Patent Grant Rates and Applicant Size*

Seung-Hyun Hong† Jorge Lemus‡ Joshua Shea§

November 2, 2026

Abstract

Why is the U.S. patent grant rate about 20 percentage points lower for small firms than for large firms? Conditional on observables, first-round rejection rates are similar, but small firms abandon applications more often and are less likely to obtain patents after amendments. These patterns arise from different channels interacting during examination. To quantify their contributions, we develop a dynamic model of patent prosecution. We estimate it using textual distances between applications and prior art, together with the references examiners cite in rejections. Prosecuting small-firm applications as if they had been filed by large firms, holding other characteristics fixed, raises their grant rate and closes 64 percent of the patent-grant gap. The remaining 36 percent reflects inventions and other differences between small and large firms. The sources of the patent-grant gap vary across technology fields, suggesting that policies aiming to raise small firms' grant rates should target different barriers in each field.

Keywords: Patent prosecution, small firms, structural estimation, text analysis.

JEL classification: L25, L53, O34, O38

*We thank workshop and conference participants at UIUC, UNSW, National Taiwan University, Northwestern University, Indiana University, Stanford SITE (2026), and EPIP (2026).

†University of Illinois at Urbana-Champaign, hyunhong@illinois.edu

‡University of Illinois at Urbana-Champaign, jalemus@illinois.edu

§University of Illinois at Urbana-Champaign, jkcshea@illinois.edu

1 Introduction

Small firms obtain patents at a lower rate than large firms even though patents are central to their innovation strategies (see, e.g., Galasso and Schankerman, 2018; Farre-Mensa et al., 2020).[1] Among utility patent applications filed in the United States in 2014, for instance, the grant rate—the ratio of granted patents to applications—was 58 percent for small firms, compared with 77 percent for large firms. This *patent-grant gap*, which persists across years, is concerning if it reflects factors unrelated to the merit of an invention. Such factors may lead small firms to abandon valuable inventions that large firms would have pursued, because inadequate protection makes ideas difficult to transfer or license (Arrow, 1962). Existing policy promotes small-firm patenting through fee discounts, but fees are not the only barrier.[2] The appropriate policy response depends on the roles of examiners, application quality, and applicants' prosecution capabilities in explaining the grant gap.

Our contribution is to disentangle the channels driving the patent-grant gap by developing and estimating a structural model of patent prosecution, which we use to evaluate counterfactual policies. Specifically, our framework models how applicants' and examiners' decisions jointly determine the grant gap through four channels: examiner allowance standards, examiner prior-art search, applicant amendment costs, and revision improvement. To estimate the model, we measure the textual distance from an application to both the broader prior-art corpus and the examiner-cited prior art. We find that prosecuting small-firm applications as if they were filed by large firms, holding application and examiner characteristics fixed, would raise their grant rate by 12.6 percentage points. This policy counterfactual closes about 64 percent of the patent-grant gap, increasing small firms' expected net private value per filed application by $51,084 on average. We also find that the relative importance of these channels varies across technology fields, suggesting that policies to narrow the grant gap should target the main barriers in each field.

[1] We use the term *small firms* to refer to applicants claiming *small-entity* or *micro-entity* status at filing, and we refer to all other applicants as *large firms*. These labels refer to USPTO fee status rather than a direct measure of employment or revenue. We also use *firms* to indicate patent *applicants*, although applicants can be individuals or institutions. Qualifying universities and nonprofit organizations may claim small-entity status; see https://www.uspto.gov/patents/apply/save-on-fees.

[2] The Unleashing American Innovators Act of 2022 raised small- and micro-entity discounts across the USPTO fee schedule from 50 to 60 percent and from 75 to 80 percent, respectively (https://www.uspto.gov/ip-policy/legislative-resources/unleashing-american-innovators-act-2022). Evidence suggests that barriers beyond fees, including legal costs, matter more for patenting outcomes (U.S. Patent and Trademark Office, 2024).

We first document that the patent-grant gap emerges primarily after the round-1 office action. After controlling for examiner and technology fixed effects, as well as application characteristics, the probability of round-1 allowance is similar across firm sizes. However, conditional on receiving a round-1 rejection, small firms abandon their applications at roughly three times the rate of large firms. Even among applicants that continue prosecution, small firms achieve a lower ultimate grant rate.

Reduced-form evidence alone cannot clarify how the different mechanisms interact to generate the patent-grant gap. In a framework with forward-looking agents, decisions at each stage anticipate subsequent equilibrium actions. If subsequent prosecution actions differ systematically by firm size, then evaluating each decision in isolation can lead to misleading policy implications. For instance, a round-1 rejection reflects not only baseline differences in application quality across firms, but also examiners' anticipation of how applicants of different sizes will respond to a rejection. These limitations motivate our structural approach.

To model examiner behavior, we focus on two patentability criteria—novelty and non-obviousness—that account for the most common grounds for rejection.[3] We construct an empirical proxy for these criteria based on textual distance to prior art using PatentBERT (Lee and Hsiang, 2020). Specifically, we compute vector embeddings for independent claims in patent applications and find the minimum distance to a corpus of over 23 million prior-art embeddings, covering U.S. patents granted between 1976 and 2013 and pre-grant publications between 2001 and 2013. In addition, we collect data on the prior art *cited by the examiner* in a rejection and use it to model prior-art search.

We simplify the prosecution process by collapsing it to three stages. In Stage 1, anticipating the applicant's response, the examiner decides whether to reject or immediately grant a patent. In Stage 2, following a rejection, the applicant amends if the expected continuation value exceeds the amendment cost (Lanjouw and Schankerman, 2004; Carley et al., 2015; de Rassenfosse and Jaffe, 2018). In Stage 3, following an amendment by the applicant, the examiner grants a patent if the application's improvement is sufficient.[4] Four primitives govern these decisions: on the examiner side, the allowance standard and the prior-art search; on the applicant side, the amendment cost and revision improvement.

[3] In practice, each application claim must satisfy the statutory requirements in Title 35 of the United States Code, including § 101 (subject-matter eligibility), § 102 (novelty), § 103 (non-obviousness), and § 112 (disclosure and claim clarity).

[4] "Improvement" captures written justifications and whether the examiner views the amended claims as narrower, clearer, or otherwise closer to satisfying the patentability requirements.

Because our model is dynamic, the examiner considers the *option value* of issuing a round-1 rejection. For a borderline application, a rejection can elicit an amendment that clarifies or narrows the claims before allowance. The value of this *reject-to-improve* mechanism depends on both the probability that the applicant will respond and the expected improvement associated with that response. We show that the examiner's round-1 strategy follows a threshold rule: applications whose novelty falls below a threshold are rejected. This threshold is determined in equilibrium and differs from the (static) allowance standard because it incorporates the option value of continued prosecution. In particular, the equilibrium threshold is higher (meaning the examiner is stricter in round 1) when an applicant is likely to respond and improve. This mechanism helps explain why large firms are rejected in round 1 at rates similar to those of small firms, even though the grant rate for large firms is higher.

We estimate the model by maximum likelihood using rejection and amendment decisions for utility patent applications filed in 2014. One challenge is that, although the examiner always evaluates an application against prior art, we observe the cited prior art only when the examiner issues a round-1 rejection. In addition, round-2 decisions are observed only for applications that received a round-1 rejection and were subsequently amended. The joint likelihood accounts for these selection issues. Examiner-cited prior art allows us to identify parameters associated with the examiner's prior-art search, whereas the applicant's decision to amend, together with the examiner's decision to reject in round 2, helps us distinguish amendment costs from the applicant's revision improvement. The private value of receiving a patent is estimated in a separate step using patent renewal decisions.

Our estimates indicate higher allowance standards and amendment costs for small firms, whereas the examiner's prior-art search is similar across firm sizes. The estimated median amendment costs are approximately \$11,900 for small firms compared with \$2,600 for large firms. Moreover, the expected revision improvement is about 7 percent lower for small firms. The expected costs and improvement prospects affect the examiner's initial rejection decision through a weaker reject-to-improve incentive for small firms. In fact, our estimates imply that this reject-to-improve mechanism accounts for 58 percent of round-1 rejections for large firms, compared with 36 percent for small firms.

We use our estimates to perform a number of counterfactual exercises. First, we simulate the prosecution of a small-firm application as if filed by a large firm. Specifically, each of the four structural channels—allowance standards, examiner's search, amendment costs, and

revision improvement—includes a small-firm indicator, and we set these indicators to zero in all four channels, while holding all other application and examiner characteristics fixed. We find that the small-firm grant rate rises by 12.6 pp, or 21.3 percent, closing approximately 64 percent of the patent-grant gap implied by the model. In terms of private economic value, this counterfactual increases the expected net private value per filed application from \$205,379 to \$256,463—a gain of \$51,084, or 25 percent. This increase closes roughly 52 percent of the baseline gap in private application value between small and large firms.

A second set of counterfactual exercises examines the four prosecution channels separately. Specifically, we set the small-firm indicator to zero for one channel at a time, holding all other channels and observed characteristics fixed. The effects of this one-at-a-time exercise are not additive because outcome probabilities are nonlinear, and each channel affects the initial screening decision as well as dynamic selection into subsequent prosecution stages. Our findings show that two channels are not the main drivers of the patent-grant gap: equalizing revision improvement increases the small-firm grant rate by only 0.7 percentage points, whereas equalizing examiner search has a negligible effect.

Applying the large-firm allowance standard to small firms raises their grant rate by 8.6 percentage points. The allowance-standard channel may capture both unobserved differences in examiner stringency and systematic differences in legally relevant application attributes not fully captured by our textual claim-based distance measure and other controls. University applicants provide an informative comparison group: while they are classified as small firms, they typically possess substantial institutional prosecution resources. We estimate a lower allowance standard for them, which is consistent with the allowance standard capturing unobserved prosecution capabilities and drafting quality.

When we remove the small-firm indicator from the amendment-cost equation, the small-firm amendment rate increases by 10 percentage points, from 79.5 to 89.5 percent. Consequently, the grant rate rises by 3.8 percentage points. Moreover, the initial rejection rate increases by 4.3 percentage points as examiners anticipate a higher likelihood of amendment, illustrating the dynamic incentives at play. Thus, lowering amendment costs helps narrow the patent-grant gap. Conversely, when we set the small-firm indicator to one for large firms, the post-rejection amendment rate for large firms falls by 7.3 percentage points, from 93.3 to 86.0 percent, and their grant rate decreases by 3.1 percentage points. Hence, large firms' losses in amendment and grant rates from a cost increase are smaller than small firms'

gains from a cost reduction of the same magnitude. This asymmetry suggests that raising large-firm fees could finance programs that reduce small-firm amendment costs.

Finally, we decompose the patent-grant gap by technology center. The results reveal that similar overall gaps across sectors can reflect different underlying mechanisms. For instance, the "Biotechnology" and "Chemical" centers feature comparable baseline patent-grant gaps, yet equalizing amendment costs has virtually no effect in the "Biotechnology" center but raises small-firm grant rates by 2.8 percentage points in the "Chemical" center.[5] Equalizing all channels closes 80 percent of the gap in the "Software" center, but only 60 percent in the "Transportation" and "Mechanical Engineering" centers.[6]

Taken together, our findings indicate that policies aimed at closing the patent-grant gap should target the post-rejection prosecution phase, where small firms face substantially higher amendment costs and allowance standards. Measures such as examiner guidance on amendments or subsidized legal assistance for responding to rejections could be financed by increasing prosecution fees for large firms. Because the main source of the gap differs across technology fields, however, such measures should be tailored to the barriers that matter in each field rather than applied uniformly.

Related literature. Patents are important for small firms. Galasso and Schankerman (2018) show that patent invalidation reduces subsequent patenting for small firms but not for large ones. Farre-Mensa et al. (2020) find that a first patent raises startups' growth and access to funding. High litigation costs disadvantage small firms in patent enforcement (Lanjouw and Schankerman, 2004). Reduced-form analysis has documented patenting gaps by race (Cook, 2014), gender (Aneja et al., 2026), nationality (de Rassenfosse and Hosseini, 2020), and legal representation (Gaudry, 2012).

Prior work documents persistent examiner heterogeneity (Lemley and Sampat, 2012; Schuster, 2023), heterogeneous evaluations (de Rassenfosse et al., 2021), and effects of limited examination time on grant decisions (Frakes and Wasserman, 2017, 2023). Tabakovic and Wollmann (2018) document that examiners may be more lenient when examining applica-

[5]Biotechnology refers to "Biotechnology and Organic Chemistry" (Technology Center 1600, or TC 1600) and Chemical refers to "Chemical and Materials Engineering" (TC 1700).

[6]Software refers to "Computer Architecture and Software" (TC 2100), Mechanical Engineering to "Mechanical Engineering, Manufacturing, and Products" (TC 3700), and Transportation to "Transportation, Construction, Electronic Commerce, and Agriculture" (TC 3600).

tions from a potential future employer. Schuett (2013) models examination as a principal-agent problem that distorts effective quality thresholds. Gaule (2018) and Farre-Mensa et al. (2020) exploit quasi-random assignment and use examiner leniency as an instrument for the downstream effects of patent protection, a design whose assumptions have since been scrutinized and refined (Righi and Simcoe, 2019; Goldsmith-Pinkham et al., 2026). These designs use examiner assignment at filing and reduced-form variation, which alone cannot separate examiner and applicant contributions during prosecution to the grant gap. Instead, we model examiner-applicant interactions as a dynamic game, in which the examiner's rejection threshold is an equilibrium object that depends on the applicant's expected response.

Dynamics matter during patent prosecution. Applicants continue when the expected return exceeds the cost of amendment (Carley et al., 2015). Higher fees increase abandonment (de Rassenfosse and Jaffe, 2018). Abandonment can also reflect strategic motives (Guellec et al., 2012; Blind et al., 2009; Cappelli et al., 2023) and procedural rules (Hegde et al., 2023; Righi and Simcoe, 2023).

Matcham and Schankerman (2026) also estimate a dynamic model of patent screening with textual distances, but do not study the patent-grant gap by firm size. Because their distance metric uses granted patent claims rather than initial applications, it misses abandoned applications and reflects text endogenously amended during prosecution. In contrast, we focus on the patent-grant gap, measure distance using application claims—observed regardless of prosecution outcome—and incorporate examiner rejection citations to study prior-art search.

Our distance measure uses transformer-based embeddings from PatentBERT (Lee and Hsiang, 2020). Ganguli et al. (2024) show that such embeddings outperform earlier text-based measures of technological proximity. Related work uses text-based distance to identify novel and important patents (Kelly et al., 2021; Arts et al., 2021).

Our analysis also relates to research on differential treatment in static decisions (Becker, 1957, 1993; Canay et al., 2024; Shea, 2026). In our dynamic model, examiners anticipate whether applicants will continue after rejection, so the initial decision is itself an equilibrium action and the pool reaching later stages is selected by applicants' earlier responses (Silveira, 2017). We do not, however, seek to identify examiner bias: allowance standards also absorb application attributes omitted by our text measure, so size-related differences in these standards cannot be interpreted solely as bias.

2 Institutional Setting: USPTO Examination Process

The patent examination process at the United States Patent and Trademark Office (USPTO) begins when an applicant files a non-provisional application and pays the required filing, search, and examination fees. The application's technology classification determines the technology center and art unit responsible for examination. Empirical work often treats examiner assignment within art units as quasi-random, although recent research has qualified this assumption (Righi and Simcoe, 2019).

Patent applications contain multiple claims, each of which must satisfy the statutory requirements in Title 35 of the United States Code, including § 101 (subject-matter eligibility), § 102 (novelty), § 103 (non-obviousness), and § 112 (disclosure and claim clarity). Our empirical analysis focuses primarily on rejections under § 102 and § 103. The Office Action data identify the references examiners cite in support of these rejections. Using these citations is a novel contribution, and our empirical framework uses them to distinguish the prior-art search channel from the allowance standard applied to the application.

Examiners are instructed to conduct a complete examination and identify all outstanding rejection grounds together. The first office action ("round 1") may *allow* the application (i.e., immediately grant a patent) but more often issues a non-final *rejection* explaining objections and citing supporting references for prior-art-based rejections. A non-final rejection does not end prosecution, but continuing requires the applicant to respond by amending claims, presenting arguments contesting the rejection, or both. Applicants may instead abandon by failing to respond or by expressly terminating prosecution.

If the applicant responds, the examiner may allow the application, issue another non-final action, or issue a "final" rejection. A final rejection changes procedural options and costs without necessarily ending prosecution. Subsequent options include an appeal, an appropriate after-final response, a continuation application, or a Request for Continued Examination (RCE), which reopens prosecution upon payment of an additional fee.

The first office action and post-rejection prosecution also differ institutionally from the examiner's perspective. The USPTO production system explicitly assigns credit to examination activities and places particular emphasis on a complete first office action and prompt resolution of applications. The system also treats first office actions and subsequent actions, including actions following RCEs, differently. This difference motivates treating the first

office action and post-rejection prosecution as distinct decision stages.

Prosecution costs vary by firm type. Since 1982, small firms have received fee reductions at the USPTO.[7] These discounts cover filing, examination, RCE, issue, and maintenance fees. In 2014, large-firm fees were $1,200 for the first RCE and approximately $960 for issuing a utility patent. Maintenance fees, required to keep patents in force, were $1,600 at 3.5 years, $3,600 at 7.5 years, and $7,400 at 11.5 years after grant. Small firms received a 50 percent reduction.

3 Data Sources and Variables

3.1 Data Description

Our primary source, the *USPTO Office Action Research Dataset for Patents* (*Office Action* data), records examiners' prosecution decisions.[8] Office actions notify applicants in writing of patentability decisions. Each records application-level information for a specific examination round: a unique identifier, claim-level rejection types (if any), and examiner-cited prior art for § 102 (novelty) or § 103 (non-obviousness) rejections.[9] These citations are key to measuring the application's distance to examiner-identified prior art, and comparing that distance with the application's distance to the closest prior art in our corpus.

We collect the text of independent claims in both granted patents since 1976 and pre-grant publications since 2001 from *USPTO PatentsView*.

We obtain application-level information from the *USPTO Patent Examination Research Dataset* (*PatEx*), which includes application, inventor, attorney, and correspondence datasets (see Graham et al., 2018, for details). Application records report the U.S. Patent Classification (USPC) class and subclass, earliest pre-grant publication number, patent number, art

[7] 35 U.S.C. § 41(h); 37 C.F.R. § 1.27. Micro entities receive a 75 percent discount and face additional eligibility requirements, including prior-application and income thresholds. 35 U.S.C. § 123; 37 C.F.R. § 1.29. The introduction discusses the larger current discounts; estimates use the historical fee regime.

