Selection in the Patent Lottery

Lorenz Gschwent*

Job Market Paper

September 25, 2026

[Click here for the most recent version]

Abstract

Patents reward inventions with exclusion rights that can block follow-on innovations. Yet patent grants at the US Patent Office are partly a lottery. The office assigns applications quasi-randomly to examiners within narrow technology fields who differ widely in how often they grant. Which applications does the lottery favor? If they are applications seeking broader rights than the invention supports, the lottery blocks later inventors. If they are higher-quality inventions the office cannot reliably assess, the lottery rewards invention. I measure invention quality from embeddings of the application's text fixed at filing: high-quality inventions differ from their field's earlier filings and resemble the field's later filings. Across 5.4 million applications, 4.7% of granted patents owe their grant to an examiner more lenient than the average examiner the application could have drawn. The share increases across filed scope – how broad a right the applicant asked for – but barely moves across quality. Better inventions request broader protection but are no likelier to be granted: the advantage of quality and the penalty of the larger request cancel almost exactly. The lottery selects on scope, creating overly broad patents, allowed only because of who examined them. In a structural model of examination as a multi-round negotiation over scope, I ask what changes when examiners are made more consistent. Holding total patents and total granted scope constant, a consistent examiner pool reallocates 11.7% of granted scope – and almost all of the reallocation is between applications of the same quality.

Keywords: patent office, patent examiners, patent quality, natural language processing
JEL Classification: O31, O34, L24, L32, K12, D73

*University of Duisburg-Essen, e-mail: *lorenz.gschwent@uni-due.de.*

I am grateful to Tobias Seidel, Philip Jung, and Martin Karlsson for their advice and encouragement. I thank Lukas Buchheim, Bernhard Ganglmaier, Daniel Gross, Ryan Hill, Hanna Hottenrott, Paul Jensen, Eugen Kovac, Matt Marx, William Matcham, Gaétan de Rassenfosse, Cesare Righi, Bhaven Sampat, Beth Webster, Galina Zudenkova, and seminar participants at the University of Duisburg-Essen, RTG Regional Disparities and Economic Policy, ZEW Mannheim, Melbourne Institute of Applied Economic and Social Research, as well as participants and discussants at the 18th RGS Doctoral Conference, 2nd Duke Strategy PhD Conference, REGIS Summer School, BSE Summer Forum, 20th EPIP Annual Conference, 8th RISE Workshop, DRUID Academy, and ICSSI Boulder for their helpful comments. This paper was previously presented under the title "Quality, Scope, and Leniency: Strategic Application Behavior at the US Patent Office". I gratefully acknowledge the computing time granted by the Center for Computational Sciences and Simulation (CCSS) of the University of Duisburg-Essen and provided on the supercomputer amplitUDE (DFG project 459398823; grant ID INST 20876/423-1 FUGG) at the Zentrum für Informations- und Mediendienste (ZIM).

1 Introduction

Which inventions receive patent protection, and how broadly, shapes both the reward for inventing and the room left for later inventors (Scotchmer, 1991; Acemoglu and Akcigit, 2012; Galasso and Schankerman, 2014). A patent is a temporary right to exclude competitors, and its claims define the boundaries of that right. Every patent starts as an application and enters a screening process. The applicant writes the claims, choosing the scope of the rights they seek. I call that the application's filed scope. The patent office decides how much of that scope, if any, survives. I call that the application's granted scope, the exclusionary rights assigned to the patent, which can be zero in the case the application is rejected. A claim to editing genes is an example of broad scope, while a claim to editing genes in mouse liver cells with one particular enzyme is a narrow one.[1]

The US Patent and Trademark Office (USPTO) screens the scope boundaries of patent applications, and it screens them inconsistently. Within narrow technology fields, applications are assigned to examiners essentially by workload, and the acceptance rates of examiners in the same field differ widely (Cockburn, Kortum, and Stern, 2002; Lemley and Sampat, 2012; Gaulé, 2018; Sampat and Williams, 2019; Farre-Mensa, Hegde, and Ljungqvist, 2020), by roughly forty percentage points between the bottom and the top percentile of examiners. Which boundary an application receives, whether the granted scope is broad or narrow or null, therefore depends partly on the examiner it draws.

Examiners may not just differ by what scope they grant, but by how they read the quality of an invention. By invention quality I mean the contribution of the invention itself, separate from the application's filed scope and separate from whether the patent is granted. A high-quality invention differs from its technology field's earlier filings and resembles the field's later filings. Examination is meant to enforce two requirements, one on the invention and one on the filed scope. The invention must clear the bar of patentability and the scope may reach no further than the invention disclosed supports.

How invention quality and filed scope affect the patent lottery is unclear. The literature has documented the lottery well, but has not characterized the applications which are granted or refused by the luck of the draw. The obstacle is measurement, as every established measure of invention quality is built on granted patents, so a refused application cannot be characterized. With no measure to settle it, the applications whose grant is decided by the lottery of the examiner draw can be read two ways. If they are *bad* – weak inventions, or claims beyond what the invention supports – the lottery grants rights that block later inventors and the office should make screening stricter (U.S. Federal Trade Commission, 2003; Boldrin and Levine, 2013; Schankerman and Schuett, 2021). If they are instead *hard to judge* – good inventions whose quality the office cannot reliably assess – stricter screening harms innovation and examiners instead need more time. The two readings imply opposite reforms, and the evidence does not settle the choice between them. Boldrin and Levine (2013) argue that too many low-quality inventions receive a patent. Yet patents granted due to the lottery carry large returns for start-ups (Farre-Mensa et al., 2020), which they should not if the inventions behind them were weak. What kind of patents does the lottery create: patents for higher-quality inventions, or for applications that asked for broader rights?

[1] The margin studied throughout is the extensive one, grant or refusal, and filed scope is the application characteristic that moves it. The intensive margin, the scope a grant retains, enters as a separate outcome.

What would the office allocate differently if every examiner in a field applied one standard?

I use the patent examiner lottery to identify what applying one standard would achieve. I characterize the applications a lenient examiner grants but a stricter one would have rejected. The examiners disagree about how much filed scope to tolerate, and hardly at all about the quality of the invention beneath it. The applications a lenient examiner grants are *bad* patents – not in the sense of weak inventions but of claims broader than the invention supports by the office's own standard. A different examiner working in the same technology field, reading the same file, would have refused the scope that was allowed. The disagreement is not a sign that the invention was hard to judge. If the draw decided good inventions the office could not assess reliably, the examiner's assessment would be noisiest where the good inventions are. I measure the assessment noise, and it decreases with invention quality. I also show that noisier assessment shrinks the set of applications the draw decides rather than adding to it. A lenient examiner lowers the grant threshold by a fixed amount, so a verdict changes only if the examiner's assessment of the application falls within that fixed amount of the threshold. Noise spreads the assessments over a wider range, so fewer of them fall within that fixed amount, and fewer verdicts change. For the applications determined by the examiner draw, examination is not a noisy screen on quality but a lottery over scope among similar applicants.

I proceed in four steps. First, I build a measure of invention quality from the application's text at filing that puts accepted and rejected applications on one scale. Second, I redraw each application's examiner from the pool of examiners who could have examined it instead. I count, at each combination of invention quality and filed scope, the grants that would not survive the redraw. Third, I identify the mechanism behind that composition. Fourth, I estimate a model of the applicant's scope request and measure what a more consistent patent office would allocate differently.

The first step builds the quality measure from embeddings of the application's own text, fixed at filing, so it cannot be rewritten in response to the decisions it is used to study. My measure extends the text-based approach of Kelly, Papanikolaou, Seru, and Taddy (2021) from granted patents to applications, and combines two comparisons: how far the application differs from what its field filed over the previous five years and how closely what the field files over the next five resembles it.[2] Among granted patents, for which established proxies exist, my quality measure is positively correlated with citations, renewals, and market value, with 38% more citations in the top quality decile than at the median. Filed scope is measured by the average number of words per claim, inverted so that a higher value is a broader request: following the literature, a longer claim is written in more specific language and therefore covers narrower scope (Feng and Jaravel, 2020). The text describing the invention, which the quality measure is built from, and the text of the claims, which the scope measure is built from, are not the same, so any correlation between quality and filed scope is a fact about applications rather than an artifact of shared text.

The two measures yield stylized facts about the screening process at the patent office. Acceptance increases only slightly in invention quality, from 71% for the lowest quality to 73% for the highest, while it decreases over most of the filed scope distribution, from 77% to 63%. In line with anecdotal evidence that any invention can receive a patent at the USPTO if the filed scope is sufficiently narrowed (Bryan and

[2] The forward comparison uses later filings, but I verify that the examiner draw does not move a field's subsequent direction, so the decisions under study do not feed back into the quality measure.

Williams, 2021), I find no acceptance threshold in the quality of inventions. Higher-quality inventions ask for more scope, give up more of it during revisions, and are not granted more of it. The highest-quality inventions go through the most revision rounds.

The second step uses the quasi-random assignment of examiners as experimental variation to show who gains and loses when examiners apply the screening process differently. Since assignment within the narrow technology fields of the patent office ("art units") follows workload, examiners who differ in leniency examine comparable applications. The underlying experiment redraws each application's examiner from the examiners who could have received it and counts the grants that would not survive the redraw, at each point of the quality-scope distribution. The applications whose outcome the redraw changes are the compliers, whether the draw granted or refused them. Among granted patents, the compliers are the ones a stricter examiner from the same pool would have refused. Their number divided by the number of all grants is the share of grants owed to the examiner draw.[3] The count is informative only if the examiner an application receives is unrelated to the application itself. I test that condition directly in Section 5.4, asking whether an application's characteristics predict the leniency of the examiner it draws within its art unit and year.

I estimate the redraw with the regression the patent-lottery literature runs as its first stage, regressing the grant outcome on the leniency of the assigned examiner. That literature uses that regression as a first step towards the effect of a grant. I use it to identify the grants the draw decides, and for that purpose leniency is measured against the examiners the application could have drawn. Leniency enters the regression as the deviation from the alternative examiners' mean leniency, in the spirit of the formula-instrument designs of Borusyak, Hull, and Jaravel (2023). Every share of compliers is therefore measured relative to the examiner the application could have expected to draw.

Across 5.4 million applications published between 2005 and 2020, 4.7% of granted patents are owed to the examiner draw. Counting those grants across invention quality and filed scope shows that the lottery selects on scope, not quality. The broadest filings are forty percent more likely than the narrowest to owe their grant to the draw, a share rising from 4.0% to 5.6%. Across the whole quality distribution that share moves by 0.4 percentage points.[4] The patent grants the draw decides can be valuable and *bad* at once, because their private value comes from the scope of the exclusionary right, and that scope is what a stricter examiner would have cut.

The third step identifies the mechanism: what exposes an application to the luck of the examiner draw is how much scope the applicant asks for. The examiner hardly grades the invention, but they test the filed scope against existing work, and reject when prior work falls inside the boundary the scope sets. A bigger request stakes out a wider boundary and is easier to reject, however good the invention inside it. Filings for gene editing with CRISPR-Cas9 show the boundary at stake. Two groups applied for patents in 2012 and 2013, one at the University of California, one at the Broad Institute of MIT and Harvard. The California application claimed the method for any cell type. The Broad application claimed it for plant

[3] The count is of compliers among granted patents, the grants owed to lenient examiners. The redraw runs the other way too: an application is refused because the examiner it drew was stricter than the pool that could have received it.

[4] Quality throughout means the composite novelty-and-impact text measure of invention quality presented in Section 3. The measure is precise enough to bound the true quality gradient at seven percent above the observed one: scoring each application twice on independent halves of the corpus gives a rank reliability of 0.94.

and animal cells only. The narrow claim was granted in 2014, while the broad one was only granted in 2018, after years of proceedings. The California application had not shown the method working in plant or animal cells, and the question was whether it could claim them anyway. That dispute was famously settled by the office's appeal board and the courts. For the ordinary application the same judgment is made once, by the examiner it draws, and examiners differ in how much scope they tolerate.

The applicant therefore faces the trade-off: asking for more raises the value of a grant and lowers the chance of getting one.[5] Better inventions file wider scope, the second stylized fact, so invention quality pulls acceptance two ways. A better invention is accepted more often at a given filed scope, and it files more scope, which is accepted less often. The two nearly cancel, which is why acceptance rises only two percentage points across quality while it falls fourteen across filed scope. Examination makes larger scope requests harder to grant by design. Nothing in its design makes acceptance flat in quality. That flatness comes from the applicants' choice of filed scope.

An examiner grants when filed scope stays below what they tolerate for the invention, and they tolerate more scope for a better invention. Invention quality and filed scope therefore matter to the verdict through one number, the probability that an application is accepted by the examiner it draws. An application with that probability near one half will be granted or refused by the luck of the draw. These are the applications decided by the draw, and those of them a lenient examiner granted are the compliers. Both are defined by the acceptance probability alone.

The share of compliers among granted patents increases across filed scope for two reasons. Broad requests put more applications at the margin where the examiner draw is decisive, which accounts for roughly two fifths of the increase. The remaining three fifths of the increase come from the denominator, the number of granted patents: acceptance decreases from 77% to 63% across filed scope, so the compliers at broad scope are divided by fewer grants.[6] Across invention quality neither part moves. This is the sense in which the lottery selects on scope: a broad request is less likely to be granted, and when it is, the chances are higher that a lenient examiner granted it.

The fourth step models the scope request, because applicants choose what to file knowing roughly how lenient examination is. The reduced-form estimates characterize the compliers among granted patents as equilibrium objects, and the model lets filed scope respond when examination changes. In the model, the applicant solves a finite-horizon optimal-stopping problem. In each revision round, they choose how much scope to file, trading the value of a broader grant against the risk of refusal, and can concede scope across rounds or abandon the application. The examiner observes filed scope directly, observes invention quality only through a noisy signal, and grants when the signal relative to filed scope clears a threshold specific to the examiner's leniency.

I estimate the model by indirect inference on the full sample of applications, matching how acceptance, filed scope, granted scope, and revision rounds vary with quality and with examiner leniency. The model reproduces the targeted acceptance and scope profiles. The model has two misfits. It misplaces the round

[5] The applicant's side is a standard endogenous choice, as when a job seeker sets a reservation wage. The screener's side is the unconventional part. Appendix B.1 develops these comparisons.

[6] Formally, that share is proportional to the inverse Mills ratio: the density of applications at the margin, over the grant probability. Proposition 1 derives it.

in which applicants abandon. Abandonment is far more common than a final rejection, but its timing carries little identifying content. And it overstates the slope of granted scope in filed scope, a caveat the counterfactual scope magnitudes inherit. The spread of the compliers over quality and filed scope, my reduced-form result, is not targeted in estimation, and the model predicts its direction.

The counterfactuals answer what applying one standard would achieve: granted scope can be reallocated with almost no effect on the private value of the patent portfolio. Limiting the number of revision rounds, by contrast, destroys value by cutting off applications that would have been granted in later rounds. Holding the number of grants and the total granted scope constant, a consistent examiner pool moves 11.7% of granted scope to different applicants. Of that scope, 98% moves between applications of essentially the same quality.

I evaluate four reforms, including bringing examiner dispersion down to the level measured among criminal-court judges (Dobbie, Goldin, and Yang, 2018; Agan, Doleac, and Harvey, 2023), which covers most of the benefits of complete consistency. I report the break-even social costs of scope rather than assigning one, since the partial-equilibrium model does not identify social costs.

A screening process is typically meant to sort on the hidden type, here the quality of the invention. My findings imply that screening at the patent office instead sorts on the action of the applicant, the size of the filed scope. Frankel and Kartik (2019) shows how agents' manipulation of a screened signal degrades what screening can learn. The implication I pursue is a different one, namely which applications the quasi-random examiner draw ends up deciding. Broad filed scope leaves a good invention and a weak one alike to the draw, and because the scope request is a choice, applicants decide themselves whether the examiner draw can influence their application. The same selection can arise wherever an applicant sets the size of the request and a screener with idiosyncratic tolerance rules on it, for example when a grant proposal chooses how ambitious its aims are and its reviewers are assigned by the program.

This paper speaks to four literatures. The first one uses the patent examiner lottery to estimate the causal effects of patent protection (Gaulé, 2018; Sampat and Williams, 2019; Farre-Mensa et al., 2020; Feng and Jaravel, 2020; Melero, Palomeras, and Wehrheim, 2020), and reading those estimates as effects for the typical invention assumes the compliers are ordinary applications. This paper is the first to characterize the compliers by the quality of the invention, and I show that they file broader scope. Feng and Jaravel (2020) establish that examiner-driven variation determines the scope a patent receives, not simply whether it is granted. My complier composition specifies which applications that variation affects.

A second literature studies the examiners themselves. Which examiner an application happens to draw shapes the outcome: more experienced examiners search less and grant more (Lemley and Sampat, 2012), a gradient that Frakes and Wasserman (2017) trace to the examination time each promotion removes, and examiner characteristics predict whether a granted patent survives a validity challenge (Cockburn et al., 2002). That work locates bad patents in the examiner's constraints and measures them as invalid patents, inventions that fail patentability (Lemley and Shapiro, 2005). de Rassenfosse, Griffiths, Jaffe, and Webster (2020) come closest to the test I use. They use twin applications filed at several offices and estimate that 1.2% of US patents are granted due to inconsistency, meaning that they would be refused if re-examined under the office's own standard. A grant that a stricter examiner from the same

pool would have refused is the same inconsistency seen from inside one office. I estimate the inconsistency at four times their share, and that the source of the inconsistency is the applicant's scope choice. The patents created by the lottery filed broad scope on inventions that clear the patentability bar, not invalid ones.

