(1-e_1)q_1]}{(e_2 + (1-e_2)(1-q_2))\Delta + l_I} = \hat{x}.$$

If (35) is not satisfied, type L does not apply.

To conclude the proof, let us check whether type H finds it profitable to invest in R&D. Using $F^H = \Delta_C$, the expression for Π_H in (29) can be rewritten as

$$\Pi^H = (e_1 + (1-e_1)q_1)\big[(1 - x(1-e_2)(1-q_2))\Delta - xl_I - \phi_P\big] - (k-\pi) - \phi_A.$$

Letting $h(\phi_A) \equiv \Pi^H/(e_1 + (1-e_1)q_1) + \phi_P$, we have that $\Pi^H \geq 0$ if and only if $h(\phi_A) \geq \phi_P$. Substituting the equilibrium values of x in the different regions, we obtain

$$h(\phi_A) = \begin{cases} h_1(\phi_A) & \text{for } \phi_A \leq \underline{\phi} \\ h_2(\phi_A) & \text{for } \underline{\phi} < \phi_A \leq \overline{\phi} \\ h_3(\phi_A) & \text{for } \phi_A > \overline{\phi}, \end{cases}$$

where

$$h_1(\phi_A) = \Delta - \frac{k - \pi + \phi_A}{e_1 + (1-e_1)q_1} \\ - \frac{(e_2 + (1-e_2)(1-q_2))\Delta_C - l_C}{(e_2 + (1-e_2)(1-q_2))\Delta + l_I}[(1-e_2)(1-q_2)\Delta + l_I]$$

$$h_2(\phi_A) = \Delta - \frac{(e_2 + (1-e_2)(1-q_2))\Delta + l_I}{(e_1 + (1-e_1)q_1)e_2\Delta}(k-\pi) \\ - \frac{[e_2(1-e_1)q_1 - e_1(1-e_2)(1-q_2)]\Delta - e_1 l_I}{(e_1 + (1-e_1)q_1)e_2\Delta}\phi_A$$

$$h_3(\phi_A) = \Delta - \frac{k - \pi + \phi_A}{e_1 + (1-e_1)q_1},$$

and

$$\underline{\phi} = \frac{(1-e_1)q_1}{e_1}\Big[k - \pi + (e_1 + (1-e_1)q_1)\big[l_C + m + (1-e_2)q_2\Delta_C \\ - (1 - \tilde{x}(1-e_2)(1-q_2))\Delta + \tilde{x}l_I\big]\Big]$$

$$= \frac{(1-e_1)q_1}{e_1}[k - \pi + (e_1 + (1-e_1)q_1)\tilde{x}e_2\Delta_C]$$

$$\overline{\phi} = \frac{(1-e_1)q_1}{e_1}(k-\pi).$$

□

We again have three regions in (ϕ_A, ϕ_P) space. For low fees (region 1), type L always applies while randomizing over the license fee to charge; for intermediate fees (region 2), type L always charges the high fee while randomizing over the application decision. In both cases, type L's randomization is done such that the competitor is indifferent between challenging and not. The competitor herself randomizes over the challenge decision so as to make type L indifferent. For large fees (region 3), type L does not apply and the competitor does not challenge. The main difference with the baseline model is that the license fee charged by the low type to avoid a challenge is now higher, namely, $F = l_C + m + (1 - e_2)q_2\Delta_C$, reflecting the fact that a challenge is not guaranteed to be successful even when the competitor is sure of facing a low type. Similarly, the high type's profit is now lower than in the baseline because his patent application can be rejected by the patent office and, if granted, can be invalidated by the court.

The comparative statics results from the basic model with respect to fees extend in a straightforward way: an increase in fees leads to (weakly) fewer bad applications (α), (weakly) higher licence fees (y), and (weakly) fewer challenges (x). With respect to examination intensity (e_1), we continue to have an ambiguous result on bad applications: as e_1 increases, we first move from region 1, where $\alpha = 1$, to region 2, where

$$\alpha = \tilde{\alpha} = \left(\frac{\lambda}{1-\lambda}\right)\left(\frac{e_1 + (1-e_1)q_1}{(1-e_1)q_1}\right)\left(\frac{l_C - (1-e_2)(1-q_2)\Delta_C}{(e_2 + (1-e_2)(1-q_2))\Delta_C - l_C}\right) < 1.$$

Within region 2, however, an increase in e_1 leads to more bad applications, $\partial\tilde{\alpha}/\partial e_1 > 0$. Finally, as e_1 continues to increase, α drops to zero in region 3. A similar ambiguous result holds for y. As for the rate of challenges, an increase in e_1 (weakly) reduces x.

With the generalized technology, we can also consider the effects of a change in the quality of the courts, as captured by e_2. As e_2 increases, the equilibrium moves from region 1 to region 2 but no further: changes in e_2 can never move the equilibrium to region 3 (this is a manifestation of our results on full screening; see Proposition 9 below). Within region 1, e_2 has no effect on α while it decreases $x = \tilde{x}$; the effect on y is ambiguous. Within region 2, e_2 has no effect on y while it continuous to reduce $x = \hat{x}$; the effect on α is ambiguous. The ambiguous effect on $\alpha = \tilde{\alpha}$ in region 2 (and similarly $y = \tilde{y} = \tilde{\alpha}$ in region 1) can be further analyzed as follows. We have

$$\frac{\partial\tilde{\alpha}}{\partial e_2} = \left(\frac{\lambda}{1-\lambda}\right)\left(\frac{e_1 + (1-e_1)q_1}{(1-e_1)q_1}\right)\left(\frac{\Delta_C((1-q_2)\Delta_C - l_C)}{((e_2 + (1-e_2)(1-q_2))\Delta_C - l_C)^2}\right) \leq 0$$

$$\Leftrightarrow \quad q_2 \geq 1 - \frac{l_C}{\Delta_C}.$$

Thus, an increase in the quality of the courts leads to fewer bad applications if and only if q_2 is sufficiently large. This result is interesting, as q_2 could be interpreted as a measure of the presumption of validity that the courts apply.

