INVENTORS AND IMPOSTORS: 
AN ANALYSIS OF PATENT EXAMINATION 
WITH SELF-SELECTION OF FIRMS INTO R&D*

Florian Schuett†
April 2013

Abstract

I present a model in which firms differing in R&D productivity choose between ambitious research projects, which are socially desirable, and unambitious ones, which are socially undesirable. The patent office must decide how rigorously to examine applications, which affects the probability of weeding out bad applications but also how firms self-select into R&D. I show that when a subset of firms is financially constrained, the patent office should examine their applications more rigorously. This generates a number of predictions that I test by exploiting the 1982 reform that introduced firm-size dependent fees in the United States.

Keywords: innovation, patent office, optimal patent policy
JEL classification numbers: O31, O38, D73, D82, L50

* This and a companion paper (Schuett [2013]) have developed out of Chapter 1 of my Ph.D. dissertation at the Toulouse School of Economics. I am grateful to Mark Schankerman and Pierre Régibeau (the editors) as well as an anonymous referee, whose comments and suggestions have greatly improved the paper. I also thank Jim Bessen, Jing-Yuan Chiou, Pascal Court, Vincenzo Denicolò, Guido Friebel, Elisabetta Iossa, Paolo Pin, Gaëtan de Rassenfosse, François Salanié, Paul Seabright, Jean Tirole, participants at the EVPAT summer school in Bologna, the “Knowledge for Growth” conference in Toulouse, the congress of the European Economic Association in Milan, the EARIE conference in Toulouse, the EPIP conference in Bern, the ASSET conference in Florence, the USPTO-Kauffman Foundation Conference on Patents, Entrepreneurship and Innovation in Washington, D.C., as well as seminar participants at the Toulouse School of Economics, the European University Institute, the University of Alicante, Ludwig Maximilian University Munich, the Centre for European Economic Research (ZEW), the University of Vienna, the Australian National University, Tilburg University, and Goethe University Frankfurt for helpful comments. All errors are mine. Financial support from the 7th European Community Framework Programme through a Marie Curie Intra European Fellowship (grant number PIF-GA-2010-275032) is gratefully acknowledged.

† TILEC & CentER, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands. Tel.: +31 13 466 4033. Fax: +31 13 466 3042. Email address: f.schuett@uvt.nl.
I Introduction

Patent law requires that patents only be granted for inventions that are novel and nonobvious, and for good reason. Granting patents for inventions that are not novel causes deadweight loss and litigation without providing any offsetting benefit to society. Granting patents for inventions that are obvious, i.e., that do not satisfy a minimum inventive step, may dampen the incentives of initial innovators or otherwise slow down the rate of technical progress (Scotchmer and Green [1990]; Scotchmer [1996]; O’Donoghue [1998]; Hunt [2004]). The patent office plays the role of a watchdog, making sure that only novel and nonobvious inventions obtain patent protection. Yet, as infringement lawsuits filed by holders of dubious patents against prominent firms such as eBay and RIM have brought to public awareness, the patent office does not reliably weed out bad patents.\(^1\) The failure of the patent office to rigorously examine patent applications is a source of concern to many observers.\(^2\) Lemley [2001], however, argues that the patent office may rationally choose to spend limited resources on examining a given application because only a tiny fraction of patents ever turn out to be commercially significant. The cost of weeding out more bad applicants might well exceed the benefit.

How rigorously should the patent office examine patent applications? What are the benefits and costs of more rigorous examination? Answering these questions requires a theory of the process through which applications are generated. In this paper, I propose a model in which firms choose between more or less ambitious research projects. Ambitious projects lead to social gains, while, if patented, unambitious projects lead to social losses. To obtain patent protection, firms have to file an application with the patent office. The patent office maximizes welfare and wields two instruments: an application fee and the intensity with which it examines applications; the combination of the two is what I define as patent policy. I first study how patent policy affects investment in R&D and characterize the optimal policy when firms are not financially constrained, so that the patent office can set application fees at any desired level. I then introduce a subset of small firms that have wealth constraints, so that they cannot afford high fees. I analyze how this affects the optimal policy and derive testable predictions that I confront with empirical evidence.

In the model presented in Section II, firms differ in their ability to produce valuable inventions (their research productivity) and choose whether to start ambitious research projects or unambitious bad projects. Research projects may or may not lead to patentable inventions and are, in expectation, socially desirable. Bad projects never lead to patentable inventions and are socially undesirable. If applications on unpatentable inventions are accepted by the patent office, they cause social losses. The private profitability of the two activities depends on firms’ productivity and on the patent office’s examination intensity, the cost of which is increasing.
and convex. More rigorous examination makes it less likely that unpatentable inventions escape detection and therefore increases the attractiveness of research projects relative to bad projects. This setup leads to self-selection of firms: under a single-crossing condition, high-productivity firms choose research projects, while low-productivity firms choose bad projects or stay inactive. Importantly, the threshold above which firms choose research decreases with the examination intensity. My formulation thus acknowledges that the patent office may have a role in encouraging R&D, as stressed by Jaffe and Lerner [2004].

I show in Section III that the optimal patent policy in the benchmark case without wealth constraints involves full deterrence of bad projects. The examination intensity pins down the threshold for research, and the application fee is set in such a way that all firms whose productivity is below the threshold remain inactive. To determine how rigorously to examine applications, the patent office equalizes the marginal benefit of patent examination with its marginal cost. The benefit consists of two parts: an ex post welfare effect due to fewer bad applications being granted and an ex ante effect whereby more rigorous examination increases the number of firms selecting into research. Although the full-deterrence result should not be interpreted too literally, as it relies on a number of simplifying assumptions, I argue that the nature of the results would be similar in a richer model.

In Section IV, I extend the model by introducing wealth constraints on the part of some firms. I first consider the case where firms do not have access to credit. Wealth constraints then put an upper bound on the application fee that the patent office can charge. Provided the patent office is able to identify which firms are wealth constrained, it can condition its patent policy on the applicant’s wealth. I show that the patent office will examine applications from wealth constrained firms more rigorously than those from other firms. Interpreting wealth constraints as being correlated with firm size, I derive predictions on differences in grant rates, patent quality, and the private value of patents between small and large firms. In particular, the model predicts a higher grant rate and a higher average value of patents for large firms, both of which find support in the available cross-sectional evidence.

Of course, such cross-sectional differences may have a variety of alternative explanations. To better test the theory, I look at the 1982 reform that introduced differentiated fees in the U.S. The reform was associated with a substantial increase in application fees for large firms, and only to a lesser extent for small firms. Using a difference-in-differences approach, de Rassenfosse [2012] provides evidence that citations and family size of patents issued to large firms increased relative to those of small firms. I complement his results by looking at renewal rates and fee-adjusted renewal rates. My findings corroborate the notion that the value of large-entity patents increased by more than that of small-entity patents, which is consistent with the predictions of the model.
In Section V, I revisit the benchmark case of no wealth constraints and derive comparative statics results concerning the optimal patent policy. Using these results, I relate the creation of a specialized patent court to a decrease in examination intensity. Such a relationship is in line with the observation that the creation of the Court of Appeals for the Federal Circuit (CAFC) in the U.S. seems to have been accompanied by a decline in the rigor of examination on the part of the USPTO. I also investigate the effect of post-grant opposition and obtain a condition for an opposition procedure to be welfare-enhancing.

I go on to consider the case where firms have access to external funding in Section VI. The project-selection aspect of the basic model combined with a financing stage à la Holmstrom and Tirole [1997] naturally gives rise to a moral-hazard problem. An entrepreneur who obtains a loan to finance a research project may be tempted to choose a bad project, which has a lower probability of success but requires a smaller investment and is therefore associated with larger private benefits. For the entrepreneur to have the incentive to choose the research project, he needs to be given a large enough share of the profit, thus limiting the size of the loan the firm can obtain. This creates an additional role for patent policy: more rigorous examination relaxes the entrepreneur’s incentive-compatibility constraint, enabling more firms to get funding for R&D. I show that financing problems make it more costly to implement a given innovation threshold.

Related literature. The paper is closely related to a small but growing literature on patent examination. Langinier and Marcoul [2009] and Atal and Bar [2010] study inventors’ incentives to search for and disclose relevant prior art to the patent office. Prior-art disclosure is not the focus of this paper. Régibeau and Rockett [2010] examine the optimal duration of patent examination as a function of the importance of an innovation. In their paper, as in mine, the patent office trades off the cost of patent examination (delay, in their paper) against its welfare benefits. In a working paper version (Régibeau and Rockett [2007]), they consider the possibility that firms choose between genuinely innovative projects and non-inventions, as I do in this paper. They assume homogeneous applicants, however, and focus on how to choose the delay of patent examination to provide incentives for firms to invest in genuine innovations. Instead, I ignore the time dimension and introduce application fees as an additional instrument. Caillaud and Duchêne [2011] and Atal and Bar [2012] study models in which, as in this paper, the patent office chooses examination intensity and application fees. In Caillaud and Duchêne [2011], valid inventions stem from successful R&D projects and invalid ones from failed projects. They focus on the positive part of the analysis, leaving normative considerations about optimal examination intensity and application fees to the side. In Atal and Bar [2012], inventors have private information about the (heterogenous) probability that
their inventions are valid. Unlike here, this probability is exogenous, however.

An assumption that all of these papers share with this one is that of a benevolent patent office maximizing welfare. By contrast, Schuett [2013] examines the agency problem within the patent office. Patent examination is modeled as a moral-hazard problem followed by an adverse-selection problem: the examiner’s incentives have to be structured so as to make the examiner exert effort searching for evidence to reject, but also to make him truthfully reveal whatever evidence he finds.

Chiou [2008] and Koenen [2011] look at how the patent office’s examination effort interacts with the incentives of private parties to bring court challenges. Chiou shows that the two types of enforcement may be complementary, and that weak patents are more likely to be settled out of court than strong patents. Koenen also looks at infringement suits brought by the patent holder and shows that in their presence increased application fees may decrease the average quality of applications.

The paper is more loosely related to the broader literature on the optimal design of the patent system. In particular, Cornelli and Schankerman [1999] and Scotchmer [1999] study the renewal mechanism, which can be interpreted as a menu of application fees and patent lengths. Hopenhayn and Mitchell [2001] add patent breadth as an additional instrument and show that breadth can be a better tool for screening than fees. Hopenhayn et al. [2006] also consider buyouts in a model of sequential innovation. In this literature, inventions differ in cost and value but do not have a validity dimension.

II The model

There is a continuum (with mass 1) of potential applicant firms. Each firm is characterized by a productivity parameter \( \theta \), which is its private knowledge and distributed according to a cumulative distribution function \( F \) with probability density \( f \) on the support \([0, \bar{\theta}]\). Assume that \( F \) is twice continuously differentiable and that \( f(\theta) > 0 \) for \( \theta \in (0, \bar{\theta}) \).

Firms choose between two activities, denoted \( R \) and \( B \), and can also remain inactive (denoted \( I \)). Activity \( R \) corresponds to a research project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activity \( B \) corresponds to a bad project that can benefit society, while activit
the patent holder to extract rents from users but does not benefit society.³

Applications resulting from activity $R$ are not easily distinguishable from applications resulting from activity $B$. The patent office needs to examine an application in order to obtain a signal about whether the claimed invention constitutes a significant advance over the prior art. The precision of the signal is determined by the patent office’s examination intensity $e \in [0, 1]$. The cost of examining an application with intensity $e$ is $\gamma(e)$, with $\gamma(0) = \gamma'(0) = 0$, $\gamma'(e) > 0$ for all $e > 0$, $\gamma'' > 0$, and $\gamma'(1) = \infty$.

**Profits and welfare.** Each firm chooses its activity from $\{R, B, I\}$. If a firm of type $\theta$ chooses $R$, its expected payoff given the patent office’s examination intensity $e$ is $\pi^R(\theta, e)$. If it chooses $B$, its payoff is $\pi^B(\theta, e)$. A firm’s payoff from inactivity ($I$) is zero. The following assumption specifies how the profit functions are related to $\theta$ and $e$. Subscripts denote partial derivatives.

**Assumption 1.** The profit functions satisfy the following conditions:

(i) $\pi^R_\theta > \pi^B_\theta \geq 0$,

(ii) $\pi^R(0, 0) < \pi^B(0, 0)$ and $\pi^B(0, e) \geq 0$,

(iii) $\pi^R(\theta, 0) > \pi^B(\theta, 0)$,

(iv) $\pi^B_e < 0$ and $\pi^B_e < \pi^R_e$.

Profits from both activities increase with $\theta$, perhaps because both ambitious and not so ambitious research projects require some of the same qualities to be brought to fruition. Profits from activity $R$ are more sensitive to $\theta$ than those from activity $B$, though. When $e = 0$, it is more profitable for firms at the lower end of the productivity distribution ($\theta = 0$) to choose activity $B$, while for firms at the upper end of the distribution ($\theta = \theta$), it is more profitable to choose $R$. Higher examination intensity negatively affects the profits from activity $B$, as the patent office is more likely to find defeating prior art. I make no assumption on how $e$ affects the profits from activity $R$. Tighter screening might lead to more rejections even for applications resulting from ambitious projects, but might also reduce the incidence of erroneous rejections (type I error; Régibeau and Rockett, 2010) or increase the value of granted patents because of higher perceived overall patent quality (Atal and Bar [2012]). I assume, however, that profits from activity $B$ decrease faster with $e$ than profits from $R$.