The patent quality measurement literature scores patent quality on granted patents, by citations, renewals, market value, and text-based novelty (Trajtenberg, 1990; Pakes, 1986; Kogan, Papanikolaou, Seru, and Stoffman, 2017; Kelly et al., 2021).[7] I extend the text-based approach of Kelly et al. (2021) and the patent-text literature (for example, deGrazia, Frumkin, and Pairolero, 2018; Arts, Hou, and Gomez, 2021) in two ways. I move from word-frequency representations to transformer-based sentence embeddings[8] (Ni, Ábrego, Constant, Ma, Hall, Cer, and Yang, 2021), and from granted patents to the application at filing, which puts rejected applications on the same scale as granted ones.

The closest paper to my work, Matcham and Schankerman (2026), prices the *level* of examination intensity, asking whether examination is too lax or too strict overall. I quantify the inefficiency in examination's cross-sectional *dispersion*, which their counterfactuals do not target. That inefficiency is the cost of the same application being examined differently depending on the examiner who receives it.[9] They too build and estimate a dynamic structural model of USPTO examination, and in both models the applicant chooses how much to ask for. They assume that choice is claim padding.[10] I show that the scope choice is the margin the patent lottery selects on. Their counterfactuals modify system-level design parameters. My counterfactuals instead ask which applications acquire scope given which examiner is drawn.

Section 2 sets out the institutional setting, Section 3 the data and the quality measure, and Section 4 the stylized facts. Section 5 gives the empirical strategy and a direct test of quasi-random assignment, and Section 6 uses the composition and the mechanism behind it to answer what kind of patents the lottery creates. Section 7 develops the model, estimates it by indirect inference, and answers what a consistent patent office would allocate with the counterfactuals. Section 8 concludes.

2 Institutional background

How applications are routed to examiners is the institutional foundation of this paper's research design. A patent application contains a title, a short abstract, a longer description, and the legal claims sought by the applicant. The description cites "prior art": earlier patents, applications, or publications, such as scientific articles, relevant to the claimed invention. The USPTO assigns each incoming application a class and subclass code under its internal Cooperative Patent Classification (CPC) and, based on the subclass, routes it to an "art unit", a group of 15–30 patent examiners with expert knowledge in certain technology fields. While there is evidence of some specialization of examiners within art units (Righi and

[7] Citations to the published application do exist for rejected applications, but they only accumulate after filing and so cannot deliver a quality measure fixed before the grant decision.

[8] Avivi (2025) conditions examiner comparisons on BERT embeddings of application text. I use embeddings to measure invention quality instead. Frequency-based similarity registers shared vocabulary, while embeddings place two documents close together when they say the same thing in different words.

[9] See also Schankerman and Schuett (2021) for the theoretical foundations of optimal screening under costly examination.

[10] Padding, in their model, is the applicant's chosen exaggeration of the claims beyond what the invention's inventive step warrants, concealing its true boundaries. Its price is a longer and costlier negotiation with the examiner.

Simcoe, 2019), the leniency literature treats the assignment of applications to examiners within an art unit as as-good-as-random (Gaulé, 2018; Sampat and Williams, 2019; Farre-Mensa et al., 2020). I test the randomness of assignment formally in Section 5.

A legal presumption of patentability, together with time constraints that can in practice be binding, shapes what examination can accomplish. Under this presumption, it is the examiner's task to show that an invention is not patentable. Although the expiry of the time allotment does not by itself lead to approval, in reality an application is likely to be granted if the examiner has not established a case for rejection within the time budget. Rejections for obviousness or lack of novelty require potentially time-consuming searches for prior-art references. Rejections for eligibility, written-description, enablement, and definiteness grounds do not, though the prior-art grounds dominate in practice. Searching for prior art beyond what the application already cites is time-intensive, the per-application time budgets bind, and truncated searches produce low-quality patents (Frakes and Wasserman, 2017, 2023). Examiners are also held to a biweekly quota of credits ("counts"), but its deadline pressure shows up in rushed first-round rejections rather than in rushed grants: an examiner short of time at the deadline issues a rejection to buy time, and when a 2011 reform smoothed work over the quota period those rejections disappeared while review accuracy did not change (Frakes and Wasserman, 2020, 2024). The time budget is the constraint that affects what is granted.[11]

Examiner experience is a further systematic driver of grant behavior. Using examiner pay scale data, Frakes and Wasserman (2017) show that experience and grant rates follow an inverse U-shape relation, with experience having the largest effect after 4-5 years as patent examiner.[12] Additionally, examiners undergo on-the-job training in their first year.[13] These patterns motivate the experience controls in my estimation strategy and imply that the term leniency bundles examiner-specific standards, time pressure, and experience effects; I follow the existing terminology throughout, but the underlying source of leniency matters for the counterfactual policies of Section 7.7.

The application process at the USPTO is often iterative. Carley, Hegde, and Marco (2015) show that first-round grants are rare, consistently below 20% between 1996 and 2005. A first-round rejection is not final, however: cumulatively over multiple rounds, almost 70% of the applications in my sample end in a grant. Applicants can revise their initial application, after which examiners can either accept, non-finally reject, or finally reject the application. A non-final rejection invites a further amendment as of right; a rejection is typically made final when the applicant's amendment does not overcome the examiner's stated grounds and necessitates no new ground of rejection, which closes the regular exchange. After a final rejection, applicants can appeal within the USPTO, file a request for continued examination, which reopens prosecution and appears in my data as further revision rounds, or file a repeat application. Marco, Sarnoff, and deGrazia (2019) find that the number of claims and their broadness decrease during the review process, matching the common understanding that applicants respond to examiners' rejections

[11] Credits are awarded for a first office action, a patent grant, or a final rejection, with at most two credits per application. The quota sets the USPTO's application throughput expectations.

[12] The authors distinguish the effects of experience and merit-based promotions (leading to a higher workload) using examiner pay scale data received via a Freedom of Information Act request. Because this information is not available for my full sample, I rely on experience proxies instead.

[13] See information provided by the U.S. Department of Commerce.

by narrowing their filed scope.

The last important institutional feature is that once such concessions have been made, they become costly to reverse. On the basis of the prosecution history estoppel principle, if an applicant narrows a claim in order to obtain a patent, they are thereby presumed to have given up the right to later recover the subject matter that was surrendered by means of the doctrine of equivalents. The presumption can be challenged and is specific to the amendment in question, and the doctrine applies after the patent has been granted rather than preventing claims from being broadened during prosecution. Together with the empirical regularity that narrowing amendments are in practice not reversed (Marco et al., 2019) it means a revision costs more than the labor of drafting it: the concession itself is a cost in foregone future profits. Claim scope is thus an endogenous outcome of a multi-round negotiation rather than a characteristic assigned at filing, which motivates the modeling framework and the empirical strategy of Section 5.

3 Data and measurement

3.1 USPTO data

The raw data cover the universe of *published* patent applications received by the USPTO from 2001, the first year for which the USPTO publishes data including rejected applications, through 2023 inclusive. Applications abandoned before the eighteen-month publication date and the small share filed with a nonpublication request never enter the public record and so cannot appear in any dataset built on it. I combine PatEx – a dataset including patent applications and relevant examiner information – with the application-text dataset to obtain a total of 6,593,022 patent applications.[14] I restrict the estimation sample to applications published between 2005 and 2020, reducing the number of observations to 5,367,642 applications. The examiner-dummy estimators of Section 5 drop a further three to stay within the largest leave-out-connected component of the examiner-cell design,[15] so estimation totals read 5,367,639. The lower bound discards a break in the claim-text parsing regime before 2005, which would otherwise contaminate the scope measure described in Section 3.2. The upper bound addresses the problem of right-censoring of prosecution outcomes at the end of the time period. The years outside this window still contribute to the construction of examiner decision histories and to the five-year text windows behind the quality measure, but the quality percentiles are re-ranked within the estimation sample. Appendix A reports summary statistics for the estimation sample.

The treatment variable throughout the paper is examiner leniency, which I measure in a realized and an expected form. Following Farre-Mensa et al. (2020), I calculate realized examiner leniency as a leave-one-out average of the acceptances among all patent applications previously examined by the same examiner. The measure is undefined for an examiner's first observed decision, and I exclude degenerate leniency values of exactly zero or one, which arise for examiners with short histories. Additionally, based on the art unit structure of the USPTO, I calculate an expected leniency. That is, for each patent application, I calculate the average leniency of examiners over all other applications decided in the same

[14] Both PatEx and patent application texts can be directly downloaded in bulk via the USPTO website.
[15] I discuss this in more detail in Appendix E.

art unit within a three-year window centered on its decision date, i.e. eighteen months on either side.[16] [17] Any examiner who made a decision in that art unit during this window could also have been assigned the application in question, so the expected leniency represents the counterfactual draw from the feasible examiner pool.[18] The deviation between realized and expected leniency isolates the random component of the examiner draw. This is the variation I use to estimate the share of "compliers" (applications that would be rejected by a stricter examiner but are accepted by a more lenient one, as defined formally in Section 5.2) in the estimation strategy of Section 5.

From the PatEx data, I additionally construct the control variables and auxiliary outcomes used throughout the analysis. These comprise the small and micro entity status of applicants, which determines the fee schedule and therefore proxies for applicant size;[19] the number of citations made to the publication and, in case of acceptance, to the patent; the number of revisions, extracted via transaction codes; an identifier for applications that are a (partial) continuation of a previous application; and, to proxy for the experience effects discussed in the previous section, the years since an examiner's first observed decision, top-coded at five, with examiners already active in the first year of the data flagged separately; the descriptive figures use the equivalent set of first-five-year experience dummies. As maintenance fees are required to renew non-design and non-plant patents every four years (up to 20 years), I use the grant date and the expiry status of a patent to calculate renewals; because many granted patents in my sample have not passed all four-year steps, I calculate realized renewals as the difference between actual and potential renewals.[20] Renewals additionally serve as a validation outcome for the quality measure.

For around 2.1 million patent applications in my sample, I can also use the Office Action Research Dataset (Lu, Myers, and Beliveau, 2017), including the reasons for rejections. For validation, I add the measure of Kogan et al. (2017) for around 1.3 million applications that end up granted and the measure of Kelly et al. (2021) for around 2.5 million.

3.2 Scope measurement

I measure the scope of patent claims using the average number of words per claim, following Feng and Jaravel (2020) and similar to Kuhn and Thompson (2019). Narrower claims use more words, because each limitation on what is protected must be stated. I therefore treat the word count as the inverse of scope. Claim length is the usual measure of scope in this literature, and two descriptives documented below show that it is related to scope rather than to the drafting style alone. On the one hand, I show

[16] According to the USPTO, the usual decision time is between 12 and 18 months.

[17] The decision date must be defined for granted and failed applications alike. I calculate it from the patent issue date, the event date of final rejections, and the event date of non-final rejections with no further registered event, which marks abandonment. All of these dates are part of PatEx.

[18] Since a more lenient examiner also results in a case being disposed of more quickly, the decision date this window centers on is not strictly fixed. In Appendix F, I reconstruct the benchmark using a pre-treatment anchor which the draw cannot move. Under this assumption, the first stage is stronger, so the conventional construction remains the baseline as the conservative choice.

[19] A small entity is an individual, a small business under the Small Business Act, or a non-profit organization; a micro entity is a small entity that is either a US institution of higher education or falls below limits on both yearly gross income and previous patent applications.

[20] This implies that the resulting variable is equal to zero for a patent that has been renewed at every possibility, while it is negative for patents that at some point were not renewed.

in Section 6.2 that the grant rate decreases sharply across the percentiles of filed scope, falling from 77% among the narrowest filings to 63% at the ninetieth percentile of scope and 58% at its extreme top. On the other hand, examiners reduce scope during prosecution, enforcing concessions that increase with prosecution length, as shown in Figure 3b. Wordiness alone would neither lead to rejections nor be the object applicants have to give up under estoppel costs.[21] I construct filed scope from the application texts and granted scope from the accepted patent texts separately, enabling a direct comparison of scope at the beginning and, if granted, the end of the prosecution process. For both filed and granted scope, I turn the raw word counts into percentile ranks by using the empirical distribution pooled over both measures within each publication year. I reverse the sign so that a higher value corresponds to wider scope. By pooling filed and granted scope, I put both measures on the same scale, which makes it possible to measure scope concessions directly. I rank the measures within publication year to make them robust to changes in claim-text data over time. Appendix I tests both dimensions of this construction: the results survive measuring scope from the first claim alone or from the total length of the claims, and the rank-based results are exactly invariant to the cardinal scale behind the ordering.

3.3 Quality measurement

A terminological distinction matters for everything in this subsection. In the patent literature (for example, Lemley and Shapiro, 2005; Galasso and Schankerman, 2014; de Rassenfosse et al., 2020; Higham, de Rassenfosse, and Jaffe, 2021), *patent quality* means the robustness of a granted patent as a legal instrument, that is, whether its claims would survive an invalidity challenge. However, legal robustness is jointly determined by the quality of the invention and the granted scope of the patent. This paper studies the first part, the quality of the *underlying invention*. Therefore, I distinguish between the technological quality of the idea seeking protection and the granted scope, which I treat as a separate, endogenous outcome. A high-quality invention forced into narrow scope yields a legally defensible patent with limited economic value. A low-quality invention receiving broad granted scope, which is what a lenient draw makes possible, yields a legally fragile patent.

I measure the quality of patent applications from their text, using transformer-based sentence embeddings (Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, and Polosukhin, 2017). An embedding maps a piece of text to an n-dimensional vector in a space that represents semantic meaning, so that similar text is closely grouped and vector distances measure the similarity of language, here between two patent applications. I use `sentence-t5-xl` (Ni et al., 2021), which outperforms both co-occurrence-based embeddings such as `GloVe` (Pennington, Socher, and Manning, 2014) and `SimCSE` (Gao, Yao, and Chen, 2022) on semantic-similarity benchmarks. I argue that for patents, where the signal of interest is an application's technical proximity to prior and subsequent work in its field, semantic similarity in the embedding space is the right object to measure. The TF-IDF word frequencies used by Kelly et al. (2021) can conflate term frequency with conceptual similarity, whereas transformer-based sentence embeddings can capture the latter directly.

[21] Claim length may still carry residual variation from technological complexity or house drafting style. The two facts establish that the component examiners screen and applicants concede is scope.

For each patent application in my sample, I first concatenate the title and the abstract. I run this combined text through sentence-t5-xl (Ni et al., 2021) to get a 768-dimensional vector representation for each application. I use this vector v_p for each patent application p, dated by its filing year t, to calculate my quality measure. Following Kelly et al. (2021), I compute within every art unit:

$$FS_p = \text{sim}\left(v_p, \bar{v}_p^+\right) = i_p \qquad i_p \in [0,1] \tag{1}$$

$$BS_p = \text{sim}\left(v_p, \bar{v}_p^-\right) = 1 - n_p \qquad n_p \in [0,1] \tag{2}$$

with the cosine similarity $\text{sim}(v,w) = \frac{vw}{\|v\|\|w\|}$, the average vector of the applications filed in the five years preceding p's filing date,[22] $\bar{v}_p^- = \frac{1}{N_p^-}\sum_{\tau=t-5}^{t-1}\sum_{i\in p_\tau} v_{i,\tau}$, and its counterpart over the five following years, $\bar{v}_p^+ = \frac{1}{N_p^+}\sum_{\tau=t+1}^{t+5}\sum_{i\in p_\tau} v_{i,\tau}$, where p_τ denotes the set of applications filed in year τ within p's art unit and N_p^-, N_p^+ count the applications in the respective window. i_p and n_p represent impact and novelty.[23] Next, I take the ratio of forward similarity over backward similarity. The ratio form means the index rewards being simultaneously novel and impactful, so an application that is unlike its predecessors but also unlike its successors scores no better than one that is unremarkable in both directions.[24] Finally, I take percentiles of the ratio to arrive at my quality measurement:

$$Q_p = P\left(\frac{FS_p}{BS_p}\right) \qquad Q_p \in [1,100] \tag{3}$$

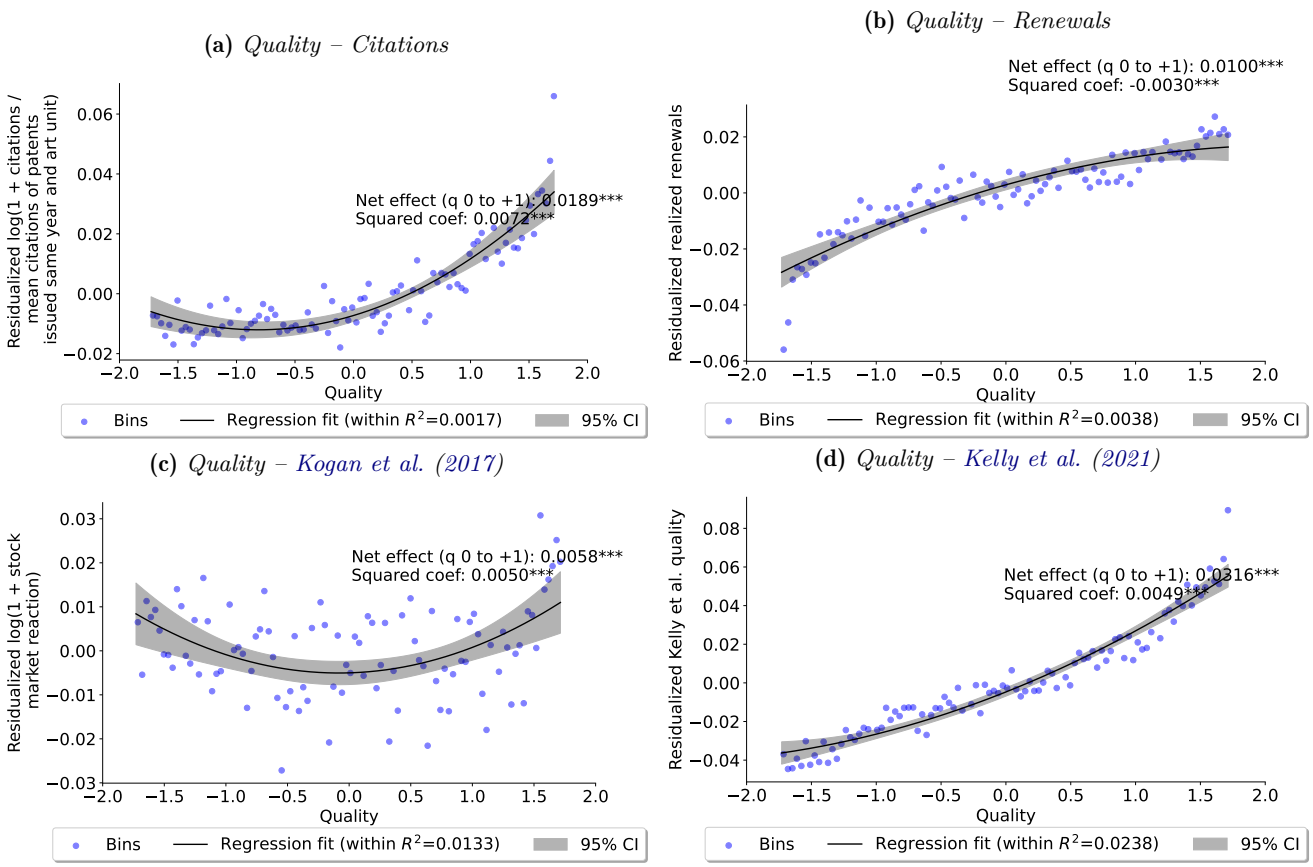
Three design choices require discussion. Appendix H discusses them in further detail. First, I embed titles and abstracts rather than patent claims. Claims are narrowed during prosecution in response to examiner objections, whereas titles and abstracts are fixed in the filing text, so the text Q_p is built from cannot have been rewritten in response to the decisions it is used to study. Second, the percentile transformation is calculated separately for each art unit $\times$ publication year cell – the same partition used as the fixed effects in the first stage in Section 5 – so Q_p ranks each application relative to the cohort it was evaluated alongside and measures quality relative to a technology field rather than on an absolute scale. Third, for my identification strategy Q_p must be orthogonal to the examiner draw. I report two tests that support this assumption. In Table H.6, I show that leniency does not move an art unit's forward technological trajectory. In Section 5.4, I formally test the balance of quality across leniency.

[22] Anchoring the comparison windows on the filing date rather than the publication date is a deliberate design choice. Filing marks the application's actual entry into the USPTO. Publication trails filing by a statutory eighteen months, and in practice by an amount that varies with the filing route, so a publication anchor would assign otherwise identical inventions systematically different comparison sets.

[23] In general, cosine similarities range over $[-1,1]$. In practice, cosine similarities between these embeddings are strictly positive, so both components lie in the unit interval as stated.

[24] The ratio does not, however, make the two near-perfect complements: the iso-score curves of FS/BS are rays through the origin, so novelty and impact substitute along them at a constant rate. Appendix H decomposes the composite into its components.

Figure 1: *Conditional correlations – Validation*



Notes: The Figure shows the conditional correlations between my quality measurement and various patent quality proxies. The sample size for the regressions in Panel 1a and 1b is 3,814,653. In Panel 1c the sample size is 1,095,179 and in Panel 1d it is 1,947,269. The control variables include small and micro entity indicators, categorical dummies for the type of continuation of a patent application, and dummies for the first five years of examiner experience. The binned scatter plots are produced using `binsreg` (Cattaneo, Crump, Farrell, and Feng, 2024) and covariates are included via auxiliary regressions and residualizing the outcome variable of interest. 95% confidence intervals based on standard errors clustered at the art unit × publication year level are represented by the shaded areas.

3.4 Validation of the quality measurement

I validate the measure against four existing quality proxies: patent citations, renewals, the stock market news measure by Kogan et al. (2017), and the text-based measure by Kelly et al. (2021). Each proxy only covers part of my sample, since all four are based on granted patents rather than on applications, and in comparison to the latter two measures I use a longer time window extending to 2023. These coverage differences reduce the sample sizes for the comparisons, but within each of the overlapping samples every proxy provides an independent benchmark.

Figure 1 shows positive net conditional correlations between Q_p and each of the four proxies. The fitted curves are not monotone throughout, declining in certain parts of the lower tail before rising, which implies that the association is based on a net effect rather than on a pointwise ordering, and that the association is stronger in the upper tail. Panel 1a shows the correlation between my quality measure and year-normalized forward citations, the most common proxy for patent value. Since citations accumulate

after the grant and only for accepted applications, they take into account both the examiner's decision and the underlying invention. The positive correlation with my pre-grant measure therefore shows validity within the sample of granted patents rather than in the full population. Panel 1b uses realized renewals. Renewals capture the holder's own revealed-preference valuation, as renewal fees are paid every four years only when the private value of the patent exceeds the fee. Panel 1c uses the Kogan et al. (2017) measure, the most demanding comparison, since it is built from abnormal stock market returns around grant announcements and is therefore independent of patent texts. The measure only exists for granted patents, and its event is the examiner-driven announcement. Panel 1d uses the Kelly et al. (2021) measure, which is also text-based but rests on TF-IDF word frequencies rather than transformer embeddings. The convergence of two structurally different text measures suggests the quality signal is present in the patent text itself rather than an artifact of one embedding model.

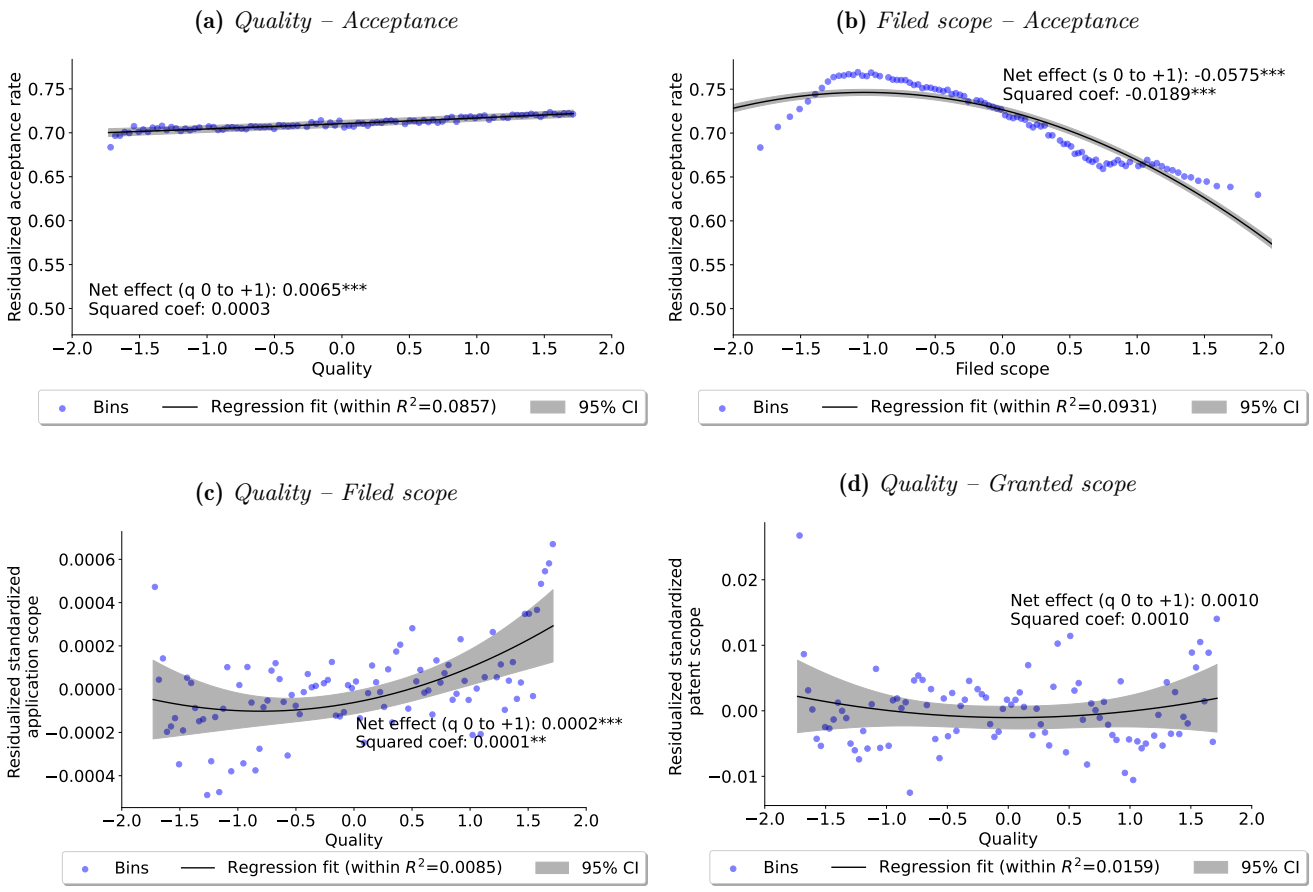
The bulk correlations are small with the index's standardized bulk slopes ranging only from 0.001 to 0.013. This is in itself a finding and not a result of the measure's construction. Q_p is estimated without reference to grant outcomes, to market value, or to the acceptance margin, so nothing in how it is built guarantees or precludes a strong association with realized value. The small slopes indicate that novelty-and-impact similarity and market value are only weakly associated in the bulk of the distribution, which reduces the usefulness of my measure as a proxy for cardinal value. When used as a screening tool, however, the appropriate test is not the bulk slope but the top of the ranking, and there the index provides useful information, the extent of which varies considerably depending on the proxy. As I show in Table H.2, relative to the median bin, the applications ranked highest by Q_p carry a citation premium rising from +38% at the top decile to +75% at the top percentile, against a market-value premium of +8% to +13% on the same bins. The steep half of the premium is carried by citations, while the market-value premium is positive but modest, which is what the nearly flat price of Appendix N reflects. I therefore read Figure 1 as establishing the direction of validity, with its magnitude carried by the tail premium in Appendix H, which develops both points.

4 Stylized facts

The filed scope of the application affects patent acceptance by an order of magnitude more than the invention's quality. Figure 2 shows conditional correlations, and the regressions behind them are in Appendix K.1. The top row highlights how quality and filed scope affect acceptance, and the bottom row shows the assocations between quality and scope. Panel 2a shows the acceptance rate increases only mildly in quality, by roughly two percentage points on a base of 71%. Importantly, the binned scatter shows no threshold of quality above which applications are materially more likely to be granted. Panel 2b puts the acceptance rate against filed scope, on an identical vertical scale. Acceptance rises over roughly the narrowest fifth of filings and falls steadily thereafter, from a peak of 77% to 63% at the right edge of the fitted window. Higher quality yields almost no additional chance of a grant, but narrower claims strongly increase the chances.

Panel 2c establishes the scope-quality relationship, which is central for the rest of the paper. The

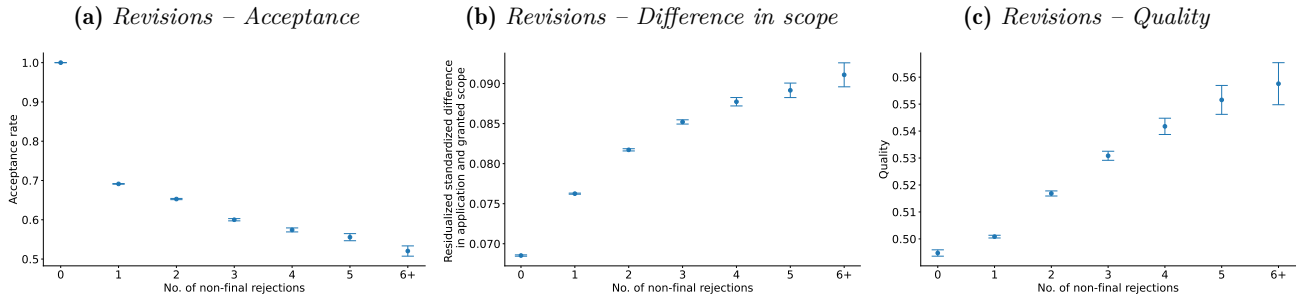
Figure 2: *Conditional correlations – Full sample*



Notes: The Figure plots conditional correlations among acceptance, invention quality, and claim scope. The sample size for the regressions in Panels 2a and 2c is 5,364,971 (the estimation window net of applications whose comparison window contains no neighbours); Panel 2b uses the full window of 5,367,642, as filed scope is defined for every application. For Panel 2d, the number of observations is 3,770,120. The control variables include small and micro entity indicators, categorical dummies for the type of continuation of a patent application, and dummies for the first five years of examiner experience. The binned scatter plots are produced using `binsreg` (Cattaneo et al., 2024) and covariates are included via auxiliary regressions and residualizing the outcome variable of interest. 95% confidence intervals based on standard errors clustered at the art unit × publication year level are represented by the shaded areas.

number of words in the average claim is standardized and transformed here, so that a higher value of scope implies less specific wording. The profile is a shallow U-shape. Filed scope is at its narrowest in the lower-middle of the quality distribution and rises through the upper half, so the net association from the middle to the top of the distribution is positive, while the least promising applications also file somewhat more broadly than the middle. Underlying this pooled profile are two different filing regimes. Applications whose claims derive from a prior document, continuations, divisionals, and national-stage entries, together 30.5% of grants, file more narrowly the higher their quality, while free-standing applications file more broadly. Section 7.3 returns to this distinction when the model reproduces this margin. Panel 2d shows that the rise in filed scope does not carry over into granted scope. Among accepted applications, granted scope bears no statistically distinguishable association with quality. Instead, the concession made during prosecution, the narrowing between filed scope and granted scope, rises in quality, as Appendix D and

Figure 3: *Patent and patent application attributes by the number of revisions*



Notes: The Figure shows the mean values of patent and patent application attributes by the number of review rounds the patent application had to go through before being accepted, abandoned, or receiving a final rejection. Because the number of patent applications going through multiple revision rounds decreases rapidly, I top-code the number of revisions at 6. The error bars represent the 95% confidence intervals based on the standard error of the means. The remaining attributes (filed scope, granted scope, forward citations, abandonment, renewals, the Kogan et al. (2017) and Kelly et al. (2021) proxies, and examiner leniency) are reported in Appendix Figure D.2.

Figure D.1 show.

Summarized, these correlations imply that better inventions ask for more scope, give up more of it during revisions, but are not granted more of it. A welfare-maximizing system in which larger quality steps warrant stronger protection (Scotchmer, 1991; Acemoglu and Akcigit, 2012) would tie granted scope to quality. I discuss the question whether the flatness observed here falls short of that optimum in Appendix N.

Across applications that end after different numbers of rounds, acceptance is lower and scope concessions larger the longer the prosecution ran. Figure 3 plots patent application and patent attributes over the number of revision rounds it takes before an application is accepted, abandoned, or receives a final rejection. The grouping variable is the realized number of rounds, an outcome of the prosecution it describes, so these profiles compare different populations rather than tracing what an additional round does to a given application. Panel 3a shows the acceptance rate declining in the number of revisions, and I show the mirror image in abandonments in Panel D.2d of Appendix Figure D.2, where only a small minority of failed applications receive a final rejection rather than being abandoned. Panel 3b replicates the finding of Marco et al. (2019) that the claims in the patent application are broader than the claims in the granted patent, and adds that this difference increases with the number of revision rounds. Filed and granted scope behave differently over prosecution length. Panel D.2a shows that filed scope rises monotonically, with no interior peak, so applications that ended after more rounds were filed broader throughout. Panel D.2b instead highlights that granted scope is hump-shaped, rising over the first two rounds and declining thereafter.[25] Longer prosecution is thus associated with both a lower probability of acceptance and larger scope concessions conditional on grant.

Higher-quality applications endure the longest prosecution. Panel 3c shows a strong positive relationship between my quality measure and the number of revision rounds. Since the measure is constructed from text fixed at filing, before any interaction with the examiner, this pattern cannot be an artifact of

[25] The location of the peak is not sharply resolved: the first and second rounds are statistically indistinguishable from one another, as are the third and later rounds, whose means are estimated on rapidly thinning cells.

examiners affecting the text. Panel D.2c shows that forward citations rise with the number of revisions as well. Since citations accumulate only after publication, they could partly reflect examiners making applications better or easier to understand, a channel the filing-fixed text of my measure excludes. The orthogonality of its forward-looking benchmark rests on the tests of Section 3. Renewals and the quality proxies of Kogan et al. (2017) and Kelly et al. (2021) follow the same profile over review length (Panels D.2e–D.2g), and, as these depend less on the examiner than citations do, they corroborate the same conclusion. Finally, Panel D.2h shows that patent applications assigned to a more lenient examiner go through fewer revision rounds. Taken together, higher-quality inventions go through a longer prosecution process, while a lenient examiner draw shortens the path to a decision.