We now extend our result on the feasibility of full screening. It turns out that whether full screening can be achieved does not depend on the existence (or quality) of the courts; it only depends on the patent office's examination intensity.

Proposition 9. *There exists a combination of fees* (ϕ_A, ϕ_P) *inducing full screening if and only if* $e_1 \geq \bar{e}$. *For a given* $e_1 \geq \bar{e}$, *any combination of fees satisfying*

$$(e_1 + (1 - e_1)q_1)(\Delta - \phi_P) - (k - \pi) \geq \phi_A \geq (1 - e_1)q_1(\Delta - \phi_P). \tag{36}$$

achieves full screening.

Proof. By Proposition 8, deterrence of type L ($\alpha = 0$) requires $\phi_A \geq ((1 - e_1)q_1)(\Delta - \phi_P)$. Investment by type H when type L is deterred (and thus, in equilibrium, $x = 0$) requires $(e_1 + (1 - e_1)q_1)(\Delta - \phi_P) - (k - \pi) \geq \phi_A$. A pre-grant fee ϕ_A satisfying both inequalities exists if and only if $(e_1+(1-e_1)q_1)(\Delta-\phi_P)-(k-\pi) \geq ((1-e_1)q_1)(\Delta-\phi_P)$, or

$$e_1 \geq \frac{k - \pi}{\Delta - \phi_P}. \tag{37}$$

Since the right-hand side increases with ϕ_P, the minimum level of e_1 required to achieve full screening is obtained by evaluating (37) at $\phi_P = 0$, yielding $e_1 \geq (k-\pi)/\Delta = \bar{e}$. □

Next, we revisit the result on the welfare-maximizing fee structure when full screening is not possible ($e_1 < \bar{e}$) from Proposition 4. With the generalized screening technology, the welfare function is

$$\begin{aligned} W(\alpha, x, y) = 2\pi + S - \lambda\big[(e_1 + (1-e_1)q_1)[(1-x(1-e_2)(1-q_2))D + x(l_C + l_I)] + k + \gamma(e_1)\big] \\ - (1 - \lambda)\alpha\Big[(1 - e_1)q_1\big([y((1 - x) + x(1 - e_2)q_2) + 1 - y]D + xy(l_C + l_I)\big) + \gamma(e_1)\Big] \end{aligned} \tag{38}$$

The threshold $\underline{e}$, which is such that region 2 is attainable for $e_1 > \underline{e}$, is now defined by

$$\begin{aligned} (1 - \underline{e})q_1[l_C + m + (1 - e_2)q_2\Delta_C] \\ = (\underline{e} + (1 - \underline{e})q_1)[(1 - \tilde{x}(1 - e_2)(1 - q_2))\Delta - \tilde{x}l_I] - (k - \pi). \end{aligned} \tag{39}$$

Proposition 10. *If* $\underline{e} < e_1 < \bar{e}$ *and* $\Delta_C > (l_C/(l_C + l_I))D$, *welfare is maximized by setting* $\phi_P = 0$ *and* ϕ_A *such that* $\Pi^H = 0$.

Proof. The proof proceeds through a series of claims.

Claim 1: $W(\alpha, y, x) \geq W(\alpha', y', x)$ *for* $\alpha y = \alpha' y'$ *if and only if* $\alpha \leq \alpha'$. Rewriting (38) as

$$\begin{aligned} W(\alpha, x, y) = 2\pi + S - \lambda\big[(e_1 + (1-e_1)q_1)[(1-x(1-e_2)(1-q_2))D + x(l_C + l_I)] + k + \gamma(e_1)\big] \\ - (1 - \lambda)\Big[\alpha xy(1 - e_1)q_1(l_C + l_I - D) + \alpha\big((1 - e_1)q_1(1 + x(1 - e_2)q_2)D + \gamma(e_1)\big)\Big], \end{aligned}$$

we have, for $\alpha y = \alpha' y'$,

$$W(\alpha, x, y) - W(\alpha', x, y') = (1-\lambda)(\alpha' - \alpha)[(1-e_1)q_1(1 + x(1-e_2)q_2)D + \gamma(e_1)] \geq 0$$

if and only if $\alpha \leq \alpha'$.