The social value accruing from a firm of type $\theta$ choosing $R$ is $w(\theta, e)$, while the social value from the same firm choosing $B$ is $-\ell(\theta, e)$. I assume $w_\theta \geq 0$, $w_e \geq 0$, $w_{ee} \leq 0$, $w_{\theta e} \leq 0$, and $\ell_e \leq 0$. That is, more productive firms create more valuable inventions; in addition,
higher examination intensity increases the expected welfare from a given firm choosing R (at a decreasing rate, and less so for firms with high productivity) and reduces the expected social loss from choosing B. I also assume that the social value of activity R weakly exceeds its private value, \( w(\theta, e) \geq \pi^R(\theta, e) \), and that activity B is socially harmful, \( \ell(\theta, e) \geq 0 \), with strict inequality for \( e < 1 \). \(^4\) The patent office maximizes welfare. Apart from the examination intensity \( e \), the only other instrument at its disposal is an application fee \( \phi \geq 0 \) that must be paid by all firms applying for patent protection.

The effect of patent policy on firm behavior. Given a patent policy \((e, \phi)\), each firm chooses the activity that maximizes its expected payoff. Firm \( \theta \) prefers activity R to B if

\[
\pi^R(\theta, e) - \phi \geq \pi^B(\theta, e) - \phi.
\]

Assumption 1 is sufficient for the existence of a unique threshold \( \hat{\theta} \), defined by

\[
(1) \quad \pi^R(\hat{\theta}, e) = \pi^B(\hat{\theta}, e),
\]

such that, in the absence of application fees, activity B is chosen for all \( \theta < \hat{\theta} \) and R is chosen for all \( \theta \geq \hat{\theta} \). The threshold depends on \( e \), i.e., \( \hat{\theta} = \hat{\theta}(e) \). Assumption 1 implies that

\[
(2) \quad \frac{\partial \hat{\theta}}{\partial e} = \frac{-\pi^R_e - \pi^B_e}{\pi^R_\theta - \pi^B_\theta} < 0.
\]

Thus, the threshold above which firms choose R decreases with \( e \), meaning that higher examination intensity leads a larger set of firms to choose ambitious R&D projects. The intuition is that stricter examination increases the relative attractiveness of ambitious research. Even though an increase in \( e \) may reduce \( \pi^R \), it reduces \( \pi^B \) even more, so that firms at the margin between B and R find it more profitable to choose R.

Provided \( \phi \leq \pi^R(\hat{\theta}(e), e) \), there is a second threshold \( \theta_0 \), defined by

\[
(3) \quad \pi^B(\theta_0, e) = \phi,
\]

such that firms with \( \theta \geq \theta_0 \) prefer B to inactivity.\(^5\) This threshold depends on both \( e \) and \( \phi \), i.e., \( \theta_0 = \theta_0(e, \phi) \). Firms with productivity below \( \theta_0 \) remain idle, firms with productivity between \( \theta_0 \) and \( \hat{\theta} \) choose B, and firms with productivity above \( \hat{\theta} \) choose R. By Assumption 1,

\[
(4) \quad \frac{\partial \theta_0}{\partial e} = \frac{-\pi^B_e}{\pi^B_\theta} > 0
\]

\[
(5) \quad \frac{\partial \theta_0}{\partial \phi} = \frac{1}{\pi^B_\theta} > 0
\]

(provided \( \pi^B_\theta > 0 \)).\(^6\) Both instruments, \( e \) and \( \phi \), increase the threshold above which firms prefer activity B to inactivity. Higher examination intensity and higher application fees lead
a smaller set of firms to choose bad projects. In summary, a patent policy \((e, \phi)\) leads to self-selection of firms between more or less ambitious R&D projects and inactivity, as illustrated in Figure 1.

III Optimal patent policy

In this section, I first derive some general features of the optimal patent policy for the benchmark case where firms do not face wealth constraints. The patent office chooses a patent policy \((e, \phi)\), consisting of an application fee and an examination intensity, to maximize welfare, given by

\[
\int_{\hat{\theta}(e)}^{\bar{\theta}(e)} w(\theta, e) dF(\theta) - \int_{\theta(e, \phi)}^{\hat{\theta}(e)} \ell(\theta, e) dF(\theta) - \gamma(e) [1 - F(\hat{\theta}(e, \phi))]
\]

subject to \(\phi \leq \pi^R(\hat{\theta}(e), e)\). The first term in (6) corresponds to the social value created by research projects (undertaken by firms whose productivity exceeds \(\hat{\theta}\)), the second term captures the expected social losses from bad projects (undertaken by firms with \(\theta \leq \theta < \hat{\theta}\)), and the third term represents the cost of examination. The constraint \(\phi \leq \pi^R(\hat{\theta}(e), e)\) ensures that \(\hat{\theta}\) and \(\theta\) are the relevant thresholds for firm behavior. The following proposition characterizes the optimal patent policy. The proof of this and all other propositions is in Appendix A.

**Proposition 1.** Suppose Assumption 1 holds. The optimal policy \((e^*, \phi^*)\) involves full deterrence of bad projects: \(\theta = \hat{\theta}\). The optimal examination intensity \(e^*\) is strictly positive and satisfies

\[
- \frac{\partial \hat{\theta}}{\partial e} w(\hat{\theta}, e^*) f(\hat{\theta}) + \int_{\hat{\theta}}^{\bar{\theta}} w_e(\theta, e^*) dF(\theta) = \gamma'(e^*) [1 - F(\hat{\theta})] - \frac{\partial \hat{\theta}}{\partial e} \gamma(e^*) f(\hat{\theta}).
\]

The optimal application fee is \(\phi^* = \pi^R(\hat{\theta}(e^*), e^*) > 0\).

The patent office chooses \(e^*\) to equalize the marginal social benefit of examination with its marginal cost and sets \(\phi^*\) so as to deter all firms with \(\theta < \hat{\theta}(e^*)\) from applying. At the optimum, only firms that given \((e^*, \phi^*)\) find it worthwhile to choose activity \(R\) apply for patents. All other firms remain inactive; no firm chooses activity \(B\). Intuitively, raising the application fee up to \(\phi = \pi^R(\hat{\theta}, e)\) does not represent a disincentive to innovation in this model: only those types of firms that would anyway find it optimal to choose bad projects are discouraged from applying for patents by such a fee. Thus, there is no loss in raising the fee up to the level where bad projects are completely deterred.
One should be cautious not to give too literal an interpretation to this full-deterrence result, as it is not robust to slight changes in modeling assumptions. For example, it relies on the fact that there is only one dimension of heterogeneity. Instead, firms could differ in more than one dimension; for example, they might differ in how well they (or their attorneys) navigate the patent application process. The result also relies on the assumption that there are only two types of projects, rather than, e.g., a continuum of projects differing in the likelihood of yielding a patentable invention. Under different assumptions, it would not be possible to achieve full deterrence with only the two instruments, $e$ and $\phi$. The full-deterrence result should therefore be seen as an approximation of the somewhat less extreme results that would obtain in these richer – but more complex – specifications.

The marginal social benefit of examination (left-hand side of (7)) has two components. The first is the effect of stricter examination on the incentive to innovate ($-\frac{\partial }{\partial e} w(\hat{\theta}, e^*) f(\hat{\theta})$), as a larger $e$ encourages more firms to choose $R$. The second is the effect of examination on the welfare generated by firms choosing $R$ ($\int w_e(\theta, e) dF(\theta)$). Stricter examination enables the patent office to weed out more applications that do not warrant patent protection. The marginal cost of examination (the right-hand side of (7)) similarly consists of a direct effect of $e$ on $\gamma$ and an indirect effect on the number of applications (by reducing $\hat{\theta}$, an increase in $e$ pushes up the number of applications).

The first-order condition of the welfare maximization problem (7) may have several solutions. While $e^*$ must be one of the solutions, (7) does not completely pin down $e^*$ in general. The next proposition provides sufficient conditions for (7) to have a unique solution.

**Proposition 2.** Suppose the distribution of $\theta$ has a nondecreasing hazard rate and that the profit functions satisfy $\pi^R_{\theta\theta} \leq \pi^B_{\theta\theta}$, $\pi^R_{ee} \leq \pi^B_{ee}$, and $\pi^R_{\theta e} \geq \pi^B_{\theta e}$. Then, equation (7) uniquely characterizes $e^*$.

### IV Wealth constraints

The optimal policy identified by Proposition 1 (deterrence of all bad projects) may require substantial application fees. A potential issue with high application fees is that small firms facing wealth constraints would be unable to afford them and would therefore forego investment in R&D altogether. To account for this issue, I now extend the model by assuming that a fraction of firms have limited wealth. In this section I consider the case where firms do not have access to external funding. In Section VI, I turn to the case where wealth-constrained firms can obtain financing from outside investors.

I modify the basic model as follows. Firms initially have wealth (or assets) $a$, which will be interpreted as cash on hand that can be used to finance the investment needed to start...
a research project. Starting a research project carries a sunk cost of $k$. A firm’s assets can be high or low, $a \in \{a_L, a_H\}$. I will refer to firms with wealth $a_L$ as small and to firms with wealth $a_H$ as large. The share of small firms is $\alpha < 1$. Assume that $\theta$ and $a$ are independently distributed and that $k < a_L < k + \phi^* < a_H$, where $\phi^*$ is the optimal fee in the absence of wealth constraints. Thus, if the application fee is equal to $\phi^*$, small firms are unable to pay for both the fee and the R&D investment.

Suppose however that the patent office observes $a$, so that it can condition application fee and examination intensity on the applicant’s wealth. The idea is that the patent office can observe whether an applicant is a small or a large firm, and that the size of the firm is correlated with the value of its assets. The patent office now chooses a policy towards small firms ($e_L, \phi_L$) and a policy towards large firms ($e_H, \phi_H$). Clearly, since large firms can afford $\phi^*$, the optimal policy towards large firms is the same as absent wealth constraints: $e_H = e^*$ and $\phi_H = \phi^*$. If the patent office wants small firms to invest in research as well, however, it can charge at most a fee of $a_L - k < \phi^*$. Provided small firms’ expected research output is sufficiently valuable, the patent office will indeed set $\phi_L = a_L - k$ (there is no reason to set a lower fee) and choose $e_L$ to maximize

$$\int_{\theta(e)}^{\theta} w(\theta, e) dF(\theta) - \int_{\theta(e)}^{\hat{\theta}(e)} \ell(\theta, e) dF(\theta) - \gamma(e) [1 - F(\hat{\theta}(e, a_L - k))]$$

subject to $e \leq \bar{e}$, where $\bar{e}$ solves $\hat{\theta}(\bar{e}) = \hat{\theta}(\bar{e}, a_L - k)$. That is, $\bar{e}$ is the value of $e$ such that bad projects are fully deterred when $\phi = a_L - k$. Let $e^{**}$ denote the value of $e$ maximizing (8). The following proposition compares the optimal policy towards small firms to the one towards large firms.\(^8\)

**Proposition 3.** Let $\Delta \equiv \phi^* + k - a_L$ and suppose $-\ell_\epsilon(\hat{\theta}(e^*), e^*) > \gamma'(e^*)$. Then, for low values of $\Delta$, the optimal policy towards small firms involves a greater examination intensity than the optimal policy towards large firms: $e^{**} > e^*$.

According to Proposition 3, when a fraction of firms is wealth constrained and the patent office can identify those firms, it will examine their applications more rigorously than those of firms that are not wealth constrained: $e_L = e^{**} > e_H = e^*$. The intuition is that because the patent office is forced to charge a lower application fee to small firms, it knows that, for a given $e$, the pool of applications from small firms will be of lower quality ($\hat{\theta} < \hat{\theta}^*$) than the pool of applications from large firms. The patent office thus has an additional incentive to tighten examination.\(^9\)

Note that $\phi_L < \phi^*$ does not necessarily mean that there will be bad projects in equilibrium. The patent office may well increase its examination intensity up to $\bar{e}$, at which point bad
projects are again fully deterred. But full deterrence requires more rigorous examination when the application fee is lower.

The result in Proposition 3 relies on the patent office being able to observe firms’ wealth and charge them wealth-dependent fees. If the patent office cannot charge wealth-dependent fees, it has two options: (a) charging a low application fee ($\phi_L$) to everybody, so that both small and large firms participate in the patent system, or (b) charging a high application fee ($\phi^*$), so that only large firms participate. Under option (a) the patent office will optimally examine all applications with intensity $e^{**}$, while under (b) it will examine all applications with intensity $e^*$. For the patent office to prefer option (a) to option (b), the share of small firms $\alpha$ must be sufficiently large.

IV(i) Predictions based on specific profit and welfare functions

To derive testable predictions from these results I now put more structure on the model. Consider the following specification of profits and welfare. Assume that activity $R$ requires an investment $k$ and leads to a major invention with probability $\nu$ and to a minor one with probability $1 - \nu$. At the time of applying for a patent, the firm does not know whether its research has generated a major or minor invention. Activity $B$ requires no investment but always leads to minor inventions. If patented, the private value of a major invention is $g\theta$ and the private value of a minor invention is $b\theta$, with $g > b > 0$. The parameters $g$ and $b$ might reflect, for example, the fact that a patent can be challenged and overturned after being granted. More generally, $g$ measures the profitability of major inventions, while $b$ measures the ability of the holder of a bad patent to extract rents from the users of the patented technology. The social value of a major invention is $S > k$, while the social value of a minor invention is $-D < 0$.