These stylized facts imply an asymmetry in the USPTO's screening process. Applicants pay costs in the form of lower acceptance probability and more revision rounds for requesting more scope, while higher invention quality yields little compensation. This sharpens the observation of Bryan and Williams (2021) that any application could be granted if its claims were narrowed sufficiently. The operative margin of examination is how much is claimed, while the greater value of wider scope pushes applicants to file broadly against it. Section 5 establishes which applications sit at the resulting margin by first formalizing the underlying incentive structure and then setting out the identification strategy.

5 Empirical framework and strategy

This section first sets out the minimal structure the empirical strategy needs by describing what determines whether an application clears the examiner's bar, and therefore which applications are close to the margin the lottery decides. I then clarify the estimation. The framework part is deliberately short. The full dynamic game that I solve and estimate is developed in Section 7. Neither the empirical strategy of this section nor the results of Section 6 depend on it.

5.1 The acceptance margin

The framework is a sequential game between a patent applicant and an examiner, in four stages. (1) Nature draws invention quality q, which neither side observes. The applicant holds a subjective prior over their own quality, the examiner a prior updated from noisy signals. (2) The applicant, holding beliefs over their own quality and over the leniency of the examiner they will draw, chooses filed scope s. (3) The examiner draws a leniency type l, receives a noisy quality signal and observes filed scope directly. The examiner grants if a latent index combining the quality signal, the filed scope, and their own leniency clears a threshold, otherwise issuing a rejection. (4) Both players update by Bayes' rule, and the applicant either revises or abandons for an outside option normalized to zero. If the examiner grants, the game ends with a patent of granted scope $\hat{s} \leq s$.

Following Scotchmer (2004), a potential applicant A is endowed with an invention of quality q and R&D costs, treated as sunk costs at the time of filing. The applicant's strategic choice is the filed scope s of the patent claims submitted to the USPTO. Wider scope raises the value of a granted patent but lowers the probability of acceptance and raises expected prosecution costs through longer expected revisions.

Scope is chosen before the examiner is assigned, so the applicant cannot condition on l. Writing w for the applicant's belief over (q, l) at the time of filing (p is reserved for the application index), the objective is the ex-ante value

$$V(s \mid w) \;=\; \underbrace{\mathbb{E}_w\big[P(q,s,l)\cdot \nu \hat{s}\big]}_{\text{expected granted value}} \;-\; \underbrace{C(s)}_{\text{filing and revision costs}}, \tag{4}$$

where $P(q, s, l)$ is the probability of acceptance, $\hat{s}$ is the granted scope conditional on acceptance, ν is the per-unit value of IP rights, and $C(s)$ captures expected prosecution costs. The term inside the expectation is an interim object, the value to an applicant who already knew both their own quality and the examiner's leniency. Only its expectation over the applicant's beliefs w is the decision criterion, a distinction that matters once rejections start signaling l across rounds. I lay down the complete problem of the applicant in Section 7.3.

An examiner E has a type of leniency $l \in [0, 1]$. The examiner observes filed scope s directly and receives a noisy signal $\xi = q + \eta$ of the application's quality, with η drawn independently of the filed scope – so there is no signal in the applicant's choice – and the examiner's own type. The examiner grants when a latent index clears a threshold,

$$D \;=\; \mathbf{1}\Big\{ \underbrace{a(q,s)}_{\text{acceptance index}} + l \;+\; \varepsilon \;\geq\; 0 \Big\}, \qquad a(q,s) \;=\; \alpha_q\, \mathbb{E}[q \mid \xi] \;-\; \Pi(s) \;-\; \mu, \tag{5}$$

where ε is an idiosyncratic assessment shock, μ the threshold, and $\Pi(s)$ a scope penalty. $a(q, s)$ denotes the acceptance index conditional on the signal ξ. α_q describes the acceptance index' dependence on quality and $\Pi(\cdot)$ captures the dependence on filed scope.[26]

Leniency enters additively and separably from quality and scope. A lenient examiner tolerates a lower index, instead of reading the same application as better.[27] Conditional on acceptance, the granted scope satisfies $\hat{s} \leq s$, and I treat any reduction as permanent.[28] The interaction of quality, scope, and leniency generates three outcome functions, which are the structural counterparts of the empirical objects estimated in the remainder of this section. Namely, these are the acceptance probability $P(q, s, l)$ as the extensive margin, the granted scope $\hat{s}(q, s, l)$ conditional on acceptance as the intensive margin, and the number of revision rounds $r(q, s, l)$.

5.2 Implications for estimation

Four features of the framework directly shape the empirical strategy. First, filed scope is predetermined with respect to the examiner draw, chosen before the examiner is assigned. Therefore, filed scope can

[26] Nothing in the empirical strategy below or in Section 6 uses the functional form of Π, which is laid out in Appendix L.1. The empirical strategy uses the acceptance-by-scope profile of Section 4 as estimated, which rises over the narrow lower tail and declines from roughly the fortieth percentile of filed scope onward, so no global sign of $\partial a/\partial s$ is assumed.

[27] This separability makes the composition analytically tractable in Section 6.3, and it is the form the estimated model of Section 7 supports.

[28] This is a modeling assumption rather than a legal necessity, whose institutional basis, prosecution history estoppel, and the empirical absence of reversals are set out in Section 2.

serve as a placebo outcome in the randomization test of Section 5.4: if assignment is as good as random within art unit × year, realized leniency must be unrelated to it. Filed scope can also serve as a dimension along which to characterize the applications whose grant status the examiner draw decides, which is how I use it throughout Section 6. I estimate the effect of leniency on grant decisions at each percentile of quality and filed scope, so filed scope enters as the axis along which the complier density is traced rather than as a control. Importantly, whether or not I hold scope fixed for these estimated objects changes the interpretation. An object estimated at fixed s reflects the *direct* effect of quality, whereas one that averages over the within-quality distribution of s, such as the descriptive conditional correlations of Section 4, reflects the *total* effect, including the component operating through the applicant's own scope choice.

Second, the source of identifying variation is examiner leniency l combined with the as-good-as-random assignment of applications to examiners within an art unit. Conditioning on art unit × year fixed effects absorbs the technology field, the competitive environment, and the annual composition of the examiner pool. The remaining variation in l, between examiners examining similar applications in the same unit and year, identifies every leniency effect estimated below.

Third, the additive separability of leniency determines the form of the empirical strategy. Differentiating the grant rule (5) with respect to l gives a first-stage sensitivity

$$\psi(q, s; \bar{l}) \;=\; \left.\frac{\partial P}{\partial l}\right|_{l=\bar{l}} \;=\; \sigma^{-1}\varphi\left(\frac{a(q,s)+\bar{l}}{\sigma}\right). \tag{6}$$

This is the slope of the acceptance curve in leniency, and separability makes it depend on (q, s) only through the index $a(q, s)$. The level at which it is evaluated is still a choice. Throughout, I evaluate at $\bar{l}$, the mean leniency of the art unit × year pool the application was drawn from, which is also the reference point of the complier estimands in Section 5.5 and of every profile plotted in Section 6. The choice of reference does not drive the paper's comparative statements. Changing $\bar{l}$ shifts every application along one common curve, so the ordering of the grant-normalized composition over (q, s), the object every profile reports, and with it the sign of every gradient, is invariant to the reference. The invariance follows from the inverse Mills ratio being strictly decreasing. Section 6.3 derives the argument. It does not extend to the level or steepness of a profile, nor to the unnormalized sensitivity ψ, which is single-peaked rather than monotone. Therefore, magnitudes carry the reference $\bar{l}$, while orderings and signs do not.

Treatment-effect heterogeneity is therefore not a collection of separate coefficients to be estimated characteristic by characteristic. It is a single object, the distribution of compliers over (q, s). This object consists of a level, the average pass-through $\bar{\psi}$, and a shape, the complier density relative to the population density. Section 5.3 estimates the two separately and multiplies them. Section 6.3 shows that this factorization makes the composition a deterministic consequence of the threshold structure.

Fourth, the framework clarifies the population of interest. The design specifies outcomes for compliers, applications granted because they were assigned to a lenient examiner but rejected if they would have been assigned to a stricter examiner. My empirical strategy is silent on always-takers and never-takers, whose grant status no examiner draw can shift. The distribution of compliers across q and s is the central

object of Section 6. Since applicants of different quality choose different levels of s, this distribution is jointly determined by quality and scope choice, which motivates the two-dimensional characterization.

5.3 The first stage and the complier composition

The object of interest in this section is the complier distribution over the (q, s) plane, which the previous Section described. The complier distribution consists of a level multiplied by a shape and this subsection describes the estimation of both parts. Section 5.5 assembles them into the complier shares reported in Section 6.

Design. $D_{p_{kt}}$ is the grant outcome (acceptance or rejection) of application p in art unit k and publication year t, and $l_{p_{kt}}$ the leniency of the assigned examiner. Z collects the full set of examiner indicators and W collects the art unit $\times$ year fixed effects together with the patent-level covariates: small and micro entity indicators, dummies for the type of continuation of the application, a linear term in examiner experience (years since the examiner's first observed decision, top-coded at five), and an indicator for examiners already active in the first year of the data. The estimation sample consists of 5,367,639 applications. These applications are allocated to 12,706 examiners within 10,270 art unit $\times$ year cells, so Z and W together carry 22,986 design parameters.

Estimator. With instruments and controls this numerous, in-sample fitted values are overfit, and the resulting first stage is biased toward the ordinary least-squares association (Bound, Jaeger, and Baker, 1995). I therefore use the unbiased jackknife instrumental-variables estimator (UJIVE; Kolesár, 2013), whose instrument $U_p = \hat{D}^{[Z,W]}_{-p} - \hat{D}^{W}_{-p}$ leaves application p out of both the examiner projection and the high-dimensional control projection. The estimator is consistent under the many-instrument and many-control asymptotics that apply here and it imposes no functional form on how the controls enter. Appendix E reports the leave-out construction in detail and shows that the honest first stage recovers 96.6% of the in-sample two-stage least-squares explained variation, so the many-instrument distortion is small at this sample size.

The level: pooled pass-through. The first component of the complier shares is the average sensitivity of the grant to the examiner draw. Collapsing the examiner identities into the single recentered index $\Delta l_p = l_p - \tilde{l}_p$, the deviation of the assigned examiner's leniency from the average leniency of the pool serving the application's art unit at the time, I estimate

$$D_{p_{kt}} = \alpha_{kt} + \bar{\psi}\,\Delta l_{p_{kt}} + X_{p_{kt}}\Theta + \upsilon_{p_{kt}}, \tag{7}$$

obtaining $\hat{\bar{\psi}} = 0.67$ by fixed-effects least squares. The leave-out leniency index averages a finite number of prior decisions, so it is a noisy proxy for an examiner's true leniency, and this estimate is attenuated.[29]

[29] Instrumenting the index with the examiner dummies removes the attenuation and raises the pass-through to 1.06. Since leniency is itself measured in grant-rate units, the mechanical benchmark for the deattenuated coefficient is one, the same one-for-one relationship the art-unit projection of Section 3 exhibits, so the 1.06 is that benchmark recovered to within six percent.

I use the attenuated 0.67 level throughout the main text, which makes every reported share of compliers among granted patents a conservative floor. I bracket the implications of the correction in Appendix E. The choice moves the level of every profile proportionally and does not affect the shape at all.

The shape: complier density by modified outcomes. The second component measures the compliers. Following Goldsmith-Pinkham, Hull, and Kolesár (2026), I use modified outcomes: for a set B of values of a predetermined characteristic $X \in \{q, s\}$, the just-identified instrumental-variables coefficient of $\mathbf{1}[X_p \in B] \cdot D_p$ on D_p, instrumenting with U and partialling out W, recovers the probability that a complier's characteristic falls in B,

$$\theta(B) \;=\; \Pr(X \in B \,|\, \text{complier})\,. \tag{8}$$

The complier share itself cancels between the coefficient's numerator and denominator,[30] so $\theta(\cdot)$ is a distribution. Evaluated on a partition, it sums to one, and it requires no auxiliary weighting scheme. Its ratio to the population density,

$$\rho(x) \;=\; \frac{f_{X|\text{complier}}(x)}{f_X(x)} \;=\; \frac{\psi(x)}{\bar{\psi}}, \tag{9}$$

is the complier representation ratio, which exceeds one where compliers are over-represented relative to the population. The second equality is the content of the factorization in Section 5.2, that complier mass at x is population mass times marginal sensitivity, and it is what allows the conditional first stage to be recovered as the product $\psi(x) = \bar{\psi} \cdot \rho(x)$ rather than estimated separately at each x.

This product is a conditional first-stage sensitivity that captures the effect of leniency on grant probability. The complier objects I present as my results further scale this sensitivity. The treated-complier mass at x multiplies $\psi(x)$ by the favorable dose $\mathbb{E}[(\Delta l)_+ \mid X = x]$, which measures how much more lenient the drawn examiner was in comparison to the rest of the eligible pool. The share among granted patents further divides the treated-complier mass by the conditional grant rate. For further analysis, I do not only investigate the favorable draws but all applications affected by the examiner draw – grants due to lenient examiners and rejections due to strict examiners. The resulting two-sided flip rate uses $\mathbb{E}[|\Delta l| \mid X = x]$ as the treatment dose. Using one pooled dose for every x is valid if the dose does not vary with x, so that all the x-dependence of complier mass runs through the sensitivity $\psi(x)$. I report the test in Appendix F, which suggests that the residual variation is negligible.[31]

The level and the shape stem from different constructions, but the two components can be combined. The level uses the recentered scalar leniency index as treatment. The resulting single coefficient is the interpretable average dose-response. The shape instead uses the full set of examiner dummies through

[30] I show the sample ratio in Appendix E.

[31] The scalar-index structure of Section 5.2 delivers the setup that the full effect of x goes through $\psi(x)$. Testing the assumption, the conditional dose is flat to within about three percent of its pooled value on both axes. Correcting for the residual variation would increase the scope gradient by roughly a tenth of its own size, so the factorization's error is modest and conservative. Running the same construction on the untreated arm, $\mathbf{1}[X_p \in B] \cdot (1 - D_p)$ on $1 - D_p$, gives a distributional balance test in every bin. This approach is the design-based analog of characterizing compliers by the distribution of idiosyncratic leniency deviations (Borusyak et al., 2023).

the leave-out instrument, which removes own-observation bias bin by bin. The pass-through cancels in the complier representation ratio, so the choice of construction shifts the shape but not the level. The construction is bounded in Appendix E.[32]

Resolution, inference, and robustness. I report $\rho(\cdot)$ one axis at a time – separately over quality or over filed scope – and jointly over both. For the one-dimensional profiles, I use a Gaussian kernel over the axis of interest. For the joint profile, I construct a quality $\times$ filed-scope plane of eight equal-mass bins per axis. The relative coarseness of the grid is deliberate. The two axes are positively correlated, so the low-quality and broad-scope cell that I discuss in Section 6.2 is sparse. Using eight bins keeps the low-quality and broad-scope cell populated enough to carry a standard error and a per-cell balance test rather than an extrapolation. The one-dimensional profiles and the joint profile answer two different questions: the one-dimensional quality profile averages over the scope choices applicants make at each quality level. This is the total object of Section 5.2, which includes both the positive direct effect of quality on acceptance and the negative indirect effect of higher-quality inventions asking for broader scope, which in turn leads to lower acceptance rates. The joint profile instead holds scope fixed, reporting the direct effect of quality. Across all reported objects, I cluster standard errors on the art unit $\times$ year cell, the level at which examiners are assigned.

5.4 Identifying assumptions and their tests

Reading the objects of Section 5.3 as a complier distribution requires two assumptions: that applications are assigned quasi-randomly to examiners within art unit $\times$ year, and monotonicity – that no application is granted by a stricter examiner but rejected by a more lenient one. I discuss both assumptions here.

Quasi-random assignment. Judge fixed-effects designs for patent-related outcomes at the USPTO rest on two premises: examiners differ in leniency, and applications are assigned to them quasi-randomly within an art unit. Under quasi-random assignment, examiners in the same unit receive, in expectation, the same mix of applications, so their differing grant rates reflect different standards rather than different caseloads. However, Righi and Simcoe (2019) raise doubts about the validity of these instruments. They find evidence for examiner specialization within technology groups. Additionally, specialized examiners are also less lenient, which could violate the exclusion restriction. To address this concern, I report two tests. The pooled regression below cannot reject the conditional independence of leniency from quality or from filed scope. A sharper bin-by-bin version of the same test does reject on the scope axis: testing the bin-conditional mean of the recentered instrument separately in each of twenty equal-mass bins, eleven of twenty filed-scope bins reject, against one of twenty quality bins, and thirteen against six when I cluster

[32] My construction is not an equivalence result, but I report supporting checks. Reading $\bar{\psi} \cdot \rho(x)$ as the level and shape of one common conditional first stage requires the two constructions to weight heterogeneous examiner effects proportionally. The appendix's aggregate-signal and profile-agreement checks support it empirically.

on art unit × publication year instead. [33] Section 6.2 and Appendix F show the rejection is mechanical, a consequence of measuring leniency as a realized grant rate over a caseload whose scope composition varies. Removing the bias steepens the scope gradient, so my baseline is conservative.[34]

The pooled regression is

$$l_{p_{kt}} = \alpha_{kt} + \beta_q^{\text{bal}}\, q_{p_{kt}} + \beta_s^{\text{bal}}\, s_{p_{kt}} + X_{p_{kt}}\Theta + \epsilon_{p_{kt}}. \tag{10}$$

The regression uses the leniency l of the examiner assigned to application p as an outcome and my quality and filed-scope measures as regressors, jointly with the covariate vector X of the first stage and art unit × publication year fixed effects α_{kt}[35]. Standard errors are two-way clustered on the applicant's USPTO customer number[36] and on the art unit. The design follows the randomization test of Dobbie et al. (2018). Table 1 shows both coefficients economically negligible and statistically indistinguishable from zero.[37] Equation (10) is a placebo test since both variables are predetermined at assignment, and maintaining conditional quasi-random allocation extends the same independence to unobservables. Appendix F reports the complementary check on the estimated complier pool itself, comparing complier means against population means characteristic by characteristic.