Claim 2: $(\partial/\partial x)W(\alpha, x, y)|_{\alpha y = \tilde{\alpha}} < 0$ if and only if $\Delta_C > (l_C/(l_C + l_I))D$. We have

$$\begin{aligned}\frac{\partial W}{\partial x} &= \lambda(e_1 + (1-e_1)q_1)[(1-e_2)(1-q_2)D - (l_C + l_I)] \\ &\quad + (1-\lambda)(1-e_1)q_1\alpha y[(1-(1-e_2)q_2)D - (l_C + l_I)] \\ &= \big[\lambda(e_1 + (1-e_1)q_1)(1-e_2)(1-q_2) + (1-\lambda)(1-e_1)q_1\alpha y(1-(1-e_2)q_2)\big]D \\ &\quad - \big[\lambda(e_1 + (1-e_1)q_1) + (1-\lambda)(1-e_1)q_1\alpha y\big](l_C + l_I).\end{aligned}$$

Using the definition of $\hat{\lambda}$ in (30), we can rewrite this as

$$\begin{aligned}\frac{\partial W}{\partial x} = \big[\lambda(e_1 + (1-e_1)q_1) + (1-\lambda)(1-e_1)q_1\alpha y\big]\Big(\big[\hat{\lambda}(1-e_2)(1-q_2) \\ + (1-\hat{\lambda})(e_2 - (1-e_2)(1-q_2))\big]D - (l_C + l_I)\Big).\end{aligned}$$

By Proposition 8, in the equilibrium in regions 1 and 2 (i.e., for $\alpha y = \tilde{\alpha}$) we have $\hat{\lambda}(1-e_2)(1-q_2)$
$+ (1-\hat{\lambda})(e_2 - (1-e_2)(1-q_2)) = l_C/\Delta_C$. Hence,

$$\frac{\partial W}{\partial x} < 0 \quad \Leftrightarrow \quad \frac{l_C}{\Delta_C}D - (l_C + l_I) < 0 \quad \Leftrightarrow \quad \Delta_C > \left(\frac{l_C}{l_C + l_I}\right)D.$$

Claim 3: $\tilde{x} \geq \hat{x}(\phi_A, \phi_P)$ for any (ϕ_A, ϕ_P) such that $l_C + m + (1-e_2)q_2\Delta_C - \phi_A/((1-e_1)q_1) \leq \phi_P$. Since $\hat{x}$ is decreasing in ϕ_A and ϕ_P, its maximum $\hat{x}^{\max}$ is attained for $l_C + m + (1-e_2)q_2\Delta_C - \phi_A/((1-e_1)q_1) = \phi_P$. We have

$$\hat{x}^{\max} = \frac{\Delta_C + m - (l_C + m + (1-e_2)q_2\Delta_C)}{(e_2 + (1-e_2)q_2)\Delta + l_I} = \frac{(e_2 + (1-e_2)(1-q_2))\Delta_C - l_C}{(e_2 + (1-e_2)(1-q_2))\Delta + l_I} = \tilde{x}.$$

Hence, $\hat{x}(\phi_A, \phi_P) \leq \tilde{x}$ for $l_C + m + (1-e_2)q_2\Delta_C - \phi_A/((1-e_1)q_1) \leq \phi_P$.

Combining these three claims with the assumption that $\Delta_C > (l_C/(l_C + l_I))D$, it follows that the welfare-maximizing combination of fees solves the following optimization problem:

$$\min_{\phi_A, \phi_P} \; \hat{x}(\phi_A, \phi_P)$$

$$\text{subject to} \begin{cases} \phi_P \geq 0 & [\mu_1] \\ \phi_A \geq ((1-e_1)q_1)[l_C + m + (1-e_2)q_2\Delta_C - \phi_P] & [\mu_2] \\ \phi_P \leq h(\phi_A) & [\mu_3] \end{cases}$$

where μ_1, μ_2, and μ_3 are the multipliers associated with the constraints and $h(\phi_A)$ was defined in the proof of Proposition 8 (recall that $\phi_P \leq h(\phi_A) \Leftrightarrow \Pi^H \geq 0$). The first-order conditions with respect to ϕ_A and ϕ_P are

$$\frac{\partial \hat{x}}{\partial \phi_A} = \mu_2 + \mu_3 h'(\phi_A) \tag{40}$$

$$\frac{\partial \hat{x}}{\partial \phi_P} = \mu_1 + \mu_2(1-e_1)q_1 - \mu_3. \tag{41}$$

We now show that at the optimum $\mu_1 > 0$ and $\mu_3 > 0$ while $\mu_2 = 0$, implying $\phi_P = 0$ and $h(\phi_A) = 0$ (which is equivalent to $\Pi^H = 0$) as claimed.

We start by establishing that $h'(\phi_A) < 0$ in region 2 when $\underline{e} < e_1 < \bar{e}$ (which holds by assumption). Note first that $h(\phi_A) = h_2(\phi_A)$ in region 2, and hence that

$$h'(\phi_A) = h_2'(\phi_A) = -\frac{[e_2(1-e_1)q_1 - e_1(1-e_2)(1-q_2)]\Delta - e_1 l_I}{(e_1 + (1-e_1)q_1)e_2\Delta}, \tag{42}$$

which is negative if and only if $[e_2(1-e_1)q_1 - e_1(1-e_2)(1-q_2)]\Delta > e_1 l_I$. Next, note that $e_1 > \underline{e}$ implies that $h(\underline{\phi}) > 0$: by definition of $\underline{\phi}$,

$$\begin{aligned} h_1(\underline{\phi}) = (1-\tilde{x}(1-e_2)(1-q_2))\Delta - \tilde{x}l_I - \frac{k-\pi+\underline{\phi}}{e_1+(1-e_1)q_1} \\ = l_C + m + (1-e_2)q_2\Delta_C - \frac{\underline{\phi}}{(1-e_1)q_1}. \end{aligned}$$