The patent office receives a signal that depends on whether the application covers a major or a minor invention. Specifically, if an invention is minor, the patent office finds evidence (prior art) showing obviousness with probability $e$, and no such evidence with probability $1 - e$. If an invention is major, the patent office can never find evidence of obviousness. Consequently, the patent office rejects an application on a minor invention with probability $e$ and never rejects applications on major inventions. With this specification, the profit functions are

\begin{align*}
\pi^R &= \theta(\nu g + (1 - \nu)(1 - e)b) - k \\
\pi^B &= (1 - e)b\theta.
\end{align*}

The expected social welfare from activity $R$ is

\begin{align*}
w &= \nu S - (1 - \nu)(1 - e)D - k.
\end{align*}
The expected social welfare from activity \( B \) is

\[
-\ell = -(1 - e)D.
\]

Note that the assumptions on the second derivatives of \( \pi^R \) and \( \pi^B \) required for Proposition 2 are all satisfied for these functional forms, as is the assumption that \( w_{\theta e} \leq 0 \). Thus, Proposition 2 applies. For the social value from activity \( R \) to be positive even for low values of \( e \), we must have \( \nu \geq (k + D)/(S + D) \), which I will assume to be satisfied in what follows.

When applied to these functional forms, Proposition 3 generates predictions on grant rates, patent quality, and the private value of patents. I start by defining these three concepts formally. For a given patent policy \((e, \phi)\) entailing cutoffs \((\tilde{\theta}, \hat{\theta})\), the grant rate (GR) is

\[
GR = \frac{(1 - F(\tilde{\theta}))[\nu + (1 - \nu)(1 - e)] + (F(\hat{\theta}) - F(\tilde{\theta}))(1 - e)}{1 - F(\hat{\theta})}.
\]

The denominator corresponds to the total number of patent applications. The numerator corresponds to research projects that either result in major inventions or, if they result in minor inventions, escape detection by the patent office, as well as bad projects that escape detection. The total grant rate in (13) is a weighted average of the conditional grant rates for research projects and bad projects. The conditional grant rate for research projects is \( \nu + (1 - \nu)(1 - e) \), while the conditional grant rate for bad projects is \( 1 - F(\theta) \).

Let us define patent quality (PQ) as the probability that an issued patent is valid, in the sense of satisfying the patentability criteria. In terms of the model, a patent is valid if it covers a major invention rather than a minor one. By Bayes’ rule, the probability that a patent resulting from a research project covers a major invention is \( \nu/(\nu + (1 - \nu)(1 - e)) \). The probability that a patent resulting from a bad project covers a major invention is zero. Letting \( A \) denote the patent office’s decision to accept an application, patent quality thus is

\[
PQ = Pr(B|A) \cdot 0 + Pr(R|A) \cdot \frac{\nu}{\nu + (1 - \nu)(1 - e)},
\]

where \( Pr(B|A) \) is the probability that a firm has engaged in activity \( B \) given that is has been granted a patent, and \( Pr(R|A) \) is defined analogously. Noting that

\[
Pr(B|A) = 1 - Pr(R|A) = \frac{(F(\hat{\theta}) - F(\tilde{\theta}))(1 - e)}{(1 - F(\hat{\theta}))[\nu + (1 - \nu)(1 - e)] + (F(\hat{\theta}) - F(\tilde{\theta}))(1 - e)},
\]

we have

\[
PQ = \frac{\nu F(\hat{\theta})}{(1 - F(\hat{\theta}))[\nu + (1 - \nu)(1 - e)] + (F(\hat{\theta}) - F(\tilde{\theta}))(1 - e)}.
\]

The expected value of a patent (EVP) is

\[
EVP = Pr(B|A)E(b|\theta \leq \tilde{\theta} < \hat{\theta}) + Pr(R|A) \left[ \frac{\nu E(g\theta|\theta \geq \hat{\theta})}{\nu + (1 - \nu)(1 - e)} + \frac{(1 - \nu)(1 - e)E(b|\theta \geq \hat{\theta})}{\nu + (1 - \nu)(1 - e)} \right].
\]
Using the fact that
\[
E(\theta|\theta \leq \theta < \hat{\theta}) = \frac{\int_{\theta}^{\hat{\theta}} \theta dF(\theta)}{F(\hat{\theta}) - F(\theta)},
\]
\[
E(\theta|\theta \geq \hat{\theta}) = \frac{\int_{\hat{\theta}}^{\theta} \theta dF(\theta)}{1 - F(\theta)},
\]
we obtain
\[
\text{(15) EVP} = \left[ \nu g + (1 - \nu)(1 - e) b \right] \int_{\theta}^{\hat{\theta}} \theta dF(\theta) + (1 - e) b \int_{\theta}^{\hat{\theta}} \theta dF(\theta)
\]
\[
\int_{\theta}^{\hat{\theta}} \frac{\theta dF(\theta)}{1 - F(\theta)} + (1 - e) \left[ F(\hat{\theta}) - \nu + (1 - \nu)(1 - e) \right].
\]

To see what Proposition 3 implies for the predicted grant rates of small and large firms, note first that because \(1 - e \leq \nu + (1 - \nu)(1 - e)\) for any \(e\), the conditional grant rate for bad projects is lower than the one for research projects. We have \(d[\nu + (1 - \nu)(1 - e)]/de \leq 0\), with strict inequality for \(\nu < 1\). Thus, irrespective of whether bad projects by small firms are fully deterred (i.e., whether or not \(e^{**} = \bar{e}\)), the model predicts that – everything else being equal – in a patent system with wealth-dependent application fees, the grant rate for small firms should be lower than that for large firms. The straightforward intuition is that large firms submit applications for major inventions at least as often as small ones, and applications submitted by small firms undergo more stringent examination, so that a larger share of minor inventions is detected.

The model’s predictions on patent quality are generally ambiguous. On the one hand, the lower fee for small firms worsens the quality of the application pool (provided bad projects are not fully deterred). On the other hand, the greater examination intensity weeds out more minor inventions. The net effect depends on parameters and on the distribution of \(\theta\). A clear prediction can be derived in the case where the optimal examination intensity towards small firms is a corner solution, \(e^{**} = \bar{e}\), so that \(\theta = \hat{\theta}\), i.e., bad projects are fully deterred. When \(\theta = \hat{\theta}\), (14) simplifies to \(PQ = \nu / [\nu + (1 - \nu) (1 - e)]\), which is increasing in \(e\). The model then predicts that small firms should hold higher-quality patents than large firms. The intuition is that when bad projects are fully deterred, the effect on the quality of the application pool vanishes, and because applications submitted by small firms are examined more rigorously, the applications that do go through are more likely to be valid.

The predictions on the private value of patents are not completely determinate either; they can be sharpened, however, by making some mild additional assumptions. In general, the lower fee for small firms invites applications from low-productivity firms, whose inventions are less valuable, but the greater examination intensity means that patents resulting from research projects are more likely to be valid and thus more valuable. To see this formally,
suppose \( e^{**} = \bar{e} \), so that there are no bad projects; the argument holds a fortiori when \( e^{**} < \bar{e} \).

Then, (15) simplifies to

\[
\text{EVP} = \left( \frac{\int_{\hat{\theta}}^{\theta} \hat{\theta} \, dF(\theta)}{1 - F(\hat{\theta})} \right) \left( \frac{\nu g + (1 - \nu)(1 - e) b}{\nu + (1 - \nu)(1 - e)} \right).
\]

The first term is the expected value of \( \theta \) conditional on doing research. It captures the effect of patent policy on value via self-selection. The second term is the expected profitability of the patent conditional on being granted. It captures the effect on value via patent quality: for a given \( \theta \), an issued patent is more likely to be valid (and thus more valuable) if it has been examined more rigorously. The first term increases with \( \hat{\theta} \) and hence decreases with \( e \).

The second term increases with \( e \): its derivative with respect to \( e \) is \( \nu(1 - \nu)(g - b)/(\nu + (1 - \nu)(1 - e))^2 \geq 0 \). Based on this derivative, we can identify two sufficient conditions for small firms to hold on average less valuable patents than large firms. The first is that \( \nu \) is not too different from 1; the second is that \( g \) is not too different from \( b \). Especially the first condition seems plausible: ambitious research projects tend to yield patentable inventions, though not always valuable ones. Note also that both of these are sufficient and not necessary conditions. In particular, they are derived under the assumption that bad projects by small firms are fully deterred. If instead there are some bad projects in equilibrium, this decreases the average value of patents held by small firms, as bad projects generate lower value.

In summary, wealth-dependent application fees should induce the patent office to examine applications from small firms more rigorously according to Proposition 3. This result leads to two predictions about a patent system in which fees depend on firm size: (1) the grant rate for small firms should be lower than for large firms; (2) under plausible assumptions, patents issued to small firms should be less valuable than those issued to large firms. Although the prediction on patent quality is ambiguous, the model is consistent with observing lower patent quality for large firms relative to small ones.

It is worth pointing out that the prediction on the private value of patents also holds if the patent office does not adjust its examination intensity to firm size, and instead uses a uniform examination intensity regardless of firm size. Note, however, that under a uniform examination intensity, patents issued to large firms should be both more valuable and of weakly higher quality than patents issued to small firms. Thus, unlike the model, a uniform examination intensity is not consistent with observing higher value but lower patent quality for large firms.
**IV(ii) Empirical evidence**

**Cross-sectional evidence.** Applicants in the U.S. receive differential treatment based on whether they are classified as small or large entities. Since the adoption of Public Law 97-247 in 1982, discussed in more detail below, small entities – individual inventors, nonprofit organizations, and small businesses – pay reduced application, issuance and renewal fees. Thus, a first test of the predictions of the model can be performed by comparing outcomes for small and large entities in the U.S. after 1982.

Frakes and Wasserman [2013] obtain data on USPTO patent processing outcomes, disaggregated by patent class and entity size, through a Freedom of Information Act request. Their sample spans the time period from 1983 to the present. They find evidence of a differential grant rate depending on whether an application is filed by a small or large entity. They argue that their estimates support the notion that the USPTO grants patents at a significantly higher rate to applicants with large-entity status, which is in line with the prediction of the model.

Bessen [2008] uses renewal data to estimate the value of a sample of U.S. patents issued in 1991, controlling for a number of patentee characteristics. Bessen reports that patents owned by patentees with small entity status are substantially less valuable than those owned by large entities. He argues that the difference remains sizeable even when accounting for the selection bias that is likely to be present because small firms tend to sell their most valuable patents to large firms. Similar results are obtained by Arora et al. [2008] using survey estimates.

The problem with cross-sectional evidence is that differences between small and large entities could have a variety of causes. For example, Frakes and Wasserman [2013] argue that the observed difference in grant rates between small and large entities may be due to the fact that the USPTO earns greater revenue from issuance and renewal fees on large-entity patents, and may therefore bias its allowance decisions in favor of large entities. The observed difference in patent value between small and large entities may have a number of alternative explanations as well. There might be a fundamental difference in the distribution of research abilities; small and large firms might have different incentives to patent (e.g., for small firms, patents may be valuable as signals to investors (Greenberg [2012])); and estimates of value based on renewal decisions may be biased downward for small firms due to the presence of wealth constraints. A better test involves adding variation across time.

**Evidence from the 1982 reform introducing differentiated fees.** Effective October 1, 1982, the USPTO began charging differentiated fees to small and large entities, and simultaneously raised fees across the board. In particular, application fees were raised from $65 to $300, issuance fees from $100 to $500, and renewal fees at ages four, eight, and twelve from 15
Small entities were granted a fee reduction of fifty percent; in essence, therefore, renewal fees remained unchanged for them, and while application and issuance fees increased, they did so considerably less than for large entities.

The low level of fees prior to the reform can be interpreted as aiming to ensure that wealth-constrained firms are able to participate in the patent system. According to the model, the low fees should have induced the USPTO to perform rigorous screening prior to the reform, and to relax its rigor somewhat afterward. The reduction in examination intensity should have been more pronounced for large entities, who faced higher application fees than small entities. The predictions of the model on the value and quality of patents thus translate as follows: patents issued to large entities should be more valuable after the reform than before, and may be of lower quality. Patents issued to small entities should be less affected by the reform; they constitute a control group that can be used to account for other factors affecting the value and quality of patents. Hence, the above predictions should be reflected in the difference in differences between large and small entities before and after the reform. Looking at the difference in differences addresses the concerns about the observed cross-sectional differences, for which there might be alternative explanations.

Very much in this vein, de Rassenfosse [2012] exploits the 1982 reform to estimate the effect of an increase in fees on patent value by using large entities as the treatment group and small entities as the control group. The idea is that any changes to the patent system occurring around the date of the reform – such as the creation of the Court of Appeals for the Federal Circuit (CAFC) – are likely to have affected the value of small and large entity patents in similar ways. de Rassenfosse measures patent value by two indicators: citations and family size. He obtains PATSTAT data on U.S. priority patents issued to U.S. owners and filed between January 1981 and June 1984. Excluding the months of September and October of 1982, for which data might be unreliable due to applicants trying to beat out the October 1 deadline for lower fees, de Rassenfosse then estimates a standard difference-in-differences specification. He finds that both his indicators of value increased more for large entities than for small entities, providing support for the relationship predicted by the model.