Monotonicity. The LATE framework requires that any application granted by a less lenient examiner would also have been granted by a more lenient examiner. The joint test of the monotonicity assumption and the exclusion restriction proposed by Frandsen, Lefgren, and Leslie (2023) requires a second-stage outcome. For my object of interest, I estimate only a first stage, so I test the weaker average monotonicity condition. Frandsen et al. (2023) show that average monotonicity implies a non-negative first-stage coefficient across all subsamples of a given dimension. The sensitivity of grant to recentered leniency is positive across all sixteen publication years and across all one hundred quality percentiles. It is also non-negative in 97.9% of art units, with significantly fewer individually significant negatives than chance

[33] The quality axis carries the paper's finer claims, so what matters here is the contrast between the two axes. Clustering two-way on customer number and art unit, the quality axis rejects in one bin of twenty, the chance rate, with a largest $|z|$ of 2.0 against the scope axis's 4.1. Clustering on art unit × publication year the quality axis rejects in six, with a largest $|z|$ of 4.1 against scope axis's 9.1. The quality axis's worst case under either clustering is no larger than the scope axis's best, and quality is if anything the more caseload-persistent of the two characteristics, so this is not a weaker test on that axis.

[34] These tests balance the within-cell rank, which is what every estimate conditions on. The raw level of the quality index behind that rank does carry a small association with the instrument. Appendix F attributes it to the gap between level and rank, and to the control set of the first stage.

[35] The balance-regression coefficients carry the bal superscript to keep them apart from the model's filing loadings β_q, β_s, and ζ is reserved for the cost type

[36] The customer number identifies the correspondence address the applicant filed with the patent office. The customer number therefore belongs to whoever prosecutes the application, which can be an external law firm or the applicant's own patent department. Clustering on it groups applications handled by the same prosecuting actor, which nests the applicant for in-house filers and is a coarser grouping otherwise.

[37] Because quality and filed scope both enter as percentile ranks on the unit interval, the reported coefficients are *full-range* effects: moving an application from the bottom to the top of its cell's quality distribution shifts assigned leniency by -0.0020, and across the distribution in filed scope by -0.0013. Against a leniency distribution with an interquartile range spanning 26.5 percentage points, the entire quality range moves assigned leniency by 0.2 percentage points. Per percentile, the implied shifts are -0.00002 and -0.00001.

Table 1: *Test of conditional randomization*

	(1) Leniency
Quality	-0.0020
	[0.0014]
Filed Scope	-0.0013
	[0.0011]
Controls	✓
Art unit x decision year FE	✓
Observations	5,367,642
Adjusted R^2	0.593

Notes: The Table presents the results of regressing examiner leniency on my quality and filed scope measures. The control variables include small and micro entity indicators, categorical dummies for the type of continuation of a patent application, and examiner experience controls. Fixed effects are art unit × publication year. Standard errors two-way clustered at the art unit level and at the applicant's USPTO customer number are reported in parenthesis.

would produce.[38] Taken together, I find support for the average monotonicity assumption.

Another related concern is that leniency might be two-dimensional, such that an examiner could be lenient on quality but strict on scope. This would generate defiers specifically on the scope margin. In Appendix F Table F.2, I tackle this concern and show that examiner's leniency ranks are robust across scope. I separately compute the leniency of an examiner on the narrow and the broad halves of their own caseloads and find that the two rankings agree to 96% of the ceiling that sampling noise allows. Scope-specific defiers are therefore rare.

5.5 From the first stage to complier shares

Under the threshold rule of Equation (5), the potential grant outcome $D_p(l)$ is almost surely non-decreasing in examiner leniency l. For each application p, I define the threshold type as $l_p^* = \inf\{l : D_p(l) = 1\}$, the minimum examiner leniency that results in a grant. An application counts as a complier if its realized and expected leniency fall on opposite sides of l_p^*. This is a strictly smaller, baseline-relative set than the one used in the judge-design literature, and it answers how many grants would be reclassified against the examiner an applicant could reasonably have expected rather than how many are sensitive to some examiner in the office. Throughout, "complier" refers to this reclassification event.

I denote the counterfactual expected leniency for application p as $\tilde{l}_p$. This is the average leniency of the examiners who decided applications in the same art unit k within an eighteen-month window on either side of p's own decision date. The construction is inspired by the recentered instruments of Borusyak et al. (2023) and the shift-share framework of Borusyak, Hull, and Jaravel (2021). The random component of the examiner assignment is the deviation $\Delta l_p = l_p - \tilde{l}_p$, which is approximately independent of application characteristics by quasi-random examiner assignment within art unit and year, and which is the treatment variable in the first stage represented in Equation (7).[39]

[38] However, not all of them are imprecise. A studentized sup-test of joint non-negativity across quality bins does not reject, while exactness in every single art unit is rejected by the most negative one, as Appendix F, Table F.1 reports in full.

[39] In the estimation sample the average *favorable* draw is $\bar{d} \equiv \mathbb{E}[(\Delta l_p)_+] = 0.050$, against a grant rate of $\Pr(D = 1) = 0.711$.

I model the conditional grant probability as a linear probability model in leniency:[40]

$$P(D = 1 \mid l,\, x) = \alpha(x) + \psi(x) \cdot l, \qquad x \in \{q,\, s\},$$

where $\alpha(x)$ absorbs the baseline grant probability and $\psi(x) = \bar{\psi} \cdot \rho(x)$ is the conditional first-stage sensitivity assembled from the pooled pass-through of Equation (7) and the representation ratio of Equation (9). The key implication is that the marginal effect of leniency on grant probability is constant across leniency levels, since it depends on the application's characteristics. The linearity is local in the observed window.[41]

Flip rate. Application p at characteristic value x is a complier if the realized leniency falls on the opposite side of l_p^* from the expected leniency, so that it is granted by an above-average examiner when the expected one would have rejected, or the reverse. Taking expectations over the instrument distribution gives the flip rate, the share of all applications at x whose grant status is decided by the examiner draw:

$$\phi_c(x) = \psi(x) \cdot \mathbb{E}\Big[\,|l_p - \tilde{l}_p| \,\Big|\, x\Big] = \bar{\psi}\, \rho(x) \cdot \mathbb{E}[|\Delta l_p|]\,. \tag{11}$$

Equation (11) multiplies a level, the sensitivity $\psi(x)$ in probability per unit of leniency, by a shape, the dose $\mathbb{E}[|\Delta l_p|]$. Only the resulting product has the units of a share of applications. Since the deviation Δl_p is independent of x under quasi-random assignment, the dose is common across applications and all of the variation is carried by $\rho(x)$, the complier representation ratio, so the profiles in Section 6 trace $\psi(x)$ up to a common scale. Because the product linearises a finite difference, they approximate the flip rate rather than counting it. Appendix E derives Equation (11) in detail.

Share among the granted. The main object of Section 6 is more specific than the flip rate. It asks, among the patents actually granted at x, what share of the grant is due to examiner draw? The answer is those applications with $l_p > l_p^* > \tilde{l}_p$, whose realized examiner leniency exceeded both the application's threshold and the expected leniency. Their mass at x is the sensitivity $\psi(x)$ times the average favorable draw $\bar{d} = \mathbb{E}[(\Delta l_p)_+]$, and expressed as a share of the patents granted at x it is

$$\phi_c^{D=1}(x) \;=\; \frac{\psi(x) \cdot \bar{d}}{P[D = 1 \mid x]} \;=\; \frac{\bar{\psi}\; \rho(x)\; \bar{d}}{P[D = 1 \mid x]}. \tag{12}$$

The three components of Equation 12 are the objects discussed in detail in Section 5.3.[42] The scaling factor is the average deviation from the within-art-unit expected leniency, not the conventional gap between the most and least lenient judge (Dahl, Kostøl, and Mogstad, 2014; Dobbie et al., 2018; Agan et al., 2023),

[40] Under a nonlinear model such as probit, the marginal effect of leniency on grant probability would be proportional to $\phi(\Phi^{-1}(P(D = 1 \mid x)))$, which attenuates toward zero when grant probabilities are far from one half. A probit specification therefore compresses the profile where grant rates are extreme. The shape of the underlying sensitivity $\psi(x)$, identified by the modified-outcome (UJIVE) estimator of Section 5.3 through the representation ratio (9), does not depend on this assumption.

[41] Over the full range of leniency, the grant probability follows the typical probit S-shape, which the counterfactuals in Section 7.7 use. Instead, under a globally linear leniency effect, compressing the dispersion of leniency around a fixed mean would leave the grant unaffected.

[42] The conditional grant rate $P(D = 1)$ is computed as a kernel average on the same grid as $\bar{\psi}$, ρ, and $\bar{d}$. Inference depends on $\rho(x)$, whose sampling variation dominates the estimation. The remaining terms shift the level of every profile, not its shape. Appendix E gives the derivation and the bookkeeping. Averaged over the population, the share collapses to $\bar{\psi}\, \bar{d}\, /\, P(D = 1)$, the aggregate reported in Section 6.1 and decomposed in Appendix E.

so the resulting share counts grants reclassified against the examiner an applicant could reasonably have expected rather than against an extreme draw. Section 6.1 quantifies the difference this convention makes. The same construction is applied with filed scope s in place of quality q, and jointly over both on the two-dimensional grid of Section 5.3.

6 Reduced-form results

6.1 Overall complier share

In my sample, 4.7% of granted patents owe their grant to an examiner more lenient than the average examiner the application could have drawn. My result is roughly four times the 1.2% that de Rassenfosse et al. (2020) estimate for the USPTO. Their figure is the share of granted US patents that would be refused if re-examined under the USPTO's own standard, recovered from twin applications filed at several different national patent offices. This is the same object as the one I estimate, inconsistency in the application of a single office's standard holding the invention constant. However, they reach this object through disagreement between offices rather than through variation in examiner permissiveness within one office. Their construction counts only disagreements detected across twin filings, so their rate is a lower bound, and Appendix E explains why their result is lower than mine in detail.

4.7% is a conservative reading of my estimation. Appendix E also brackets it against measurement error in the leave-out leniency index. Measurement error can raise the share, to 6.0% once both the pass-through and the leniency dose are corrected for noise, and to as much as 7.5% when only the pass-through is corrected. I use the conservative lower bound throughout.

The natural external comparison is with the judge fixed-effects designs in the economics of crime, which report complier shares of 8 to 13%. Those shares measure a different object.[43] Estimating a similar object, I find that examiner dispersion at the USPTO is far wider than the judge dispersion. The difference in grant probability between the ninety-ninth and the first percentile of my recentered leniency distribution is roughly forty percentage points under the baseline, attenuated pass-through, three to four times their shares.[44]

I focus on the share of compliers among granted patents, but the mirror-image also exists: the share of applications rejected only due to the draw of a strict examiner. The favorable draw mass, applications affected by the examiner draw, is 3.4% of all filings. Divided by the granted pool, this gives the 4.7% of

[43] The conventions differ in both the contrast and the denominator. In the economics of crime literature, the contrast is a full-sample difference in treatment probability between the most and least lenient decision-maker, conventionally the ninety-ninth and first percentiles of the leniency distribution. My contrast is the favorable-draw mass against the draw an applicant could have expected, reported as a share of granted patents. Dobbie et al. (2018) identify 13% of their sample as compliers under a local linear first stage and 11% under a linear one, and Agan et al. (2023) approximately 10%, and $8-10$% across the cutoffs and specifications they report. An extreme-versus-extreme contrast mechanically counts more marginal cases than an expected-draw contrast, so a smaller number here is what the definitions predict, not a like-for-like difference in complier prevalence.

[44] A structural asymmetry distinguishes the two settings as well. A defendant cannot shape the case a judge rules on. The treatment is assigned by the judge, and defendants cannot affect the "quality" of their case at the time the judge makes their decision. USPTO applicants choose filed scope before the draw, and that choice concentrates them near the acceptance margin, a mechanism with no analog in pretrial detention.

compliers among granted. Since the leniency deviation is recentered to mean zero, the mass of applications granted on a favorable draw exactly equals the mass rejected on an unfavorable one. Dividing this number by the rejected pool, which is only 28.9%, gives 11.6%, so roughly one rejection in nine is a lottery rejection against one grant in twenty-one. The two figures are one fact reported over two denominators, not two effects of different size.[45]

6.2 Rising in scope, flat in quality

The share of compliers among granted patents increases strongly in filed scope and is nearly flat in invention quality. The quality profile is the total object of Section 5.2, averaging over the scope applicants choose at each quality level. Marginal grants are broader applications at every level of quality and quality does not protect applications from the lottery. Depending on the choice of scope, the lottery can affect any segment of the quality distribution. Figure 4 plots the conditional complier share across both quality and filed scope. Panel 4a shows the quality profile, which moves only within half a percentage point of its mean across the whole distribution. It is slightly elevated at the lowest percentiles, reaches a minimum of approximately 4.6% through the middle, and rises to approximately 5.0% around the eightieth percentile before decreasing again at the very top. This shape does not reflect a gradient in the share of compliers across quality, but an interior lift.[46] In Panel 4b, I show the scope profile. The share rises from approximately 4.0% among the narrowest filings to approximately 5.6% among the broadest, an increase of roughly 40%.

On the joint surface the two margins reinforce one another rather than substituting for one another, and the scope increase is not a mechanical reflection of quality. Quality and filed scope are related across applications, though not monotonically, as established by the shallow-U profile in Section 4. However, the increase in scope survives on the two-dimensional surface, which holds quality fixed. Within every quality band, the share of compliers among granted patents rises monotonically across the scope bins, so the concentration on ambitious claims holds independent of the underlying quality of the invention. Holding scope fixed instead gives the direct effect of quality, which is negative, as the share within every scope band is higher at the bottom of the quality distribution than at the top. However, the scope margin has the stronger effect. The flat quality profile above is therefore a total effect, in which the direct decline is offset by the broader scope that stronger applicants file. I plot the joint surface in Appendix Figure D.5 and discuss the two-dimensional surface in more detail in Appendix D.

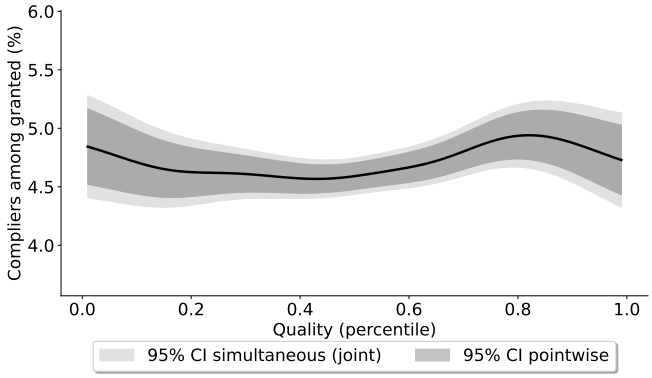
Appendix J shows what the composition of compliers implies for existing work that uses the patent

[45] For an individual application, neither share is the relevant object. The relevant measure is the probability the draw decides its outcome at all, counting both directions. The grant a lenient draw creates and the rejection a strict one forces, which makes it roughly twice the favorable-side mass: it runs from 6.4% in the cell combining the highest quality with the narrowest claims to 7.2% in the cell combining the lowest with the broadest. The per-application exposure therefore moves by less than a point where the share among granted patents spreads over 2.4 points, which anticipates Proposition 1 of Section 6.3: the larger part of the gradient in the composition is the denominator $\Phi(t)$, and the smaller part is the density of applications at the margin. Along the scope margin alone, the split is roughly two fifths from the density and three fifths from the denominator, as stated in the introduction.

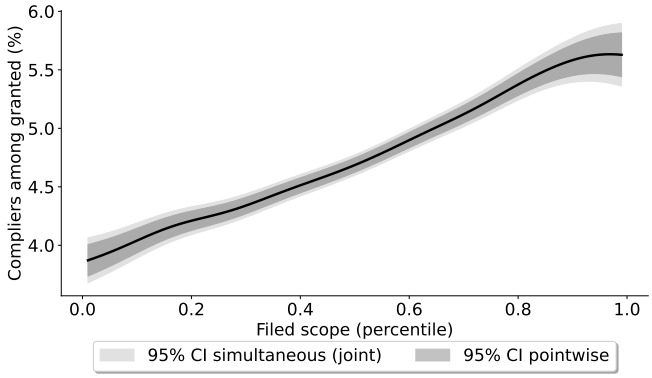
[46] The endpoints are statistically indistinguishable, while a simultaneous test rejects exact flatness of the complier density ($p = 0.03$ for the overall share of compliers, $p = 0.05$ for the share of compliers among granted patents), driven by the interior lift. In Section 6.3, I show that lift is the scope choices of strong applicants, not a quality effect.