Replacing $\underline{\phi}$ on the right-hand side we have that

$$\begin{aligned} h(\underline{\phi}) = h_1(\underline{\phi}) = l_C + m + (1-e_2)q_2\Delta_C - \frac{e_1+(1-e_1)q_1}{e_1}\Bigg[l_C + m + (1-e_2)q_2\Delta_C \\ - (1-\tilde{x}(1-e_2)(1-q_2))\Delta + \tilde{x}l_I + \frac{k-\pi}{e_1+(1-e_1)q_1}\Bigg] > 0, \end{aligned}$$

or

$$(1-e_1)q_1[l_C+m+(1-e_2)q_2\Delta_C] < (e_1+(1-e_1)q_1)[(1-\tilde{x}(1-e_2)(1-q_2))\Delta - \tilde{x}l_I] - (k-\pi)$$

$$\Leftrightarrow \quad e_1 > \underline{e}.$$

Next, replacing $\underline{\phi}$ in $h_1(\underline{\phi})$ yields

$$\begin{aligned} h_1(\underline{\phi}) &= \Delta - \frac{k-\pi}{e_1+(1-e_1)q_1} - \tilde{x}[(1-e_2)(1-q_2)\Delta + l_I] \\ &\quad - \frac{(1-e_1)q_1}{e_1(e_1+(1-e_1)q_1)}[k-\pi-(e_1+(1-e_1)q_1)\tilde{x}e_2\Delta] \\ &= \Delta - \frac{k-\pi}{e_1} + \frac{\tilde{x}}{e_1}[(e_2(1-e_1)q_1 - e_1(1-e_2)(1-q_2))\Delta - e_1 l_I]. \end{aligned}$$

Since $\Delta - (k-\pi)/e_1 < 0$ by the assumption that $e_1 < \bar{e}$ and we have just established that $h_1(\underline{\phi}) > 0$ when $e_1 > \underline{e}$, it must be that $(e_2(1-e_1)q_1 - e_1(1-e_2)(1-q_2))\Delta > e_1 l_I$, and hence that $h'(\phi_A) < 0$.

Combining $h'(\phi_A) < 0$ with the fact that $\partial\hat{x}/\partial\phi_A < 0$, (40) implies $\mu_3 > 0$. Using (40) to replace μ_2 in (41), we have

$$\mu_1 = \frac{\partial\hat{x}}{\partial\phi_P} - (1-e_1)q_1\left(\frac{\partial\hat{x}}{\partial\phi_A} - \mu_3 h'(\phi_A)\right).$$

Noting that $\partial\hat{x}/\partial\phi_P = (1-e_1)q_1\partial\hat{x}/\partial\phi_A$ and substituting for $h'(\phi_A)$ from (42), we obtain

$$\begin{aligned}\mu_1 &= \mu_3\left[1 - \frac{[e_2(1-e_1)q_1 - e_1(1-e_2)(1-q_2)]\Delta - e_1 l_I}{(e_1+(1-e_1)q_1)e_2\Delta}\right] \\ &= \mu_3\frac{e_1[(e_2+(1-e_2)(1-q_2))\Delta + l_I]}{(e_1+(1-e_1)q_1)e_2\Delta} > 0.\end{aligned}$$

We conclude that $\phi_P = 0$ and $h(\phi_A) = \phi_P = 0$ at the optimum. □

As Proposition 10 shows, our results on the welfare-maximizing structure of fees are robust to the introduction of patent office and court errors.

Appendix C Simulations

C.1 Homogeneous-good Cournot model with two-sided litigation costs

We assume a linear inverse demand: $P(Q) = a - Q$ where $Q = q_I + q_C$. The invention reduces unit production cost from c to $c' = (1-s)c$ where $s \in (0,1)$.

Let $q(c_i, c_j)$ denote the Cournot equilibrium output when firm i's cost is c_i and its rival's cost is c_j. We have

$$q(c_i, c_j) = \frac{a - 2c_i + c_j}{3}.$$

Let $\pi_I(r)$ and $\pi_C(r)$ denote the inventor's and competitor's equilibrium profits, respectively, when the competitor accepts a royalty rate of r. We have

$$\begin{aligned}\pi_I(r) &= [a - q(c', c'+r) - q(c'+r, c') - c']q(c', c'+r) + rq(c'+r, c') \\ \pi_C(r) &= [a - q(c', c'+r) - q(c'+r, c') - c' - r]q(c'+r, c')\end{aligned}$$

Note that $\pi_I(0) = \pi_C(0) = (\frac{a-c'}{3})^2$. Moreover, it can be shown that

$$\begin{aligned}\pi_I(r) &= \pi_I(0) + 5\delta(r) \\ \pi_C(r) &= \pi_C(0) - 4\delta(r),\end{aligned}$$

where

$$\delta(r) \equiv \frac{1}{9}r(a - (1-s)c - r).$$

Notice that this implies

$$4[\pi_I(r) - \pi_I(0)] = 5[\pi_C(0) - \pi_C(r)]. \tag{43}$$

In region 1, the type-L inventor proposes a royalty r_H with probability y and a royalty r_L with probability $1-y$, while the competitor challenges with probability x when being offered the high royalty rate r_H. Let us first determine r_H and r_L. r_H is the royalty charged by the type-H inventor, namely $r_H = sc$ (Fauli-Oller and Sandonis, 2002). We then have