To complement de Rassenfosse’s analysis, I use data from the USPTO’s maintenance fee events database, which I merge with NBER patent citations data, to compute renewal rates (Schankerman and Pakes [1986]; Schankerman [1998]) and truncation-adjusted citation counts (Trajtenberg [1990]; Hall et al. [2001]) for patents filed around the date of the reform. From the maintenance fee events database, I obtain data on U.S. utility patents filed between December 12, 1980 and August 20, 1984. The start date was dictated by data availability: patents filed before December 12, 1980 were exempt from renewal fees. The end date was chosen to have an equal-sized window on both sides of the date of the reform. I exclude
patents issued after 1992 as well as reissue patents. I also exclude patents recorded as having been renewed at age $t \in \{4, 8, 12\}$ more than once or as having been renewed at age 8 (12) without having been renewed at age 4 (8). This leaves me with 235,065 patents, which I match with the NBER citations database to obtain information on the number of citations received until 2006. There were 9 patents that could not be found in the NBER data, so the final data set consists of 235,056 patents.

For each of these patents, I have information on whether they were renewed at age $t$ and whether the renewal fees paid were small-entity fees, large-entity fees, or undifferentiated fees. I also have information on the current entity status of the patent holder and on changes in the entity status over time. Unfortunately, I do not have information on the entity status at the time of filing. To infer the likely entity status at filing, I use information on status changes and the payment of small-entity or large-entity renewal fees. Details of the methodology I use to infer the entity status at filing are provided in Appendix B.

Inferring the entity status at filing is important because of selection effects. Patents issued to small firms can be acquired by large firms and vice versa. The main issue is that the most valuable patents are probably more likely to change owners than the rest. To see why this is problematic for estimating the difference in differences of patent value, suppose that, for some reason unrelated to the change in application fees (e.g., the creation of the CAFC) the value of both small and large-entity patents increases after 1982. Suppose also that a constant fraction of all patents whose value is above some threshold changes owners, and that this threshold is above the mean of both small and large entity patents. Because of the increase in value after 1982, there will thus be more ownership changes taking place after 1982 than before. Then, using the current entity status to categorize patents whose status changed from small to large will overstate the increase in value for large-entity patents and understate the increase for small-entity patents. Conversely, using the current entity status to categorize patents whose status changed from large to small has the opposite effect. This would be less of a problem if both types of changes occurred equally often, so that they might cancel out. But there is evidence that small firms transfer their patents substantially more often than large firms (Serrano [2010]). Hence, neglecting status changes would likely lead one to overestimate the effect of the increase in application fees on the value of large-entity patents.

A different issue that arises is specific to the use of renewal data and concerns comparing renewal rates across time. Such a comparison may be problematic for two reasons. First, renewal fees vary over time, as the fee schedule undergoes frequent changes. Second, and more importantly, patents filed before August 27, 1982 had to pay undifferentiated renewal fees, which were set at the level of small-entity fees (before the dependence on the application date was abolished in 1991). Consider two large-entity patents granted on the same date, and
of which one was applied for before August 27, 1982 and the other one after. For a renewal occurring before 1991, the latter patent is typically due double the renewal fee of the former.\textsuperscript{16} Not accounting for these differences in renewal fees biases the comparison of renewal rates and is likely to lead one to underestimate the effect of the increase in application fees on the value of large-entity patents.

To address this concern, I use information on the renewal fee schedule over time to infer the amount of renewal fees each patent was due at various ages based on its grant date.\textsuperscript{17} For patents which were renewed, I know whether small or large-entity fees were paid. For patents which were not renewed, I do not know whether they were due small or large-entity fees because I cannot be sure of their entity status at the time of renewal. To impute the amount they were due, I use the inferred entity status at filing for renewal at age 4 and the current entity status for ages 8 and 12.\textsuperscript{18} I deflate the inferred renewal fees using the Consumer Price Index. To filter out the part of the renewal decisions that is explained by renewal fees, I then regress the renewal decision at age $t$ on the deflated renewal fee due. The residual from this regression gives me a measure of the fee-adjusted renewal decision.\textsuperscript{19} The regression is performed separately for small and large entities, using a linear probability model. For renewal at age $t \in \{8, 12\}$, only patents renewed at age $t - 4$ are included.

Table I presents means and differences in means before and after the reform by inferred entity status at filing, as well as differences in differences. The sample used to compute these estimates covers only patents filed at least six months before and six months after the October 1, 1982 reform. This minimizes any bias from inventors rushing their applications to the patent office to avoid paying increased fees. Table II shows the number of patents in each category on which the means in Table I are based.

In line with the cross-sectional evidence presented previously, renewal rates at all ages are higher for large entities than for small entities, both before and after the reform. For large entities, renewal rates at ages 4 and 12 increase after the reform, but the renewal rate at age 8 decreases. For small entities, only the renewal rate at age 4 increases after the reform; renewal rates at ages 8 and 12 decrease. When adjusting for renewal fees, the decrease in renewals at age 8 for large entities is reversed; after the reform, fee-adjusted renewal is higher at all ages. For small entities, the increase in renewal at age 4 is no longer significant; after the reform, fee-adjusted renewal is lower at all ages. Citations (both raw and truncation-adjusted) increase for both small and large entities. Both before and after the reform, small-entity patents are cited slightly more often than large-entity patents.
The observed differences in renewal rates and fee-adjusted renewal rates all support the prediction that the value of large-entity patents increased after the reform relative to the value of the control group of small-entity patents. As expected, adjusting for renewal fees increases the size of the effect. This is especially visible at age 8. At age 12, adjusting for renewal fees has almost no effect for large entities. This should also be expected, as 12-year renewal fees do not depend on the filing date. (The average deflated renewal fee due by large entities at age 12 is almost the same for patents filed before and after reform.)

The observed differences in citations paint a different picture. Although citations increased for both small and large entities, the difference in differences is insignificant; if anything, the increase in citations seems to be more pronounced for small entities. In the light of the model, a tentative explanation for this observation is that citations reflect a mixture of the quality and the commercial value of the patent. If the reform led to an increase in the value of large-entity patents but to a contemporaneous decrease in their quality, as the model predicts if the examination intensity was chosen to achieve full deterrence of bad projects both before and after the reform, then an indicator that picks up a mixture of both quality and value may well fail to exhibit a significant difference. Unfortunately, the data does not allow me to isolate patent quality from value. An interesting avenue for future research would be to investigate how the reform affected patent quality and grant rates.

The reported findings hold up under a variety of changes in the empirical specification. They are robust to different time windows: excluding only applications filed between August 27 and November 5, 1982, excluding all applications filed in 1982, and extending the sample up to 1986 does not change the qualitative results. The results are also robust to the use of a different definition of inferred entity status at filing (only reclassifying those patents that have status changes from small to large), to including technology categories in the difference-in-differences regression and the regression of renewal decisions on renewal fees, and to replacing the linear probability model by a probit.

V Comparative statics in the absence of wealth constraints

In this section, I abandon the assumption of wealth constraints to derive some additional results and policy implications for the specification of profit and welfare functions from Subsection IV(i). I start by providing comparative statics for the optimal patent policy. I then discuss implications for specialized patent courts and post-grant opposition.

Proposition 4. Suppose profit and welfare functions are as described in equations (9)-(12) and that the distribution of \( \theta \) has a nondecreasing hazard rate. Letting \( \dot{\theta}^* \equiv \dot{\theta}(e^*) \), we have the following comparative statics results:
\( (i) \) \( \partial e^*/\partial g < 0 \) and \( \partial \hat{\theta}^*/\partial g < 0 \),
\( (ii) \) \( \partial e^*/\partial b > 0 \) and \( \partial \hat{\theta}^*/\partial g > 0 \),
\( (iii) \) \( \partial \hat{\theta}^*/\partial k > 0 \),
\( (iv) \) \( \partial e^*/\partial S > 0 \), \( \partial \hat{\theta}^*/\partial S < 0 \), and \( \partial \phi^*/\partial S < 0 \).

Recall that \( g \) is a parameter that shifts the private returns to innovation without changing the social returns. Similarly, the parameter \( b \) shifts the payoff from a bad patent. According to Proposition 4, an increase in \( g \) leads the patent office to examine applications less rigorously (\( \partial e^*/\partial g < 0 \)). Nevertheless, a higher \( g \) leads to more research as some firms switch from activity \( B \) to activity \( R \) (\( \partial \hat{\theta}^*/\partial g < 0 \)). An increase in \( b \) should result in a tightening of examination by the patent office (\( \partial e^*/\partial b > 0 \)), yet would lead to a decline in research (\( \partial \hat{\theta}^*/\partial b > 0 \)).

Proposition 4 also shows that an increase in the cost of research (\( k \)) is associated with a reduction in the amount of research induced by the optimal patent policy, while an increase in the social value of innovation (\( S \)) leads to more research, tighter examination, and lower application fees.

V(i) Implications for specialized patent courts

The comparative statics for the parameters \( g \) and \( b \) have implications for the introduction of a specialized court for patent disputes. In 1982, the U.S. Congress established the Court of Appeals for the Federal Circuit (CAFC), which replaced the decentralized circuit courts in hearing appeals of patent cases. The European Union is currently considering a similar measure. How does the establishment of a specialized patent court affect the optimal patent policy? This depends on the effect of specialization on \( g \) and \( b \). Formally,
\[
d e^* = \frac{\partial e^*}{\partial g} dg + \frac{\partial e^*}{\partial b} db.
\]

One possibility is that a specialized court is able to appoint judges with greater expertise in the field of specialization. If more expertise on the part of judges reduces mistakes, a specialized court should lead to an increase in \( g \) and a decrease in \( b \) (i.e., \( dg > 0 \) and \( db < 0 \)). Another possibility is that a specialized court reduces judicial uncertainty by making outcomes more predictable. Galasso and Schankerman [2010] estimate the effect of the CAFC on the duration of patent disputes and find that it led to disputes being settled more quickly, a result they attribute to an increase in the predictability of court decisions. Assuming that, at the time of litigation, the patentee knows whether he has a major or a minor invention and that
$g \geq 1/2 \geq b$, a policy change that increases $g$ and decreases $b$ is associated with a decrease in the variance of the judicial outcome for both types of inventions.\textsuperscript{21}

Proposition 4 then implies that $d e^* < 0$. Thus, the introduction of the CAFC should have led the patent office to optimally decrease its examination intensity. Katznelson [2007] shows that the USPTO grant rate rose from 60 percent in the early 1980s to 76 percent in 1998. One possible interpretation of this increase in the grant rate is that it reflects a decline in the rigor of patent examination, as several observers have suggested (see, e.g., Jaffe and Lerner [2004]).\textsuperscript{22} Such an interpretation is consistent with the model.

Another possibility is that a specialized patent court leads to a pro-patent shift, raising both $g$ and $b$ (i.e., $dg > 0$ and $db > 0$), which some commentators claim was the effect of the creation of the CAFC (Quillen [1993]; Jaffe and Lerner [2004]).\textsuperscript{23} In that case, the implications for $e^*$ are generally ambiguous. Suppose that $dg$ is small relative to $db$. Then, a pro-patent CAFC should have led the patent office to examine applications more rigorously, rather than less.

V(ii) Implications for post-grant opposition

We can also use the model to assess the effects of post-grant opposition, an administrative procedure allowing third parties to challenge patents after they are granted. The European Patent Office (EPO) has long had a well-functioning opposition procedure, whereas the USPTO equivalent (called re-examination) has been rarely used (Graham \textit{et al.} [2002]). Strengthening the re-examination procedure was one of the most prominent measures in the U.S. patent reform act of 2011.

Post-grant opposition is likely to reduce the costs of challenging a patent. Private parties may be knowledgeable about prior art that is relevant to the assessment of the novelty and nonobviousness of a claimed invention. Making it easier for these parties to bring such prior art to the attention of the authorities is the primary goal of an opposition procedure. By reducing the costs of challenging patents, however, opposition may also facilitate challenges against genuine inventions. Indeed, critics have expressed concerns that opposition would lead to abusive challenges and could thus reduce the incentive to innovate.

This discussion suggests that we should expect an opposition procedure to reduce both $g$ and $b$: both good and bad patents will be challenged (and revoked) more often. It may be reasonable to think, however, that an effective opposition procedure would decrease $b$ by more than $g$. The following definition formalizes this idea.

\textbf{Definition.} Implementing an effective system of post-grant opposition leads to changes in $g$ and $b$ satisfying $db < dg \leq 0$. 
The model of self-selection into R&D presented in this paper shows that what matters for the incentive to innovate is not the absolute return from R&D but rather the return from R&D relative to the return from bad projects. Therefore, post-grant opposition may well enhance the incentive to innovate even though it reduces the return from R&D. In the absence of patent examination \((e = 0)\), it is clear that \(\hat{\theta}(0) = k/(\nu(g - b))\) decreases following the introduction of an effective opposition procedure, in the sense defined above:

\[
\frac{d\hat{\theta}(0)}{dg} dg + \frac{d\hat{\theta}(0)}{db} db = - \frac{k}{\nu (g - b)} (dg - db) < 0.
\]

Nevertheless, as the following proposition shows, in the presence of patent examination, \(db < dg\) is not sufficient for post-grant opposition to be welfare-enhancing.