[8] The USPTO created the initial datasets by extracting information from 4.4 million office actions mailed between 2008 and mid-2017 using natural language processing and machine learning (Lu et al., 2017). We obtained data for rejections after mid-2017 through the USPTO's Office Action APIs.

[9] Unique alphanumeric Image File Wrapper (IFW) identifiers establish the order of office actions. Transaction data from the Patent Examination Research Dataset record correspondence between applicants and the USPTO, but are available only at the application level and lack rejection information.

unit, assigned examiner, and status as of September 2023, allowing us to identify abandoned and allowed applications. Using these records, we measure examiner experience by the total number of applications previously processed by the examiner and leniency by the grant rate of applications assigned to that examiner in 2005–2013. We also construct measures of experience for inventors and correspondence contacts using applications filed before 2014.

The *USPTO Patent Maintenance Fee Events* dataset contains maintenance-fee events for all U.S. patents granted since 1981, which we use to estimate patent values (Pakes, 1986).

We use applications filed in 2014 for two reasons. First, key provisions of the America Invents Act (AIA) went into effect in 2012 and 2013, introducing substantial changes to the U.S. patent system. Using applications from 2014 reduces the overlap with this transition, although earlier priority claims can place applications under different statutory regimes. Second, 2014 filings leave enough time to observe the first two maintenance decisions (due 3.5 and 7.5 years after grant, modeled as ages 4 and 8; Online Appendix OA.6.1).

3.2 Textual Embeddings

We represent claim text as numerical vectors using PatentBERT, a language model adapted to patent text that encodes meaning from word context (Lee and Hsiang, 2020). Each claim maps to a 768-dimensional vector (an "embedding"), with similar text closer together. Textual proximity provides a continuous proxy for technological similarity but need not preserve distinctions that determine legal patentability.

We compute embeddings for the independent claims of three sets of documents: pre-grant publications of applications in the 2014 filing cohort, earlier pre-grant publications in the prior-art corpus, and patents granted since 1976. We use these embeddings to measure semantic similarity between applications and prior art, including examiner citations, and to compare claims across applicant types. These embeddings provide a data-driven, quantitative measure of technological content that would otherwise be difficult to incorporate into economic models of patent examination.

We compute about 12.5 million embedding vectors for all independent claims of around 5 million patents granted in 1976–2013, and about 11 million embedding vectors for all

independent claims of around 3.6 million pre-grant publications in 2001–2013.[10] Thus, our corpus of "prior art" is a set of about 23.5 million embedding vectors.

For each claim i in a 2014 application, we compute the distance to prior art as

$$D_i = \min_{j \in \text{Prior art}(i)} d(i, j), \tag{1}$$

where j is a claim in our corpus of U.S. patent and pre-grant-publication prior art available before application i's priority date, and $d(i, j)$ is the cosine (chord) distance between the vector embeddings for claims i and j.[11] The corpus measure uses claims in prior-art documents. Legal examination, however, can rely on the full disclosure of a reference, including its description and drawings. Thus, our claim-based distances measure one aspect of similarity to prior art. For § 102 or § 103 claim rejections, whenever the examiner cites a patent granted in 1976–2013 or a pre-grant publication from 2001–2013, we compute the distance between claim i and the document as

$$P_i^* = \min_{k \in \text{Cited document}} d(i, k), \tag{2}$$

where k indexes the claims of the cited document.

Embedding full patent descriptions is computationally demanding: descriptions typically exceed PatentBERT's 512-token context window and require "chunking," which can lose context across segments. For this reason, we use claim embeddings for prior-art distances. Online Appendix OA.1 uses a small sample to compare distances based on prior-art claims and descriptions. Their positive but imperfect rank correlation (0.58) supports using claims as a tractable proxy and documents measurement limitations.

Table 1 shows that our distance measure is systematically related to examination outcomes. Conditional on examiner and USPC class–subclass fixed effects, claims farther from prior art are more likely to receive an initial allowance and less likely to receive a §§ 102–103 rejection in round 1. Moreover, among claims rejected under § 102 or § 103, greater distance to the closest prior art in our corpus is associated with greater distance to the prior art cited by the examiner. These patterns support using D_i as a continuous proxy for information relevant to examination and are consistent with recent work that uses textual distance to prior art

[10] Online Appendix OA.3 reports summary statistics on embeddings.

[11] Specifically, chord distance $= \sqrt{2(1 - \cos(\theta))}$. Results using Euclidean distance are similar.

Table 1: Examiner's Decisions and Distance to Prior Art: Linear Regression

	Dependent variable:		
	Dummy for initial allowance	Dummy for 102–103 rejection in round 1	Distance to cited prior art
	(1)	(2)	(3)
Distance to prior art	0.053	−0.185	0.390
	(0.003)	(0.006)	(0.002)
Observations	889,723	889,723	446,734

Notes: The table uses the sample of independent claims of utility patent applications filed in 2014 that are matched to pre-grant publications from 2014–2016 (the matched-text sample). Column 3 uses the sample of independent claims with prior art (either prior U.S. patents or pre-grant publications) cited by the examiner to reject the claims based on 35 U.S.C. § 102 or § 103 in round 1 of examination. All specifications include examiner and USPC class–subclass fixed effects. Standard errors are in parentheses.

to operationalize novelty and non-obviousness (see, e.g., Matcham and Schankerman, 2026).

Our subsequent analysis uses three nested samples. The *2014 application universe*, used for the descriptive grant-rate calculations, consists of all utility applications filed in 2014. The *matched-text sample* additionally requires a match to a pre-grant publication from 2014–2016; approximately 95 percent of 2014 utility applications satisfy this requirement.[12] The *structural-estimation sample* further requires round-1 rejections to be based on §102 or §103 so that we can compute the distance to examiner-cited prior art.[13]

4 Descriptive Statistics

4.1 The Patent-Grant Gap and the Prosecution Process

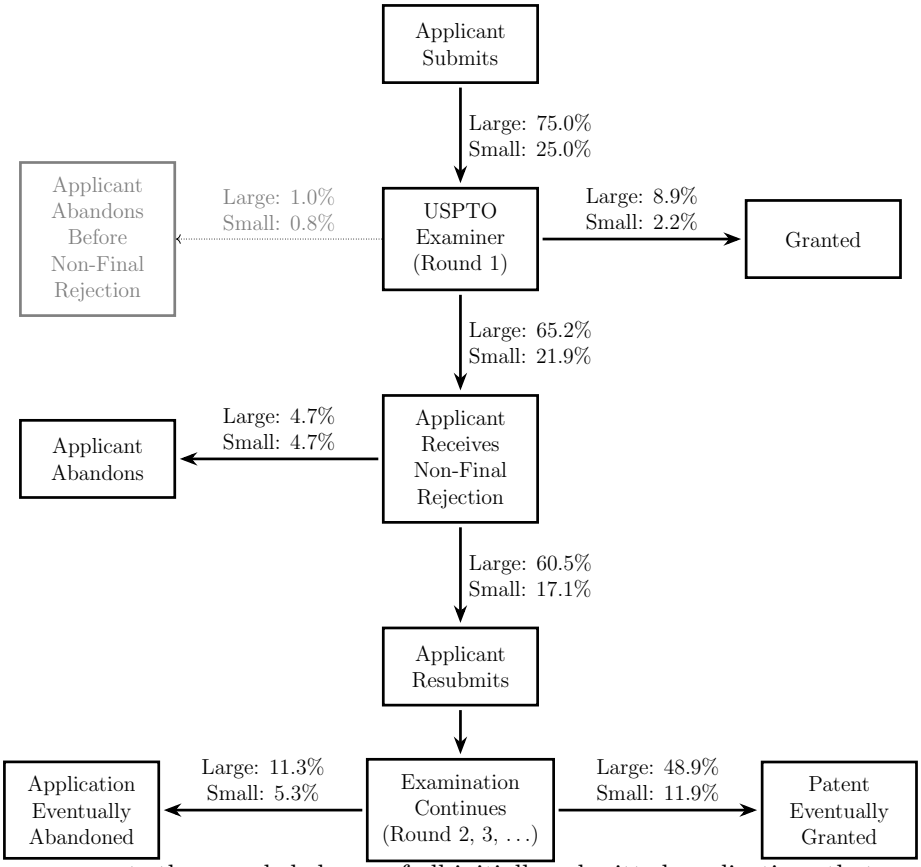
We begin by documenting the patent-grant gap between large and small firms. For the 2014 application universe, the patent-grant rate is 77.0 percent for large firms and 58.2 percent for small firms. Although we focus on applications filed in 2014 for the reasons discussed in Section 3, the patent-grant gap is not specific to that cohort.[14] In the matched-text sample,

[12] We use application claims in pre-grant publications, rather than in issued patents, because numbering and claim scope can change during prosecution, and patent claims are missing when an application is not granted. Only about 0.4 percent of 2014 filings have pre-grant publications that fall outside the 2014–2016 window; the remaining applications without a pre-grant publication were abandoned or granted early (or subject to a non-publication request) before reaching the pre-grant publication stage.

[13] This sample contains 250,160 applications satisfying the restrictions described in Appendix A.

[14] The patent-grant gap persists across 2010–2018 filing cohorts (Table OA.1, Online Appendix OA.2).

Figure 1: Grant and Abandonment through the Patent Application Process



Notes: Each arrow reports the rounded shares of all initially submitted applications that are large or small and follow that path. For small and large firms, respectively, round-1 allowance rates are $2.2/25 \approx 8.8\%$ and $8.9/75 \approx 11.9\%$; abandonment rates conditional on a round-1 non-final rejection are $4.7/21.9 \approx 21.5\%$ and $4.7/65.2 \approx 7.2\%$; and eventual grant rates conditional on continuing past round 1 are $11.9/17.1 \approx 69.6\%$ and $48.9/60.5 \approx 80.8\%$. Small firms' share of all grants is $(2.2 + 11.9)/(2.2 + 11.9 + 8.9 + 48.9) \approx 19.6\%$.

the grant rate is 76 percent for large firms and 56 percent for small firms, similar to the rates above, and small firms account for approximately one quarter of applications, as in the 2014 application universe.

Using the 2014 application universe, Figure 1 summarizes the patent examination process and documents the patent-grant gap at several prosecution stages. Small firms account for 25.0 percent of initial filings but only about 19.6 percent of granted patents. Three differences stand out in the raw data. Initial allowance rates are lower for small firms than for large firms (8.8 versus 11.9 percent). After a round-1 non-final rejection, small-firm applications are more likely to be abandoned (21.5 versus 7.2 percent). Among applications that continue past round 1, eventual grant rates are also lower for small firms (69.6 versus 80.8 percent).

These differences do not directly identify mechanisms. First, they confound behavior with composition: lower initial allowance may reflect differences in novelty, differences in examiner characteristics, or technology-field composition across firm sizes. The regressions below show that composition accounts for much of this raw gap. Second, regressions cannot separate mechanisms even after controlling for composition. Higher abandonment after rejection may reflect differences in patent value, amendment costs, or rejection severity. Lower grant rates conditional on amendment may reflect both the extent of improvement needed to get a grant and systematic differences in revision improvement by firm size. These sequential decisions jointly determine the observed patterns, motivating the use of a structural model.

4.2 Rejection Grounds and Abandonment

Novelty and non-obviousness account for most rejections among both small and large firms, motivating our focus on §§ 102–103 rejections of independent claims.[15] Small firms face substantially more § 112 rejections—concerning written description, enablement, and definiteness—especially in round 1. Their greater prevalence is consistent with differences in drafting or prosecution resources. However, most applications receiving a § 112 rejection also receive § 102 or § 103 rejections.

Abandonment decisions reinforce our focus on novelty and non-obviousness. Table OA.3 in Online Appendix OA.2 reports the results of regressing abandonment on the type of rejection received in round 1, separately for large and small firms. A round-1 §§ 102–103 rejection is associated with a 12.8 percentage-point increase in abandonment for small firms, compared with 0.8 points for large firms. The corresponding increases for § 112 rejections are 3.4 and 1.4 points. Although § 101 rejections are far less common than §§ 102–103 rejections, the results for these rejections closely resemble those for §§ 102–103.

[15] Table OA.2 in Online Appendix OA.2 documents little difference between *any rejection* and *§§ 102–103* rejection rates across all rounds. We pool § 102 and § 103, so embedding distance to a single document differs somewhat in meaning across grounds. The model compares found-art distance to a standard for both grounds, which are frequently issued together and also pooled by Matcham and Schankerman (2026).

Table 2: Reduced-Form Regression Results

	Dependent variable:						
		Not rejected for		Allowance in		Amend af.	Dist. to
	Patented	102–103	any	round 1	round 2+	round 1	prior art
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Small	−0.132	0.008	0.006	−0.003	−0.084	−0.112	0.015
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.000)
Observations	346,974	346,974	346,974	346,974	272,131	306,536	346,974

Notes: All regressions include examiner, USPC class–subclass fixed effects, and a 768-dimensional vector embedding as controls. All regressions use the first independent claims in the matched-text sample. The dependent variable varies across columns. In column 1, it is a dummy for whether the application of the claim is patented (i.e., allowance in any round). In columns 2–3, it is a dummy for whether the claim is not rejected in round 1 based on §§ 102–103 (column 2) or any rejection criteria (column 3). In columns 4–5, it is a dummy for whether the application of the claim receives an allowance in round 1 (column 4) or in round 2 or later (column 5). In column 6, it is whether the application of the claim is amended and resubmitted after rejections in round 1. In column 7, it is the distance to prior art. Standard errors are in parentheses.

4.3 Reduced-Form Evidence on the Patent-Grant Gap

The raw evidence points to three patterns: small-firm applications are less likely to receive an initial allowance; conditional on a round-1 rejection, small firms abandon at much higher rates, particularly after §§ 102–103 rejections; and small-firm applications are ultimately granted patents at substantially lower rates.

We examine which of these raw patterns survive after controlling for examiner and technology composition. Table 2 reports the results from regressing rejection and amendment decisions on an indicator for whether an application is submitted by a small firm. The unit of observation is the first independent claim of a utility patent application in the matched-text sample.[16] All regressions include examiner and USPC class–subclass fixed effects, as well as the claim's 768-dimensional vector embedding, as controls.

The regressions establish that the patent-grant gap itself survives: small-firm claims are about 13 percentage points less likely to be eventually patented (column 1), even after controlling for examiner and technology heterogeneity. Round-1 examiner decisions, by contrast, show no such gap. Small-firm claims are marginally *less* likely to be rejected in round 1 (columns 2–3), and the conditional round-1 allowance gap is only 0.3 percentage points (column 4). This gap is substantially smaller than the raw gap in Figure 1, indicating that examiner and technology composition account for much of that raw difference. The

[16] Results using all independent claims are nearly identical; see Table OA.4 in Online Appendix OA.2.

allowance gap in round 2 or later (column 5) is much larger, but it cannot be interpreted as a difference in post-rejection examiner behavior alone, because reaching post-rejection prosecution is endogenous to round-1 examiner decisions and applicant amendment decisions.

The amendment gap also survives: conditional on reaching the amendment decision, small-firm applicants are approximately 11 percentage points more likely to abandon (column 6). This gap could be driven by patent value, amendment costs, the severity of the rejection received, or expected prospects of eventual allowance. Lastly, small-firm claims are, on average, *farther* from the closest prior art as measured by the distance in (1), so measured distance does not provide an immediate explanation for the lower grant rate (column 7).

The reduced-form results point to post-rejection prosecution as the main source of the patent-grant gap. Conditional on examiner and technology composition, small-firm applications are slightly less likely to be rejected in round 1, and their claims are farther from prior art under our textual measure. However, small firms amend much less often after rejection and have substantially lower final grant rates. The patterns in Table 2 do not identify the underlying mechanisms because prosecution stages are jointly determined. The decision to amend depends on rejection severity and whether a revision will result in allowance. Terminal outcomes are therefore a selected sample: they are observed only for applicants who amend. Examiners may anticipate these choices and outcomes when deciding whether to reject in round 1. Hence, separate stage regressions cannot disentangle these mechanisms.

To disentangle the mechanisms behind the patent-grant gap, the dynamic model below uses the joint distribution of round-1 rejection, amendment, and terminal outcomes to distinguish the allowance standard, search gap, amendment costs, and revision improvement.

5 Theoretical Framework

We collapse the patent examination process into three stages: the examiner's initial decision to allow or reject, the applicant's decision to amend or abandon after a rejection, and the ultimate allowance decision conditional on continued prosecution. Stage 1 is the first office action (round 1); Stages 2 and 3 correspond to all post-rejection prosecution, labeled round 2. We model these stages as a dynamic game, which we solve by backward induction.

An application i is characterized by its distance $D_i \geq 0$ to the closest prior art. A smaller

distance indicates greater similarity to prior art. The examiner does not observe D_i directly, but draws prior art at distance

$$P_i^* = D_i + S_i$$

from the application, where $S_i \geq 0$ represents a *search gap*, implying $P_i^* \geq D_i$. The examiner draws an application-specific *allowance standard* T_i (henceforth, the "standard"), which governs whether the application's assessed patentability is sufficient for eventual allowance.[17]

Examiners understand that their search may not find the closest prior art. Using examiner and application characteristics, as well as the found prior art, the examiner forms an assessment $\widehat{D}_i$ of the application's actual minimum distance from the closest relevant prior art. An application falls short of the standard under the examiner's initial assessment when the *patentability gap* $\widetilde{G}_i$ is positive, where

$$\widetilde{G}_i \equiv T_i - \widehat{D}_i. \tag{3}$$

Conditional on application characteristics, the amendment cost and improvement draws are independent. The examiner knows their distributions but does not observe the cost realization. When the applicant submits a revision, the application's improvement e_i may close the patentability gap. The timing of the game is as follows:[18]

1. The examiner observes (P_i^*, T_i), forms $\widetilde{G}_i$, and chooses immediate allowance ($R_{i,1} = 0$) or non-final rejection ($R_{i,1} = 1$).
2. Conditional on $R_{i,1} = 1$, the applicant observes $\widetilde{G}_i$ by reading the rejection, and the amendment cost $C_i^R \sim F_c$, and chooses to abandon ($a_i = 0$) or amend ($a_i = 1$).
3. An amendment ($a_i = 1$) yields a stochastic improvement $e_i \sim H_i$ with support $[0, \infty)$. The examiner observes e_i and allows the application ($R_{i,2} = 0$) or not ($R_{i,2} = 1$).