$$\pi_I(r_H) - \pi_I(0) = \frac{5}{9}(a-c)sc = \Delta.$$

r_L must be such that the competitor is indifferent between challenging and not, given that the inventor is sure to be bad. That is, r_L solves

$$\pi_C(0) - l_C = \pi_C(r_L).$$

Then, (43) implies

$$\pi_I(r_L) - \pi_I(0) = \frac{5}{4}[\pi_C(0) - \pi_C(r_L)] = \frac{5}{4}l_C.$$

Solving explicitly for r_L yields

$$r_L = \frac{1}{2}\left[a - (1-s)c - \sqrt{(a-(1-s)c)^2 - 9l_C}\right].$$

With these expressions, we can compute the cost of the patent to the competitor, $\Delta_C = \pi_C(0) - \pi_C(r_H) = \frac{4}{9}(a-c)sc$, and the incremental profit from monopoly power via royalty licensing, which (in the Cournot model) depends on whether the high or low royalty is charged. With a slight abuse of notation, we denote by $m(\Delta_C)$ the incremental profit when r_H is charged and by $m(l_C)$ the incremental profit when r_L is charged. We have

$$m(\Delta_C) = \Delta - \Delta_C = \pi_I(r_H) - \pi_I(0) - (\pi_C(0) - \pi_C(r_H)) = \frac{1}{9}(a-c)sc = \frac{1}{4}\Delta_C$$
$$m(l_C) = \pi_I(r_L) - \pi_I(0) - (\pi_C(0) - \pi_C(r_L) = \frac{1}{4}l_C.$$

Thus, for $z \in \{\Delta_C, l_C\}$, $m(z) = \frac{1}{4}z$.

Also note that the price elasticity, evaluated at the equilibrium with r_H is given by

$$\eta = -\frac{a + (2-s)c}{2a - (2-s)c}.$$

C.2 Embedding the theoretical model into a Cournot setting

We next determine x. The competitor chooses x so as to make the type-L inventor indifferent between r_H and r_L. Note that, in the event of a challenge, the L type's patent will be invalidated; hence the royalty rate will be $r = 0$ and his profit $\pi_I(0)$. Thus x solves

$$\pi_I(r_L) = x[\pi_I(0) - l_I] + (1-x)\pi_I(r_H).$$

Solving for x and plugging in the values computed above we obtain

$$x = \frac{\Delta_C + m(\Delta_C) - (l_C + m(l_C))}{\Delta_C + m(\Delta_C) + l_I} = \frac{(5/4)[\Delta_C - l_C]}{(5/4)\Delta_C + l_I} = \frac{4(a-c)sc - 9l_C}{4(a-c)sc + (36/5)l_I}.$$

Finally, we determine y. The low type inventor chooses y so as to make the competitor indifferent between challenging and not when observing r_H:

$$\hat{\lambda}\pi_C(r_H) + (1-\hat{\lambda})\pi_C(0) - l_C = \pi_C(r_H),$$

which yields

$$(1-\hat{\lambda})[\pi_C(0) - \pi_C(r_H)] = l_C$$

where

$$\hat{\lambda} = \frac{\lambda}{\lambda + (1-\lambda)(1-e)y}.$$

is the competitor's posterior probability that the patent is of type H. Note that because the competitor's belief is correct in equilibrium, $\hat{\lambda}$ is also the probability that a challenged patent is upheld by courts, and thus equals the patentee win rate (VR). Since $\Delta_C = \pi_C(0) - \pi_C(r_H)$ we can write

$$\hat{\lambda} = \frac{\Delta_C - l_C}{\Delta_C} = VR.$$

We thus have

$$\hat{\lambda} = 1 - \frac{9l_C}{4(a-c)sc}.$$

Using the two equations for $\hat{\lambda}$ and solving for y yields

$$y = \frac{\lambda}{1-\lambda}\frac{9l_C}{(1-e)(4(a-c)sc - 9l_C)}.$$

With these results, we can express the challenge credibility constraint as

$$\left(1 - \frac{\lambda}{\lambda + (1-\lambda)(1-e)}\right)\frac{4}{9}(a-c)sc \geq l_C.$$

For each simulation run, we confirm that the output (a, c, l_C, λ, e) satisfies this constraint.

The grant rate is

$$GR = \lambda + (1-\lambda)(1-e)$$

and the observed litigation rate is

$$LR = \frac{x[\lambda + (1-\lambda)(1-e)y]}{\lambda + (1-\lambda)(1-e)} = \frac{x\lambda}{GR\,\hat{\lambda}}.$$

Assuming identical litigation costs, $l_I = l_C$, and using the fact that, in equilibrium, $l_C = (1-\hat{\lambda})\Delta_C$, we can simplify the expression for x as

$$x = \frac{5\hat{\lambda}}{9 - 4\hat{\lambda}}.$$

Plugging this into LR, using $\hat{\lambda} = VR$, and solving for λ yields

$$\lambda = \frac{LR \cdot GR(9 - 4VR)}{5}.$$

Finally, from the grant rate equation we then obtain e as

$$e = 1 - \frac{GR - \lambda}{1 - \lambda} = \frac{5(1 - GR)}{5 - LR \cdot GR(9 - 4VR)}.$$

R&D Equation

Expected returns to innovation relative to no invention for types L and H are:

$$\Pi_H = \pi_I(r_H) - xl_I - \phi_A - \phi_P - \pi_I^{NI}$$

$$\Pi_L = e\pi_I(0) + (1 - e)\big[y(x\pi_I(0) + (1 - x)\pi_I(r_H)) + (1 - y)\pi_I(r_L) - xyl_I - \phi_P\big] \\ - \phi_A - \pi_I^{NI}.$$

Setting observed R&D expenditures, R (including a rate of return) equal to expected returns to innovation:

$$(1 + \tau)R = \lambda\Pi_H + (1 - \lambda)\Pi_L,$$

where τ is the private rate of return to R&D, estimated by Bloom, Schankerman and Van Reenen (2013) at 0.25.