**Proposition 5.** An effective system of post-grant opposition improves total welfare if and only if \((1 - e)db \leq dg\).

Proposition 5 says that the decrease in \(g\) brought about by the introduction of opposition must be sufficiently small relative to the decrease in \(b\) to ensure that opposition is beneficial for welfare. Rewriting the condition, the ratio of the changes must satisfy

\[
\frac{|dg|}{|db|} \leq 1 - e.
\]

When \(e\) is low, i.e., the patent office does a poor job weeding out bad applications, effective opposition is likely to be welfare-enhancing. Part of the intuition for this result comes from how the patent office adapts its policy following the introduction of post-grant opposition. We know from Proposition 4 that the patent office tends to cushion the effect of a reduction in \(g\) on the incentive to innovate by increasing \(e\); similarly, it reacts to a reduction in \(b\) by decreasing \(e\) to save on examination costs.

Proposition 5 also provides grounds to be cautious about post-grant opposition, however. When \(e\) is large, i.e., the patent office does a good job of eliminating bad applications, even an effective opposition procedure may reduce welfare. The reason is that a reduction in \(b\) has a smaller impact on firms’ incentives to innovate than \(g\). This is easy to see from \(\hat{\theta} = k/(\nu(g - (1 - e)b))\). In the denominator, \(b\) is multiplied by \(1 - e\) because it only matters if the firm escapes detection by the patent office. By contrast, \(g\) matters independently of \(e\). Therefore the reduction in \(g\) can outweigh the reduction in \(b\), in which case the introduction of an opposition procedure results in lower innovation and welfare.

**VI Patent policy and firms’ financing problem**

An obvious objection to the analysis of wealth constraints in Section IV is that in the presence of well-functioning financial markets, firms’ wealth should not matter. We know that credit
markets do not work perfectly, however, and this is particularly salient in the financing of innovation. This section explicitly models the firms’ financing problem. I build on Section IV by introducing a financing stage for firms whose assets are insufficient to fund the R&D investment. This allows me to study how patent policy affects firms’ access to financing.

It is well known from the corporate finance literature that an entrepreneur needs to have sufficient pledgeable income to obtain a loan from outside investors (Holmstrom and Tirole [1997]). The reason that is generally invoked is moral hazard: after obtaining the loan, the entrepreneur may have to choose between projects that differ in their probability of success and in the private benefits they procure the entrepreneur. In order for the entrepreneur to choose the efficient project, rather than the one maximizing his private benefits, he has to be given a sufficient share of the profits.

The model of self-selection into R&D from Section II fits this description: firms – which we can interpret as being run by entrepreneurs in this context – must choose between two kinds of projects, one of which (B) has a lower probability of success (because it is more likely to be rejected by the patent office or revoked by the courts) but requires a lower investment and may thus be associated with larger private benefits. In what follows, I will use the terms “firm” and “entrepreneur” interchangeably.

In the basic model, firms with productivity close to \( \hat{\theta} \) make very little profit, as \( \phi^* = \pi^R(\hat{\theta}, e^*) \). As the preceding discussion suggests, in the presence of a financing problem, these firms may not be able to secure a loan. To induce them to invest in research, the patent office must leave them with sufficient “rent” by setting \( \phi < \pi^R(\hat{\theta}, e^*) \). I will now formalize this idea.

Assume that wealth constrained firms have assets \( a_L \) satisfying \( 0 \leq a_L < k \). Thus, in order to finance the R&D investment required for activity R, firms need outside financing. By contrast, activity B requires outside financing only if the application fee exceeds agents’ assets, \( \phi > a_L \). In order to make a profit from either kind of project, firms need to obtain a patent. The filing of an application to the patent office is publicly observable.

There is a competitive financial market where investors can provide funding to firms. The investors have an alternative investment (outside option) whose rate of return is normalized to zero. The outcome of a project is uncertain: a research project succeeds with probability \( p \), while a bad project succeeds with probability \( pb \), where \( 0 < b \leq 1 \). In case of success, the project of entrepreneur \( \theta \) – be it a research project (R) or a bad project (B) – yields \( \theta/p \). In case of failure, the project yields zero. For simplicity, I assume that investors observe a firm’s productivity \( \theta \) but not its choice of project.\(^{25}\)

If an entrepreneur obtains a loan of size \( \sigma \) but does not spend its entirety on R&D and application fees, he can divert the remainder to other uses, which for simplicity is assumed
to procure him a private benefit equal to the amount of money that is diverted. This private benefit is only realized if the investors do not shut down the project before its outcome is realized.

The timing is as follows:

1. The entrepreneur makes a take-it-or-leave-it offer to investors, specifying a loan of size $\sigma$ and a repayment in case of success $\rho$. Investors accept or reject.

2. The entrepreneur chooses $R$ (in which case he invests $k$) or $B$. He then applies for a patent and pays the application fee $\phi$. The patent office examines applications with intensity $e$ and grants or rejects. Research projects are never rejected while bad projects are rejected with probability $1 - e$. If an application is rejected the entrepreneur obtains zero.

3. The project succeeds or fails. In case of success, the entrepreneur pays back $\rho$.

I solve the model backwards starting from the moral-hazard stage.

**Stage 2: choice of activity.** An entrepreneur who has obtained a loan of size $\sigma < k + \phi - a_L$ cannot choose $R$ and therefore decides between $B$ and $I$. $B$ is more profitable if and only if

$$ (1 - e)pb(\theta/p - \rho) - \phi + \sigma \geq 0. $$

Because patent applications are observable, the investors can shut down the project if the firm does not apply for a patent, so the right-hand side is zero.

An entrepreneur who has obtained a loan of size $\sigma \geq k + \phi - a_L$ chooses among $R$, $B$, and $I$. He prefers $R$ to $B$ if and only if

$$ p(\theta/p - \rho) - k - \phi + \sigma \geq (1 - e)pb(\theta/p - \rho) - \phi + \sigma. $$

While a research project $R$ has a higher probability of success, it also requires a larger investment ($k$). Expression (18) can be rewritten as

$$ \rho \leq \frac{\theta}{p} - \frac{k}{p(1 - (1 - e)b)}. $$

The entrepreneur prefers $R$ to $I$ if and only if $p(\theta/p - \rho) - k - \phi + \sigma \geq 0$.

**Stage 1: loan proposal.** The entrepreneur will propose the loan contract $(\sigma, \rho)$ that allows investors to just break even. An entrepreneur who anticipates choosing $R$ will propose a contract such that $\sigma = k + \phi - a_L = pp$. Investors only accept this contract if it indeed induces the entrepreneur to choose $R$. If the contract were to induce $B$, the probability of
repayment would be too low for investors to break even. This places an upper bound on \( \rho \).

Replacing \( \rho = (k + \phi - a_L)/\rho \) in (18), we obtain

\[
(20) \quad k + \phi - a_L \leq \theta - \frac{k}{1 - (1 - e)b} \iff \theta \geq \frac{k(2 - (1 - e)b)}{1 - (1 - e)b} + \phi - a_L \equiv \tilde{\theta}.
\]

If this condition is not met, even the lowest possible reimbursement (such that investors just break even in expectation) cannot ensure that the entrepreneur chooses \( R \). According to this definition, \( \tilde{\theta} \) is the threshold above which firms have the ability to innovate, i.e., they can obtain financing to do research.

An entrepreneur who anticipates choosing \( B \) and requires a loan (i.e., if \( a_L < \phi \)) will propose \( \sigma = \phi - a_L = (1 - e)pb\rho \). By proposing \( \sigma < k + \phi - a_L \), the entrepreneur effectively commits to choosing \( B \). Because patent applications are observable, there is no moral hazard, so all such loan contracts are accepted. If \( a_L \geq \phi \), the entrepreneur does not require a loan. In both cases, (17) implies that it is profitable to choose \( B \) if and only if

\[
(21) \quad (1 - e)b\theta - \phi \geq 0 \iff \theta \geq \theta = \frac{\phi}{(1 - e)b}.
\]

We are now ready to describe the effect of patent policy on the behavior of wealth-constrained firms. Given a patent policy \( (e, \phi) \), firm \( \theta \) prefers activity \( R \) to activity \( B \) if

\[
p\theta/p - k - \phi \geq (1 - e)b\phi/p - \phi \iff \theta \geq \tilde{\theta} = \frac{k}{1 - (1 - e)b}.
\]

The firm’s moral-hazard problem, however, implies that it can only obtain financing if \( \theta \geq \tilde{\theta} \). We have \( \tilde{\theta} > \hat{\theta} \) if and only if

\[
\frac{k(2 - (1 - e)b)}{1 - (1 - e)b} + \phi - a_L > \frac{k}{1 - (1 - e)b} \iff a_L < k.
\]

Thus, if small firms’ wealth is lower than the investment required for R&D, then regardless of the patent policy \( (e, \phi) \) there exists a nonempty set of firms that would find it profitable to do research but cannot obtain financing.

The threshold \( \tilde{\theta} \), above which firms have both the incentive and the ability to innovate, depends on \( e \) and \( \phi \) as follows:

\[
\frac{\partial \tilde{\theta}}{\partial e} = -\frac{bk}{(1 - (1 - e)b)^2} < 0
\]

\[
\frac{\partial \tilde{\theta}}{\partial e} = 1 > 0.
\]

Unlike in the model without financing, the threshold above which firms innovate also depends on \( \phi \); \( \tilde{\theta} = \tilde{\theta}(e, \phi) \) decreases in \( e \) and increases in \( \phi \).
Patent policy thus has an effect on firms’ ability to obtain funding for research. Raising the examination intensity improves a given firm’s chance to get funding. The intuition is that more rigorous examination relaxes the entrepreneur’s incentive-compatibility constraint (18): it decreases the success probability of activity $B$, making deviations less attractive for the entrepreneur. This increases the amount of income that can be pledged to investors and hence encourages investors to provide the entrepreneur with funds for research.

As previously, the threshold $\theta$ above which firms prefer activity $B$ to inactivity increases in both instruments, $\partial \theta / \partial e = \phi / (b(1 - e)^2) > 0$ and $\partial \theta / \partial \phi = 1 / ((1 - e)b) > 0$. Again, the model generates self-selection of firms according to their productivity. Firms with $\theta < \tilde{\theta}$ remain inactive, firms with $\theta \leq \tilde{\theta} < \tilde{\theta}$ choose $B$, and firms with $\theta \geq \tilde{\theta}$ choose $R$.

An implication of this analysis is that if the patent office implements the patent policy that is optimal in the basic model (where $\tilde{\theta} = \hat{\theta}$), wealth-constrained entrepreneurs with $\theta \in [\tilde{\theta}, \tilde{\theta})$ are credit rationed: they cannot obtain a loan covering $k + \phi$. As a result they are forced to choose bad projects. Although they would like to invest in research projects, they cannot obtain the required funding. By contrast, they can always obtain a loan for a bad project whenever it is profitable to do so. Since $\tilde{\theta} = \hat{\theta}$, activity $B$ is profitable for all entrepreneurs that are credit rationed. This suggests that the patent office has to adapt its patent policy to account for firms’ financing problems either by lowering fees or raising its examination intensity, or both. The following proposition makes this claim precise.

**Proposition 6.** In the presence of financing problems, achieving a given level of innovation $\tilde{\theta} \geq 2k - a_L$ requires a lower application fee and greater examination intensity than in the basic model. In choosing the application fee, the patent office trades off research projects against bad projects: $\partial (\tilde{\theta} - \theta) / \partial \phi < 0$.

The proposition shows that when firms are wealth-constrained, inducing a given level of innovation is costlier than absent wealth constraints: the patent office has to set a lower fee and higher examination intensity. Unlike in the basic model, where raising the application fee up to $\pi^R(\hat{\theta}, e)$ has no effect on innovation, here the patent office faces a tradeoff. On the one hand, raising $\phi$ reduces the number of firms choosing bad projects (firms with $\theta \in [\theta, \hat{\theta})$). On the other hand, it pushes up the threshold above which firms choose research projects, thus diminishing innovation. To keep the level of innovation constant, an increase in $\phi$ must be accompanied by an increase in $e$.

Note that it is still possible, and may be desirable, for the patent office to fully deter bad projects. In the presence of financing problems, this requires choosing $e$ and $\phi$ such that $\tilde{\theta}(e, \phi) = \hat{\theta}(e, \phi)$. Since $\tilde{\theta} > \hat{\theta}$ for any $\phi$, achieving full deterrence necessitates a larger $e$ and is thus costlier than in the basic model. Depending on parameters, it will sometimes be
preferable to allow some bad projects and save on examination costs.

VII Conclusion

I present a model of patent examination in which firms self-select into R&D depending on their productivity. Specifically, firms choose between more or less ambitious research projects and then apply for patent protection. Ambitious research projects are socially desirable and often yield patentable inventions. By contrast, unambitious projects are socially harmful and never result in patentable inventions. The patent office, charged with the verification of patentability, must separate the wheat from the chaff. It has two instruments at its disposal: the intensity with which it examines applications, and an application fee. Higher examination intensity allows the patent office to weed out more bad applications (an ex post effect), but the self-selection feature of the model implies that it also affects the set of firms that choose ambitious research projects (an ex ante effect). Both the ex post and ex ante welfare effects need to be taken into account in determining the optimal examination intensity.