5.1 Equilibrium Characterization

Let $\delta_i \in [0, 1)$ denote the examiner's weight on an outcome realized through continued prosecution rather than immediate allowance (hereafter, the discount factor), which may reflect workload, delay, and institutional production incentives. Let V_i denote the applicant's

[17] Online Appendix OA.4.2 provides a microfoundation for the examiner's static payoff.

[18] Ties are resolved in favor of rejection at Stage 1, amendment at Stage 2, and allowance at Stage 3.

expected private value of a patent at grant (hereafter, "patent value").

Conditional on amending ($a_i = 1$), the application's assessed position improves to $\widehat{D}_i + e_i$. The examiner allows the application at Stage 3 if and only if $e_i \geq \widetilde{G}_i$. Thus, the Stage-3 allowance probability conditional on amending is:

$$\varphi_i(\widetilde{G}_i) \equiv \Pr(e_i \geq \widetilde{G}_i) = 1 - H_i(\widetilde{G}_i).$$

When $\widetilde{G}_i \leq 0$, eventual allowance conditional on amending is guaranteed, $\varphi_i(\widetilde{G}_i) = 1$. When $\widetilde{G}_i > 0$, eventual allowance conditional on amending is no longer guaranteed, $\varphi_i(\widetilde{G}_i) < 1$, which reduces the incentive to amend. Anticipating $\varphi_i(\widetilde{G}_i)$, the applicant amends at Stage 2 if and only if the expected benefit of amending exceeds the cost: $V_i\varphi_i(\widetilde{G}_i) \geq C_i^R$. The equilibrium amendment probability is:

$$q_i(\widetilde{G}_i) = F_c\Big(V_i\varphi_i(\widetilde{G}_i)\Big).$$

In Stage 1, the forward-looking examiner balances immediate allowance against the option value of continued prosecution.

Proposition 1. *The game has a unique subgame-perfect equilibrium characterized by decision rules* $(R_{i,1}^*, a_i^*, R_{i,2}^*)$*:*

1. *Round-1 Rejection Decision:* $R_{i,1}^* = \mathbf{1}\{\widetilde{G}_i \geq \kappa_i\}$*, where*

$$\kappa_i \equiv \frac{-\delta_i\, F_c(V_i)\, \mathbb{E}[e_i]}{1 - \delta_i\, F_c(V_i)}. \tag{4}$$

2. *Applicant Amendment Decision:* $a_i^* = \mathbf{1}\Big\{V_i\varphi_i(\widetilde{G}_i) \geq C_i^R\Big\}$.
3. *Terminal Non-Allowance Indicator:* $R_{i,2}^* = \mathbf{1}\{e_i < \widetilde{G}_i\}$.

Proof. See Online Appendix OA.4.1. ∎

Economic interpretation. The round-1 cutoff $\kappa_i \leq 0$ measures the *option value* of a non-final rejection because a non-final rejection may induce the applicant to revise and improve the application. Consequently, an examiner may issue a non-final rejection to elicit an amendment that clarifies or narrows the claims before allowance. Figure 2 illustrates how

Figure 2: Dynamic Examination Incentives, κ_i.



the patentability gap shapes the examiner's incentive. If the examination process did not give applicants the opportunity to amend, the examiner would allow whenever $\widetilde{G}_i \leq 0$. The dynamic incentive triggers rejections when $\widetilde{G}_i \in [\kappa_i, 0]$. From (4), this option value expands ($|\kappa_i|$ increases) when: (i) patents generate larger private value, increasing the amendment probability (larger $F_c(V_i)$); (ii) expected revision improvement $\mathbb{E}[e_i]$ is higher; or (iii) the examiner is more forward-looking (higher δ_i).

Proposition 1 highlights why observed round-1 rejection rates cannot be interpreted as direct measures of examination strictness. Fixing $\widetilde{G}_i$, round-1 rejections are determined not only by an examiner's allowance standard, but also by the expected improvement, amendment costs, patent value, and the examiner discount factor. Thus, interpreting round-1 rejections as a measure of examiner stringency is misleading.

6 Empirical Framework

For each patent application $i = 1, \ldots, N$, the key empirical variables are

$$(R_{i,1},\, a_i,\, R_{i,2},\, D_i,\, P_i,\, X_i). \tag{5}$$

$R_{i,1} \in \{0, 1\}$ indicates a round-1 rejection; $a_i \in \{0, 1\}$ indicates an amendment after a round-1 rejection; and $R_{i,2} \in \{0, 1\}$ is the terminal non-allowance indicator for an amended application, equal to one if it is *ultimately* rejected or abandoned and to zero if it is ultimately allowed. Thus, $R_{i,2} \in \{0, 1\}$ is a terminal outcome that includes the applicant's decision to discontinue, rather than a literal second examiner decision. Some of these outcomes are missing: a_i is observed only if $R_{i,1} = 1$, and $R_{i,2}$ only if $R_{i,1} = a_i = 1$. D_i is the closest distance between the application and the prior-art corpus of Section 3.2, P_i is the distance to examiner-cited prior art, and X_i collects application and examiner characteristics.[19]

[19]For estimation, we rescale distances to make them comparable across technology centers while ensuring non-negativity. This rescaling maps raw center z-scores to a lower-truncated normal distribution.

The econometrician observes D_i but the examiner does not because comparing all prior art with an application is generally too time-consuming. The examiner observes the distance P_i^* to the prior-art documents identified during the examination, but the econometrician observes it only when it is cited in a rejection. Thus, the econometrician observes:

$$P_i = \begin{cases} P_i^*, & R_{i,1} = 1, \\ \text{missing}, & R_{i,1} = 0. \end{cases}$$

Using the data in (5), the goal is to recover the distributions of the standard T_i, the search gap S_i, the amendment cost C_i^R, and the revision improvement e_i, defined in Section 5.

6.1 Estimation Specification

We specify the unobservables as

$$T_i = \tau_i + \sigma_\tau u_i, \tag{6}$$

$$S_i^* = \lambda_i + \sigma_s v_i, \tag{7}$$

$$\ln C_i^R = c_i + \sigma_c \varepsilon_i, \tag{8}$$

$$e_i \sim \text{Exp}(\omega_i), \tag{9}$$

$$\delta_i = \Lambda(\alpha_i), \tag{10}$$

where, before imposing the restriction $S_i^* \geq 0$, (u_i, v_i) is standard bivariate normal with correlation $\rho_{s\tau}$, ε_i is standard normal, and $\Lambda(x) = (1 + e^{-x})^{-1}$ is the logistic function. The shocks (u_i, v_i), ε_i, and e_i are mutually independent conditional on X_i, and the indices $(\tau_i, \lambda_i, c_i, \ln \omega_i, \alpha_i)$ are linear in covariates X_i. Here τ_i and λ_i are location parameters for the latent standard and search gap before truncation. The index c_i locates the log amendment-cost distribution, and ω_i is the exponential rate parameter for revision improvement.

Three features of this specification deserve comment. First, the realized search gap is the latent gap truncated at zero, $S_i = S_i^* \mid S_i^* \geq 0$, so $P_i^* = D_i + S_i \geq D_i$. The parameter $\rho_{s\tau}$ permits correlation between the latent search and standard shocks. Under the joint truncation used in the likelihood,

$$\mathbb{E}[T_i \mid X_i] = \tau_i + \rho_{s\tau}\sigma_\tau \frac{\phi(\lambda_i/\sigma_s)}{\Phi(\lambda_i/\sigma_s)}.$$

Second, D_i is an empirical proxy for legally relevant distance, so τ_i absorbs systematic gaps between measured and effective distance as well as examiner stringency. For this reason, application characteristics enter τ_i alongside examiner characteristics (see Section 6.2). We call τ_i the *allowance-standard index*: it locates the allowance standard relative to measured distance, without identifying whether a residual difference originates with the examiner or with application attributes omitted from D_i.

Third, $\mathbb{E}[e_i \mid X_i] = 1/\omega_i$, so a larger ω_i means that the applicant is *less* effective at improving the application after a rejection. Because $R_{i,2}$ is a terminal outcome, e_i should be interpreted as a reduced-form measure of the application's improvement over the entire prosecution path after the first rejection, including later amendments, arguments, examination, and decisions to discontinue.

The examiner does not observe D_i. We approximate the information structure of Section 5 with the rule

$$\widehat{D}_i \equiv P_i^* - \bar{s}_i, \qquad \bar{s}_i \equiv \mathbb{E}[S_i \mid X_i] = \lambda_i + \sigma_s \frac{\phi(\lambda_i/\sigma_s)}{\Phi(\lambda_i/\sigma_s)}, \tag{11}$$

which treats the found prior art as an imperfect signal of the closest prior art and corrects its distance by the expected search gap given observables. The rule is a location correction rather than a Bayesian posterior mean for D_i: it requires no prior over D_i and no updating, while retaining the model's key feature that the examiner acts on a noisy assessment of the closest distance. The assessed patentability gap is

$$\widetilde{G}_i \equiv T_i - \widehat{D}_i = (T_i - D_i) - (S_i - \bar{s}_i),$$

the gap relative to measured corpus distance, $T_i - D_i$, minus a correction $S_i - \bar{s}_i$ with mean zero conditional on X_i.

Under (9), the probability that revision improvement overcomes a gap $\widetilde{G}$ is

$$\varphi_i(\widetilde{G}) = \Pr(e_i \geq \widetilde{G} \mid X_i) = \exp(-\omega_i \widetilde{G}^+), \qquad \text{where } \widetilde{G}^+ \equiv \max\{\widetilde{G}, 0\}.$$

The applicant amends when the amendment cost falls below the expected value of continuing prosecution, $V_i \varphi_i(\widetilde{G}_i)$, where V_i is the patent value. We estimate this value separately with the patent-renewal model in Online Appendix OA.6. By (8), the amendment probability

conditional on the gap is

$$q_i(\widetilde{G}) = \Pr\big(C_i^R \le V_i \varphi_i(\widetilde{G})\big) = \Phi\left(\frac{\ln V_i - c_i - \omega_i\, \widetilde{G}^{+}}{\sigma_c}\right),$$

which equals $q_i^0 \equiv \Phi\big((\ln V_i - c_i)/\sigma_c\big)$ for $\widetilde{G} \le 0$ and declines in $\widetilde{G}$ when $\widetilde{G} > 0$.

The three observed decisions are

$$R_{i,1} = \mathbf{1}\{\widetilde{G}_i \ge \kappa_i\}, \tag{12}$$

$$a_i = \mathbf{1}\{C_i^R \le V_i \varphi_i(\widetilde{G}_i)\}, \tag{13}$$

$$R_{i,2} = \mathbf{1}\{e_i < \widetilde{G}_i\}. \tag{14}$$

Under this specification, the round-1 cutoff is weakly negative and given by

$$\kappa_i = \frac{-\delta_i q_i^0}{(1 - \delta_i q_i^0)\omega_i}. \tag{15}$$

In terms of the found distance, (12) states that $R_{i,1} = 1$ if and only if $P_i^* \le T_i + \bar{s}_i - \kappa_i$: conditional on (T_i, X_i), closer found prior art makes a round-1 rejection more likely.

6.2 Covariates and Exclusion Restrictions

The five structural indices in (6)–(10) are

$$\begin{aligned}\tau_i = {} & \beta_0 + \beta_s\,\text{small}_i + \beta_a\,\text{leniency}_i + \beta_n\,\text{experience}_i + \beta_u\,\text{univ}_i + \beta_h\,\text{hedge}_i + \beta_f\,\text{foreign}_i \\ & + \beta_p\,\text{rep}_i + \beta_{ca}\,\text{lnAppC}_i + \beta_{cr}\,\text{rateC}_i + \beta_{ia}\,\text{lnAppI}_i + \beta_{ir}\,\text{rateI}_i \\ & + \text{small}_i\big(\beta_{su}\,\text{univ}_i + \beta_{sh}\,\text{hedge}_i + \beta_{sp}\,\text{rep}_i\big) + \text{small}_i\big(\beta_{s,ca}\,\text{lnAppC}_i + \beta_{s,cr}\,\text{rateC}_i\big) \\ & + \text{small}_i\big(\beta_{s,ia}\,\text{lnAppI}_i + \beta_{s,ir}\,\text{rateI}_i\big) + \textstyle\sum_j \beta_j^{tc}\,\text{TC}_{ij},\end{aligned} \tag{16}$$

$$\lambda_i = \lambda_0 + \lambda_s\,\text{small}_i + \lambda_n\,\text{experience}_i + \lambda_a\,\text{leniency}_i + \textstyle\sum_j \lambda_j^{tc}\,\text{TC}_{ij}, \tag{17}$$

$$c_i = c_0 + c_s\,\text{small}_i + c_m\,\text{micro}_i + c_r\,\text{attRatio}_i + \textstyle\sum_j c_j^{tc}\,\text{TC}_{ij}, \tag{18}$$

$$\ln \omega_i = \gamma_0 + \gamma_s\,\text{small}_i + \gamma_D\, D_i + \gamma_{sD}\,\text{small}_i D_i + \gamma_r\,\text{attRatio}_i, \tag{19}$$

$$\alpha_i = \alpha_0 + \alpha_a\,\text{leniency}_i + \alpha_n\,\text{experience}_i + \textstyle\sum_j \alpha_j^{tc}\,\text{TC}_{ij}. \tag{20}$$

Here small_i and micro_i indicate small- and micro-entity status; leniency_i and experience_i are examiner leniency and experience, standardized to mean zero and unit variance; univ_i indicates an academic or research-institution assignee; hedge_i indicates hedged or relative terminology in the independent claims; foreign_i indicates a foreign applicant; rep_i indicates professional representation at filing; lnAppC_i and rateC_i are the log of one plus the number of prior applications listing the same correspondence contact and the share of those applications that resulted in a patent. For inventors, lnAppI_i is the log of one plus their mean prior application count, and rateI_i is their mean prior patent count divided by their mean prior application count. The attorney ratio attRatio_i is the share of patent attorneys among attorneys and patent agents of record in the most recent PatEx snapshot, centered at its sample mean; TC_{ij} indicates technology center j. Appendix B provides detailed definitions of these variables.

Examiner characteristics enter the standard, search-gap, and discount-factor indices. They affect terminal outcomes through the assessed gap and selection into continuation, but do not enter the improvement-rate index ω_i directly. University affiliation, hedging language, foreign status, representation at filing, and the correspondent and inventor experience measures enter τ_i only. University affiliation may reflect institutional prosecution resources or invention selection, while hedging is a text control: generic qualifying language may change embedding distance without a corresponding change in legally relevant patentability.

Micro-entity status enters the cost index only. The attorney ratio enters both post-rejection indices but not the assessment stage. The distance to prior art, D_i, may also affect the applicant's ability to improve, so it enters the improvement parameter ω_i. We allow this relationship to differ by applicant size. Finally, the discount-factor covariates mirror the determinants of examiner credit under the USPTO count system, namely seniority and technology complexity (Matcham and Schankerman, 2026, Appendix E), and exclude applicant size because δ_i is an attribute of the examination environment.

We estimate the parameter vector $\big(\{\beta\};\, \{\lambda\};\, \{c\};\, \{\gamma\};\, \{\alpha\};\, \sigma_s,\, \sigma_c,\, \rho_{s\tau}\big)$ by maximum likelihood. This vector collects the coefficients in (16)–(20) together with $\sigma_s > 0, \sigma_c > 0$, and $\rho_{s\tau} \in (-1, 1)$. The standard dispersion is calibrated at $\sigma_\tau = 5$. This calibration constrains latent heterogeneity and affects the counterfactuals. The calibrated dispersion is constant across applications. Section 6.3 motivates the calibration, and all standard errors are conditional on it. Online Appendix OA.5 presents the derivation of the likelihood function used

for the estimation.

Our choice of covariates across indices follows the prosecution process: examiner characteristics and attributes fixed at filing enter the round-1 indices, variables tied to the cost and revision improvement enter the post-rejection indices, and the discount-factor index includes covariates relevant to examiner credits. Our results are robust to alternative specifications (see Online Appendix OA.7).

6.3 Identification

We discuss identification in the order in which the data discipline the parameters: the search process and the allowance standard, the amendment-cost and revision-improvement parameters, and then the dynamic rejection incentive, whose separation relies on the previously identified objects. We close with the interpretation of the allowance-standard coefficients and the calibrated dispersion.

Search and the initial assessment. For rejected applications, the distance to examiner-cited prior art determines the search gap, $S_i = P_i - D_i$. However, P_i is missing when $R_{i,1} = 0$, and selection into rejection depends on the latent standard, the search gap, and the equilibrium cutoff, κ_i. Consequently, the distribution of S_i is not identified from the empirical distribution of $P_i - D_i$ among rejected applications alone. Instead, $(\lambda_i, \sigma_s, \rho_{s\tau})$ are jointly pinned down by the density of the search gaps observed among rejected applications and by the probability of round-1 rejection under the maintained joint distribution of (T_i, S_i). Conditional on this search process, shifts in τ_i change the location of the distribution of $\widetilde{G}_i$ and hence the likelihood of round-1 rejection and the distribution of $\widetilde{G}_i - e_i$ in round 2.

Amendment costs and revision improvement. Conditional on an assessed patentability gap G, the two post-rejection outcomes have distinct structural probabilities:

$$q_i(G) = \Phi\left(\frac{\ln V_i - c_i - \omega_i G^+}{\sigma_c}\right),$$
$$p_{2i}(G) \equiv \Pr(R_{i,2} = 1 \mid a_i = 1, G, X_i) = 1 - \exp(-\omega_i G^+).$$

The cost index c_i affects the amendment margin directly, whereas ω_i governs terminal success and also affects amendment through the expected return to amendment. Variation in V_i provides additional separation because V_i enters the amendment payoff but is excluded from the standard, search-gap, and improvement equations. Identification of c_i and ω_i also relies on the conditional independence of the amendment-cost and improvement shocks.

The dynamic rejection incentive. The round-1 decision depends on the dynamic cutoff κ_i in (15), so round-1 rejections directly identify the composite dynamic rejection incentive κ_i. Once q_i^0 and ω_i are pinned down by the post-rejection decisions, we can back out δ_i from κ_i. Therefore, separate identification of δ_i depends on variation in q_i^0 and ω_i generated by covariates excluded from the standard equation.

The allowance standard index. Coefficients in τ_i measure differences in assessed patentability relative to measured distance. They may combine genuine differences in examination standards with systematic differences between measured and legally relevant patentability that remain after conditioning on the included application characteristics. Our counterfactuals retain this interpretation. The standard dispersion σ_τ is calibrated rather than estimated: although this parameter is identified in theory, it is only weakly identified in our data. We set $\sigma_\tau = 5$ to impose that the average revision improvement does not dramatically change the application. Appendix C documents the σ_τ choice in more detail.