Given the earlier analysis, the right hand side of this equation can be expressed as a function of $(l, a, c, \lambda, s, \eta, VR, GR)$.

C.3 Computing investment costs, R&D and welfare

Note that $K = \lambda\kappa_H + (1 - \lambda)\kappa_L$, where K denotes observed investment expenditure per patent application. In the model we normalize $\kappa_L = 0$ but for the simulations we need estimates of both κ_L and κ_H. To obtain these, we exploit the following relationships:

$$\kappa_H = \frac{K - (1 - \lambda)\kappa_L}{\lambda}$$

and

$$\kappa_L \leq \pi_I(0) - \pi_I^{NI} \leq \kappa_H \leq \pi_I(r_H) - xl_I - \phi_A - \phi_P - \pi_I^{NI}$$

where $\pi_I(0)$ is Cournot profit for the innovator if there is no patent (zero royalty), $\pi_I(r_H)$ is innovator profit at the high royalty and π_I^{NI} is profit if there is no cost-reducing invention (the outside option payoff). The first equation defines a locus of (κ_L, κ_H) consistent with observed total development cost. The first two inequalities in the second equation say that low type inventions would be developed without patent protection but high types would not. The last inequality ensures that there is sufficient incentive with a patent to develop the high type. Among all (κ_L, κ_H) on the locus that satisfy the inequalities, we use the average. Nothing of substance changes if we use other feasible values.

Finally, to compute welfare, note that industry output in equilibrium is

$$Q(c_i, c_j) = q(c_i, c_j) + q(c_j, c_i) = \frac{2(a - c') - r}{3}.$$

Hence, consumer surplus, as a function of the royalty rate r, is

$$CS(r) = \frac{1}{2}\left(\frac{2(a - c') - r}{3}\right)^2.$$

Let $w(r) = \pi_I(r) + \pi_C(r) + CS(r)$. Then, expected total welfare is

$$\begin{aligned} W = \lambda\Big[& w(r_H) - x(l_C + l_I) - \gamma(e) - \kappa_H\Big] \\ &+ (1-\lambda)\Big[\alpha\big[ew(0) + (1-e)\big(y(xw(0) + (1-x)w(r_H)) + (1-y)w(r_L) \\ &- xy(l_C + l_I)\big) - \gamma(e)\big] + (1-\alpha)w(0) - \kappa_L\Big]. \end{aligned}$$

C.4 Calibration

Examination cost per patent application, $\gamma(e)$

We assume constant returns to scale in examination intensity per application, $e : \gamma(e) = \psi e$. Given the simulated value of e, say $\hat{e}$, we compute the unit cost ψ using the examination/search cost per patent reported by the USPTO as \$3,660. Hence $\psi = \frac{3660}{\hat{e}}$, and in percentage-point terms, $MCE = \psi/100$. For each policy experiment, we use the new simulated examination intensity, e^*, and then compute the counterfactual cost of processing an application as $\gamma(e^*) = \psi e^*$.

Pre-grant (application) fees, ϕ_A

Patent office pre-grant fees include filing, search, examination and processing fees. This yields $\phi_A = \$1,740$ (www.uspto.gov/learning-and-resources/fees-and-payment/uspto-fee-schedule). We exclude filing fees for excess independent claims (above three) and total claims (above 20) because the average numbers of claims fall below these thresholds (Dennis Crouch, "The Rising Size and Complexity of the Patent Document," University of Missouri School of Law, http://dx.doi.org/10.2139/ssrn.1095810). We also include the the cost of preparing the patent application. External estimates (www.ipwatchdog.com/2015/the-cost-of-obtaining-a-patent-in-the-us) fall in the range $\$10,000-\$20,000$. We use \$15,000 as the baseline, but results are very similar for other values in the range.

Post-grant (activation) fees, ϕ_P

To compute post-grant fees, we include the issuance fee and (large entity) maintenance fees at ages 4, 8 and 12 (www.uspto.gov/learning-and-resources/fees-and-payment/uspto-fee-schedule). Assuming maintenance to full term, the computed post-grant fees (ignoring discounting) are $\phi_P = \$13,560$. We do not include fees for ex parte re-examination, supplemental examination or various appeals.

R&D per patent, R

The worldwide R&D expense per patent application in 2008 is reported as $2.4 million (*InfoBrief NSF13-207, National Center for Science and Engineering Statistics, U.S. National Science Foundation*). We use this figure as the baseline, but simulation results are similar using $2, 3 and 4 million.

Investment cost, K

We obtain the fraction of private R&D costs accounted for by development expenditures, taken from annual U.S. National Science Foundation surveys (www.nsf.gov/statistics/2016/nsf16301/pdf/tab17.pdf). We take the average value of this ratio for the period 1995-2000, which is 0.80, and multiply it by the R&D per patent as described above.