Figure 4: *Conditional share of patents accepted due to the examiner*

(a) *Conditional share of patents accepted due to the examiner by quality*



(b) *Conditional share of patents accepted due to the examiner by filed scope*



Notes: The Figure plots the share of compliers among granted patents, $\phi_c^{D=1}$, across the percentiles of quality and of filed scope, respectively. Estimates come from the examiner-dummy UJIVE modified-outcome estimator of Section 5.3: the complier representation ratio $\rho(x)$ is identified from the full set of examiner instruments, multiplied by the pooled first-stage pass-through $\hat{\bar{\psi}} = 0.67$ and the average favorable leniency draw, and divided by the conditional grant rate, as in Equation 12. The shaded areas show 95% pointwise (dark gray) and 95% simultaneous (light gray) confidence bands; the outermost percentiles (below 0.03 and above 0.97) are trimmed as kernel-edge regions.

lottery, re-examining the settings of Sampat and Williams (2019) and Farre-Mensa et al. (2020) with marginal treatment effect curves. The integrated average effect lies below the LATE the paper's own specification delivers in every Farre-Mensa et al. (2020) outcome, and is statistically indistinguishable from it in Sampat and Williams (2019).

Robustness. I test the near-flatness of the quality profile across multiple decisions in the construction underlying the result. In Appendix H, particularly Table H.3, I show that the near-flatness of the composite index is the result of offsetting components rather than an absent signal. Decomposing the quality measure into its novelty (backward-looking) and impact (forward-looking) parts, shows that novelty decreases the chances of a grant and impact increases it. Therefore, estimating the share of compliers on novelty alone, the profile rises where the composite does not. In Appendix H, I show that rebuilding the index under eight alternative construction choices – window lengths and gaps, comparison pools, and the anchor date – leaves the profile in the same narrow band around its mean, with no variant rejecting flatness. Appendix H also bounds the attenuation from measurement noise, showing that noise cannot hide a steep true gradient.[47]

Apart from questions about the construction of my measure, I also test the robustness of the compliers profiles to sample composition. Continuation applications inherit prosecution history from a parent application, which could bias the estimates if their acceptance probability differs from that of original applications at the same quality, or inflate the effective leniency variation through the within-examiner correlation between parent and child filings. Re-estimating on non-continuation applications leaves the

[47] I score each application twice on independent halves of the text corpus, which results in a full-corpus rank reliability of 0.94. This rank reliability bounds the attenuation so that the true gradient can exceed the observed one by at most seven percent, against an observed profile that moves within half a percentage point of its mean.

share rising in filed scope and flat in quality. Restricting the sample to the art unit × year cells in which the USPTO allocated applications by random terminal docket digit (Feng and Jaravel, 2020) reproduce the scope rise on a fifth of the sample, with a quality profile that declines rather than flattening. In Appendix G, I discuss both sample restrictions in detail, particularly that marginal grants are not located in the high-quality tail, so my main conclusion on the profile of the share of compliers across quality remains intact.

Finally, an examiner who is handed an unusually broad caseload would have a lower grant rate without necessarily being stricter. Since grant probability falls steeply in filed scope, measured leniency carries a caseload component. In Appendix F, I show that purging this component from the measure only steepens the observed gradient of the complier share in scope, while not affecting the flat share across quality. Appendix F states the argument formally and signs the bias.

6.3 Mechanism: the complier composition is an inverse Mills ratio

The composition documented above follows from a single feature of the grant decision. Compliers are the applications close to the grant margin, the ones an examiner draw can flip, and that margin is a threshold in an index that weights claimed scope more than measured quality. The asymmetry is institutional. The main rejection reasons require examiners to assess whether a claim covers a novel and non-obvious step, and whether a claim covers more than the application discloses. Scope therefore sets an application's distance to the margin, and quality enters only through the weight the acceptance index places on the examiner's assessment. One object, the sensitivity of the equilibrium grant probability to the examiner draw, then delivers all three findings: the rise in filed scope, the flat total quality profile that declines at fixed scope, and the near-flatness of the flip rate.[48] This needs only the threshold structure of Section 5.1. Section 7 develops the model that generates it in equilibrium and that I estimate to price the inconsistency.

Take the threshold rule (5) of Section 5.1, in which the deterministic acceptance index $a(q,s) = \alpha_q\,\mathbb{E}[q] - \Pi(s) - \mu$ collects the quality premium and the scope penalty. Its derivative in leniency is the first-stage sensitivity of Section 5, $\psi = \sigma^{-1}\varphi((a+l)/\sigma)$, the density of applications sitting at the grant margin. Dividing the treated-complier mass $\psi \cdot \bar{d}$ by the grant rate P from Equation 12 gives the share of compliers among granted patents as, up to the leniency dose $\bar{d}$,

$$\phi_c^{D=1}(q,s) \;\propto\; \frac{\varphi\big((a+l)/\sigma\big)}{\Phi\big((a+l)/\sigma\big)} \;=\; \lambda\big((a(q,s)+l)/\sigma\big), \tag{13}$$

the *inverse Mills ratio* of the acceptance index. Because $\lambda(\cdot)$ is strictly decreasing, the entire complier composition is a monotone transformation of where an application is located in the grant distribution. The reference leniency $\bar{l}$ enters as a constant added equally to every application's index. A common shift cannot reorder applications, so the ordering of the composition over (q,s), and with it the sign of every gradient, depends on $a(q,s)$ alone. This is the invariance asserted in Section 5.2. It stops at the ordering because λ is nonlinear: a shift common to every argument is not a shift common to every value, so a

[48] Appendix B.1 states the same structure in four canonical threshold models.

change in $\bar{l}$ moves the level and the steepness of a profile.

Under a latent threshold-crossing model, by construction, the first-stage derivative is the density of the index at the margin, which is the marginal treatment effect weight of Heckman and Vytlacil (2007). Kline and Walters (2019) show that such a model and the instrumental variables estimator are numerically the same object. My following proposition adds a restriction: it forces the three gradients above to be readings of a single scalar index, so that any pattern violating that ordering falsifies the threshold structure.

Proposition 1 (Complier composition)**.** *Under the threshold grant model with additively separable leniency and Gaussian noise, the share of compliers among granted patents $\phi_c^{D=1}(q,s)$ is (i) strictly increasing in any application characteristic that lowers the acceptance index $a(q,s)$ and invariant to any characteristic that leaves it unchanged; and (ii) the unconditional flip rate* $\Pr(\textit{complier} \mid q,s) \propto \varphi((a+l)/\sigma)$ *is single-peaked in the acceptance index, attaining its maximum where the grant probability equals one half.*

Proof in Appendix B.

Proposition 1 turns the two stylized facts of Section 4 directly into the two complier facts of Section 6.2. Along the scope margin, acceptance decreases strongly, with the conditional grant rate dropping from 77% at the tenth percentile of filed scope to 63% at the ninetieth. Therefore, the acceptance index falls, and by part (i) the complier share rises from approximately 4.0% to 5.6% between the same percentiles.

Along the quality margin, the logic runs through both the direct and the indirect channel, highlighting that the direct complier share and the total complier share in Section 6.2 answer different questions. Holding filed scope fixed, as the two-dimensional surface does, the quality premium ($\alpha_q > 0$) raises the acceptance index, so by part (i) the complier share must fall in quality. The surface shows a decrease from 4.9% to 4.1% within every scope band. The profile of Panel 4a is the total share of compliers: it averages over the scope choices high-quality applicants actually make, and those choices run against the direct premium. As shown in Section 4, filed scope rises with quality through the upper half of the distribution, and wider claims lower the acceptance index, so the indirect scope channel raises the complier share exactly where the direct premium lowers it. The same offset shows up in acceptance itself, which rises only mildly in quality, one to two percentage points across the full distribution: part of the direct premium is spent on scope. Part (i) therefore predicts a total profile much flatter than the direct one, with its deviation concentrated where strong applicants file broadest, which is the shape Panel 4a shows: near-flat, with an interior lift around the eightieth percentile.[49]

Part (ii) disciplines the flip rate. Every cell of the two-dimensional surface has a grant probability above one half, in a range of 56% to 79%, so the data lie entirely on one side of the peak at $P = \frac{1}{2}$. The flip rate should therefore rise as the acceptance index falls, but never reach its maximum. I show this in Appendix Figure D.5. The flip rate changes by less than a percentage point between the corner cells, while the share among granted patents, which is divided by $\Phi(t)$, moves by over 2.4 points between the same cells and peaks in the low-quality, broad-scope corner at 6.5%. Across the surface, the flip rate

[49] Compare the direct-versus-total remark in Section 5.2.

declines with the cell grant rate[50] and is highest in the cells closest to a grant probability of one half. Therefore, the larger part of the gradient in the inverse Mills ratio comes from the denominator $\Phi(t)$, and the smaller part from the density of applications at the margin. Along the scope margin, where the introduction states the split, the density accounts for roughly two fifths of the log change in the share from 4.0% to 5.6%, and the grant rate, falling from 77% to 63%, for the remaining three fifths.

The same mechanism speaks to the "hard to judge" reading, which has two parts: that marginal grants are good inventions, and that the office cannot assess them reliably. The first part would require the complier pool to concentrate at the top of the quality distribution. The model cannot deliver it, because high-quality applications carry a high acceptance index $a(q, s)$, and $\lambda(\cdot)$ is a decreasing function. Therefore, by part (i) a high index implies a lower complier share, not a higher one. At fixed filed scope the share must fall in quality, and Section 6.2 reports that fall at every point of the surface's scope range. Two further facts point the same way. Compliers average 0.6 percentile points below the population mean quality.[51] The total profile does not peak at the top either: the interior lift sits below the ninetieth percentile and the share decreases above that.

The second is a statement about the variance of the examiner's assessment. In Equation (5), I hold σ fixed across applications, but the design does not necessitate that. A probit of the grant outcome on the leniency deviation, run within cells of quality and filed scope, has slope β/σ, where β is the pass-through of leniency into the acceptance index. The art unit $\times$ year pool effect enters that index additively rather than through β, so its coefficient in the same probit varies with σ alone. The two slopes together identify β and σ separately. Appendix B.2 reports the estimates in Table B.1. Assessment noise falls by about a sixth across quality and rises by about a sixth across filed scope, and weighting applications by the complier density of Equation (11) puts the compliers' assessment noise 0.32% above the population mean.

Assessment noise is not only too small to create the observed complier distribution, its effect even points in a different direction. The complier density is $\varphi(z)/\sigma$ at $z = (a + l)/\sigma$, so $\partial\psi/\partial\sigma$ carries the sign of $z^2 - 1$. Noise therefore adds to the margin only more than one standard deviation from the threshold, which means a grant probability outside of $[0.16, 0.84]$. Every cell of the surface reported above lies between 0.56 and 0.79. Over the range the data cover, a noisier assessment thins the complier margin rather than supplying it, so being hard to assess and being sensitive to the draw are different properties and in this sample they pull in opposite directions. Instead, the data supports the reading that the marginal grants are applications with claims beyond what the invention supports.

7 A structural model of prosecution

The reduced-form composition of Section 6 is a fact about equilibrium behavior: it reflects both sides of a negotiation over a boundary. The applicant chooses how much to claim, knowing prosecution is a one-way ratchet in which claims can be narrowed but never broadened, so a broad filing is the option-preserving opening position and its price is distance to the grant margin. The examiner polices the claimed boundary under a threshold rule tilted by the leniency draw. To separate these two sides, and to run counterfactuals,

[50] The flip rate and the grant rate have a between-cell Spearman rank correlation of -0.74.
[51] The difference in means is statistically significant with $z = 4.5$.

a model is needed that takes into account both. The model serves three purposes. First, it allows me to rank consistency reforms against alternative levers of prosecution intensity, like a cap on the number of revision rounds. Second, it prices the scope these reforms reallocate against its unobserved social cost, which I assign based on an external benchmark. Third, it distinguishes between the misallocation the patent office could actually remove from level effects, which a partial-equilibrium model cannot price.

I take the acceptance margin of Section 5.1 to its dynamic form and estimate it. The model embeds the grant technology of Section 5.1 in the multi-round prosecution game described in Section 2. Two design choices keep the exercise linked to the first half of the paper. First, leniency enters through draws from its empirical distribution rather than a parametric family, so the model inherits the observed dispersion. Second, the conditional complier shares of Section 6 are withheld from estimation, so the model must reproduce them through its mechanism without ever being shown them, a restriction the data could have rejected, which is the check reported in Section 7.6. Sections 7.1–7.4 describe the model and its equilibrium, Section 7.5 takes it to the data, and Section 7.7 runs the counterfactuals.

7.1 Environment, states, and timing

Time within an application is discrete, indexed by revision rounds $r = 0, 1, \ldots, R-1$, with $R = 12$, and a common discount factor $\beta = 0.95$.[52] R is a finite-horizon approximation chosen to be well above the observed round distribution, not a statutory limit. [53]

An application is a triple (q, ζ, l). Invention quality q and the applicant's revision-cost type ζ are drawn at filing. Examiner leniency l is drawn at assignment. Quality enters as a rank, $q \in [0.01, 1]$, the percentile $Q_p \in [1, 100]$ of Section 3 rescaled to the unit interval, alongside scope $s \in [0.01, 1]$. Leniency enters in the units of the Section 5 measure itself, the leave-one-out grant rate: l is drawn from a kernel-density estimate $\hat{F}_l$ of that measure, clipped to the grid support $[0.01, 0.99]$, so it is a probability-scale object rather than a rank, and the dispersion-compression counterfactuals operate on that scale.

Importantly, the information structure is asymmetric. Neither party observes q. Each round, the examiner draws a noisy signal and forms a posterior. The applicant also does not observe q. Instead, they hold a subjective Gaussian density over own quality, centered at q up to a bias q_{err} and with dispersion q_{unc}.[54] The applicant knows their own cost type ζ, but observes neither l nor the examiner's signal draws.

[52] β discounts *rounds*, so its calendar interpretation follows from the cadence of the office-action cycle. In the transaction records of the estimation sample, the median gap between successive office actions is 5.9 months across 5.5 million applications. The gap is stable between 5.3 and 6.4 months across the calendar years and between 5.6 and 6.8 across technology centers. $\beta = 0.95$ per round therefore corresponds to an annual discount factor of roughly $0.95^{12/5.9} \approx 0.90$. Measuring instead the span from first office action to disposal divided by the number of rounds, which adds allowance processing to the timeframe, the gap is 10.3 months and the implied annual factor 0.94. The queue before the first office action is excluded throughout, as it precedes the negotiation the model discounts over.

[53] US prosecution has no fixed cap on rounds, and continuations can potentially extend it indefinitely. The distinction matters because the three- and four-round cap counterfactuals of Section 7.7 are measured against R. These counterfactuals are reductions relative to a modeling horizon. In Appendix L.1, I re-solve the model over horizons $R \in \{3, 4, 6, \ldots, 16\}$, which shows the baseline horizon is slack. No simulated application is still active in the last round, and the aggregate grant rate stops adjusting before twelve rounds. The caps therefore bite because three and four rounds are short, not because twelve sits near an edge.

[54] This applicant-side prior is the counterpart of the prior in Section 5.1, and is distinct from the examiner's prior $\mathcal{N}(\bar{q}, \sigma_q^2)$ of Section 7.2. The examiner-prior moments are calibrated ($\bar{q} = 0.5$, $\sigma_q = 1/\sqrt{12}$). The applicant prior combines the estimated dispersion q_{unc} with the fixed normalization $q_{\text{err}} = 0$.

On each rejection, applicants update a joint posterior over (q, l) and decide whether to revise or abandon. Each round proceeds as follows: (i) the applicant files a scope s_r, (ii) the examiner draws their quality signal, updates, and grants the patent or issues a non-final rejection, (iii) on rejection, both sides update and the applicant either revises to s_{r+1} or abandons for an outside value which is normalized to zero. The game ends at grant, abandonment, or horizon R, whichever happens first. A granted application of filed scope s receives granted scope $\hat{s} \leq s$. Granted-scope concessions are permanent, and amendments are bounded at the original filed scope, $s_r \leq s_0$.

7.2 The examiner's problem

Learning. The examiner holds a conjugate Gaussian prior $q \sim \mathcal{N}(\bar{q}, \sigma_q^2)$, calibrated to the mean and standard deviation of quality. Through round r, they have observed $r+1$ signals $\xi_j = q + \eta_j$, $\eta_j \sim \mathcal{N}(0, \sigma_\eta^2)$, i.i.d. across rounds. The examiner's posterior mean is a weighted average of the signal mean $\bar{\xi}_r = (r+1)^{-1} \sum_{j \leq r} \xi_j$ and the prior, $\mathbb{E}_r[q] = f_r\, \bar{\xi}_r + (1 - f_r)\, \bar{q}$, with precision weight

$$f_r \;=\; \frac{(r+1)/\sigma_\eta^2}{1/\sigma_q^2 + (r+1)/\sigma_\eta^2} \qquad \Longrightarrow \qquad \mathbb{E}\big[\mathbb{E}_r[q] \,\big|\, q\big] \;=\; f_r\, q \;+\; (1 - f_r)\, \bar{q}. \tag{14}$$

The right-hand expression is the signal-integrated conditional mean, the posterior an examiner holds on average given true quality q, and it is the object that enters the solution.