7 Estimation Results

We estimate the structural model by maximum likelihood on the structural-estimation sample described in Appendix A. This sample contains 250,160 applications: 209,413 round-1 rejections and 40,747 allowances. The variables used in the estimation are described in Appendix B. In the estimation, technology-center fixed effects are relative to TC 2800 (Semiconductors, Electrical and Optical Systems and Components). The patent value, V_i, is estimated separately, and the estimates are reported in Appendix F. Model parameters are estimated using the joint likelihood (Online Appendix OA.5).

Table 3 reports the maximum likelihood estimates. Panels A and C report estimates for the examiner's initial assessment (the allowance standard and the prior-art search), and Panels

Table 3: Main Structural Parameter Estimates

Parameter	Estimate	SE	Parameter	Estimate	SE
Panel A: allowance standard (τ_i)			*Panel C: search gap (λ_i)*		
β_0 (intercept)	9.0633	0.1776	λ_0 (intercept)	2.1548	0.0077
β_s (small)	0.3656	0.1633	λ_s (small)	−0.1874	0.0079
β_a (leniency)	−3.2249	0.0380	λ_n (experience)	0.0583	0.0041
β_n (experience)	−0.6299	0.0270	λ_a (leniency)	0.0604	0.0042
β_u (university)	−0.8429	0.1106			
β_h (hedging language)	−0.3643	0.0370	*Panel D: amendment-cost index (c_i)*		
β_{su} (small × university)	−0.6240	0.1382	c_0 (intercept)	−0.5512	0.1415
β_{sh} (small × hedging)	−0.2769	0.0739	c_s (small)	1.2964	0.0625
β_f (foreign)	−0.3813	0.0769	c_m (micro-entity)	0.5656	0.0522
β_p (representation at filing)	−0.5008	0.0434	c_r (attorney ratio)	−0.9310	0.0503
β_{ca} (correspondent log apps.)	−0.0480	0.0071			
β_{cr} (correspondent patent rate)	−4.8958	0.1133	*Panel E: revision improvement ($\ln \omega_i$)*		
β_{ia} (inventor log applications)	−0.0711	0.0118	γ_0 (intercept)	−1.2174	0.0501
β_{ir} (inventor patent rate)	−1.7262	0.0643	γ_s (small)	−0.2417	0.0322
β_{sp} (small × repr. at filing)	−0.3823	0.0785	γ_D (D_i)	−0.0469	0.0021
$\beta_{s,ca}$ (small × corr. log apps.)	0.0860	0.0140	γ_{sD} (small × D_i)	0.0517	0.0049
$\beta_{s,cr}$ (small × corr. patent rate)	1.4120	0.1937	γ_r (attorney ratio)	−0.0331	0.0114
$\beta_{s,ia}$ (small × inv. log apps.)	0.0588	0.0270			
$\beta_{s,ir}$ (small × inv. patent rate)	0.2061	0.1116	*Panel F: discount factor (α_i)*		
			α_0 (intercept)	0.6360	0.0741
Panel B: dispersion and correlation			α_a (leniency)	0.1606	0.0069
σ_s (search std. dev.)	1.3086	0.0028	α_n (experience)	−0.0159	0.0046
$\rho_{s\tau}$ (search–standard corr.)	0.1183	0.0055			
σ_c (amendment-cost std. dev.)	3.0040	0.1006			

Notes: $N = 250{,}160$. Technology-center fixed effects are included but not reported. Examiner leniency and experience are standardized. Standard errors are computed from the numerical Hessian and are conditional on the calibrated σ_τ and on estimated V_i.

D and E for the applicant's response to a rejection (the amendment cost and the revision improvement). Panel B presents estimates for the dispersion parameters (σ_s, σ_c) and the correlation coefficient ($\rho_{s\tau}$). Panel F reports the parameters of the examiner's discount factor, δ_i. Most parameters are precisely estimated. The model closely reproduces the three prosecution margins by firm type and technology center. Overall prediction errors are 0.002 for round-1 rejection and smaller than 0.001 in absolute value for amendment conditional on rejection and for grant conditional on amendment. Appendix D reports the fit by technology center. We organize the discussion around the four channels through which applicant size enters the model, since these are the components the counterfactuals in Section 8 replace, and we close with the round-1 cutoff that combines them.

Allowance standard and search gap. Examiner characteristics move the standard in the expected direction. A one-standard-deviation increase in examiner leniency lowers τ_i by 3.2 units, about two-thirds of the calibrated dispersion $\sigma_\tau = 5$; a one-standard-deviation increase in experience lowers it by 0.6 units. Measures of prosecution resources are associated with a lower standard, and more so for small firms: representation at filing lowers τ_i by 0.5 units for large-firm applications and by 0.9 for small-firm applications, and university affiliation lowers it by 0.8 and 1.5 units, respectively. Correspondents' and inventors' prior patenting rates also lower the standard, although by less for small firms. The standalone size coefficient, $\beta_s = 0.37$, is not the size difference in the standard: that difference depends on the interactions and on the covariate distributions of the two groups, and the counterfactuals in Section 8.3 measure it directly. These coefficients combine examiner stringency with application attributes that D_i does not capture. Examiners may be more lenient toward large firms for several reasons, including getting hired by them later (Tabakovic and Wollmann, 2018).

Search varies little with applicant size. The latent search-gap location is 0.19 units lower for small firms ($\lambda_s = -0.19$), against an intercept of 2.15 and a dispersion of $\sigma_s = 1.31$: the prior art examiners cite against small-firm applications is slightly closer to the closest art in our corpus. The correlation between the search and standard shocks is positive but small ($\rho_{s\tau} = 0.12$).

Amendment costs and revision improvement. Applicant size matters most on the cost side, which can reflect access to better legal counsel (de Rassenfosse et al., 2023). The amendment-cost index is 1.30 log points higher for small firms ($c_s = 1.30$), with a further 0.57 for micro entities, and it falls with the share of patent attorneys among counsel of record ($c_r = -0.93$). The conditional median cost at each group's average index is about \$2,600 for large-firm and \$11,900 for small-firm applications.

These figures measure the full burden of continuing prosecution—attorney fees, applicant time, and organizational costs—not USPTO fees alone (Appendix E). Together with the patent values reported in Appendix F, they imply a baseline amendment probability q_i^0—the probability of amending when any amendment guarantees allowance—of 0.953 for large firms and 0.873 for small firms.

Revision improvement differs less by applicant size. Expected improvement, $1/\omega_i$, averages

4.6 units for large-firm and 4.3 for small-firm applications, about 7 percent lower for small firms and in both cases comparable to one standard deviation of the latent standard. The size difference depends on prior-art distance: because $\gamma_D < 0$ and $\gamma_{sD} \approx -\gamma_D$, expected improvement rises with D_i for large firms but is flat for small firms.[20]

The round-1 cutoff. The cost and improvement estimates determine the examiner's option value of a non-final rejection. The implied discount factor averages 0.65, and the cutoff κ_i in (15) averages -7.8 for large-firm and -5.5 for small-firm applications. Because small firms amend less often and improve less when they do, an examiner gains less from rejecting a borderline small-firm application. The difference matters for initial screening: 58 percent of model-implied round-1 rejections of large-firm applications fall in the reject-to-improve region $\kappa_i < \widetilde{G}_i \leq 0$, where the application would have been allowed at $\kappa_i = 0$, compared with 36 percent for small firms. This asymmetry helps reconcile the descriptive evidence in Section 4: the stronger reject-to-improve motive for large firms pushes their round-1 rejection rate toward that of small firms, even though large firms obtain patents more often.

8 Counterfactual Analysis

We use the estimated model to quantify how a change in the prosecution environment affects small-firm grant rates. Throughout, we hold fixed the observed application characteristics, predicted patent values, and examiner characteristics, and change only selected size-dependent components of the structural model. The examiner's round-1 cutoff κ_i is recomputed in every counterfactual because changes in applicants' amendment decisions or revision improvement affect the examiner's round-1 decision.

8.1 Large Firms Prosecuting Small-Firm Inventions

Applicant size enters the model through the allowance standard τ_i, the search-gap location λ_i, the amendment-cost index c_i, and the improvement rate ω_i. In our first counterfactual, we set the small-firm indicator and its interactions to zero in all four indices for each small-

[20] One interpretation is that greater distance from prior art leaves more scope to differentiate claims, but exploiting it requires prosecution expertise that small firms are less likely to have.

Table 4: Joint Counterfactual: Large-Firm Parameters for Small-Firm Applications

Outcome	Small Baseline	Joint Counterfactual	Large Baseline
Round-1 rejection	0.8546	0.8659	0.8292
Amendment | rejection	0.7951	0.9118	0.9334
Grant | amendment	0.6567	0.7391	0.7969
Final grant	0.5916	0.7176	0.7875
Expected net private value ($000)	205.4	256.5	303.7

Notes: The joint counterfactual replaces the small-firm indicator and its interactions in all four prosecution indices for small-firm applications. All remaining observed characteristics are held fixed, and the examiner's cutoff κ_i is recomputed in every counterfactual. Entries are model-implied probabilities averaged over applications in the indicated group; conditional rates are ratios of the corresponding averaged joint probabilities (Appendix I.1). Expected net private value is conditional on filing and net of expected future amendment costs; see Appendix I.2.

firm application in the structural-estimation sample and recompute the examiner's round-1 cutoff κ_i. Everything else stays at its baseline value. Hence, the counterfactual changes the prosecution environment for a fixed set of applications. We refer to it as the joint counterfactual, in contrast to the one-channel counterfactuals in Section 8.3. Appendix I.1 gives implementation details, and Online Appendix OA.8 reports the mirror-image counterfactual for large firms.

Table 4 reports model-implied outcome probabilities averaged over applications in each group. Conditional rates are ratios of averaged joint probabilities, so they reflect composition changes across prosecution stages. In the structural-estimation sample, the model-implied grant rate is 59.2 percent for small firms and 78.8 percent for large firms, a gap of 19.6 percentage points. Under the joint counterfactual, the small-firm grant rate rises to 71.8 percent, closing 12.6 points, or about 64 percent, of the gap; the increase is about 21 percent of the small-firm baseline. The remaining 7 points reflect everything held fixed: differences between the two groups in prior-art distance, patent value, representation, prosecution histories, claim language, examiner assignment, and technology, together with the size terms left at baseline. The counterfactual does not apportion this residual among those sources.

The gain comes from both post-rejection margins. The amendment rate conditional on rejection rises from 79.5 to 91.2 percent, and the grant rate conditional on amendment rises from 65.7 to 73.9 percent. The round-1 rejection rate also rises, from 85.5 to 86.6 percent. This increase is a net effect: the lower standard reduces initial rejections, but

better continuation prospects raise the examiner's option value of a non-final rejection by more. Hence, applicant gains coexist with more initial rejections.

8.2 Expected Private Value

We next express the prosecution outcomes in dollars. For each application we compute its expected net private value, W_i: the value of a patent, weighted by the probability of allowance, net of expected amendment costs, conditional on the application having been filed. The calculation uses the patent values estimated in Appendix F and the prosecution probabilities implied by the model; Appendix I.2 gives the formula. W_i is an applicant-side measure, not social welfare. It excludes filing and examination costs, the cost of any intervention that would bring about the counterfactual change, and effects on competitors, consumers, and follow-on innovation. It also holds each application's value conditional on grant fixed, so it does not capture changes in value from narrower claims or longer prosecution, and it takes the set of filed applications and their claims as given.

At baseline, expected net private value averages \$205,379 per small-firm application and \$303,722 per large-firm application, a gap of \$98,343, or 32 percent of the large-firm value. The gap is far larger than the 6-percent difference in expected patent value at grant (Appendix F) because small-firm applications are granted less often and cost more to prosecute: small firms realize about 54 cents of expected net private value per dollar of expected patent value at grant, compared with 75 cents for large firms.

Under the joint counterfactual, expected net private value for small firms rises to \$256,463, a gain of \$51,084 per application, or 25 percent of baseline. The gain closes about 52 percent of the dollar gap, less than the 64 percent of the grant-rate gap, mainly because the lower patent value of small-firm applications is held fixed and scales every percentage point of additional grant probability.

8.3 One-Channel Counterfactuals

Which of the four channels drives the joint counterfactual? Table 5 reports counterfactuals that set the small-firm terms in one index to zero at a time, holding the other three indices and all covariates at baseline and recomputing κ_i. Equalizing the allowance standard raises

Table 5: One-Channel Counterfactuals for Small-Firm Applications

Outcome	Small Baseline	Small firms with large-firm component			
		Allowance Standard	Search Gap	Amendment Cost	Revision Improvement
Round-1 rejection	0.8546	0.7825	0.8539	0.8978	0.8767
Amendment | rejection	0.7951	0.8129	0.7943	0.8951	0.8040
Grant | amendment	0.6567	0.7231	0.6538	0.6566	0.6739
Final grant	0.5916	0.6774	0.5896	0.6298	0.5983
Expected net private value ($000)	205.4	234.4	205.7	225.3	207.5

Notes: Each counterfactual replaces one size-dependent structural component with its large-firm counterpart, holding remaining observed characteristics fixed and recomputing κ_i. These effects are not additive shares of Table 4 and do not identify the causal origin of a channel. Entries are model-implied probabilities averaged over applications in the indicated group; conditional rates are ratios of the corresponding averaged joint probabilities (Appendix I.1). Expected net private value is conditional on filing and net of expected future amendment costs (Appendix I.2).

the small-firm grant rate by 8.6 percentage points and expected net private value by $29,088 per application; equalizing amendment costs raises them by 3.8 points and $19,913. Revision improvement and the search gap contribute little: 0.7 points and $2,162, and −0.2 points and $317, respectively. The one-channel effects sum to about 13 points, close to the joint effect of 12.6, but they are not an additive decomposition, because changing one channel alters selection into later prosecution stages and the examiner's round-1 cutoff. The mirror-image counterfactual for large firms yields the same ranking (Online Appendix OA.8).

The two main channels operate at different stages. The allowance standard shifts the distribution of the assessed patentability gap, so it affects both the round-1 screen and the shortfall that an amendment must close: equalizing it lowers the round-1 rejection rate by 7.2 points and raises the grant rate conditional on amendment from 65.7 to 72.3 percent. This channel collects the standalone size coefficient and the size interactions with university affiliation, claim language, representation, and prosecution histories in equation (16); as discussed in Section 6.3, it combines differences in examination with application attributes that measured distance omits. University applicants are classified as small firms, but they often have access to institutional prosecution resources. Assigning university affiliation to all small-firm applications lowers their allowance standard and raises their grant rate by 7.0 percentage points, closing about 35 percent of the gap. We interpret these results with caution because the university association may reflect examiner treatment, invention selection, or unobserved quality (Appendix G).

The amendment-cost channel acts on the amendment decision, raising the amendment rate from 79.5 to 89.5 percent while leaving the grant rate conditional on amendment unchanged. Its effect is not confined to post-rejection prosecution: a higher amendment probability raises the examiner's option value of a non-final rejection, and the round-1 rejection rate rises by 4.3 points. Equalizing revision improvement raises it by 2.2 points for the same reason. These offsetting round-1 responses are why rejection rates alone reveal neither the allowance standard nor the consequences of better continuation prospects. The cost channel also accounts for a larger share of the dollar gain than of the grant-rate gain: \$19,913 of the \$51,084 joint gain (39 percent) versus 3.8 of the 12.6 points (30 percent). The difference arises because lower costs also benefit applicants who would amend even at baseline, whose grant outcomes do not change.

Lastly, the channel composition varies across technology fields. Appendix H repeats the one-channel counterfactuals within each technology center, re-estimating the model with technology-center interactions in the allowance-standard and amendment-cost equations. The model-implied gap ranges from 9.8 percentage points in biotechnology to 25.8 in software. Equalizing amendment costs has virtually no effect in biotechnology but raises the small-firm grant rate by 2.8 points in chemicals and 6.5 in software, while the allowance-standard contribution ranges from 4.4 points in chemicals to 13.8 in software. The channels jointly close 71 to 88 percent of the within-center gap, except in transportation and mechanical engineering, where small firms are most concentrated and about 40 percent of the gap remains. Hence, similar aggregate gaps can reflect different mechanisms at play, so the same intervention may have different effects across fields.

8.4 Policy Implications

The counterfactuals carry two messages for policy. First, the amendment-cost channel maps to interventions that exist. Equalizing it raises the small-firm grant rate by 3.8 percentage points and expected net private value by about \$20,000 per application. Conversely, imposing the small-firm cost index on large firms lowers their grant rate by 3.1 points (Online Appendix OA.8). This suggests that revenue from higher large-firm fees could help finance assistance that lowers small-firm response costs, at a comparatively modest cost to large firms. Matcham and Schankerman (2026) do not distinguish firms by their size, but in their counterfactual exercise, a \$25,000 per-round fee lowers total social costs, and the

gains grow if the revenue is reinvested in more intensive examination. Our results suggest that the reinvestment should be redistributive, reducing small-firm fees and subsidizing legal services. Second, the other important channel, the allowance standard, is not a direct policy lever as estimated. It combines examiner stringency with application attributes that measured distance omits, and its size interactions with representation, university affiliation, and prosecution histories suggest that part of it reflects prosecution resources rather than examination.

Fee discounts alone cannot close the cost gap. Small firms already paid half of USPTO fees in 2014, including $600 rather than $1,200 for a first RCE, yet their estimated median amendment cost exceeds that of large firms by about $9,300. The estimated cost is the full burden of continuing prosecution, including legal expenses, applicant time, and organizational effort, which the model does not separate. The USPTO's fee study reaches a similar conclusion for filing decisions, finding that legal costs rather than fees are the main barrier for small applicants (U.S. Patent and Trademark Office, 2024). The more promising policies are therefore interventions that lower the cost of an effective response, such as subsidized or pro bono counsel and assistance in preparing responses.[21] Although the model does not evaluate any specific program, it provides a benchmark against which program costs can be weighed: equalizing amendment costs raises small firms' expected net private value by about $20,000 per application, with gains varying across technology fields.

The model also has implications for how such programs should be evaluated. Because examiners anticipate applicants' responses, an intervention that improves amendment prospects raises the option value of a non-final rejection: in the cost counterfactual, round-1 rejections rise by 4.3 points even as final grants rise. An evaluation that stops at the first office action would thus misread the effect. Applications must be followed to final disposition, with claim scope and program cost measured alongside grants. A higher grant rate is not by itself a welfare gain, which also depends on competition and follow-on innovation.