Litigation rate

The litigation rate is set at 0.015, computed as the ratio of patent suits to the total number of patent grants (Lemley, 2001). However, studies of renewal data show that most patents have very little value, and thus are unlikely to satisfy the challenge credibility constraint in the model. Thus we adjust the overall litigation rate to correspond to those patents that satisfy this constraint in order to compute LR used in the simulations. To do this, we compute the minimum value that satisfies the constraint, denoted by Δ_C^*, given an assumption on the minimum litigation costs $l_{\min}$. That is, Δ_C^* solves:

$$\left(1 - \frac{\lambda}{\lambda + (1-\lambda)(1-e)}\right)\Delta_C^* = l_{\min}.$$

Substituting for λ and e derived in Appendix C.2, we can express Δ_C^* as

$$\Delta_C^*(p) = \frac{5l_{\min}}{4\left(1 - \frac{0.015(9-4VR)}{5(1-p)}\right)}$$

where p is the percentile of the lognormal distribution of the value of patent rights that corresponds to Δ_C^*. We can therefore solve for p.[37] Then we compute the relevant litigation rate for patents that satisfy the challenge credibility constraint in the model

[37] To solve for p, we use the estimated parameters of the lognormal distribution of the value of patent rights for Germany (Schankerman and Pakes, 1986), rescaled for inflation and relative values of GDP of the U.S. and Germany for 2014 (in this we assume that the value of patent rights is proportional to GDP, which is roughly consistent with the findings in their paper). Details are available on request.

as $LR = \frac{0.015}{p}$. In the baseline simulations with perfect courts, we find $p = 0.107$ and thus $LR = 0.14$. For the simulations with imperfect courts it is very similar.

We set $l_{\min} = \$350,000$, which is the minimum litigation cost to the end of discovery for cases where the value at stake is less than \$1 million, as reported by the American Intellectual Property Law Association (2011).

Validation rate

We use the fraction of patent challenge cases in which the validity of the patent is upheld by the court, taken from Allison, Lemley and Schwartz (2014). Using all cases filed in U.S. district courts for 2008-09 (decided 2009-13), they compute that the *challenger* wins in about 43% of the invalidity suits. As an approximation, we set the validation rate at 0.60.

Grant rate

The patent grant rate is measured by the number of patent grants divided by the number of applications (including continuation and divisional filings). Carley, Hegde and Marco (2015) compute this grant rate over the period 1996-2005, which they refer to as the 'family allowance rate'. We set the grant rate at 0.75, which is roughly the average over the period 1991-2000.

C.5 Imperfect courts

In this section we extend the Cournot model to a setting with imperfect courts. We consider a special case of the generalized screening technology introduced in Appendix B where $e_1 = e$, $e_2 = \beta + (1-\beta)e$ with $\beta \in [0,1]$, $q_1 = 1$, and $q_2 = \hat{\lambda}$. That is, the courts are more accurate than the patent office (the higher β, the closer the courts are to being perfect), and while the patent office accepts all applications for which it does not find evidence either way, the courts apply a presumption of validity that depends on what happens at the previous stages: if no evidence is found by the court, patents are held valid with probability $\hat{\lambda}$, which corresponds to the Bayesian posterior that the inventor is of type H. In region 1, we have

$$\hat{\lambda} = \frac{\lambda}{\lambda + (1-\lambda)(1-e_1)y}. \tag{44}$$

Thus (in region 1) q_2 depends endogenously on e_1, λ and y.

The high royalty rate charged by the type-H inventor continues to be $r_H = sc$. The low royalty rate r_L, which must be such that the competitor is indifferent between challenging and not given that the inventor is sure to be bad, now solves

$$(1-e_2)\hat{\lambda}\pi_C(r_H) + (e_2 + (1-e_2)(1-\hat{\lambda}))\pi_C(0) - l_C = \pi_C(r_L).$$

Subtracting $\pi_C(0)$ from both sides and rearranging, we obtain

$$\begin{aligned}\pi_C(0) - \pi_C(r_L) &= l_C + (1-e_2)\hat{\lambda}[\pi_C(0) - \pi_C(r_H)] \\ &= l_C + 4(1-e_2)\hat{\lambda}\delta(r_H) \\ &= l_C + \frac{4}{9}(1-e_2)\hat{\lambda}(a-c)sc.\end{aligned}$$

Then, (43) implies

$$\pi_I(r_L) - \pi_I(0) = \frac{5}{4}[\pi_C(0) - \pi_C(r_L)] = \frac{5}{4}l_C + \frac{5}{9}(1-e_2)\hat{\lambda}(a-c)sc.$$

One can solve explicitly for r_L, yielding

$$r_L = \frac{1}{2}\left[a - (1-s)c - \sqrt{(a-(1-s)c)^2 - 4(1-e_2)\hat{\lambda}(a-c)sc - 9l_C}\right].$$

Thus, for any $e_2 < 1$, r_L is higher than in the baseline model, where $e_2 = 1$. Intuitively, the low type can ask for a higher royalty because the competitor is not sure to win in court even when she is sure of facing a low type.