I then introduce a subset of small firms that are wealth constrained, so that they cannot afford the application fee that would be optimal in the absence of wealth constraints. I analyze how wealth constraints change the optimal policy depending on whether small firms have access to external financing. I show that when firms do not have access to financing and the patent office can identify wealth-constrained firms, it should implement a differentiated patent policy whereby applications submitted by small firms pay a lower fee but are examined more rigorously. Based on specific profit and welfare functions, I derive predictions on differences in grant rates, patent quality, and the private value of patents between small and large firms. These predictions are in line with the available empirical evidence. To complement the existing evidence, I use data on renewals and citations to assess the effect of the 1982 reform introducing differentiated application fees in the U.S. My empirical findings support the predictions of the model.

To analyze the case where small firms have access to financial markets, I extend the model by adding a financing stage. Firms’ choice between projects then leads to a moral hazard problem. I show how patent policy needs to be adjusted to relax the incentive constraints this creates. To attain a given level of innovation, the patent office needs to tighten examination or reduce fees, compared to the case without financing problems. Finally, I provide comparative statics for the optimal policy absent wealth constraints and derive implications for the creation of a specialized patent court and for post-grant opposition.
Notes

For example, RIM (Research In Motion), the maker of BlackBerry mobile devices, was sued by patent-holding company NTP, and settled out of court for a reported $612.5 million, even though on re-examination the U.S. Patent and Trademark Office (USPTO) revoked all of the patents NTP had asserted against RIM. See Time Magazine, “Patently Absurd”, April 2, 2006, available online at http://www.time.com/time/magazine/article/0,9171,1179349,00.html.

See, e.g., Jaffe and Lerner [2004], Farrell and Shapiro [2008] and Bessen and Meurer [2008].

In practice, the possibility of challenging a patent in court mitigates this problem, but does not eliminate it if the court decision is uncertain. There may also be too little challenging of questionable patents because of the public good nature of these challenges (Chiou [2006]; Farrell and Shapiro [2008]).

The assumption that \( w(\theta, e) \geq \pi^R(\theta, e) \) is not crucial but simplifies the exposition.

If the profit from activity \( B \) does not depend on a firm’s productivity \( \theta \), i.e., \( \pi^B(\theta, e) = \pi^B(e) \) for all \( \theta \), then
\[
\theta = \begin{cases} 
0 & \text{if } \phi < \pi^B(e) \\
\hat{\theta} & \text{if } \phi = \pi^B(e) 
\end{cases}
\]

If \( \pi^B(\theta, e) = \pi^B(e) \) for all \( \theta \), \( \hat{\theta} \) is discontinuous in \( \phi \) with a jump at \( \phi = \pi^B(e) \); see footnote 5.

The proof of Proposition 1 shows that it is indeed optimal to set \( \phi \leq \pi^R(\hat{\theta}(e), e) \) under the assumption that social value exceeds private value, implying \( w(\hat{\theta}, 0) \geq 0 \), and that \( w_{\theta e} \leq 0 \), implying \( w(\hat{\theta}(e), e) > \gamma(e) \) at the optimal \( e \). The intuition is that deterring ambitious research projects through higher fees cannot be optimal as long as their social value \( w(\theta, e) \) exceeds the cost of examination \( \gamma(e) \).

Note that I am implicitly assuming that the welfare from the optimal policy is greater than the welfare from shutting down small firms completely (zero). This is sure to be the case as long as \( \Delta \) is not too large.

A sufficient condition for this result is that more rigorous examination does not increase examination costs by more than the benefits it generates in terms of weeding out bad patents. As shown in the proof, the assumption that \( -\ell_e(\hat{\theta}(e^*), e^*) > \gamma'(e^*) \), which says that the marginal reduction in the welfare loss from bad patents is greater than the marginal cost of examination at \( e^* \), ensures that intensifying examination is indeed worthwhile, at least for low values of \( \Delta \).

With these functional forms, \( w_\theta = 0 \) because \( S \) and \( D \) are independent of \( \theta \). More generally, \( S \) and \( D \) could be functions of \( \theta \). For \( w_\theta \geq 0 \) and \( w_{\theta e} \leq 0 \), it is then sufficient that \( S'(\theta) \geq 0 \) and \( D'(\theta) \leq 0 \).
Note that while the concepts of quality and value can be clearly separated in the model, this is more difficult empirically.

This is not a fully convincing explanation. Until the America Invents Act of 2011, part of the patent office’s fee revenue was regularly cordoned off by Congress. Thus, it is not clear that the patent office had the incentive to maximize revenue.

See 47 Fed. Reg. 41256 implementing Public Law 97-247. Note that all patents filed after August 27, 1982 paid differentiated renewal fees, but only those filed after October 1, 1982 paid differentiated application fees. Patents filed before August 27, 1982 paid undifferentiated renewal fees until 1991, when this dependence on the application date was abolished.

Note that because data on rejected applications were not publicly available prior to 1999, I cannot estimate the effect of the reform on grant rates.

The maintenance fee events database is available at http://www.google.com/googlebooks/uspto-patents-maintenance-fees.html and the NBER citations data at https://sites.google.com/site/patentdataproject/.

Because no renewal at age 12 could occur before 1991, this concerns renewal at ages 4 and 8.

I am grateful to Jim Bessen, who kindly provided me with the information on the fee schedule.

For age 8, I tried both the status at filing and the current status. I retained the current entity status because it yields the more conservative estimate of the difference in differences.

I thank Mark Schankerman for suggesting this method.

Katznelson [2007] argues that other factors could also account for the increased grant rate. He cites an increase in the number of claims per application and a reduction in the scope of patents.

Henry and Turner [2006] find empirical support for such a pro-patent shift. They document a structural break associated with the creation of the CAFC that reduced the incidence of invalidity decisions.

Compared to the specific model from Section IV(i), here I have normalized \( g \) to 1 and...
assume for simplicity that research always yields patentable inventions (ν = 1).

25This assumption is made for reasons of tractability; it avoids complications arising from signaling.

Appendix A  Proofs

Proof of Proposition 1. Let us first show that the constraint \( \phi \leq \pi^R(\hat{\theta}(e), e) \) must be binding, implying \( \theta = \hat{\theta} \). Let \( \mu \) be the multiplier associated with the constraint. Suppose first \( \pi^B > 0 \), so that \( \hat{\theta} \) is differentiable in \( \phi \). Differentiating (6) with respect to \( \phi \), we have

\[
(22) \quad \frac{\partial \theta}{\partial \phi} f(\hat{\theta}, e) = \frac{\ell(\theta, e)}{\mu} = 0.
\]

Since \( \frac{\partial \theta}{\partial \phi} > 0 \), \( \mu > 0 \), so indeed \( \theta = \hat{\theta} \). Suppose instead \( \pi^B = 0 \), so that \( \theta \) takes the form described in footnote 5. If the planner sets \( \phi = \pi^B(e) \), we again have \( \theta = \hat{\theta} \). If the planner sets \( \phi < \pi^B(e) \), then \( \theta = 0 \). The optimal policy \((e^*, \phi^*)\) cannot entail \( \theta = 0 \) and \( \hat{\theta}(e^*) > 0 \). Suppose it did. Then, the second term in (6) would be

\[
- \int_{\hat{\theta}(0)}^{\hat{\theta}(e^*)} \ell(\theta, e^*)dF(\theta) < 0.
\]

The planner could raise the second term to 0 and raise the third term (i.e., decrease examination costs) without changing the first term by setting \( \phi = \pi^B(\hat{\theta}(e^*), e^*) \), contradicting the optimality of \((e^*, \phi^*)\).

Therefore, \( \phi = \pi^R(\hat{\theta}(e), e) \), which is strictly positive. We obtain (7) by differentiating (6) with respect to \( e \), substituting for \( \mu \) from (22) and using the fact that \( \theta = \hat{\theta} \). Next I show that \( e^* \) is indeed interior. Evaluating the left-hand side of (7) at \( e = 0 \) yields

\[
(23) \quad - \frac{\partial \hat{\theta}}{\partial e} w(\hat{\theta}(0), 0)f(\hat{\theta}(0)) + \int_{\hat{\theta}(0)}^{\hat{\theta}} w(\theta, 0)dF(\theta),
\]

while the right-hand side is zero because \( \gamma(0) = \gamma'(0) = 0 \). Assumption 1 implies \( \hat{\theta}(0) > 0 \) and thus \( f(\hat{\theta}(0)) > 0 \). We know from (2) that \( \partial \hat{\theta}/\partial e < 0 \). By the definition of \( \hat{\theta} \), \( \pi^R(\hat{\theta}(0), 0) = \pi^B(\hat{\theta}(0), 0) > 0 \). Hence, \( w(\hat{\theta}(0), 0) \geq \pi^R(\hat{\theta}(0), 0) > 0 \). Moreover, by assumption \( w_e \geq 0 \). Therefore, expression (23) is strictly positive, implying that welfare is strictly increasing in \( e \) at 0 and thus \( e^* > 0 \).

When evaluated at \( e = 1 \), the left-hand side of (7) is finite while the right-hand side tends to infinity because \( \gamma'(1) = \infty \), implying that welfare is strictly decreasing in \( e \) at 1 and thus \( e^* < 1 \). As \( e^* \) is interior and welfare is continuous, (7) is a necessary condition for optimality.

Finally, I establish that setting \( \phi \leq \pi^R(\hat{\theta}(e), e) \) is optimal provided \( w(\theta, e) \geq \pi^R(\theta, e) \) and \( w_{\theta e} \leq 0 \). By Assumption 1 and the definition of \( \hat{\theta} \), \( \pi^R(\hat{\theta}, 0) = \pi^R(\hat{\theta}, 0) \geq 0 \), implying \( w(\hat{\theta}, 0) \geq 0 \). The first-order condition (7) implies that sign(\( w - \gamma \)) = sign(\( \gamma'(1 - F) - \int w_e dF \)). I now show that this is incompatible with \( w - \gamma \leq 0 \) when \( w_{\theta e} \leq 0 \). Define \( e_0 \) as the value of
Moreover, it must be the case that $\gamma(0) = 0$ and $w_{\varepsilon \varepsilon} - \gamma'' < 0$, $w(\hat{\theta}, \varepsilon) - \gamma(\varepsilon) > 0$ if and only if $\varepsilon < \varepsilon_0$. Moreover, it must be the case that $\varepsilon_0 > \varepsilon_1$ defined by $w(\hat{\theta}, \varepsilon_1) - \gamma'(\varepsilon_1) = 0$. We have

$$
\int_\theta^{\varepsilon_0} w_e(\theta, \varepsilon) dF(\theta) = w_e(\hat{\theta}, \varepsilon)(1 - F(\hat{\theta})) + \int_\theta^{\varepsilon_0} (w_e(\theta, \varepsilon) - w_e(\hat{\theta}, \varepsilon)) dF(\theta),
$$

implying that $\varepsilon_1 > \varepsilon_2$ defined by $f_\theta^{} \! w_e(\theta, \varepsilon_2) dF(\theta)/(1 - F(\hat{\theta})) - \gamma'(\varepsilon_2) = 0$, and $\int_\theta^{\varepsilon_0} w_e(\theta, \varepsilon) dF(\theta) - \gamma'(\varepsilon)(1 - F(\hat{\theta})) < 0$ if and only if $\varepsilon > \varepsilon_2$. Thus, $w - \gamma$ and $\gamma'(1 - F) - \int w_e dF$ being of the same sign requires $\varepsilon_2 < \varepsilon < \varepsilon_0$. But this implies $w(\hat{\theta}(\varepsilon^*), \varepsilon^*) - \gamma(\varepsilon^*) > 0$.

Let $\theta^R$ be defined as the solution to $\pi^R(\theta^R, \varepsilon) = \phi$. For $\phi > \pi^R(\hat{\theta}(\varepsilon), \varepsilon)$, there are some firms for which $R$ is preferred to $B$ but is not profitable, so that $\theta^R$ becomes the relevant threshold. Welfare becomes

$$
\int_{\theta^R(\varepsilon, \phi)} w(\theta, \varepsilon) dF(\theta) - \gamma(\varepsilon)[1 - F(\theta^R(\varepsilon, \phi))].
$$

Differentiating welfare with respect to $\phi$ yields

$$
(24) \quad -\frac{\partial \theta^R}{\partial \phi} f(\theta^R)[w(\theta^R, \varepsilon) - \gamma(\varepsilon)].
$$

By the implicit function theorem, $\partial \theta^R/\partial \phi = 1/\pi^R_\theta > 0$. Combined with the fact that $w(\hat{\theta}(\varepsilon^*), \varepsilon^*) - \gamma(\varepsilon^*) > 0$, this implies that (24) is negative at $\varepsilon^*$. Hence, at the optimal effort, increasing $\phi$ above $\pi^R(\hat{\theta}(\varepsilon), \varepsilon)$ reduces welfare.