Finally, AI tools now used by examiners and practitioners could move several of the model's primitives. Examiner search tools change the references located (λ_i); applicant tools lower response costs (c_i), improve responses (ω_i), or change the application filed, which lies outside

[21] Examples include expanding the USPTO's Patent Pro Bono Program and Law School Clinic Certification Program. Both programs can represent applicants through prosecution, but they are small, together filing about 370 applications per year, and serve mainly under-resourced individual inventors (U.S. Patent and Trademark Office, 2023). Expanding them would require funding beyond what volunteers and law schools now provide, which could come from higher large-firm fees.

our fixed-application counterfactuals. Our finding that the size-related search gap contributes little to the grant gap is not an estimate of a common improvement in search technology. The 2014 cohort predates these tools, so the estimates do not identify their effects or the current size of the gap. Whether the tools substitute for scarce expertise or complement experienced counsel is a question for future work, for which our model provides the building blocks.[22]

9 Conclusion

Small firms obtain patents on 58 percent of their applications, large firms on 77 percent. We model prosecution as a dynamic game in which the examiner's initial decision anticipates the applicant's response, and estimate it on the 2014 cohort using textual distance to prior art and to examiner-cited references.

We find that two channels account for nearly all of the counterfactual reduction in the patent-grant gap. The first is the allowance standard, which combines examiner stringency with application attributes that our text-based measure fails to capture. The second is amendment costs, which lead small firms to abandon rejected applications more often than large firms. These differences affect initial examination decisions: examiners reject borderline large-firm applications to elicit amendments, a motive that accounts for 58 percent of their round-1 rejections, compared with 36 percent for small firms. This asymmetry helps explain why round-1 rejection rates look similar even though grant rates do not.

Prosecuting small-firm applications under the parameters estimated for large firms increases their grant rate by 12.6 percentage points, closing 64 percent of the gap, and raises expected net private value by $51,084 per application. Equalizing the allowance standard yields the largest gains in both outcomes, followed by amendment costs. Each channel's importance varies across technology fields.

Two implications follow (Section 8.4). First, the cost gap is far too large for fee discounts to close, so policy should focus on lowering the cost of preparing an effective response. Because examiners also anticipate responses, a program that improves amendment prospects will raise

[22] USPTO, "Guidance on Use of Artificial Intelligence-Based Tools in Practice Before the United States Patent and Trademark Office," April 2024, identifies cost and access as potential benefits of AI tools.

initial rejections as well as final grants, so it must be judged at final disposition. Second, the model yields a prediction for AI-assisted search. If examiners could locate the closest prior art for every application, their cited references would coincide with the closest prior art in the corpus. Because examiners' search differs little by firm size, the model suggests that more complete search alone would do little to narrow the grant gap. AI tools may also change how examiners apply the allowance standard and how applicants prepare and amend their claims, but the model does not specify how AI adoption would affect these channels. Re-estimating the model on cohorts examined after AI adoption would show which channels have changed and whether the gap has narrowed relative to the 2014 cohort.

References

Aneja, A., O. Reshef, and G. Subramani (2026). Attrition and the Gender Patenting Gap. *Review of Economics and Statistics 108*(4), 1017–1026.

Arrow, K. J. (1962). Economic welfare and the allocation of resources for invention. In R. R. Nelson (Ed.), *The Rate and Direction of Inventive Activity: Economic and Social Factors*, pp. 609–626. Princeton University Press.

Arts, S., J. Hou, and J. C. Gomez (2021). Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures. *Research Policy 50*(2), 104144.

Becker, G. S. (1957). *The Economics of Discrimination*. University of Chicago Press.

Becker, G. S. (1993). Nobel Lecture: The Economic Way of Looking at Behavior. *Journal of Political Economy 101*(3), 385–409.

Blind, K., K. Cremers, and E. Mueller (2009). The influence of strategic patenting on companies' patent portfolios. *Research Policy 38*(2), 428–436.

Canay, I. A., M. Mogstad, and J. Mountjoy (2024). On the Use of Outcome Tests for Detecting Bias in Decision Making. *Review of Economic Studies 91*(4), 2135–2167.

Cappelli, R., M. Corsino, K. Laursen, and S. Torrisi (2023). Technological competition and patent strategy: Protecting innovation, preempting rivals and defending the freedom to operate. *Research Policy 52*(6), 104785.

Carley, M., D. Hegde, and A. Marco (2015). What is the probability of receiving a US patent? *Yale Journal of Law and Technology 17*, 203–223.

Cook, L. (2014). Violence and Economic Activity: Evidence from African American Patents, 1870-1940. *Journal of Economic Growth 19*(2), 221–257.

de Rassenfosse, G., W. E. Griffiths, A. B. Jaffe, and E. Webster (2021). Low-quality patents in the eye of the beholder: Evidence from multiple examiners. *Journal of Law, Economics, and Organization 37*(3), 607–636.

de Rassenfosse, G. and R. Hosseini (2020). Discrimination against foreigners in the US patent system. *Journal of International Business Policy 3*(4), 349–366.

de Rassenfosse, G. and A. B. Jaffe (2018). Are patent fees effective at weeding out low-quality patents? *Journal of Economics & Management Strategy 27*(1), 134–148.

de Rassenfosse, G., P. H. Jensen, T. Julius, A. Palangkaraya, and E. Webster (2023). Is the patent system an even playing field? The effect of patent attorney firms. *The Journal of Industrial Economics 71*(1), 124–142.

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

Frakes, M. D. and M. F. Wasserman (2017). Is the time allocated to review patent applications inducing examiners to grant invalid patents? Evidence from microlevel application data. *Review of Economics and Statistics 99*(3), 550–563.

Frakes, M. D. and M. F. Wasserman (2023). Investing in ex ante regulation: Evidence from pharmaceutical patent examination. *American Economic Journal: Economic Policy 15*(3), 151–183.

Galasso, A. and M. Schankerman (2018). Patent rights, innovation, and firm exit. *RAND Journal of Economics 49*(1), 64–86.

Ganguli, I., J. Lin, V. Meursault, and N. F. Reynolds (2024). Patent text and long-run innovation dynamics: The critical role of model selection. Working Paper 32934, National Bureau of Economic Research.

Gaudry, K. S. (2012). The lone inventor: Low success rates and common errors associated with pro-se patent applications. *PLoS ONE 7*(3), e33141.

Gaule, P. (2018). Patents and the success of venture-capital backed startups: Using examiner assignment to estimate causal effects. *The Journal of Industrial Economics 66*(2), 350–376.

Goldsmith-Pinkham, P., P. Hull, and M. Kolesár (2026). Leniency designs: An operator's manual. *Journal of Economic Perspectives 40*(3), 213–240.

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

Guellec, D., C. Martinez, and P. Zúñiga (2012). Pre-emptive patenting: securing market exclusion and freedom of operation. *Economics of Innovation and New Technology 21*(1), 1–29.

Hegde, D., K. Herkenhoff, and C. Zhu (2023). Patent publication and innovation. *Journal of Political Economy 131*(7), 1845–1903.

Kelly, B., D. Papanikolaou, A. Seru, and M. Taddy (2021). Measuring technological innovation over the long run. *American Economic Review: Insights 3*(3), 303–320.

Lanjouw, J. O. and M. Schankerman (2004). Protecting intellectual property rights: are small firms handicapped? *Journal of Law and Economics 47*(1), 45–74.

Lee, J.-S. and J. Hsiang (2020). Patent classification by fine-tuning BERT language model. *World Patent Information 61*, 101965.

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

Lu, Q., A. F. Myers, and S. Beliveau (2017, November). USPTO patent prosecution research data: Unlocking office action traits. USPTO Economic Working Paper 2017-10, United States Patent and Trademark Office.

Matcham, W. and M. Schankerman (2026). Screening property rights for innovation. *Econometrica 94*(5), 1849–1886.

Pakes, A. (1986). Patents as options: Some estimates of the value of holding European patent stocks. *Econometrica 54*(4), 755–784.

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

Righi, C. and T. Simcoe (2023). Patenting inventions or inventing patents? Continuation practice at the USPTO. *The RAND Journal of Economics 54*(3), 416–442.

Schuett, F. (2013). Patent quality and incentives at the patent office. *RAND Journal of Economics 44*(2), 313–336.

Schuster, W. M. (2023). The patent examiner sweepstakes. *American Business Law Journal 60*(3), 599–650.

Shea, J. (2026). Testing for racial bias in police traffic searches. *University of Illinois, Champaign Urbana, USA*.

Silveira, B. S. (2017). Bargaining with asymmetric information: An empirical study of plea negotiations. *Econometrica 85*(2), 419–452.

Tabakovic, H. and T. G. Wollmann (2018). From revolving doors to regulatory capture? Evidence from patent examiners. Working Paper 24638, National Bureau of Economic Research.

U.S. Patent and Trademark Office (2023). Unleashing American Innovators Act of 2022: Study of the patent pro bono programs. Report to Congress, USPTO.

U.S. Patent and Trademark Office (2024). Unleashing American Innovators Act of 2022: Fee study summary report. Report to Congress, USPTO.

A Structural-Estimation Sample

The structural-estimation sample contains applications filed in 2014 whose first independent claim is rejected under §§ 102–103 and for which the distance to examiner-cited prior art is available, $R_{i,1} = 1$, together with applications receiving no round-1 rejection on any claim, $R_{i,1} = 0$.

We exclude three groups: applications whose first independent claim is not rejected although another claim is rejected; applications whose first independent claim is rejected under §§ 102–103 but for which the distance to examiner-cited prior art cannot be constructed (e.g., when the relevant reference falls outside the U.S. patent and pre-grant-publication corpus used to construct our embeddings); and applications whose first-independent-claim rejection is based only on other statutory grounds, such as § 101 or § 112.

The distance measures refer to the first independent claim, whereas patent applications typically contain multiple claims and may receive rejections on multiple statutory grounds. Incorporating the excluded cases would require a claim-level prosecution model and additional prior-art information not available in our data. The descriptive patterns underlying the grant gap remain present in the structural-estimation sample. In particular, the grant rates for small and large firms are 58 percent and 78 percent, respectively, in the structural-estimation sample. They are 58 percent and 77 percent in the 2014 application universe and 59 percent and 76 percent among applications in the 2014 application universe excluded from the structural-estimation sample.

B Variable Definitions

This section defines the covariates used in the structural model; the outcomes and distances ($R_{i,1}$, a_i, $R_{i,2}$, D_i, P_i) are defined in Section 6. For examiner characteristics, we use leniency_i and experience_i, as defined in Section 3.1, each standardized to mean zero and unit variance.

The following variables are used for applicant characteristics. small_i indicates small firms (small or micro entities; see footnote 1), and micro_i indicates micro-entity status. foreign_i equals one if the application's correspondence address is outside the United States. rep_i equals one if an attorney or patent agent is of record at filing. attRatio_i is the ratio of patent attorneys among all patent agents and attorneys of record for the application in the

most recent PatEx snapshot;[23] it is set to zero when there is no patent attorney or agent and centered at its sample mean. lnAppC_i is the log of one plus the number of applications filed before 2014 that list the same correspondence contact, computed from PatEx, and rateC_i is the share of the correspondence contact's pre-2014 applications that resulted in a patent.[24] lnAppI_i is the log of one plus the mean number of pre-2014 applications across the application's inventors, computed from PatEx, and rateI_i is the ratio of the mean number of patents to the mean number of pre-2014 applications across the inventors. univ_i indicates a university-affiliated applicant.[25] Finally, hedge_i equals one if the independent claims contain hedged or indefinite qualifying language.[26]

C Calibrated Dispersion: σ_τ

Table 6 profiles the likelihood over σ_τ. The log likelihood improves monotonically in σ_τ, but the gains beyond $\sigma_\tau = 5$ are modest. The implied mean improvement $\mathbb{E}[e_i]$ is non-monotone and attains its minimum near $\sigma_\tau = 5$, at about 4.6 for large-firm and 4.3 for small-firm applications, below the sample mean of D_i of about six. As σ_τ increases from 5 to 20, $E[e_i]$ roughly triples and κ_i falls from about -7.8 for large-firm applications to -33.0. Higher values of σ_τ thus yield implausibly large estimates for both $\mathbb{E}[e_i]$ and κ_i, so we set $\sigma_\tau = 5$.

Moreover, our counterfactual results are robust to different values of σ_τ. When $\sigma_\tau = 5$, the small-firm grant rate rises to 0.718 under the joint counterfactual (Table 4) and to 0.677 under the allowance-standard channel counterfactual (Table 5). When we re-estimate the model using alternative values of σ_τ, the corresponding grant rates are 0.728 and 0.663 for $\sigma_\tau = 3$; 0.716 and 0.679 for $\sigma_\tau = 6$; and 0.712 and 0.676 for $\sigma_\tau = 10$.

[23] For 2014 applications, this reflects representation over the course of prosecution rather than at filing.

[24] The correspondence contact is the person or organization designated to receive USPTO communications about the application, such as the applicant, a patent attorney or agent, or a law firm.

[25] The indicator equals one if any assignee listed on the application is classified as an academic or research institution, based on keyword matching of disambiguated assignee organization names: universities (including non-English variants), colleges, institutes, polytechnics, academies, research foundations and research centers, national laboratories, university governance entities (e.g., "Regents of the University of ..."), and named research organizations such as the Fraunhofer and Max Planck Societies.

[26] These include "substantially," "approximately," "about," "essentially," "generally," "preferably," "optionally," "at least one of," or "and/or," which are detected by pattern matching on claim text. The USPTO's examination manual groups such terms under relative terminology in its definiteness guidance (MPEP § 2173.05(b)); they are not inherently indefinite, so the indicator proxies for drafting style rather than a finding of indefiniteness.

Table 6: Likelihood Profile over σ_τ

		$\mathbb{E}[e_i]$		Median cost ($)		Mean κ_i	
σ_τ	NLL	Large	Small	Large	Small	Large	Small
1	580,010.1	10.72	6.61	3,366	18,553	−0.51	−0.23
3	573,930.0	6.72	5.47	3,091	15,744	−3.56	−2.25
4	572,530.6	4.74	4.23	2,785	13,250	−5.89	−4.07
5	571,754.8	4.61	4.26	2,588	11,881	−7.82	−5.53
6	571,319.2	4.89	4.59	2,497	11,198	−9.62	−6.87
10	570,685.6	7.37	7.00	2,395	10,545	−16.35	−11.76
20	570,429.1	14.18	13.49	2,355	10,211	−32.97	−23.76

Notes: Each row re-estimates the model at the indicated σ_τ. NLL is the negative log likelihood. $\mathbb{E}[e_i]$ is the implied mean improvement in standardized distance units, and the median cost is the implied conditional median amendment cost at each group's average index. Mean κ_i is the average dynamic rejection cutoff.

D Model Validation: Goodness of Fit

Table 7 compares observed and predicted prosecution outcomes by firm type and technology center. The model closely matches the three margins. In the full structural-estimation sample, the differences between observed and predicted rates are 0.002 for round-1 rejection and smaller than 0.001 in absolute value for amendment conditional on rejection and for grant conditional on amendment. No firm-size-specific discrepancy exceeds 0.004. The model also fits these outcomes well across technology centers.

E Amendment Costs: Scale and Interpretation

Under $\ln C_i^R = c_i + \sigma_c \varepsilon_i$, with $\varepsilon_i \sim \mathcal{N}(0,1)$, amendment probabilities imply $c_i = \ln V_i - \sigma_c \Phi^{-1}(q_i^0)$, so the scale determines the cost levels needed to fit the observed amendment behavior. The estimated scale passes an external plausibility check. With V_i measured in \$10,000 units, the conditional median cost is $10{,}000 \exp(c_i)$, yielding approximately \$2,600 for large-firm and \$11,900 for small-firm applications at their sample-average indices.

For comparison, Matcham and Schankerman (2026) estimate per-round legal costs of \$2,143–\$3,737 at the mean and \$3,201–\$6,002 at the ninetieth percentile, depending on technological complexity. Our large-firm median falls within their mean range, consistent with direct costs dominating absent financing constraints; the small-firm median exceeds both ranges, consistent with additional non-legal burdens. Our cost measure covers the aggregated continuation decision, so differences from per-round estimates need not reflect different attorney prices.

Table 7: Goodness of Fit by Technology Center

Technology center	N	Actual	Predicted	Gap
Panel A. Round-1 rejection rate				
TC 2800: Semiconductors, Electrical and Optical Systems	58,704	0.7498	0.7487	+0.0011
TC 1600: Biotechnology and Organic Chemistry	14,833	0.8054	0.8038	+0.0016
TC 1700: Chemical and Materials Engineering	29,603	0.8690	0.8671	+0.0019
TC 2100: Computer Architecture and Software	20,201	0.8797	0.8774	+0.0023
TC 2400: Networking, Multiplexing, Cable, and Cybersecurity	26,659	0.8882	0.8865	+0.0018
TC 2600: Communications Technology	30,224	0.8456	0.8430	+0.0027
TC 3600: Transportation, Agriculture	32,040	0.8657	0.8631	+0.0027
TC 3700: Mechanical Engineering and Medical Devices	37,896	0.8703	0.8671	+0.0032
Overall	250,160	0.8371	0.8350	+0.0021
Large firms (small = 0)	193,153	0.8308	0.8292	+0.0016
Small firms (small = 1)	57,007	0.8585	0.8546	+0.0039
Panel B. Amendment rate among rejected applications				
TC 2800: Semiconductors, Electrical and Optical Systems	44,017	0.9198	0.9192	+0.0006
TC 1600: Biotechnology and Organic Chemistry	11,946	0.8647	0.8632	+0.0015
TC 1700: Chemical and Materials Engineering	25,724	0.8943	0.8928	+0.0015
TC 2100: Computer Architecture and Software	17,770	0.9190	0.9206	−0.0016
TC 2400: Networking, Multiplexing, Cable, and Cybersecurity	23,679	0.9277	0.9289	−0.0012
TC 2600: Communications Technology	25,558	0.9281	0.9296	−0.0015
TC 3600: Transportation, Agriculture	27,738	0.8669	0.8688	−0.0019
TC 3700: Mechanical Engineering and Medical Devices	32,981	0.8710	0.8715	−0.0005
Overall	209,413	0.9007	0.9011	−0.0004
Large firms (small = 0)	160,471	0.9332	0.9334	−0.0001
Small firms (small = 1)	48,942	0.7939	0.7951	−0.0012
Panel C. Grant rate conditional on amendment				
TC 2800: Semiconductors, Electrical and Optical Systems	40,488	0.8464	0.8451	+0.0013
TC 1600: Biotechnology and Organic Chemistry	10,330	0.6551	0.6579	−0.0028
TC 1700: Chemical and Materials Engineering	23,005	0.7042	0.7096	−0.0054
TC 2100: Computer Architecture and Software	16,331	0.7598	0.7581	+0.0017
TC 2400: Networking, Multiplexing, Cable, and Cybersecurity	21,967	0.7917	0.7910	+0.0007
TC 2600: Communications Technology	23,721	0.8052	0.8037	+0.0015
TC 3600: Transportation, Agriculture	24,045	0.7289	0.7287	+0.0002
TC 3700: Mechanical Engineering and Medical Devices	28,725	0.7395	0.7373	+0.0022
Overall	188,612	0.7683	0.7680	+0.0003
Large firms (small = 0)	149,758	0.7972	0.7969	+0.0003
Small firms (small = 1)	38,854	0.6569	0.6567	+0.0002

Notes: Gap is Actual minus Predicted. Positive values indicate underprediction by the model; negative values indicate overprediction. The sample sizes in Panels B and C are, respectively, round-1 rejections and amended applications.