The precision weight f_r is increasing in r, so later rounds include a better-informed examiner, but at the calibrated signal noise the weight stays small in every round the data cover.[55] The examiner's quality signal is therefore close to uninformative over the range the data cover, and $\mathbb{E}_r[q]$ stays near the prior mean throughout. This is a maintained feature of the specification and not a finding. As discussed further in Section 7.5, σ_η is not separately identified from the data and is fixed rather than estimated. The consequence for the fitted quality profile is picked up in Section 7.6.

Grant rule. The examiner grants in round r if a latent index clears zero,

$$\underbrace{\alpha_q\, \mathbb{E}[\mathbb{E}_r[q] \mid q]}_{\text{quality premium}} - \underbrace{\Pi(s_r)}_{\text{scope penalty}} + \underbrace{(l - \mu)}_{\text{leniency}} - \underbrace{b_r\, r}_{\text{round drift}} - \underbrace{b_0\, \mathbf{1}[r = 0]}_{\text{first-round barrier}} + \varepsilon_r \;\geq\; 0, \quad \varepsilon_r \sim \mathcal{N}\big(0,\, \sigma_{\text{tot}}^2(r)\big), \tag{15}$$

with $\Pi(s) = \alpha_1\, s + \alpha_2\, s^2 + \alpha_3\, s^3$ the scope penalty of Section 5.1, now given its estimated cubic form.[56] Quality enters at its signal-integrated value from Equation (14).[57] Here, μ is the grant threshold, and $l - \mu$ makes explicit that it varies with the specific examiner's leniency. b_r captures a drift against long-running applications, and b_0 is a reluctance term against allowing in the first round. Together, they reproduce

[55] At $\sigma_\eta = 3.0$ and $\sigma_q = 1/\sqrt{12}$ (see Appendix Table L.1), f_r runs from 0.009 at $r = 0$ to 0.053 at $r = 5$, and $f_r = \frac{1}{2}$ would require on the order of a hundred signals against a round count top-coded at six.

[56] Π is a fitted polynomial and is not monotone on all of $[0, 1]$. As discussed in Appendix L.1, nothing in the design-based results of Sections 5–6 requires that it be.

[57] Pairing the realized $\mathbb{E}_r[q]$ with the total variance would count the dispersion $\sigma_{\text{tot}}^2(r)$ already carries twice.

the empirical rise in grant hazard across revision rounds.[58]

The shock variance $\sigma^2_{\text{tot}}(r)$ combines the variance of the examiner's posterior around its mean, which increases across the observed rounds, with an idiosyncratic case-specific component σ_{grant} that does not vanish as $\alpha_q \to 0$. The case-specific component keeps the grant probability interior when quality carries little weight. Appendix L states the decomposition, its magnitudes, and the accuracy of the independent-shock approximation the simulation uses.

Granted scope. Conditional on a grant, the granted scope is the filed claim scaled by the examiner's *offer ratio*,

$$\hat{s}_r \;=\; \min\Big\{\, s_r,\; s_r \cdot \pi(l, s_r, q) \cdot \gamma_{\text{E}}^{\,r} \,\Big\}, \qquad \pi(l,s,q) \;=\; \gamma_s + \gamma_l\,(l - \tfrac{1}{2}) + \gamma_{s\text{-sl}}\,(s - \tfrac{1}{2}) + \gamma_q\,(q - \tfrac{1}{2}), \tag{16}$$

$\gamma_{\text{E}} \le 1$ a per-round decay of the offer. γ_s is the base offer ratio, evaluated at the grid midpoint, and the three gradients are symmetric deviations from it. γ_l is the deviation in leniency, $\gamma_{s\text{-sl}}$ in filed scope, which decouples the filed-to-granted slope from the level γ_s pins, and γ_q in quality. The quality gradient is a reduced-form device, the first of three in the model (the filing loadings β_s and β_q of Equation (19) share its logic). The offer conditions on q itself rather than on the examiner's round-r posterior: γ_q loads the realized concession schedule across quality directly, without modeling the belief path behind it. The estimate of γ_q should be read as that schedule's slope. This device is not a shortcut past an estimable alternative, and the model-based counterfactuals do not depend on it either. Since only filed scope and granted scope – for granted patents – are observed, but no in-between points, an offer conditioned on the examiner's evolving posterior differs from the loaded schedule in exactly one testable implication, the grant-round profile of the concession's quality gradient. I show in Appendix L.3 that the observed profile rejects such a belief path at every signal noise and offer decay, and that the constant form does not match it either: on endpoint data, the pooled slope is the only object either form identifies, γ_q carries it, and the round profile is a stated limitation of both. Since γ_q enters neither the grant rule nor the applicant's problem, the counterfactuals of Section 7.7 cannot respond through the concession's quality gradient and do not depend on its value as I show in more detail in Appendix L.3.[59]

On the estimated support, the offer does not censor scope. $\pi(\cdot)\,\gamma_{\text{E}}^{\,r} < 1$ holds throughout, so granted scope is $s_r\,\pi(\cdot)\,\gamma_{\text{E}}^{\,r}$ at every state, and mean granted scope is a clean multiple of mean filed scope. This linearity lets the level and the filed-minus-granted wedge be matched jointly without censoring bias, and it is the property the repricing argument of Appendix L.3 inverts. The offer applies to granted scope only: the applicant's value function below is computed against the full filed claim, with the level of the concession absorbed by the estimated multiplier ν, which I expand upon in Appendix L.

[58] Quality enters the grant rule only through the posterior mean, so learning could steepen the quality profile around the prior mean but never shift it uniformly. Any common trend in generosity has to come from b_r and b_0, and at the calibrated precision weights, the steepening via learning is invisible in the fitted profiles. Appendix L derives the difference between learning and the drift parameters.

[59] γ_q is a gradient of the offer ratio, exactly parallel to γ_l, and appears nowhere in the grant index in Equation (15). It is estimated, and small: the fitted value reproduces the mild decline of residualized granted scope across quality that the moment basis of Section 7.5 records and Appendix L.1 expands on, a slope an order of magnitude too small to be visible in the unconditional profile of Section 4.

7.3 The applicant's problem

The applicant solves a finite-horizon optimal-stopping problem. In each round, the applicant chooses how broad the filed scope should be, or whether to abandon the application. They do not observe q or l, so every term is taken in expectation against the round-r joint belief w_r over (q, l). Write

$$\bar{P}_r(s) \;=\; \int P_r(q, s, l)\, \mathrm{d}w_r(q, l) \tag{17}$$

for the belief-averaged grant probability implied by Equation (15). The value of entering round r with scope s is

$$V_r(s) \;=\; \bar{P}_r(s)\, \nu\, s \;-\; \kappa\, s^{\chi} \;+\; (1 - \bar{P}_r(s)) \cdot \underbrace{\max\Big\{\, 0, \;\; \beta \max_{s' \le s_0} V_{r+1}(s') \;-\; C(s, r) \,\Big\}}_{\text{continue or abandon}}, \tag{18}$$

where νs is the value of a patent issued at scope s, κs^{χ} the scope-acquisition cost with curvature χ, and $C(s, r) = (c_0 + c_s s + c_r r)\, e^{\zeta}$ the revision costs in round r, scaled by the applicant's cost type e^{ζ} with $\zeta \sim \mathcal{N}(0, \sigma_c^2)$.

The outer $\max\{0, \cdot\}$ is the abandonment decision. An applicant for whom the discounted continuation does not cover the revision cost exits for the outside option, so the model generates endogenous abandonment and a distribution of stopping rounds $T^* = \inf\{r : \text{grant or abandonment}\}$ whose mean is the empirical counterpart of the revision-round outcome $r(q, s, l)$ in Section 5. The inner maximization runs over $s' \le s_0$, imposing the within-prosecution upper bound of Section 7.1. The bound makes broad filing the option-preserving choice for the applicant. Scope conceded can never be recovered, so the initial claim is an opening position, not a gamble. Beliefs enter only through w_r, updated on each rejection as described in Section 7.4.

Two filing regimes. The optimal stopping problem does not reproduce the filed-scope behavior the reduced form isolates, and the reason is the split Section 4 introduced. Applications whose claims derive from a prior document file more narrowly the higher their quality, free-standing applications file more broadly, and the pooled slope the estimator matches is an average of opposite signs that describes neither regime. Appendix Table L.2 reports both on the estimator's own residualised basis. The model has one regime: an application choosing scope from a continuous set, with no parent document and no restriction requirement in its state space. I let filed scope be displaced from the optimum,[60] and read the displacement as what it is, the aggregate of two filing technologies the model carries as one. Filed scope is the optimum s_r^* displaced by

$$\Delta_r \;=\; \big[\, \sigma_s\, \xi_s \;+\; \beta_s\, (l - \tfrac{1}{2}) \;+\; \beta_q\, (q - \tfrac{1}{2}) \,\big]\, \gamma_{\mathrm{A}}^{r}, \tag{19}$$

[60] Solving for an optimum and estimating a parametric departure from it is the structure of Hortaçsu, Luco, Puller, and Zhu (2019), who embed a cognitive-hierarchy model in a structural auction model to price firms' deviations from Nash bidding. There, the departure is an attribute of the bidder. Here it is not, since absorbing applicant fixed effects strengthens the gradient rather than removing it, so the displacement captures a property of applications and not of who files them.

where ξ_s is an i.i.d. filing taste shock scaled by σ_s, β_q is the quality gradient of filed scope, β_s is the loading of filed scope on the examiner's leniency, and $\gamma_A \in [0,1]$ is the round-by-round *concession rate* at which the displacement decays back toward the optimum as prosecution proceeds. This adjusted scope remains on the same grid as the optimum and, after the first round, is capped at the original filing. The displacement therefore inherits the upper bound of Equation (18). The applicant's concession decay γ_A and the examiner's offer decay γ_E produce the narrowing of claims across rounds observed in the data together.

The β_s and β_q drift terms are reduced-form devices in the same sense as the offer's quality gradient γ_q. Read literally, they say the applicant files broader when the drawn examiner is lenient, and the invention is good. However, the applicant observes neither l nor q before choosing scope. Instead of a choice, the drift terms reproduce a correlation in the data. The model compresses both channels into the initial filing, so it places a correlation at $r = 0$, which is created after assignment, and is a limitation of the model.

7.4 Optimal policy and solution

The examiner is not given an objective function and does not optimize. The threshold rule in Equation (15) and the offer in Equation (16) are a *grant technology*, taken as given and disciplined empirically by the moments of Section 7.5. I solve the applicant's finite-horizon decision problem against that technology.

The examiner's beliefs nonetheless evolve, but exogenously to the applicant's strategy. After rejection, the examiner updates through Equation (14) via their private signal, which is drawn independently of the applicant's scope choice, so the applicant forecasts the examiner's posterior as an exogenous stochastic process. The applicant updates a joint posterior over (q, l) that is truncated by the observed rejection and smoothed by a Gaussian kernel of bandwidth σ_{upd}, capturing that revision decisions respond to accumulating bad news about leniency. This decoupling reduces the problem to a single-agent dynamic program.

Proposition 2 (Existence of an optimal policy)**.** *Fix the grant technology* (15)–(16) *and the belief dynamics above. Then for every round r and belief state w_r the applicant's problem* (18) *admits a maximizer, and the value function V_r is bounded and continuous in w_r. The resulting policy maps the applicant's type (q, ζ) into a filed-scope and stopping rule which, together with the examiner's draw l, determines (a) the probability of acceptance, (b) the number of revision rounds, and (c) the granted scope $\hat{s}$ conditional on acceptance. Beliefs are updated by Bayes' rule on the path.*

Proof in Appendix C.[61]

The policy together with the threshold rule delivers a grant probability $P(q, s, l)$ for each type, the same object whose leniency-sensitivity generated the inverse Mills structure of Section 6.3. That probability governs which applications end up on the margin, so the reduced-form composition and the

[61] Existence does not require quasiconcavity. The objective in (18) is continuous on a compact action set, and, in the solved model, on a finite grid. $\partial V/\partial s$ is not monotone, since the value of a grant and the scope conceded both rise in s while the probability of obtaining one falls, so the balance can reverse across the domain. The local optima this creates are resolved by the grid search, and the solved policy is pure.

structural model are statements about the same primitive. Two features of the data bound what the scope block can be asked to do. The intra-prosecution path is never observed, as the data only contains filed scope at entry and granted scope at issue, so the concession decay γ_A and the offer decay γ_E are identified only through their endpoints. Filed scope is set via a continuation relative to a prior document for roughly a third of applications, which the state space does not represent.

7.5 Estimation

Indirect inference. The model has no closed-form likelihood, so I estimate θ by indirect inference (Gouriéroux, Monfort, and Renault, 1993; Smith, 1993) on auxiliary statistics computed identically in the data and in the simulation: binned conditional-mean curves of acceptance, filed scope, granted scope, and revision rounds against quality, filed scope, and leniency, together with dispersion profiles. On the quality and round axes, these are the descriptive objects of Section 4, recomputed on simulated data. The conditional complier shares are deliberately excluded and evaluated as untargeted moments. Stacking the data statistics in m^{data} and their simulated counterparts, computed from N_{sim} applications drawn with the empirical quality and leniency distributions, in $m^{\text{sim}}(\theta)$, the estimator is

$$\hat{\theta} \;=\; \arg\min_{\theta\in\Theta} \; \big[\, m^{\text{data}} - m^{\text{sim}}(\theta) \,\big]' \, W \, \big[\, m^{\text{data}} - m^{\text{sim}}(\theta) \,\big], \tag{20}$$

over the bounds Θ (see Appendix Table L.1). Identification requires that the map $\theta \mapsto m^{\text{sim}}(\theta)$ be locally injective at $\hat{\theta}$, which the sensitivity analysis below checks. W is diagonal rather than efficient. Each profile is standardized by its own cross-bin amplitude, so the criterion targets the shape of the conditional-mean curve, and a small set of across-profile weights keeps shape-carrying moments from being swamped, as I further discuss in Appendix L.2.

Identification. Equation (20) jointly identifies θ, but each parameter is disciplined primarily by a distinct feature of the data. The reduced form measures leniency based on examiner grant rates. The acceptance-by-leniency profile, the moment which is assigned the most weight, therefore calibrates the model's grant index. With respect to leniency, acceptance in the threshold model is a probit setup. The slope of the acceptance-by-leniency profile determines the grant noise σ_{grant}. The more acceptance increases with the draw, the smaller the idiosyncratic noise present in the grant decision. The level of the acceptance-by-leniency profile identifies the threshold μ as the probit intercept. With the scale of the grant index fixed, the acceptance-by-filed-scope profile[62] targets the scope penalty Π, the acceptance price of a marginal unit of claimed scope. The peak of acceptance by filed scope locates the cubic term α_3. The grant hazard across rounds disciplines the drift terms b_r and b_0.

The level of filed scope determines the value-cost pair (ν, κ). β_q is the slope of filed scope in quality and β_s is the slope of filed scope in leniency, while the dispersion of filed scope within a given quality bin determines the taste shock σ_s. The level of the granted-to-filed ratio pins the base offer γ_s, its slope in filed scope the pass-through $\gamma_{s\text{-sl}}$, and the narrowing across rounds the two decays, the examiner's offer decay

[62] I residualize the moment both in the data and in the model by quality and leniency.

γ_E and the applicant's concession rate γ_A, whose separation Appendix L.2 details. Prosecution length identifies the revision costs. The costs govern how far an applicant pursues revisions after a rejection before abandoning, so the level and dispersion of the rounds distribution, and their profiles across quality, separate the cost components. The moment set is least informative about how fast applicants learn. With the signal noise calibrated large, the belief pair loads on the objective's two flattest directions, as the inference below records. Appendix L.2 states the mapping parameter by parameter in Table L.7, gives the block-level reasoning in Table L.8, and checks the assignment against the objective's Jacobian.

Inference. Under standard indirect-inference regularity conditions (Gouriéroux et al., 1993), $\hat{\theta}$ is asymptotically normal around the minimizer of the population criterion in Equation (20). Since the over-identification test rejects (Appendix Table L.4), I read that minimizer as the best-fitting member of the parametric family under the stated W, not a true parameter the data were generated from. The estimand is therefore defined relative to W, so the across-profile weights are part of it, not only its efficiency.

I estimate the variance of the empirical auxiliary statistics by a block bootstrap that resamples examiners, the level at which the instrument varies, and combine it with the binding-function Jacobian in the sandwich of Appendix L.2, which stays valid when the model is an approximation and the weighting is not efficient. Standard errors follow for the twenty-five parameters the sandwich inverts in Appendix Table L.3, twenty of which lie more than two standard errors from zero. The quality loading α_q is reported without one, as the pinned normalization described below. Five parameters sit on a bound, where a symmetric interval would misstate the sampling distribution, and the parameters in the objective's flattest directions carry standard errors several times their estimates.

Calibrated constants. Six quantities are fixed rather than estimated as reported in Table L.1. The signal noise σ_η shapes the objective only through the examiner's posterior, so it moves the criterion only jointly with the loadings that scale the posterior. The scope-cost curvature χ enters only through the cost κs^χ over the scope range prosecution realizes, where a change in curvature is nearly a change in the level κ. The prior moments are the mean and standard deviation of a uniform distribution on the unit interval, so the examiner's prior matches the support of the quality rank by construction.[63] The offer's leniency gradient γ_l is pinned by the round-zero granted-to-filed ratio, as discussed in Appendix L.2,[64] and the belief parameter q_{err} is set to zero because it is collinear with q_{unc}. A seventh quantity, the quality loading α_q, is estimated but reported as a pinned normalization. Since the examiner's posterior mean averages one half at every round, (α_q, μ) move the acceptance index along a single direction, so the pair is identified jointly and no standard error is reported for α_q. I expand on the argument in Appendix L.2.