**Proof of Proposition 2.** What needs to be shown is that the second-order condition for a maximum holds at any $\varepsilon$ that solves (7), implying that there cannot be a local minimum, which in turn implies that there must be a unique maximum. Dropping the arguments of functions and integral bounds for brevity, the second-order condition is

$$
(25) \quad -(w - \gamma) \left[ \frac{d^2 \theta}{d \varepsilon^2} f + \left( \frac{\partial \theta}{\partial \varepsilon} \right)^2 f' \right] - \left( \frac{\partial \theta}{\partial \varepsilon} \right)^2 w_\theta f + \int w_{\varepsilon \varepsilon} dF - \gamma''(1 - F) + 2(\gamma' - \varepsilon_0) \frac{\partial \hat{\theta}}{\partial \varepsilon} f < 0.
$$

If $\varepsilon$ solves (7), then $-(\partial \theta/\partial \varepsilon)(w - \gamma) = (\gamma'(1 - F) - \int w_e dF)/f$, which we can use to rewrite (25) as

$$
(26) \quad -(w - \gamma) \frac{d^2 \theta}{d \varepsilon^2} f + \frac{\partial \hat{\theta}}{\partial \varepsilon} 2(\gamma' - \varepsilon_0) f^2 + \left[ \gamma'(1 - F) - \int w_e dF \right] f' - \left( \frac{\partial \theta}{\partial \varepsilon} \right)^2 w_\theta f + \int w_{\varepsilon \varepsilon} dF - \gamma''(1 - F) < 0.
$$
We know from (2) that $\partial \hat{\theta}/\partial e < 0$. Moreover,
\[
\frac{d^2 \hat{\theta}}{de^2} = \left( \frac{\pi^R_e - \pi^e_e}{\varpi^R_e - \varpi^e_e} \right) \left[ \frac{\partial^2 \hat{\theta}}{\partial \varpi^R_e \partial \varpi^e_e} (\pi^R_R - \pi^R_{\theta\theta}) + \frac{\varpi^R_e - \pi^R_{\theta e}}{\varpi^R_e - \pi^R_{\theta e}} \right] \geq 0,
\]
where the inequality follows from Assumption 1 and the assumption that $\pi^R_{\theta\theta} \leq \pi^R_{\theta e},$ $\pi^R_{\theta e} \leq \pi^R_{ee},$ and $\pi^R_{ee} \geq \pi^R_{\theta e}$.

We know from the proof of Proposition 1 that $w_{\theta e} \leq 0$ implies $w(\hat{\theta}(e^*), e^*) - \gamma(e^*) > 0$, $\int w_e dF \leq w_e (1 - F)$, and $\gamma' - w_e > 0$. Thus, a sufficient condition for (26) is
\[
-(w - \gamma) \frac{d^2 \hat{\theta}}{de^2} f + \frac{\partial \hat{\theta}}{\partial e} (\gamma' - w_e) \frac{2f^2 + (1 - F)f'}{f} - \left( \frac{\partial \hat{\theta}}{\partial e} \right)^2 w_{\theta e} f + \int w_{ee} dF - \gamma''(1 - F) < 0.
\]
The last three terms are negative by assumption. Owing to the fact that $w - \gamma > 0$, the first term is also negative. Finally, combined with the fact that $\gamma' - w_e > 0$, a nondecreasing hazard rate again suffices for the second term to be negative.

**Proof of Proposition 3.** A necessary condition for the claimed result is that $\bar{e} > e^*$ for $\Delta > 0$. This condition holds because, by Proposition 1, $\bar{e} = e^*$ for $\Delta = 0$ (we have $\hat{\theta}(e^*) = \hat{\theta}(e^*, \phi^*)$) and
\[
\frac{\partial \bar{e}}{\partial \Delta} = \frac{\partial \theta / \partial \phi}{\partial \theta / \partial e - \partial \theta / \partial e} > 0,
\]
where the inequality follows from (2), (4) and (5).

Differentiating (8) with respect to $e$ and rearranging yields
\[
(27) \quad - \frac{\partial \hat{\theta}}{\partial \Delta} w(\hat{\theta}, e) f(\hat{\theta}) + \int_0^{\bar{\theta}} w_e(\theta, e) dF(\theta) - \gamma'(e)[1 - F(\hat{\theta})] + \varpi^R_e \gamma(e) f(\hat{\theta})
\]
\[- \left( \frac{\partial \hat{\theta}}{\partial e} (\ell(\hat{\theta}, e) + \gamma(e)) f(\hat{\theta}) - \frac{\partial \hat{\theta}}{\partial e} (\ell(\hat{\theta}, e) + \gamma(e)) f(\hat{\theta}) \right) - \int_0^{\bar{\theta}} [\ell_e(\theta, e) + \gamma'(e)] dF(\theta).
\]
By the first-order condition determining $e^*$, equation (7),
\[- \frac{\partial \hat{\theta}}{\partial \Delta} w(\hat{\theta}, e^*) f(\hat{\theta}) + \int_0^{\bar{\theta}} w_e(\theta, e^*) dF(\theta) - \gamma'(e^*)[1 - F(\hat{\theta})] + \varpi^R_e \gamma(e^*) f(\hat{\theta}) = 0.
\]
Thus, evaluating (27) at $e^*$ yields
\[
\left( \frac{\partial \hat{\theta}}{\partial \Delta} - \frac{\partial \hat{\theta}}{\partial e} \right) (\ell(\hat{\theta}, e^*) + \gamma(e^*)) f(\hat{\theta}) - \int_0^{\bar{\theta}} [\ell_e(\theta, e^*) + \gamma'(e^*)] dF(\theta).
\]
The first term is strictly positive for any $\Delta > 0$ by (2) and (4). Thus, a sufficient condition for (27) to be positive at $e^*$ is
\[
(28) \quad \int_{\theta(e^*, \phi^* - \Delta)}^{\theta(e^*)} [-\ell_e(\theta, e^*) - \gamma'(e^*)] dF(\theta) \geq 0.
\]
By assumption, \(-\ell_e(\hat{\theta}(e^*), e^*) > \gamma'(e^*)\). There are thus two cases to consider. If \(\ell_{e\theta} \geq 0\), then 
\(-\ell_e(\theta, e^*) > \gamma'(e^*)\) for all \(\theta \leq \hat{\theta}(e^*)\), so (28) is satisfied for all \(\Delta\). If instead \(\ell_{e\theta} < 0\), then there exists \(\Delta^* > 0\) such that (28) is satisfied for all \(\Delta \leq \Delta^*\). This is because \(\theta(e^*, \phi^* - \Delta) \rightarrow \hat{\theta}(e^*)\) as \(\Delta \rightarrow 0\). Hence, a sufficient condition for welfare to be increasing in \(e\) at \(e^*\) is that \(\Delta\) is sufficiently small. It follows that \(e^{**}\) must be strictly greater than \(e^*\) for low values of \(\Delta\).

Since \(\ddot{e} > e^*\) for \(\Delta > 0\), the result holds irrespective of whether \(e^{**}\) is an interior solution or a corner solution. \(\Box\)

**Proof of Proposition 6.** Suppose the patent office wants to implement an innovation threshold of \(\theta'\). For this to be feasible with financing, \(\theta'\) must be larger than \(\hat{\theta}\) evaluated at \(\phi = 0\) and \(e = 1\), i.e., \(\theta' \geq 2k - a_L\). In the basic model, implementing \(\theta'\) requires

\[
\hat{\theta}(e) = \frac{k}{1 - (1 - e)b} \leq \theta' \iff e \geq 1 - \frac{\theta' - k}{b\theta'} \equiv \bar{e}
\]

and \(\phi \leq \theta' - k \equiv \bar{\phi}\). With financing, implementing \(\theta'\) requires

\[
\hat{\theta}(e, 0) = \frac{k(2 - (1 - e)b)}{1 - (1 - e)b} - a_L \leq \theta' \iff e \geq 1 - \frac{\theta' - 2k + a_L}{b(\theta' - k + a_L)} \equiv \tilde{e}
\]

and

\[
\hat{\theta}(1, \phi) = \frac{k(2 - b)}{1 - b} + \phi - a_L \leq \theta' \iff \phi \leq \theta' - k + a_L - \frac{k}{1 - b} \equiv \tilde{\phi}.
\]

Straightforward computations show that \(\ddot{e} > \bar{e}\) if and only if \(k(a_L - k) < 0\) and that \(\ddot{\phi} < \tilde{\phi}\) if and only if \(a_L - k/(1 - b) < 0\), both of which are implied by the assumption that \(a_L < k\).

For the second claim, taking the derivative of

\[
\ddot{\theta} - \hat{\theta} = \frac{k(2 - (1 - e)b)}{1 - (1 - e)b} + \phi - a_L - \frac{\phi}{(1 - e)b}
\]

with respect to \(\phi\) yields \(\partial(\ddot{\theta} - \hat{\theta})/\partial \phi = -(1 - (1 - e)b)/(1 - e)b < 0\), where the inequality follows from \(b \leq 1\). \(\Box\)

**Proof of Proposition 4.** Consider general profit and welfare functions \(\pi^R, \pi^B,\) and \(w\). Let \(e(\hat{\theta})\) denote the implicit function defined by \(\pi^R(\hat{\theta}, e) = \pi^B(\hat{\theta}, e)\) (i.e., \(e(\hat{\theta})\) is the inverse function of \(\hat{\theta}(e)\)). Assume \(w_\theta = 0\) and let \(W(\hat{\theta}, e) = (w(e) - \gamma(e))[1 - F(\hat{\theta})]\). Let \(m\) be a parameter affecting profits or welfare.

**Lemma 1.** Suppose \(\partial \theta/\partial m \neq 0\) and \(\text{sign}(\partial^2 \hat{\theta}/\partial m^2 e) = -\text{sign}(\partial \hat{\theta}/\partial m)\). If one of the following holds, then \(\text{sign}(\partial e^*/\partial m) = \text{sign}(\partial \hat{\theta}/\partial m)\):

(a) \(w_m = 0\),

(b) \(w_{em} = 0\) and \(\text{sign}(w_m) = \text{sign}(\partial \hat{\theta}/\partial m)\).
\[(c) \text{sign}(w_m) = \text{sign}(w_{em}) = \text{sign}(\partial \hat{\theta}/\partial m)\].

Suppose instead \(\partial \hat{\theta}/\partial m = 0\). If \(w_{em} = 0\) or sign\(w_{em}\) = sign\(w_m\), then \(\text{sign}(\partial e^*/\partial m) = \text{sign}(w_m)\).

**Proof.** Let \(\hat{W} \equiv W(\hat{\theta}(e), e)\). By Proposition 1, \(e^* = \arg \max_e \hat{W}\) is an interior solution, implying that at \(e^*, \partial \hat{W}/\partial e = 0\) and \(\partial^2 \hat{W}/\partial e^2 < 0\). Thus,

\[
\frac{de^*}{dm} = -\frac{\partial^2 \hat{W}/\partial e \partial m}{\partial^2 \hat{W}/\partial e^2}
\]

has the sign of \(\frac{\partial^2 \hat{W}}{\partial e \partial m}\). We have

\[
\frac{\partial^2 \hat{W}}{\partial e \partial m} = \frac{\partial^2 \hat{\theta}}{\partial m \partial e} \frac{\partial W}{\partial \hat{\theta}} + \frac{\partial \hat{\theta}}{\partial m} \left[ \frac{\partial \hat{\theta}}{\partial e} \frac{\partial^2 W}{\partial \hat{\theta} \partial e} + \frac{\partial^2 W}{\partial e \partial \hat{\theta} \partial m} + \frac{\partial^2 W}{\partial e \partial \hat{\theta} \partial m} \right]
\]

\[
= -\frac{\partial^2 \hat{\theta}}{\partial e \partial m}(w - \gamma)f + \frac{\partial \hat{\theta}}{\partial m} \left[ -\frac{\partial \hat{\theta}}{\partial e}(w - \gamma)f + (\gamma' - w_e)f \right] - \frac{\partial \hat{\theta}}{\partial e} w_m f + (1 - F)w_{em}
\]

\[
= -\frac{\partial^2 \hat{\theta}}{\partial e \partial m}(w - \gamma)f + \frac{\partial \hat{\theta}}{\partial m} (\gamma' - w_e)(1 - F)f' + f^2 - \frac{\partial \hat{\theta}}{\partial e} w_m f + (1 - F)w_{em}.
\]

By the assumption on the hazard rate, the fraction is nonnegative. We know from the proof of Proposition 2 that \(w_{\theta} = 0\) implies that \(w - \gamma\) and \(\gamma' - w_e\) are positive. Recalling that \(\partial \hat{\theta}/\partial e < 0\), the result is immediate. \(\square\)

**Lemma 2.** Suppose \(\partial e/\partial m \neq 0\) and either \(\partial^2 e/\partial m \partial \hat{\theta} = 0\) or sign\(\partial^2 e/\partial m \partial \hat{\theta}\) = -sign\(\partial e/\partial m\).

If one of the following holds, then \(\text{sign}(\partial \hat{\theta}^*/\partial m) = \text{sign}(\partial e/\partial m)\):

(a) \(w_m = 0\),

(b) \(w_{em} = 0\) and sign\(w_m\) = -sign\(\partial e/\partial m\),

(c) sign\(w_m\) = sign\(w_{em}\) = -sign\(\partial e/\partial m\).