The cost C_i^R captures the effective burden of continuing prosecution: counsel fees, inventors' and managers' time, the opportunity cost of committed funds, and the shadow cost of external finance. The model does not separately identify these components or financing constraints. The higher small-firm cost reproduces their lower amendment rate; it does not imply higher fees for identical legal work.

F Patent Value Estimates

Table 8 reports estimates from the patent value model described in Online Appendix OA.6. The model specifies latent annual flow value as lognormal and estimates its parameters from renewal decisions. A coefficient η_k implies that a unit increase in the covariate changes flow value by $(e^{\eta_k} - 1) \times 100$ percent.

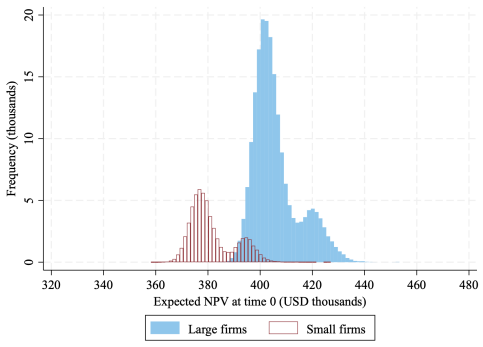
Table 8: Patent Value Model

Parameter	Estimate	SE
η_0 (intercept)	1.0135	0.0010
η_s (small)	−0.0656	0.0010
η_d (prior-art distance D_i)	−0.0179	0.0010
σ_v	0.6207	0.0006
ψ (effective decay rate)	0.0459	0.0002

Notes: The specification includes WIPO technology-field effects. Standard errors are based on 500 bootstrap samples. Table OA.7 in the Online Appendix reports all the estimates for all the parameters.

Small-firm patents have estimated flow values 6.3 percent below large-firm patents ($\eta_s = -0.066$). Flow value declines with measured prior-art distance: a one-standard-deviation increase in the standardized distance D_i lowers it by 1.8 percent ($\eta_d = -0.018$). Private values are highly dispersed: a one-standard-deviation shock raises value by 86 percent. The estimated decay rate $\psi = 0.046$ implies a 4.5 percent probability of the payoff stream ending within a year and a half-life of $\ln 2/\psi \approx 15$ years; because the value model contains no separate discount rate, ψ is an effective decay parameter rather than a separately identified rate of technological obsolescence.

Figure 3: Distribution of Patent Value by Firm Type



Notes: Ex-ante expected net present private value of a patent at age zero, defined in the Online Appendix, equation (OA.5).

Figure 3 shows substantial overlap in patent values across firm types. Mean expected NPV at grant is $405,913 for large-firm and $381,472 for small-firm applications, a difference of about 6 percent.

G University Affiliation

University affiliation is associated with a lower allowance standard, especially among small applicants. This association may reflect institutional prosecution infrastructure, selection of inventions, unmeasured application quality, or examiner treatment. Nevertheless, we report the magnitude of the association under the strong assumption that the entire university coefficient is portable to other small applicants.

We set the university indicator to one for all small-firm applications and recompute every index in which university terms appear. Under the estimated specification, university affiliation enters only equation (16). Treated applications receive a 1.47-unit reduction in the allowance standard. Applications already affiliated with universities (11.7 percent of the small-firm sample) and all large-firm applications remain unchanged.

Table 9: University-Affiliation Counterfactual

	Small, all		Small, non-university		Large
	Baseline	Counterfactual	Baseline	Counterfactual	Baseline
Round-1 rejection	0.8546	0.8010	0.8681	0.8074	0.8292
Amendment | rejection	0.7951	0.8111	0.7903	0.8082	0.9334
Grant | amendment	0.6567	0.7116	0.6433	0.7062	0.7969
Final grant	0.5916	0.6613	0.5732	0.6534	0.7875
Expected net private value ($000)	205.3	229.3	199.0	226.1	303.7

Notes: The counterfactual sets the university indicator to one for all small-firm applications, holding all other covariates and estimated parameters at baseline values; large firms are unchanged. "Small, non-university" is the treated subgroup. The reported later-stage rates are evaluated on the observed rejection and amendment samples, as in Table 7. Patent value is computed with common simulation draws. The exercise switches a covariate, not a structural primitive, and is not additive with Table 4.

Table 9 reports the results. The small-firm final grant rate rises by 7.0 percentage points, from 0.592 to 0.661, narrowing the grant gap from 19.6 to 12.6 percentage points, about 35 percent of the baseline gap; among the treated subgroup the increase is 8.0 points, and expected net private value rises by $27,148 per treated application.

H Decomposition by Technology Center

Table 10 repeats the one-channel counterfactuals of Section 8.3 within each technology center. For these technology-level results, the model is re-estimated with the technology-center dummies interacted with the small-firm indicator in equations (16) and (18), so that each center has its own size effects in the two key channels. The aggregate gap closes by 0.63 versus 0.64 in the baseline specification of Table 4.

Table 10: Final Grant Rate by Technology Center: Channel Counterfactuals for Small Firms

TC	Area	Small share	Small baseline	Allowance standard	Amend. cost	Revision improv.	All	Large baseline	Gap (pp)	Share of gap closed
TC 2800	Semiconductors	0.146	0.7217	0.7976	0.7613	0.7221	0.8318	0.8523	13.1	0.84
TC 1600	Biotech	0.407	0.5969	0.6672	0.5966	0.6020	0.6699	0.6945	9.8	0.75
TC 1700	Chemical	0.206	0.5979	0.6419	0.6255	0.6043	0.6735	0.7037	10.6	0.71
TC 2100	Software	0.140	0.5122	0.6504	0.5768	0.5234	0.7187	0.7703	25.8	0.80
TC 2400	Networking	0.139	0.5779	0.6830	0.6330	0.5875	0.7413	0.7944	21.6	0.75
TC 2600	Communications	0.153	0.6245	0.7483	0.6708	0.6309	0.7924	0.8159	19.1	0.88
TC 3600	Transportation	0.344	0.5301	0.6218	0.5762	0.5399	0.6726	0.7638	23.4	0.61
TC 3700	Mechanical	0.372	0.5647	0.6432	0.6011	0.5738	0.6845	0.7653	20.1	0.60
Overall		0.228	0.5917	0.6773	0.6278	0.5988	0.7148	0.7876	19.6	0.63

Notes: Model-implied final grant rates for small-firm applications by technology center, from the specification with technology-center dummies interacted with the small-firm indicator in the allowance-standard and amendment-cost equations. "Small share" is the number of small-firm applications divided by all applications in the technology center in the structural-estimation sample. The four counterfactual columns correspond to those of Table 4: each channel column sets the small-firm components of that channel, including its technology-center interactions, to their large-firm values, holding the others fixed; "All" sets all four simultaneously, including the search-gap channel, which is not reported separately because its effect is negligible in every technology center. "Gap" is the large-firm baseline minus the small-firm baseline within the technology center, in percentage points, and "Share of gap closed" is the all-channel increase divided by this gap. Full technology-center names are given in Table 7.

Allowing technology-specific size effects matters for the cross-technology comparison. The model-implied gap now ranges from 9.8 percentage points in biotechnology to 25.8 points in software, and the channel composition varies: the allowance-standard contribution runs from 4.4 points (chemical) to 13.8 points (software), and the amendment-cost contribution from essentially zero (biotechnology) to 6.5 points (software). The share of the gap the four channels close is 0.71 to 0.88 in six of the eight technology centers, with no systematic relation to the size of the gap. The exceptions are TC 3600 and TC 3700, where the channels close only about 60 percent of gaps of 23.4 and 20.1 points. Small-firm applications are also concentrated there, 34 and 37 percent of applications versus 14 to 21 percent in most other centers (TC 1600, at 41 percent, is the exception). Transportation and mechanical engineering are therefore the fields in which small firms are most prevalent and the residual composition component is largest. In the remaining centers, the prosecution channels account

for most of the gap. Because each share is computed relative to the model-implied gap within the technology center, these comparisons inherit the model's fit at the technology-center level.

I Counterfactual Construction

I.1 Counterfactual Implementation

The joint counterfactual sets the small-firm indicator to zero when evaluating the four prosecution indices for small-firm applications. All terms involving that indicator, including its interactions, are removed. The exercise leaves the other covariates fixed. Predicted patent value is also held fixed, including its estimated relationship with applicant size.

The examiner discount factor δ_i remains at baseline because size does not enter its equation. For every counterfactual, we recompute q_i^0 and the round-1 cutoff κ_i. Changes in costs and revision improvement feed back into initial examination.

The population outcome probabilities implied by a counterfactual can be written as

$$\begin{aligned} p_i^R &= \Pr(\widetilde{G}_i \geq \kappa_i), \\ p_i^A &= \mathbb{E}\Big[\mathbf{1}\{\widetilde{G}_i \geq \kappa_i\} q_i(\widetilde{G}_i)\Big], \\ p_i^{AG} &= \mathbb{E}\Big[\mathbf{1}\{\widetilde{G}_i \geq \kappa_i\} q_i(\widetilde{G}_i)\varphi_i(\widetilde{G}_i)\Big], \\ p_i^G &= 1 - p_i^R + p_i^{AG}. \end{aligned}$$

Here p_i^A is the joint probability of rejection and amendment, and p_i^{AG} is the joint probability of rejection, amendment, and eventual allowance. All expectations use the counterfactual indices and latent distribution. For a group of applications, the unconditional grant rate is the mean of p_i^G. Amendment conditional on rejection is $\sum_i p_i^A / \sum_i p_i^R$, and grant conditional on amendment is $\sum_i p_i^{AG} / \sum_i p_i^A$. These ratios account for changes in the composition reaching each stage.

Outcome probabilities are nonlinear and the cutoff adjusts endogenously, so the effect of switching several channels is not the sum of the corresponding one-at-a-time effects. Online Appendix OA.8 reports the reverse exercise for large-firm applications.

I.2 Expected Net Private Value

The expected net private value of a filed application, W_i, weights patent value V_i by the probability of allowance and subtracts expected future amendment costs. Let $\varphi_i(g) = \exp(-\omega_i g^+)$ denote the probability that continued prosecution ultimately succeeds, and let $q_i(g) = \Phi\left(\frac{\ln V_i - c_i - \omega_i g^+}{\sigma_c}\right)$ denote the probability of amendment. Conditional on X_i, define the truncated first moment of the amendment-cost distribution by

$$M_i(x) \equiv \mathbb{E}\left[C_i^R \mathbf{1}\{C_i^R \leq x\} \mid X_i\right].$$

For $\ln C_i^R \sim \mathcal{N}(c_i, \sigma_c^2)$,

$$M_i(x) = \exp\left(c_i + \frac{\sigma_c^2}{2}\right) \Phi\left(\frac{\ln x - c_i - \sigma_c^2}{\sigma_c}\right).$$

Expected net private value can therefore be written as

$$W_i = V_i \Pr(\widetilde{G}_i < \kappa_i) + V_i \mathbb{E}\left[\mathbf{1}\{\widetilde{G}_i \geq \kappa_i\} q_i(\widetilde{G}_i) \varphi_i(\widetilde{G}_i)\right] - \mathbb{E}\left[\mathbf{1}\{\widetilde{G}_i \geq \kappa_i\} M_i\left(V_i \varphi_i(\widetilde{G}_i)\right)\right].$$

The first (second) term is the expected patent value from immediate (eventual) allowance. The third term subtracts expected amendment costs, which are incurred only following rejection and only when the realized cost is low enough for the applicant to amend. The expectations integrate over latent search-gap and standard shocks. We approximate them using $K = 500$ draws per application and use common simulation draws across counterfactuals. Patent values and costs are expressed in units of \$10,000 in estimation and rescaled to dollars for reporting.

At baseline, mean W_i is \$205,379 for small firms and \$303,722 for large firms. Equalizing all four structural size components raises the small-firm mean to \$256,463. Table 4 reports the joint counterfactual; Table 5 reports the associated one-channel counterfactuals.

Online Appendix

Patent Grant Rates and Applicant Size

by Seung-Hyun Hong, Jorge Lemus, and Joshua Shea

Supplemental Material – Intended for Online Publication

OA.1 Comparing Claim and Description Embeddings

We compare distances based on claims and descriptions in prior-art documents, keeping the focal application's claim text fixed.

The exercise computes the distance between 1,000 granted patents with priority year 2012 (denoted j) and 100 published applications from 2014 (denoted i). To construct description-based embeddings for j, we split each description into 400-word chunks with 50-word overlap to accommodate the model's 512-token limit, yielding approximately 11,000 description chunks. For the claim-based measure, we use our existing claim-level embeddings.

Embeddings for each 2014 application i are computed using claim text, because examiners evaluate the patentability of the application's claims. For each patent–application pair, we compute the minimum cosine (chord) distance across all description chunks and, separately, the minimum cosine distance across all claims.

Figure OA.1: Claim-Text and Description-Text Distances

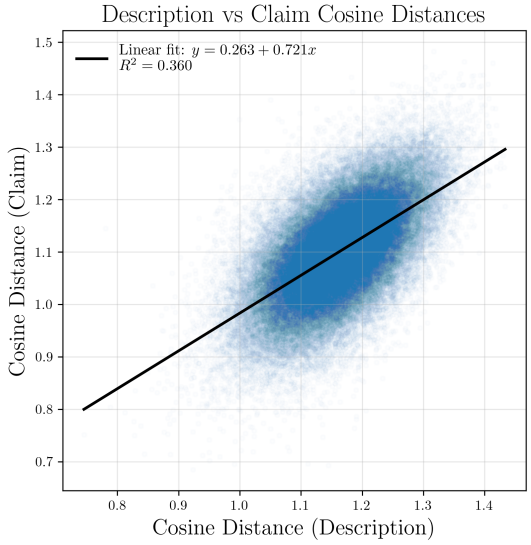


Figure OA.1 plots the claim-based distance against the description-based distance across pairs. The two measures are positively related, with a Spearman rank correlation of 0.58. The correlation is substantial but not tight, indicating that some pairwise rankings differ depending on whether similarity is measured using claims or descriptions. Overall, the figure supports using claim-based distances as a practical, computationally tractable proxy. It does not establish that either measure reproduces the legal assessment of the full disclosure.

OA.2 Additional Descriptive Statistics

This section reports additional descriptive evidence supporting Section 4 of the paper.

OA.2.1 Patent-Grant Gap Across Filing Cohorts

Table OA.1 reports the number of utility patent applications and grant rates by firm type for filing cohorts from 2010 to 2018. Large firms file roughly three times as many applications as small firms. Grant rates generally rise across these cohorts before declining in 2018, while the gap narrows from 20.2 to 14.4 percentage points and remains substantial.

Table OA.1: Applications and Grant Rates Across Filing Cohorts

	Large firms		Small firms	
Filing year	Applications	Grant rate	Applications	Grant rate
2010	242,311	0.738	81,808	0.536
2011	258,134	0.737	84,102	0.542
2012	277,413	0.750	87,461	0.560
2013	289,498	0.757	95,763	0.568
2014	295,739	0.770	102,127	0.582
2015	296,982	0.793	101,132	0.611
2016	297,463	0.808	101,439	0.645
2017	303,607	0.820	105,012	0.665
2018	304,070	0.790	108,472	0.646

Notes: The table reports the number of utility patent applications filed in each year and the fraction granted as of September 2023. Because later cohorts have had less time to complete prosecution, their observed grant rates may be affected by right censoring. Source: USPTO PatEx application data.

OA.2.2 Rejection Grounds

Table OA.2 reports rejection rates by statutory ground for large and small firms, in round 1 and across all rounds. An application is coded as rejected on a given ground if at least one independent claim receives a rejection on that ground; results based on any claim, independent or dependent, are nearly identical. Two patterns are relevant. First, the difference between any rejection and a §§ 102–103 rejection is narrow, so novelty and non-obviousness screening drives the bulk of prosecution. Second, § 112 rejections are substantially more prevalent for small firms, particularly in round 1, a pattern that the novelty and non-obviousness

model does not separately explain. Most applications receiving § 112 rejections, however, also receive a § 102 or § 103 rejection.

Table OA.2: Application-Level Fractions of Rejection Types

	Round 1		All rounds	
Rejection type	Large	Small	Large	Small
§ 101	0.12	0.14	0.14	0.16
§ 102	0.43	0.46	0.46	0.48
§ 103	0.44	0.46	0.56	0.56
§ 112	0.21	0.31	0.30	0.39
Any rejection	0.86	0.89	0.87	0.89
§§ 102–103	0.78	0.79	0.79	0.81

Notes: The sample consists of utility applications filed in 2014 and matched to pre-grant publications from 2014–2016 (the matched-text sample). An application is coded as rejected on a ground if at least one independent claim receives that rejection in the indicated examination round; results based on any claim are nearly identical. "Any rejection" includes §§ 101, 102, 103, and 112, double patenting, and objections. "§§ 102–103" indicates a rejection under either provision. "All rounds" indicates that the application receives the rejection at least once in any round.

In the raw data, small-firm applications receive slightly more round-1 rejections across most statutory grounds. This difference does not survive conditioning on examiner and technology composition: as shown in Table 2 of the paper, small-firm claims are slightly less likely to be rejected once these controls are included. Raw rejection rates therefore cannot by themselves identify differences in examination standards.

OA.2.3 Abandonment by Rejection Ground

Table OA.3 regresses an indicator for abandonment after round 1 or eventual abandonment on the type of rejection received in round 1, separately for large and small firms. The rejection-type indicators, their interactions with the small-firm indicator, and an intercept are entered jointly in a single linear probability model. Because an application may receive more than one rejection ground, the indicators are not mutually exclusive, and each coefficient measures the marginal association of that ground while holding the others fixed. Abandonment depends strongly on rejection type for small firms but less so for large firms. After a round-1 §§ 102–103 rejection, abandonment is 12.8 percentage points higher for small firms and 0.8 points higher for large firms, relative to applications not receiving that rejection. For § 112 rejections, the corresponding increments are 3.4 and 1.4 points. Applications receiving both grounds carry the sum of the corresponding increments. The patterns are

similar for eventual abandonment.