Now let us determine x. The competitor chooses x so as to make the bad type of inventor indifferent between r_H and r_L. Note that, in the event of a challenge, the low type's patent is upheld with probability $(1-e_2)\hat{\lambda}$ and invalidated with probability $(e_2 + (1-e_2)(1-\hat{\lambda}))$, while the high type's patent is upheld with probability $(e_2 + (1-e_2)\hat{\lambda})$ and invalidated with probability $(1-e_2)(1-\hat{\lambda})$. Thus x solves

$$\pi_I(r_L) = x\big[(1-e_2)\hat{\lambda}\pi_I(r_H) + (e_2 + (1-e_2)(1-\hat{\lambda}))\pi_I(0) - l_I\big] + (1-x)\pi_I(r_H).$$

Subtracting $\pi_I(0)$ from both sides yields

$$\pi_I(r_L) - \pi_I(0) = x\big[(1-e_2)\hat{\lambda}[\pi_I(r_H) - \pi_I(0)] - l_I\big] + (1-x)[\pi_I(r_H) - \pi_I(0)].$$

Solving for x and plugging in the values computed above we obtain

$$\begin{aligned}x &= \frac{\pi_I(r_H) - \pi_I(0) - [\pi_I(r_L) - \pi_I(0)]}{(1-(1-e_2)\hat{\lambda})[\pi_I(r_H) - \pi_I(0)] + l_I} \\ &= \frac{\frac{5}{9}(1-(1-e_2)\hat{\lambda})(a-c)sc - \frac{5}{4}l_C}{\frac{5}{9}(1-(1-e_2)\hat{\lambda})(a-c)sc + l_I}.\end{aligned}$$

Finally, let us determine y. The low type of inventor chooses y so as to make the competitor indifferent between challenging and not when observing r_H. That is, y solves

$$\begin{aligned}&\hat{\lambda}\big[(e_2 + (1-e_2)\hat{\lambda})\pi_C(r_H) + (1-e_2)(1-\hat{\lambda})\pi_C(0)\big] \\ &\quad + (1-\hat{\lambda})\big[(1-e_2)\hat{\lambda}\pi_C(r_H) + (e_2 + (1-e_2)(1-\hat{\lambda}))\pi_C(0)\big] - l_C = \pi_C(r_H)\end{aligned}$$

$$\Leftrightarrow \quad \big[(1-e_2)(1-\hat{\lambda})+e_2(1-\hat{\lambda})\big][\pi_C(0)-\pi_C(r_H)] = (1-\hat{\lambda})[\pi_C(0)-\pi_C(r_H)] = l_C,$$

which is the same expression as in the baseline model. We have

$$\pi_C(0)-\pi_C(r_H) = \frac{4}{9}(a-c)sc.$$

Thus, in equilibrium

$$\hat{\lambda} = 1 - \frac{9l_C}{4(a-c)sc}. \tag{45}$$

Using (44) to solve for y yields

$$y = \frac{\lambda}{1-\lambda}\frac{9l_C}{(1-e_1)(4(a-c)sc-9l_C)}.$$

The challenge credibility constraint becomes

$$\begin{aligned}&\underline{\lambda}\big[(e_2+(1-e_2)\underline{\lambda})\pi_C(r_H)+(1-e_2)(1-\underline{\lambda})\pi_C(0)\big]\\ &\qquad +(1-\underline{\lambda})\big[(1-e_2)\underline{\lambda}\pi_C(r_H)+(e_2+(1-e_2)(1-\underline{\lambda}))\pi_C(0)\big]-l_C \geq \pi_C(r_H)\\ &\qquad\qquad \Leftrightarrow \quad (1-\underline{\lambda})\,\frac{4}{9}(a-c)sc \geq l_C,\end{aligned}$$

where

$$\underline{\lambda} = \frac{\lambda}{\lambda+(1-\lambda)(1-e_1)}.$$

Inferring λ and e from VR, GR, and LR.

The validation rate is

$$VR = \hat{\lambda}[e_2+(1-e_2)\hat{\lambda}]+(1-\hat{\lambda})(1-e_2)\hat{\lambda} = \hat{\lambda}.$$

The grant rate is

$$GR = \lambda+(1-\lambda)(1-e_1).$$

The litigation rate is

$$LR = \frac{x[\lambda+(1-\lambda)(1-e_1)y]}{\lambda+(1-\lambda)(1-e_1)} = \frac{x\lambda}{GR\cdot\hat{\lambda}}.$$

Using $l_I = l_C$, $l_C = (1-\hat{\lambda})\frac{4}{9}(a-c)sc$, $e_1 = e$, $e_2 = \beta+(1-\beta)e$, and $VR = \hat{\lambda}$, we can write

$$\begin{aligned}x &= \frac{\frac{5}{9}(1-(1-\beta)(1-e)VR)(a-c)sc-\frac{5}{9}(1-VR)(a-c)sc}{\frac{5}{9}(1-(1-\beta)(1-e)VR)(a-c)sc+\frac{4}{9}(1-VR)(a-c)sc}\\ &= \frac{5VR(1-(1-\beta)(1-e))}{9-VR(5(1-\beta)(1-e)+4)}.\end{aligned}$$

Substituting for x and $\hat{\lambda}$ in the litigation rate equation yields

$$LR = \frac{5\lambda(1-(1-\beta)(1-e))}{GR[9-VR(5(1-\beta)(1-e)+4)]}.$$

Together with the grant rate equation (replacing $e_1 = e$), this can be used to determine e and λ.