Suppose instead \(\partial e/\partial m = 0\). If \(w_{em} = 0\) or sign\(w_{em}\) = sign\(w_m\), then \(\text{sign}(\partial \hat{\theta}^*/\partial m) = -\text{sign}(w_m)\).

**Proof.** Let \(\hat{W} \equiv W(\hat{\theta}(e), e)\). The optimal threshold \(\hat{\theta}^*\) can be defined as \(\hat{\theta}^* = \hat{\theta}(e^*)\) or, alternatively, as \(\hat{\theta}^* = \arg \max_{\hat{\theta}} \hat{W}\), which must satisfy \(\partial \hat{W}/\partial \hat{\theta} = 0\) and \(\partial^2 \hat{W}/\partial \hat{\theta}^2 < 0\). Thus,

\[
\frac{d\hat{\theta}^*}{dm} = -\frac{\partial^2 \hat{W}/\partial \theta \partial m}{\partial^2 \hat{W}/\partial \theta^2}
\]
has the sign of $\partial^2 W / \partial \hat{\theta} \partial m$. We have

$$
\frac{\partial^2 \hat{W}}{\partial \hat{\theta} \partial m} = \frac{\partial^2 e}{\partial \hat{\theta} \partial m} \frac{\partial W}{\partial e} + \frac{\partial e}{\partial m} \left[ \frac{\partial e}{\partial \hat{\theta}} \frac{\partial^2 W}{\partial \hat{\theta} \partial e} + \frac{\partial^2 W}{\partial \hat{\theta} \partial e} \right] + \frac{\partial^2 W}{\partial \hat{\theta} \partial e} + \frac{\partial e}{\partial \hat{\theta}} \frac{\partial^2 W}{\partial e \partial m} \\
= -\frac{\partial^2 e}{\partial \hat{\theta} \partial m} (\gamma' - w_e)(1 - F) + \frac{\partial e}{\partial m} \left[ \frac{\partial e}{\partial \hat{\theta}} (w_e - \gamma'')(1 - F) + (\gamma' - w_e)f \right] \\
- w_m f + \frac{\partial e}{\partial \hat{\theta}} (1 - F) w_{em}.
$$

By the inverse function theorem, $\partial e / \partial \hat{\theta} = 1 / (\partial \hat{\theta} / \partial e) < 0$. We know from the proof of Proposition 2 that $w_{\hat{\theta}} = 0$ implies $\gamma' - w_e > 0$. Moreover, $w_e e - \gamma'' < 0$ by assumption. Combining these yields the result.

I now apply Lemmata 1 and 2 to the specific functional forms set out in (9) through (12). Note that

$$
\hat{\theta}(e) = \frac{k}{\nu(g - (1 - e)b)} \quad e(\hat{\theta}) = \frac{k}{\nu b \hat{\theta} - g - b}.
$$

Claim (i). We have $\partial \hat{\theta} / \partial g = -k / (\nu(g - (1 - e)b)^2) < 0$, $\partial^2 \hat{\theta} / \partial g \partial e = 2bk / (\nu(g - (1 - e)b)^3) > 0$, $\partial e / \partial g = -1 / b < 0$, $\partial^2 e / \partial g \partial \hat{\theta} = 0$, and $w_g = 0$. By Lemma 1, $\partial e / \partial g < 0$. By Lemma 2, $\partial \hat{\theta}^* / \partial g < 0$.

Claim (ii). We have $\partial \hat{\theta} / \partial b = (1 - e)k / (\nu(g - (1 - e)b)^2) > 0$, $\partial^2 \hat{\theta} / \partial b \partial e = -k(g + (1 - e)b) / (\nu(g - (1 - e)b)^3) < 0$, $\partial e / \partial b = -(k - \nu g \hat{\theta}) / (\nu b^2 \hat{\theta}) > 0$, $\partial^2 e / \partial b \partial \hat{\theta} = k / (\nu b^2 \hat{\theta}^2) > 0$, and $w_b = 0$. By Lemma 1, $\partial e^* / \partial b > 0$. By Lemma 2, $\partial \hat{\theta}^* / \partial b > 0$.

Claim (iii). We have $\partial e / \partial k = 1 / (\nu b \hat{\theta}) > 0$, $\partial^2 e / \partial k \partial \hat{\theta} = -1 / (\nu b^2 \hat{\theta}^2) < 0$, $w_k = -1$, and $w_{ek} = 0$. By Lemma 2, $\partial \hat{\theta}^* / \partial k > 0$.

Claim (iv). We have $\partial \hat{\theta} / \partial S = \partial e / \partial S = 0$, $w_S = v > 0$, and $w_{eS} = 0$. By Lemma 1, $\partial e^* / \partial S > 0$. By Lemma 2, $\partial \hat{\theta}^* / \partial S < 0$. Finally, recalling from Proposition 1 that $\phi^* = \pi_B(\hat{\theta}^*, e^*) = (1 - e^*)b \hat{\theta}^*$, the previous results imply

$$
\frac{\partial \phi^*}{\partial S} = b \left[ (1 - e^*) \frac{\partial \hat{\theta}^*}{\partial S} - \hat{\theta}^* \frac{\partial e^*}{\partial S} \right] < 0.
$$

Proof of Proposition 5. Total welfare under an optimal patent policy is

$$
W^* \equiv \max_e (w(e) - \gamma(e))[1 - F(\hat{\theta}(e))].
$$
By the envelope theorem,

\[ \frac{\partial W^*}{\partial g} = - \frac{\partial \hat{\theta}}{\partial g} (w(e) - \gamma(e)) f(\hat{\theta}) \]

\[ \frac{\partial W^*}{\partial b} = - \frac{\partial \hat{\theta}}{\partial b} (w(e) - \gamma(e)) f(\hat{\theta}). \]

Therefore,

\[ dW^* = \frac{\partial W^*}{\partial g} dg + \frac{\partial W^*}{\partial b} db = -(w(e) - \gamma(e)) f(\hat{\theta}) \left( \frac{\partial \hat{\theta}}{\partial g} dg + \frac{\partial \hat{\theta}}{\partial b} db \right) \]

(29)

\[ = -(w(e) - \gamma(e)) f(\hat{\theta}) k(db(1-e) - dg) \]

\[ \nu(g - (1-e)b)^2. \]

Hence, a necessary and sufficient condition for \( dW^* \geq 0 \) is \( db(1-e) \leq dg \). □

**Appendix B  Inferring entity status at filing**

For each patent filed on or after December 12, 1980, the USPTO’s maintenance fee events database contains information on the patent holder’s entity status as currently on record with the office. In addition, it contains information on changes in entity status over time and about the payment of renewal fees at various ages. To infer the likely entity status at filing, I use the following methodology.

In the absence of recorded status changes, I set the patent’s entity status at filing equal to the current entity status, with two exceptions:

- if the current entity status is large but the patent is recorded as having paid small-entity renewal fees the first time that differentiated fees were paid, I classify the patent as filed by a small entity;
- if the current entity status is small but the patent is recorded as having paid large-entity renewal fees the first time that differentiated fees were paid, I classify the patent as filed by a large entity.

In both cases, it is likely that a status change occurred but was not recorded in the database. Note that I can distinguish the differentiated (small-entity/large-entity) renewal fees required for patents filed after August 27, 1982 from the undifferentiated fees required for patents filed before August 27, 1982 (prior to 1991).

For patents that have at least one change in entity status recorded in the database, I use the first status change to infer the entity status at filing. If the first status change recorded is from small to large, I classify the patent as filed by a small entity. It is likely that the vast majority of these patents were either filed by small businesses and subsequently acquired
by large firms, or that the small businesses that filed them grew into large ones, thus losing small-entity status. If the first status change recorded is from large to small, I classify the patent as filed by a large entity, except in the following cases:

- If a patent is filed before August 27, 1982 (or is recorded as having paid undifferentiated fees initially) and the first status change occurs before age 12 and after October 9, 1991, I classify the patent as filed by a small entity. Until December 12, 1991, patents filed before August 27, 1982 had to pay undifferentiated renewal fees. This distinction in the renewal fee due was eliminated by a bill introduced on October 9, 1991 (H.R. 3531); see 56 Fed. Reg. 65142. The idea here is the following. By default the patent office classifies patent holders as large entities. Small entities having filed their patents before August 27, 1982 had no reason to notify the patent office of their small-entity status because they did not have to pay the increased fees that large entities were due based on their filing date. When the bill that abolished the distinction based on the filing date was introduced, those of them who still had to pay renewal fees at least once probably notified the patent office of their small-entity status, so that the database records a status change from large to small.

- If the first status change occurs after age 4, but the patent is recorded as having paid small-entity renewal fees at age 4, I classify the patent as filed by a small entity.

Note that this methodology errs on the side of misclassifying large-entity patents as small-entity patents. As explained in the text, this should lead to more conservative estimates of the difference-in-differences in patent value.

References


Table I: Means of renewals and citations by entity status before and after '82 reform

<table>
<thead>
<tr>
<th>Entity</th>
<th>Renewal, age 4</th>
<th>Fee-adj. R4</th>
<th>Renewal, age 8</th>
<th>Fee-adj. R8</th>
<th>Renewal, age 12</th>
<th>Fee-adj. R12</th>
<th>Citations</th>
<th>Adj. citations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
<td>After</td>
<td>Diff.</td>
<td>Diff-in-diff.</td>
<td>Before</td>
<td>After</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td>0.848</td>
<td>0.882</td>
<td>0.034*** (0.002)</td>
<td>0.021*** (0.004)</td>
<td>10.088</td>
<td>10.783</td>
<td>0.695*** (0.082)</td>
<td>10.598</td>
</tr>
<tr>
<td>Small</td>
<td>0.734</td>
<td>0.747</td>
<td>0.013** (0.004)</td>
<td>0.022*** (0.004)</td>
<td>0.000</td>
<td>-0.001</td>
<td>0.017*** (0.003)</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.021*** (0.002)</td>
<td>0.022*** (0.004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td>-0.012</td>
<td>0.008</td>
<td>-0.001</td>
<td>-0.001</td>
<td>0.063</td>
<td>-0.032</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>0.739</td>
<td>0.700</td>
<td>0.039*** (0.003)</td>
<td>0.027*** (0.006)</td>
<td>0.063</td>
<td>-0.032</td>
<td>0.017*** (0.003)</td>
<td>0.112*** (0.006)</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>-0.001</td>
<td>0.021*** (0.002)</td>
<td>0.022*** (0.004)</td>
<td>0.063</td>
<td>-0.032</td>
<td>0.017*** (0.003)</td>
<td>0.112*** (0.006)</td>
</tr>
<tr>
<td>Large</td>
<td>0.595</td>
<td>0.659</td>
<td>0.065*** (0.003)</td>
<td>0.087*** (0.008)</td>
<td>10.172</td>
<td>10.913</td>
<td>0.741*** (0.148)</td>
<td>10.725</td>
</tr>
<tr>
<td>Small</td>
<td>0.633</td>
<td>0.610</td>
<td>0.022** (0.007)</td>
<td>0.101*** (0.008)</td>
<td>10.172</td>
<td>10.913</td>
<td>0.741*** (0.148)</td>
<td>10.725</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.022** (0.007)</td>
<td>0.101*** (0.008)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td>-0.032</td>
<td>0.032</td>
<td>0.064*** (0.003)</td>
<td>0.046 (0.169)</td>
<td>10.598</td>
<td>11.666</td>
<td>1.068*** (0.088)</td>
<td>10.598</td>
</tr>
<tr>
<td>Small</td>
<td>0.026</td>
<td>-0.011</td>
<td>0.037*** (0.007)</td>
<td></td>
<td>10.598</td>
<td>11.666</td>
<td>1.068*** (0.088)</td>
<td>10.598</td>
</tr>
<tr>
<td></td>
<td>1.068*** (0.088)</td>
<td></td>
<td></td>
<td></td>
<td>1.091*** (0.159)</td>
<td></td>
<td>-0.023 (0.181)</td>
<td></td>
</tr>
</tbody>
</table>

Standard errors in parentheses. * p < 0.05, ** p < 0.01, *** p < 0.001

Entity: inferred entity status at filing. Before: patent filed before April 1, 1982. After: patent filed after April 1, 1983. Renewal, age 4/8/12: renewal decision at age \( t \in \{ 4, 8, 12 \} \) conditional on renewal at age \( t - 4 \). Fee-adj. R4/R8/R12: residual from regression of renewal decision at age 4/8/12 on renewal fee due (linear probability model, separately estimated for small and large entities). Citations: number of citations received until 2006. Adj. citations: citations adjusted for truncation using Hall-Jaffe-Trajtenberg correction (Hall et al. [2001]).
Table II: Number of patents in the sample by entity status and filing date

<table>
<thead>
<tr>
<th>Entity</th>
<th>Before</th>
<th>After</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large</td>
<td>65,742</td>
<td>63,286</td>
<td>129,028</td>
</tr>
<tr>
<td>Small</td>
<td>17,018</td>
<td>25,246</td>
<td>42,264</td>
</tr>
<tr>
<td>Total</td>
<td>82,760</td>
<td>88,532</td>
<td>171,292</td>
</tr>
</tbody>
</table>

42
Figure 1: Self-selection of firms according to productivity $\theta$