Table OA.3: Rejection Grounds and Abandonment

	Dependent variable: abandonment	
	After round 1	Eventually
Large × § 101 rejection	0.008	0.108
	(0.002)	(0.003)
Small × § 101 rejection	0.133	0.243
	(0.003)	(0.004)
Large × §§ 102–103 rejection	0.008	0.111
	(0.002)	(0.002)
Small × §§ 102–103 rejection	0.128	0.298
	(0.002)	(0.003)
Large × § 112 rejection	0.014	0.035
	(0.001)	(0.002)
Small × § 112 rejection	0.034	0.055
	(0.002)	(0.003)
Observations	324,103	324,103

Notes: Each rejection variable indicates whether any independent claim receives the specified rejection in round 1. All coefficients included in the regression, except for the constant, are reported in the table. Standard errors are in parentheses.

OA.2.4 Reduced-Form Regressions Using All Independent Claims

Table OA.4 replicates Table 2 of the paper using all independent claims rather than only the first independent claim. The estimates are nearly identical.

Table OA.4: Reduced-Form Results Using All Independent Claims

	Dependent variable:							
		Not rejected for		Allowance in		Amend af.	Dist. to	Embed.
	Patented	102–103	any	round 1	round 2+	round 1	prior art	controls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	
Small	−0.140	0.020	0.014	−0.003	−0.087	−0.119	0.023	×
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	
Small	−0.131	0.018	0.014	−0.004	−0.082	−0.112	0.014	✓
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	
Observations	889,723	889,723	889,723	889,723	714,198	795,298	889,723	

Notes: All specifications include examiner and USPC class–subclass fixed effects. "Embedding controls" indicates whether the specification additionally includes the 768-dimensional claim embedding. The sample contains all independent claims of utility applications filed in 2014 and matched to pre-grant publications from 2014–2016 (the matched-text sample). Dependent variables are defined as in Table 2 of the paper. Standard errors are in parentheses.

OA.3 Embeddings Summary Statistics

Table OA.5: Pre-Grant Publication Embeddings by Year

Year	PGPub	Claims	Mean	S.D.	Max	Year	PGPub	Claims	Mean	S.D.	Max
2001	56,412	211,561	3.750	3.766	145	2009	328,206	916,516	2.793	2.038	67
2002	199,001	811,182	4.076	6.373	393	2010	333,165	904,689	2.715	2.126	382
2003	237,079	1,029,595	4.343	9.528	565	2011	321,145	854,407	2.661	1.945	169
2004	268,359	1,047,386	3.903	4.402	421	2012	331,544	868,086	2.618	1.936	404
2005	289,577	979,107	3.381	3.299	531	2013	347,119	896,236	2.582	1.670	89
2006	294,512	928,215	3.152	3.077	581	2014	380,416	972,119	2.555	1.653	196
2007	300,164	908,838	3.028	2.477	121	2015	382,445	955,474	2.498	1.859	522
2008	320,469	927,109	2.893	2.229	139	2016	381,746	945,646	2.477	1.679	172

Notes: "PGPub" is the number of pre-grant publications with embeddings in each year, and "Claims" is the total number of independent claims. "Mean," "S.D.," and "Max" are the mean, standard deviation, and maximum of the number of independent claims per publication. The minimum is one in every year and is omitted.

Table OA.6: Patent Embeddings by Year

Year	Patents	Claims	Mean	S.D.	Max	Year	Patents	Claims	Mean	S.D.	Max
1976	70,602	141,436	2.003	1.749	39	2000	158,770	466,501	2.938	2.494	73
1977	65,767	133,267	2.026	1.756	48	2001	167,247	498,203	2.979	2.595	119
1978	66,592	136,819	2.055	1.856	77	2002	168,996	514,795	3.046	2.668	122
1979	49,267	101,146	2.053	1.776	53	2003	170,507	534,484	3.135	2.753	120
1980	62,199	129,235	2.078	1.901	70	2004	165,711	525,543	3.171	2.722	130
1981	66,311	138,533	2.089	1.902	69	2005	144,855	461,512	3.186	2.749	78
1982	58,304	123,286	2.115	1.947	79	2006	175,490	541,258	3.084	2.569	62
1983	57,415	119,605	2.083	1.813	46	2007	158,893	471,541	2.968	2.411	81
1984	67,711	142,144	2.099	1.776	57	2008	159,678	448,468	2.809	2.188	55
1985	72,176	152,114	2.108	1.756	42	2009	168,926	461,327	2.731	2.059	67
1986	71,513	151,650	2.121	1.747	43	2010	221,778	591,087	2.665	1.928	138
1987	83,619	181,633	2.172	1.763	67	2011	226,729	597,105	2.634	2.069	248
1988	78,766	175,002	2.222	1.850	58	2012	255,319	667,356	2.614	1.980	140
1989	96,616	223,722	2.316	1.941	48	2013	280,150	722,348	2.578	1.899	187
1990	91,242	214,647	2.353	1.992	95	2014	303,344	773,084	2.549	1.706	203
1991	97,312	236,111	2.426	2.082	81	2015	300,969	750,507	2.494	1.564	103
1992	98,259	242,907	2.472	2.108	80	2016	305,788	745,925	2.439	1.460	100
1993	99,287	246,271	2.480	2.072	65	2017	321,667	772,565	2.402	1.366	79
1994	102,609	260,773	2.541	2.127	76	2018	310,590	738,765	2.379	1.371	86
1995	102,238	265,491	2.597	2.155	63	2019	357,805	853,019	2.384	1.382	135
1996	110,391	295,659	2.678	2.234	84	2020	355,677	863,154	2.427	1.337	106
1997	112,764	307,266	2.725	2.286	94	2021	330,668	789,861	2.389	1.293	128
1998	148,488	420,504	2.832	2.369	84	2022	326,250	778,279	2.386	1.293	158
1999	154,498	445,866	2.886	2.431	70	2023	314,816	753,898	2.395	1.260	108

Notes: "Patents" is the number of granted patents with embeddings in each year, and "Claims" is the total number of independent claims. "Mean," "S.D.," and "Max" are the mean, standard deviation, and maximum of the number of independent claims per patent. The minimum is one in every year and is omitted.

OA.4 Equilibrium Proof and Payoff Microfoundation

OA.4.1 Proof of the Equilibrium Characterization

The proof proceeds by backward induction.

1. *Stage 3.* The examiner allows if and only if

$$U_{E,2}(A) \geq U_{E,2}(R) \quad \Longleftrightarrow \quad e_i - \widetilde{G}_i \geq 0,$$

   where $U_{E,t}(\cdot)$ denotes the examiner's payoffs defined in Online Appendix OA.4.2, which yields $R^*_{i,2} = \mathbf{1}\{e_i < \widetilde{G}_i\}$.
2. *Stage 2.* The applicant expects payoff $V_i \Pr(e_i \geq \widetilde{G}_i) - C_i^R = V_i \varphi_i(\widetilde{G}_i) - C_i^R$. Amendment is optimal if and only if $C_i^R \leq V_i \varphi_i(\widetilde{G}_i)$, establishing a_i^* and $q_i(\widetilde{G}_i)$.
3. *Stage 1.* The examiner compares $U_{E,1}(A) = -\widetilde{G}_i$ with the expected payoff from rejection,

$$U_{E,1}(R) = \delta_i q_i(\widetilde{G}_i) \mathbb{E}[\max\{e_i - \widetilde{G}_i, 0\}].$$

   For $\widetilde{G}_i > 0$, $U_{E,1}(A) < 0 \leq U_{E,1}(R)$, so rejection is strictly optimal. For $\widetilde{G}_i \leq 0$, $\varphi_i(\widetilde{G}_i) = 1$, $q_i(\widetilde{G}_i) = F_c(V_i)$, and $\mathbb{E}[\max\{e_i - \widetilde{G}_i, 0\}] = \mathbb{E}[e_i] - \widetilde{G}_i$. The examiner is indifferent at $\widetilde{G}_i = \kappa_i$, where

$$-\kappa_i = \delta_i F_c(V_i)(\mathbb{E}[e_i] - \kappa_i) \quad \Longrightarrow \quad \kappa_i = -\frac{\delta_i F_c(V_i) \mathbb{E}[e_i]}{1 - \delta_i F_c(V_i)}.$$

   On $\widetilde{G}_i \leq 0$, $U_{E,1}(A) - U_{E,1}(R)$ is strictly decreasing in $\widetilde{G}_i$. Together with strict rejection for positive gaps and the stated tie-breaking rule, this establishes rejection if and only if $\widetilde{G}_i \geq \kappa_i$.

OA.4.2 Microfoundation of the Examiner's Static Payoff

Suppose an application is patentable under the applicable standard ("good") with probability π and non-patentable ("bad") otherwise. The examiner receives v_H from allowing a good application and v_L from allowing a bad one, and receives r_H and r_L, respectively, from rejecting them. Assume $v_H > r_H$ and $r_L > v_L$, so correct dispositions are preferred.

Allowance is strictly optimal when

$$\pi v_H + (1-\pi)v_L > \pi r_H + (1-\pi)r_L,$$

which is equivalent to

$$\pi > \frac{r_L - v_L}{(v_H - r_H) + (r_L - v_L)} \equiv \tau^*. \tag{OA.1}$$

Suppose the probability that the application is good increases in its distance to prior art, $\pi = G(P)$, and take G to be the uniform CDF on $[0, \bar{P}]$, so $G(P) = P/\bar{P}$ for P in that interval. Equation (OA.1) becomes

$$P > T \equiv \tau^* \bar{P}.$$

The examiner allows exactly when the distance to prior art exceeds a threshold. The comparative statics are intuitive. The threshold falls when allowing good applications becomes more valuable ($v_H \uparrow$) or rejecting them becomes more costly ($r_H \downarrow$), and rises when allowing bad applications becomes more costly ($v_L \downarrow$ or $r_L \uparrow$) or when the same distance implies a lower probability of patentability ($\bar{P} \uparrow$). The model in Section 5 takes this reduced form as primitive: the examiner's payoff from allowance increases in assessed distance to prior art relative to an allowance standard T_i, and heterogeneity in T_i across examiners, technology centers, and applications summarizes heterogeneity in the underlying payoffs (v_H, v_L, r_H, r_L) and in the informativeness of distance.

Let the examiner's underlying payoff from disposition $y \in \{\text{Allow}, \text{Reject}\}$ for an application with true distance D_i and allowance standard T_i be

$$u_E(\text{Allow}) = D_i - T_i, \qquad u_E(\text{Reject}) = 0.$$

If the assessment is a Bayesian posterior mean, then $\widehat{D}_i = \mathbb{E}[D_i \mid P_i^*, T_i]$ and the expected allowance payoff is $\widehat{D}_i - T_i = -\widetilde{G}_i$. The empirical model instead specifies $\widehat{D}_i$ by the location-correction rule in equation (11) and uses the same payoff as a function of that assessment. This rule is not derived as a Bayesian posterior.

At Stage 3, conditional on improvement e_i, the updated assessment is $\widehat{D}_i + e_i$, yielding terminal payoffs $U_{E,2}(A) = e_i - \widetilde{G}_i$ and $U_{E,2}(R) = 0$. Eventual allowance therefore requires $e_i \geq \widetilde{G}_i$.

OA.5 Derivation of the Likelihood Function

Let $S_i^* = \lambda_i + \sigma_s v_i$ and $S_i = S_i^* \mid S_i^* \geq 0$. Define $z_i(s) = \frac{s-\lambda_i}{\sigma_s}$.

The density of the realized search gap is

$$f_S(s \mid X_i) = \frac{\phi(z_i(s))}{\sigma_s \Phi(\lambda_i/\sigma_s)} \mathbf{1}\{s \geq 0\}. \tag{OA.2}$$

Because (u_i, v_i) are jointly standard normal with correlation $\rho_{s\tau}$, conditional on a realized search gap $S_i = s$,

$$T_i \mid S_i = s, X_i \sim \mathcal{N}\big(\tau_i + \rho_{s\tau}\sigma_\tau z_i(s), \sigma_\tau^2(1-\rho_{s\tau}^2)\big)\,.$$

Under the assessment rule $\widehat{D}_i = D_i + S_i - \bar{s}_i$, the assessed patentability gap conditional on $S_i = s$ is

$$\widetilde{G}_i \mid S_i = s, D_i, X_i \sim \mathcal{N}\big(m_i(s), \sigma_\tau^2(1-\rho_{s\tau}^2)\big)\,,$$

where

$$m_i(s) = \tau_i - D_i - s + \bar{s}_i + \rho_{s\tau}\sigma_\tau \frac{s-\lambda_i}{\sigma_s}. \tag{OA.3}$$

For an application allowed in round 1, S_i is unobserved and its likelihood contribution is

$$L_i^{A1} = \int_0^\infty f_S(s \mid X_i)\, \Pr(\widetilde{G}_i < \kappa_i \mid S_i = s, D_i, X_i)\, ds.$$

For a rejected application, $s_i = P_i - D_i$ is observed. Its contribution contains both the density of that particular search gap and the probability that the associated assessed patentability gap lies above the round-1 cutoff. Consequently, search gaps observed among rejected applications are not a random sample from equation (OA.2).

This selection is also the source of information about $\rho_{s\tau}$. Because P_i is not observed for round-1 allowances, $\rho_{s\tau}$ is not identified from an unconditional empirical covariance between S_i and T_i. Rather, it is parametrically identified from how the distribution of search gaps among selected rejections covaries with the probability of rejection.

For a rejected application with search gap s_i, let $f_G(g \mid s_i, D_i, X_i)$ denote the normal density implied by equation (OA.3). The amendment probability and terminal non-allowance

probability conditional on g are

$$q_i(g) = \Phi\left(\frac{\ln V_i - c_i - \omega_i g^+}{\sigma_c}\right),$$
$$p_{2i}(g) \equiv \Pr(R_{i,2} = 1 \mid a_i = 1, g, X_i) = 1 - \exp(-\omega_i g^+).$$

The likelihood contribution of a rejected application is therefore

$$L_i^R = f_S(s_i \mid X_i) \int_{\kappa_i}^{\infty} f_G(g \mid s_i, D_i, X_i) \ell_i^{\text{post}}(g)\, dg,$$

where

$$\ell_i^{\text{post}}(g) = \begin{cases} 1 - q_i(g), & a_i = 0, \\ q_i(g)[p_{2i}(g)]^{R_{i,2}}[1 - p_{2i}(g)]^{1-R_{i,2}}, & a_i = 1. \end{cases}$$

Depending on the observed post-rejection outcome, L_i^R equals L_i^{RN}, L_i^{RA}, or L_i^{RR}, whose closed forms are derived below.

These expressions summarize the main sources of parametric identification. Search-gap parameters affect both the distribution of observed search gaps and selection into rejection. Parameters of the allowance standard shift the conditional distribution of the assessed patentability gap. Amendment-cost parameters primarily affect the amendment margin, whereas improvement parameters affect both amendment incentives and the probability of grant through their effect on the examiner's payoff. The dynamic parameter enters the round-1 margin through κ_i. Estimation exploits these restrictions jointly.

Before conditioning on the rejection decision, the assessed patentability gap conditional on the latent found distance $P_i^* = p$ is normal:

$$\widetilde{G}_i \mid P_i^* = p \sim \mathcal{N}(m_i(p), \sigma_{g|p}^2),$$

where

$$m_i(p) \equiv \tau_i - p + \bar{s}_i + \rho_{s\tau}\frac{\sigma_\tau}{\sigma_s}(p - D_i - \lambda_i),$$
$$\sigma_{g|p} \equiv \sigma_\tau \sqrt{1 - \rho_{s\tau}^2}.$$

The marginal density of the latent distance of the examiner-found prior art implied by the

truncated-normal search process is

$$f_{P_i^*}(p \mid D_i, X_i) = \frac{\phi((p - D_i - \lambda_i)/\sigma_s)}{\sigma_s \Phi(\lambda_i/\sigma_s)}, \qquad p \geq D_i.$$

Round-1 allowance. For applications allowed in round 1, P_i is unobserved and must be integrated out. Before truncating the latent search gap at zero, define

$$\begin{aligned}
\widetilde{m}_i &\equiv \tau_i - D_i - \lambda_i + \bar{s}_i, \\
\sigma_G^2 &\equiv \sigma_\tau^2 + \sigma_s^2 - 2\rho_{s\tau}\sigma_s\sigma_\tau, \\
r_{sG} &\equiv \frac{\rho_{s\tau}\sigma_\tau - \sigma_s}{\sigma_G}.
\end{aligned}$$

Thus, $\widetilde{m}_i$ and σ_G^2 are the mean and variance of $\widetilde{G}_i$ before truncation, and r_{sG} is the correlation between the latent search gap and $\widetilde{G}_i$. Because the realized search gap is conditioned on $S_i^* \geq 0$, the likelihood contribution of an application allowed in round 1 is

$$L_i^{A1} = \frac{\Phi\left(\frac{\kappa_i - \widetilde{m}_i}{\sigma_G}\right) - \Phi_2\left(-\frac{\lambda_i}{\sigma_s}, \frac{\kappa_i - \widetilde{m}_i}{\sigma_G}; r_{sG}\right)}{\Phi(\lambda_i/\sigma_s)},$$

where $\Phi_2(\cdot, \cdot; \rho)$ denotes the standard bivariate-normal CDF with correlation ρ. The numerator is the probability that the application clears the round-1 cutoff, $\widetilde{G}_i < \kappa_i$, intersected with $S_i^* \geq 0$; the denominator rescales by the probability of truncation.

Rejected applications. For a rejected application with observed $P_i = p$, define

$$z_{0,i} \equiv -\frac{m_i(p)}{\sigma_{g|p}}, \qquad z_{\kappa,i} \equiv \frac{\kappa_i - m_i(p)}{\sigma_{g|p}}, \qquad S_{q,i} \equiv \sqrt{\sigma_c^2 + \omega_i^2 \sigma_{g|p}^2}, \qquad r_{q,i} \equiv \frac{\omega_i \sigma_{g|p}}{S_{q,i}}.$$

Here $z_{0,i}$ and $z_{\kappa,i}$ standardize the boundaries at zero and at the round-1 cutoff, while $S_{q,i}$ and $r_{q,i}$ arise from convolving the lognormal cost shock with the conditional gap distribution in the amendment rule.

Because $\kappa_i \leq 0$, the probabilities of the two relevant rejected regions are

$$\begin{aligned}
\Delta_i^- &\equiv \Pr(\kappa_i \leq \widetilde{G}_i \leq 0 \mid P_i^*) = \Phi(z_{0,i}) - \Phi(z_{\kappa,i}), \\
\Delta_i^+ &\equiv \Pr(\widetilde{G}_i > 0 \mid P_i^*) = 1 - \Phi(z_{0,i}).
\end{aligned}$$

For $\widetilde{G}_i \leq 0$, amendment occurs with probability q_i^0 and leads to allowance with probability one. For positive gaps, define

$$Q_i \equiv \int_0^\infty q_i(g) f_{\widetilde{G}_i|P_i^*}(g)\, dg,$$
$$A_i \equiv \int_0^\infty q_i(g) e^{-\omega_i g} f_{\widetilde{G}_i|P_i^*}(g)\, dg.$$

Thus, Q_i is the probability mass of positive-gap applications that continue, conditional on the latent found distance $P_i^* = p$, while A_i is the probability mass that both continue and are eventually allowed.

We can compute both integrals analytically; they are the same function of a single argument. For any m, define

$$\Psi_i(m) \equiv \Phi\left(\frac{\ln V_i - c_i - \omega_i m}{S_{q,i}}\right) - \Phi_2\left(-\frac{m}{\sigma_{g|p}}, \frac{\ln V_i - c_i - \omega_i m}{S_{q,i}}; r_{q,i}\right).$$

Evaluating Ψ_i at the conditional mean gives Q_i; evaluating it at the exponentially tilted mean and rescaling gives A_i:

$$Q_i = \Psi_i(m_i(p)),$$
$$A_i = K_i \Psi_i\big(m_i(p) - \omega_i \sigma_{g|p}^2\big),$$
$$K_i \equiv \exp\big(-\omega_i m_i(p) + \tfrac{1}{2}\omega_i^2 \sigma_{g|p}^2\big).$$

Because the normal family is closed under exponential tilting, the factor $e^{-\omega_i g}$ shifts the conditional mean down by $\omega_i \sigma_{g|p}^2$ and leaves the variance unchanged, at the cost of the constant $K_i = \mathbb{E}[e^{-\omega_i \widetilde{G}_i} \mid P_i^*]$.

Observed post-rejection outcomes. The three post-rejection outcomes partition the rejected sample. An application that is rejected and not amended contributes

$$L_i^{RN} = f_{P_i^*}(P_i)\left[(1 - q_i^0)\Delta_i^- + (\Delta_i^+ - Q_i)\right].$$

An application that continues and is eventually allowed contributes

$$L_i^{RA} = f_{P_i^*}(P_i)\left[q_i^0 \Delta_i^- + A_i\right],$$

and an application that continues but ultimately terminates without a patent contributes

$$L_i^{RR} = f_{P_i^*}(P_i)[Q_i - A_i].$$

The three rejected-application contributions sum to

$$f_{P_i^*}(P_i)\Pr(\widetilde{G}_i \geq \kappa_i \mid P_i^*) = f_{P_i^*}(P_i)[\Delta_i^- + \Delta_i^+],$$

as required.

The full sample log likelihood is

$$\sum_{i=1}^{N} \big[(1 - R_{i,1}) \ln L_i^{A1} + R_{i,1}(1 - a_i) \ln L_i^{RN} + R_{i,1} a_i (1 - R_{i,2}) \ln L_i^{RA} + R_{i,1} a_i R_{i,2} \ln L_i^{RR}\big]. \tag{OA.4}$$

We maximize equation (OA.4) with respect to the model parameters. The search-gap dispersion σ_s and correlation $\rho_{s\tau}$ are estimated subject to $\sigma_s > 0$ and $\rho_{s\tau} \in (-1, 1)$. The maximized log likelihood is $-571{,}754.8$.

Standard errors are computed from the inverse of a numerical approximation to the observed information matrix, with the delta method applied to transformed parameters. They are conditional on two components not jointly estimated with the structural model. First, we fix $\sigma_\tau = 5$. Second, the patent value V_i is replaced by its first-stage estimate from Appendix F and treated as known in the structural likelihood.

We report the estimates in Section 7.

OA.6 Patent-Value Model

We model the private annual flow value of patent i as a draw from a lognormal distribution,

$$v_i \sim \text{Lognormal}(\eta_i, \sigma_v^2), \qquad \ln v_i = \eta_i + \sigma_v \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(0, 1), \qquad \eta_i = Z_i' \eta.$$

Here Z_i contains the covariates of the value model: measured prior-art distance, firm type, and WIPO technology-field indicators. This covariate set differs from the prosecution-model covariates X_i.

The renewal model represents the end of the patent's payoff stream by $O_i \sim \text{Exp}(\psi)$, with $\Pr(O_i > t) = \exp(-\psi t)$. We assume O_i is independent of v_i. There is no separate financial discount rate, so ψ is an effective decay parameter rather than a separately identified rate of technological obsolescence. When the flow payoff is constant at v_i, the expected payoff over the next l years is

$$\mathbb{E}\left[\int_0^{\min(O_i,l)} v_i \, dt\right] = v_i \mathbb{E}[\min(O_i, l)] = v_i \left(\frac{1 - \exp(-\psi l)}{\psi}\right).$$

OA.6.1 Renewal Decisions

The model approximates the first two maintenance decisions by renewal at ages 4 and 8 after grant. At age 4, payment of $c_{4,i}$ preserves the option to renew at age 8; payment of $c_{8,i}$ then secures the remaining payoff stream. Fees depend on firm type. Let $\zeta_i > 0$ denote the modeled remaining life beyond age 8, and let $y_{t,i}$ indicate renewal at age $t \in \{4, 8\}$. The model contains two renewal decisions and abstracts from a separate third maintenance decision.

Decision at age 8. Conditional on renewal at age 4 and $O_i > 8$, the age-8 decision is

$$y_{8,i} = \mathbf{1}\left[v_i \left(\frac{1 - \exp(-\psi\zeta_i)}{\psi}\right) - c_{8,i} \geq 0\right] = \mathbf{1}\{v_i \geq b_{8,i}\},$$

where

$$b_{8,i} \equiv \frac{\psi c_{8,i}}{1 - \exp(-\psi\zeta_i)}.$$

Decision at age 4. Anticipating the age-8 option, the continuation value at age 4 is

$$V_{4,i} \equiv v_i \left(\frac{1 - \exp(-4\psi)}{\psi}\right) + \exp(-4\psi) \max\{v_i - b_{8,i}, 0\} \left(\frac{1 - \exp(-\psi\zeta_i)}{\psi}\right) - c_{4,i}.$$

For low values, $v_i \leq b_{8,i}$, the patentee does not plan to renew at age 8, and

$$y_{4,i} = \mathbf{1}\{v_i \geq b_{4L,i}\}, \qquad b_{4L,i} \equiv \frac{\psi c_{4,i}}{1 - \exp(-4\psi)}.$$

For high values, $v_i \geq b_{8,i}$, the patentee plans to renew at age 8, and

$$y_{4,i} = \mathbf{1}\{v_i \geq b_{4H,i}\}, \qquad b_{4H,i} \equiv \frac{\psi c_{4,i} + b_{8,i}\exp(-4\psi)[1-\exp(-\psi\zeta_i)]}{1-\exp[-\psi(\zeta_i+4)]}.$$

Equivalently, conditional on $O_i > 4$, renewal at age 4 occurs when

$$b_{4L,i} \leq v_i < b_{8,i} \quad \text{or} \quad \max\{b_{8,i}, b_{4H,i}\} \leq v_i.$$

OA.6.2 Renewal Probabilities and Likelihood

Let $\overline{\Phi}(a) = 1 - \Phi(a)$. For lognormal v_i,

$$\Pr(v_i \leq a) = \Phi\left(\frac{\ln a - \eta_i}{\sigma_v}\right), \qquad \Pr(v_i > a) = \overline{\Phi}\left(\frac{\ln a - \eta_i}{\sigma_v}\right).$$

Define

$$z_{8,i} = \frac{\ln b_{8,i} - \eta_i}{\sigma_v}, \qquad z_{4L,i} = \frac{\ln b_{4L,i} - \eta_i}{\sigma_v},$$
$$z_{4H,i} = \frac{\ln b_{4H,i} - \eta_i}{\sigma_v}, \quad z_{\max,i} = \frac{\ln \max\{b_{8,i}, b_{4H,i}\} - \eta_i}{\sigma_v}.$$

The observable outcomes have probabilities

$$\begin{aligned}
p_{11,i} &\equiv \Pr(y_{4,i}=1, y_{8,i}=1) = \exp(-8\psi)\overline{\Phi}(z_{\max,i}),\\
p_{10,i} &\equiv \Pr(y_{4,i}=1, y_{8,i}=0)\\
&= \exp(-4\psi)\max\{\Phi(z_{8,i}) - \Phi(z_{4L,i}), 0\} + [\exp(-4\psi) - \exp(-8\psi)]\overline{\Phi}(z_{\max,i}),\\
p_{00,i} &\equiv \Pr(y_{4,i}=0) = 1 - p_{11,i} - p_{10,i}.
\end{aligned}$$

The maximum operator assigns zero probability to the low-value renewal region when $b_{4L,i} \geq b_{8,i}$. For patents with both renewal outcomes observed, the log likelihood is

$$\begin{aligned}
\ell(\eta, \sigma_v, \psi) = \sum_{i=1}^{N} \big[&\mathbf{1}\{y_{4,i}=1, y_{8,i}=1\}\ln p_{11,i} + \mathbf{1}\{y_{4,i}=1, y_{8,i}=0\}\ln p_{10,i}\\
&+ \mathbf{1}\{y_{4,i}=0\}\ln p_{00,i}\big].
\end{aligned}$$

OA.6.3 Ex-Ante Value at Grant

The net present value of a patent at grant is

$$\mathbb{E}[V_{0,i}] = \mathbb{E}[v_i]\left(\frac{1-\exp(-4\psi)}{\psi}\right) + \exp(-4\psi)\mathbb{E}[\max\{V_{4,i},0\}],$$

where $\mathbb{E}[v_i] = \exp(\eta_i + \sigma_v^2/2)$.

The value at age 4 is positive in two disjoint regions. In Region 1, the patent is renewed at age 4 but not at age 8: $v_i \in [b_{4L,i}, b_{8,i})$ when $b_{4L,i} < b_{8,i}$. Its contribution is

$$B_i^{\text{region 1}} = \left(\frac{1-\exp(-4\psi)}{\psi}\right)\int_{b_{4L,i}}^{\max\{b_{4L,i},b_{8,i}\}} (v_i - b_{4L,i})\, dF(v_i).$$

In Region 2, the patent is renewed at both ages: $v_i \geq \max\{b_{8,i}, b_{4H,i}\}$. Its contribution is

$$B_i^{\text{region 2}} = \left(\frac{1-\exp[-\psi(\zeta_i+4)]}{\psi}\right)\int_{\max\{b_{8,i},b_{4H,i}\}}^{\infty} (v_i - b_{4H,i})\, dF(v_i).$$

Hence,

$$\mathbb{E}[\max\{V_{4,i},0\}] = B_i^{\text{region 1}} + B_i^{\text{region 2}},$$

and

$$V_i \equiv \mathbb{E}[V_{0,i}] = \mathbb{E}[v_i]\left(\frac{1-\exp(-4\psi)}{\psi}\right) + \exp(-4\psi)\left(B_i^{\text{region 1}} + B_i^{\text{region 2}}\right). \tag{OA.5}$$

For lognormal v_i, the truncated expectations admit closed forms. For bounds $0 < A \leq B$,

$$\int_A^B (v_i - K)\, dF(v_i) = \exp\left(\eta_i + \frac{\sigma_v^2}{2}\right) A_1 - K A_2,$$

where

$$A_1 = \Phi\left(\frac{\ln B - \eta_i - \sigma_v^2}{\sigma_v}\right) - \Phi\left(\frac{\ln A - \eta_i - \sigma_v^2}{\sigma_v}\right),$$
$$A_2 = \Phi\left(\frac{\ln B - \eta_i}{\sigma_v}\right) - \Phi\left(\frac{\ln A - \eta_i}{\sigma_v}\right).$$

When $B = \infty$, the corresponding normal CDF terms equal one.

OA.6.4 Full Patent-Value Estimates

Table OA.7: Patent-Value Model: Full Parameter Estimates

Parameter	Estimate	SE	Parameter	Estimate	SE
Baseline covariates (log-value index)			*WIPO technology fields (cont.)*		
η_0 (intercept)	1.0135	0.0010	Food chemistry	0.0332	0.0010
η_s (small)	−0.0656	0.0010	Furniture, games	−0.0147	0.0010
η_d (min. distance D_i)	−0.0179	0.0010	Handling	−0.0064	0.0009
			IT methods for management	0.0210	0.0010
Distribution and decay parameters			Machine tools	−0.0033	0.0010
$\ln \sigma_v$	−0.4769	0.0010	Macromolecular chemistry	0.0087	0.0010
$\ln \psi$	−3.0810	0.0039	Materials, metallurgy	0.0021	0.0010
σ_v (delta method)	0.6207	0.0006	Measurement	0.0213	0.0010
ψ (delta method)	0.0459	0.0002	Mechanical elements	−0.0263	0.0010
			Medical technology	0.0414	0.0009
WIPO technology fields			Micro/nano-technology	0.0198	0.0010
Basic communication proc.	0.0117	0.0010	Optics	0.0002	0.0010
Basic materials chemistry	−0.0145	0.0010	Organic fine chemistry	−0.0579	0.0010
Biotechnology	0.0065	0.0010	Other consumer goods	−0.0348	0.0010
Chemical engineering	−0.0171	0.0010	Other special machines	0.0021	0.0010
Civil engineering	−0.0034	0.0010	Pharmaceuticals	−0.0026	0.0010
Computer technology	−0.0278	0.0010	Semiconductors	−0.0027	0.0010
Control	0.0262	0.0010	Surface tech., coating	0.0001	0.0010
Digital communication	0.0033	0.0010	Telecommunications	−0.0021	0.0010
Electrical machinery	0.0026	0.0010	Textile, paper machines	0.0205	0.0010
Engines, pumps, turbines	−0.0231	0.0011	Thermal processes	0.0090	0.0010
Environmental technology	0.0044	0.0010	Transport	0.0101	0.0010

Notes: SE is the bootstrap standard error. For σ_v and ψ, standard errors are obtained using the delta method.

OA.7 Robustness to Covariate Assignment

This section examines whether our results are robust to alternative specifications. A first check adds the small-firm indicator to the discount-factor index. The coefficient is -2.1×10^{-6} (s.e. 0.059), and all estimates and counterfactuals are identical to the baseline at reported precision, so we find no evidence that examiners' dynamic rejection incentives vary by applicant size.

Table OA.8 then examines the covariate placements. The first specification (Baseline) reports key results from the main model in the paper. The second specification (Reassigned covariates) relaxes four of them at once: micro-entity status and the attorney ratio, excluded from the allowance standard in the baseline, now enter it, and university affiliation and hedging language, restricted to the standard in the baseline, now also enter the rate parameter and the cost index. The key structural objects barely move. Expected revision improvement, the discount factor, and the cutoff change by one to three percent for both size groups, and the median amendment cost gap remains large, \$11,347 against \$2,561 for small

Table OA.8: Robustness of estimates and counterfactuals to covariate reassignment

	Baseline		Reassigned covariates	
	Small firm	Large firm	Small firm	Large firm
Panel A: Key structural objects				
Expected improvement $\mathbb{E}[e_i]$	4.26	4.61	4.15	4.48
Discount factor δ_i	0.642	0.657	0.651	0.665
Cutoff κ_i	−5.53	−7.82	−5.63	−7.88
Median amendment cost ($)	11,881	2,588	11,347	2,561
Panel B: Counterfactual grant rates				
Baseline (observed size)	0.592	0.788	0.592	0.788
Allowance standard switched (τ)	0.677	0.695	0.679	0.694
Amendment cost switched (c)	0.630	0.757	0.632	0.755
All channels switched	0.718	0.659	0.721	0.656

Notes: The first specification (Baseline) reports key results from the main model in the paper. The second specification relaxes four maintained placements jointly: micro-entity status and the attorney ratio enter the allowance standard, and university affiliation and hedging language enter the rate parameter and the cost index. Panel A reports sample means by applicant size. Panel B reports mean predicted final grant rates by sample: Small firm columns apply the large-firm indices to small-firm applications channel by channel ($s \to 0$); Large firm columns apply the small-firm indices to large-firm applications ($s \to 1$). Additional cost covariates in the second specification enter at observed values in all counterfactuals.

and large firms. The counterfactuals are equally stable: equalizing the allowance standard raises the small-firm grant rate by 8.7 percentage points (8.6 in the baseline), equalizing amendment costs by 4.0 points (3.8), and equalizing all channels by 12.9 points (12.6). The mirror-image counterfactuals for large firms show the same stability: imposing the small-firm cost index lowers the large-firm grant rate by 3.3 percentage points against 3.1 in the baseline, preserving the asymmetry between large-firm losses and small-firm gains discussed in Section 8.4. Because the second specification relaxes the placements jointly, it allows the largest departure from the baseline; the stability of all reported objects thus suggests that no individual placement is driving the results.

OA.8 Treating Large Firms as Small

Table OA.9 reports the mirror exercise for the large-firm subsample. This exercise illustrates how the results depend on the population whose covariates are held fixed; it is not an independent validation of the model. The reverse exercise yields the same ranking in the one-channel counterfactuals. Imposing the small-firm allowance-standard component lowers

Table OA.9: Reverse Counterfactual: Large Firms with Small-Firm Components

		Large firms with small-firm component				
Outcome	Large Baseline	Allowance Standard	Search Gap	Amendment Cost	Revision Improvement	All
Round-1 rejection	0.8292	0.9050	0.8300	0.7667	0.8018	0.8406
Amendment | rejection	0.9334	0.9232	0.9336	0.8604	0.9311	0.8405
Grant | amendment	0.7969	0.7178	0.7993	0.7934	0.7860	0.7074
Final grant	0.7875	0.6948	0.7894	0.7567	0.7851	0.6592
Expected net private value ($000)	303.7	269.7	303.5	284.0	302.5	246.8

Notes: Each column replaces the indicated large-firm structural component with its small-firm counterpart. All other observed characteristics are held fixed, and κ_i is recomputed in every counterfactual.

the large-firm final grant rate by 9.3 percentage points, the amendment-cost component by 3.1 points, and the revision-improvement component by 0.2 points, while the search-gap component is negligible. Switching all four lowers the grant rate from 78.8 to 65.9 percent.