7.6 Estimates and fit

The parameter estimates, their bounds, and the moment-sensitivity map are collected in Appendix L, together with the fit of all fifteen targeted moment sets in Appendix L.3, Table L.4. I report seven of them

[63] $\bar{q} = 0.5$ and $\sigma_q = 1/\sqrt{12}$.

[64] No auxiliary statistic in Equation (20) is satisfied by calibration alone: the round-zero ratio is not itself a moment, and granted scope by leniency remains a weighted target the estimates must fit.

as Figures L.2 and L.3. The model reproduces the level moments the offer mechanism targets, as well as the sign and shape of the two scope-margin profiles of the reduced-form estimation. Two misfits stand out. The larger by far is the abandonment hazard by round: late in prosecution, the model abandons about half of the applications still live, against under three in ten in the data. It carries 84 percent of the over-identification statistic and almost none of the identifying content, so the rejection above is dominated by a confined failure. The model misses the timing of exit, but this does not reach the parameters the counterfactuals rest on. The same overshoot extends mildly to the mean of the round distribution. The second bears directly on the claims below: the model overstates the slope of granted scope in filed scope, which is the source of the scope-channel caveat I return to in Section 7.7. The model also reproduces the acceptance profile's near-flatness in quality, a reproduction that follows from the calibrated signal noise rather than from the grant rule, as set out below. Granted scope by quality, by contrast, is fit through a single estimated parameter, the offer's quality gradient γ_q, which reproduces the moment's mild decline almost exactly.

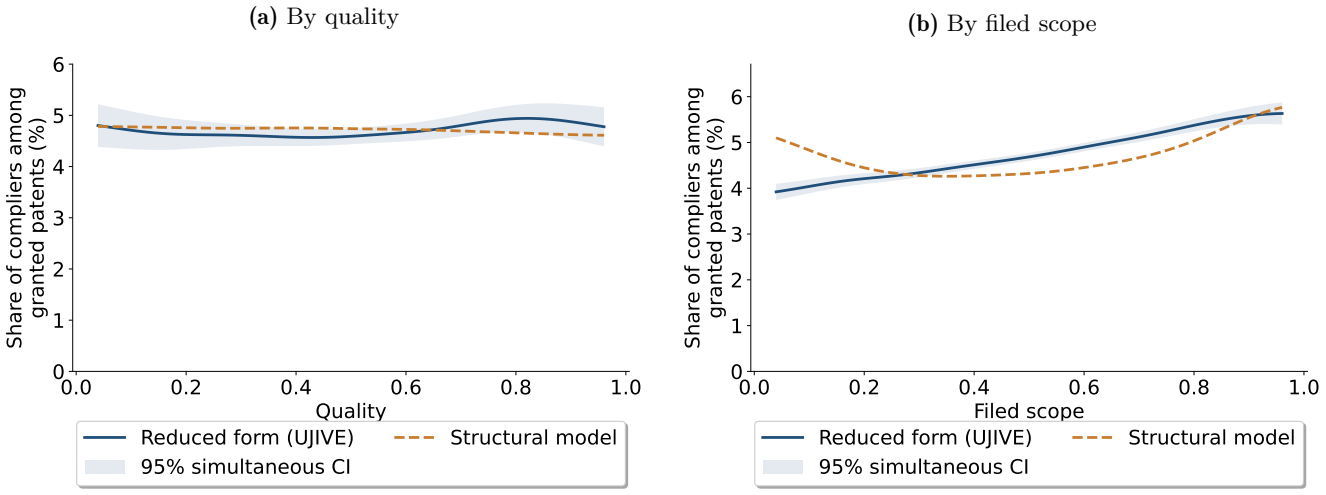
The fit above is targeted, but four available moments the objective never saw are available as checks. The conditional complier shares, held out of estimation for this purpose, carry the most weight and are taken up below. Of the remaining three, two are quality profiles reported in Table L.9, granted scope in unresidualized form and the scope concession,[65] and the third is the joint acceptance surface in quality and leniency, depicted in Appendix Figure L.4. The surface establishes two claims. What the surface tests is the *absence of an interaction*. The grant rule carries no quality-by-leniency term, so if the quality profile had tilted across leniency quintiles in the data, the model would have had no parameter to follow it. It does not tilt in either, which is the separability Proposition 1 requires. The surface does not test the near-flatness of the profile itself. Under the calibrated signal noise the quality-premium term is nearly constant in quality by construction, varying by under a hundredth of σ_{grant} across the whole quality range, so the model was not built to earn the data's mild acceptance-quality gradient through the grant rule. The case for a nearly quality-neutral grant rule rests on the design-based evidence of Section 4, not on this fit.

An untargeted check: the complier share. The share of grants attributable to the examiner draw is not among the estimation targets, though not literally an out-of-sample check. The targeted acceptance profiles constrain the complier share through the threshold structure, so the check is informative about joint shape rather than independent of the fit. Figure 5 plots the model-implied share of compliers among granted patents against the reduced-form UJIVE estimates of Section 6 (Figure 4), along the quality dimension in Panel a and the filed-scope dimension in Panel b.

The model reproduces the near-flatness of the quality dimension, though the calibration partly guarantees it for the reason above, so the informative untargeted test is the scope dimension. There, the model reproduces the direction but not the shape. Model levels below are stated under the figure's display convention, which converts the simulated shares to the reduced form's pass-through and dose conventions (see the note to Figure 5). In quality, the model's share of granted patents due to the examiner draw

[65] The residualized profile is itself a weighted target of the objective (Section 7.5).

Figure 5: *Share of compliers: model versus reduced form*



Notes: Solid line, the reduced-form share of compliers among granted patents from the examiner-dummy UJIVE modified-outcome estimator (Section 6, Figure 4), with its 95% simultaneous band. Dashed line, the model-implied share at the converged estimate. The dashed line is produced by applying the same estimator to the simulated panel: the examiner's leniency draw plays the role of the examiner identifier, the leave-one-out instrument and the kernel-smoothed modified-outcome ratio are computed exactly as in Section 5.5, and the bandwidth and evaluation grid are those of the reduced form. Both curves therefore share the same estimand–the share of grants in a percentile bin that would not have been granted by an examiner of mean ("expected") leniency. Panel (a) conditions on quality, Panel (b) on the within-sample percentile of filed scope. The reduced-form curve is anchored on the fixed-effects OLS first-stage slope ($\hat{\psi} = 0.67$), the conservative of the two conventions discussed in Appendix E. The model has no leniency measurement error, so its pass-through is one by construction, and its simulated leniency dose is 0.069 against the estimation sample's 0.050; the dashed line is therefore converted to the reduced form's conventions by the combined ratio of the two, 2.05, computed from the run's own diagnostics and applied exactly. The complier share is not a targeted moment.

decreases from 4.8% to 4.6% across the distribution, against a reduced-form profile that moves within half a percentage point of its 4.7% mean. In filed scope, broader filings are more likely to owe their grant to the examiner draw in model and data alike, but they reach the higher endpoints of the curves differently. The reduced form rises throughout, 4.0% to 5.6%. The model starts elevated at 5.1%, falls to an interior minimum of 4.3% near the thirty-sixth percentile, and rises from there to 5.8%. Over the upper two-thirds of the scope distribution, the model reproduces the gradient that carries the paper's central mechanism without being shown it, but the non-monotonicity over the lower third is a property of the model that the data exclude: over that range the model curve leaves the reduced form's simultaneous band at most grid points, by up to a percentage point.

7.7 Counterfactuals

With the model estimated and its untargeted fit verified, I turn to the counterfactuals. Reducing leniency dispersion stands for a class of real reforms. The USPTO could adjust examiner standards through the examination time allotted per application (Frakes and Wasserman, 2017), through the count system that credits finished cases, and through additional training or second review of allowances before they issue. The equalization counterfactual tests the upper bound for the entire class, and the examination-

time experiment prices its best-documented member. The judge-benchmark counterfactual asks how the patent office would work if it were possible to bring dispersion in human verdicts down to the level of another public institution.

The results below are not all equally conditional, and the order in which I present them reflects that. The least conditional result is the budget-neutral reallocation share. When I hold the number of grants and the total granted scope fixed at their status quo values, a consistent examiner reassigns 11.7% of the granted-scope budget, almost all of it within quality cells. In other words, the reallocation is horizontal. That number needs no price for scope, because the social cost of scope cancels when aggregate scope does not change.[66] The budget-neutral reallocation is a statement about who receives scope rather than the value of scope.

Everything else in this subsection is conditional on external estimates. First, a cardinalization of scope, which is a rank within-cell percentile, so that changes in the level of granted scope, $\Delta\hat{S}$, and the break-evens computed from them are normalization-dependent and should be read as magnitudes under this normalization rather than as quantities of scope. Second, a price for scope, δ. I report the break-even δ^* at which each reform's verdict changes sign and then locate it against an external benchmark. Appendix N collects the limitations.

Portfolio value and the cost of scope. I evaluate each counterfactual against the value of the granted patent portfolio. Each granted application of quality q and granted scope $\hat{s}$ is priced at the mean market value $v(q, \hat{s})$ of patents in its $(q, \hat{s})$ cell, estimated from the stock-market-based value measure of Kogan et al. (2017). Because claim scope imposes a social cost that private value does not internalize, since broader rights foreclose more follow-on paths (Galasso and Schankerman, 2014), I net a social cost of scope from portfolio value:

$$W(\delta) \;=\; \sum_i D_i \left[\, v(q_i, \hat{s}_i) \;-\; \delta\, \bar{v}\, \hat{s}_i \,\right], \tag{21}$$

where D_i is the grant indicator, $\bar{v}$ the mean granted value, and δ is the social cost of one unit of granted scope, expressed as a fraction of the average patent's private value. Two features of the data discipline this object. First, private value is essentially flat in granted scope, with the correlation between $\hat{s}$ and market value being approximately zero, so the value channel operates through the quality composition of grants rather than through their scope. Excess scope carries a social cost but no offsetting private value, and is treated as deadweight throughout, a valuation assumption Appendix N returns to. Second, the within-quality value-scope slope in the price map is contaminated by selection, because only broad applications that survive prosecution are ever priced. I therefore report every magnitude under two pricings, the full $(q, \hat{s})$ map and a quality-only pricing that removes the scope slope entirely. The second pricing imposes a zero rather than testing one, and I state in Appendix N what that leaves unidentified.

[66] It is not free of the cardinalization, since a share of the scope budget sums ranks. Appendix I measures that exposure directly: the share moves between 10.6% and 12.3% across the alternative scales. Appendix N returns to it.

Counterfactual experiments. Section 6 counts 4.7% of grants as owed to the draw, and the natural reading of equalization is a reduction of grants. Remove the dispersion, and the lottery grants disappear, beginning with the marginal broad grants of the most lenient examiners. The simulated reform shows the opposite result, as compressing leniency dispersion raises the grant rate. Pulling the strict examiners' under-granted applications up toward the by-the-book benchmark outweighs trimming the lenient examiners' over-granted ones, so both aggregate grants and aggregate scope increase ($\Delta\hat{S} > 0$). The source of the asymmetry is the curvature of the grant rule rather than the shape of the leniency distribution.[67] The applicants' filing and concession responses, the two margins the model prices, then determine which applications carry the additional scope.

I report the change in the portfolio value, $W(0)$, before including the scope cost for three experiments. Table 2 collects each with its break-even. Since value is nearly flat in granted scope, the two pricings give the same counterfactual magnitudes, so the table shows a single set of numbers rather than one per pricing. Capping the maximum number of revision rounds at three reduces the portfolio value by 5.47%, because it denies grants that would have happened only in later rounds. Equalizing examiner leniency at the sample median raises portfolio value by 7.16%, but that figure is not a pure dispersion effect.[68] The sample's median examiner is 3.1 percentage points more lenient than the mean, so assigning every examiner the median combines a change in levels with a change in dispersion. As a comparison, I run the level-neutral experiment, which collapses the distribution to its own mean. This counterfactual raises portfolio value by 3.36%. The residual 3.81 percentage points are the level shift.

Compressing dispersion to the level observed among judges and prosecutors in the criminal-IV literature (Dobbie et al., 2018; Agan et al., 2023) while holding the mean fixed raises the portfolio value by 2.9 to 3.1%, between 86 and 92% of the 3.36% dispersion change with complete consistency. The portfolio value is moved by the cross-sectional dispersion of leniency itself, as the median-leniency experiment's larger figure adds a level shift in where the grant bar sits, not a consistency effect, and the examination-time channel isolated next is at most a twentieth of the dispersion effect.

A second externally anchored compression isolates the part of that dispersion the existing literature has identified. Frakes and Wasserman (2017) document a gradient in the grant rates by examiner grade driven by examination hours. Examiners on a higher grade receive less allotted time and issue fewer time-intensive rejections, so heterogeneity across grades generates part of the leniency dispersion, which would be removed by equalizing examiner time. Shrinking the dispersion by exactly that component while holding the mean fixed raises the portfolio value by 0.18%. The implementable slice of the consistency gain is therefore about a twentieth of what full equalization delivers, so examination time is a real component of examiner inconsistency but small. Since private value is flat in scope and each of these experiments changes the number of grants, $\Delta W(0)$ tracks the change in grant count almost one-for-one, so the allocative

[67] Every cell of the acceptance surface sits above the Φ function's inflection at a grant probability of one half (the observed range is 56 to 79%), so the grant probability is concave over the leniency values the cells realize, and a mean-preserving compression raises the average grant rate, Jensen's inequality supplying the intuition and the simulated counterfactual the sign, since individual draws below the inflection are not ruled out by the cell range alone.

[68] Counterfactual magnitudes carry Monte-Carlo noise, because the simulated leniency population is redrawn between runs, and the measured drift at identical parameters is about 0.01 percentage points. Therefore, second decimals are run-specific, but no conclusion here depends on a second decimal.

Table 2: *Counterfactual policy experiments and break-even social cost of scope*

Reform	$\Delta W(0)$	$\Delta\hat{S}$	Δ grants	δ^*
Budget-neutral consistency rule	+0.09	+0.00	+0.09	*δ-free*
Median-leniency equalisation	+7.16	+7.00	+7.16	2.40
Judge-dispersion compression	+3.03	+2.23	+3.03	3.19
Full equalisation (all examiners at the pool mean)	+3.36	+2.48	+3.35	3.17
F&W time-dispersion compression	+0.18	+0.12	+0.18	3.40
Cap revision rounds at 3	−5.47	−5.86	−5.47	2.19
Cap revision rounds at 4	−1.05	−1.10	−1.05	2.24

Notes: Each reform's effect on portfolio value before netting the scope cost ($\Delta W(0)$), on aggregate granted scope ($\Delta\hat{S}$) and on grant count, all as percentages of the status-quo portfolio, with the break-even social cost of scope δ^* at which $\Delta W(\delta) = \Delta W(0) - \delta\bar{v}\Delta\hat{S}$ crosses zero. δ^* prices marginal grants at the mean patent value. The budget-neutral rule fixes aggregate scope by construction, so its verdict is δ-free; its grant count is matched by root-finding the common threshold to a 0.2% tolerance, so that row's residual entries are the solver's stated tolerance rather than an allocative gain. Because private value is close to flat in granted scope and rises only six percent from the median quality bin to the top decile of the measure, the price a granted patent receives is nearly constant across the plane the model moves: $\Delta W(0)$ accordingly tracks Δ grants almost exactly, and δ^* is to three decimal places the ratio of the two proportional changes divided by mean granted scope. Values are essentially identical under the full $(q, \hat{s})$ and quality-only pricings because private value is flat in scope; the $(q, \hat{s})$ pricing is shown ($\bar{v} = 11.07$). Estimated model, $N = 7{,}000{,}000$ simulated applications.

content of these experiments lies in the scope-cost netting and the budget-neutral reallocation developed next.

The social cost of scope and the break-even threshold. Whether a value-changing reform is welfare-improving depends on the social cost of scope δ, which I do not observe.[69] Rather than impose a value, I report the break-even δ^* at which a reform's net welfare change $\Delta W(\delta) = \Delta W(0) - \delta\,\bar{v}\,\Delta\hat{S}$ crosses zero, where $\Delta\hat{S}$ is the induced change in the portfolio's aggregate granted scope.

The scope-cost term splits the reforms by the sign of $\Delta\hat{S}$. The dispersion reforms expand aggregate scope, so a higher cost term reduces the welfare. The dispersion reforms increase welfare with social costs below their break-even values, which run from 2.4 to 3.4 across the four scenarios in Table 2. The round cap contracts both value and scope, so a higher cost term increases its welfare: denying late-round grants pays off above its break-even of 2.2.

Without a value for δ, these numbers settle a ranking, not a sign. Capping prosecution length leaves the dispersion of leniency untouched, so it is the least promising lever: it destroys value without improving the allocation. The ranking is an accounting of the value side alone – $W(\delta)$ counts the cap's denied grants but not the prosecution costs a shorter path would save on both sides of the exchange, so the criterion itself works against the cap. Nor is it unconditional in δ: with $\Delta\hat{S}$ of opposite signs, the cap's net welfare

[69] "Welfare" here means $W(\delta)$ of (21), the private market value of the granted portfolio net of a social cost of scope. Only granted applications enter, so applicant surplus and prosecution costs are not included, and the model carries no general-equilibrium feedback from the stock of granted rights back to innovation. Aggregate magnitudes in this subsection are therefore statements about allocation under a maintained externality rather than social welfare accounts.

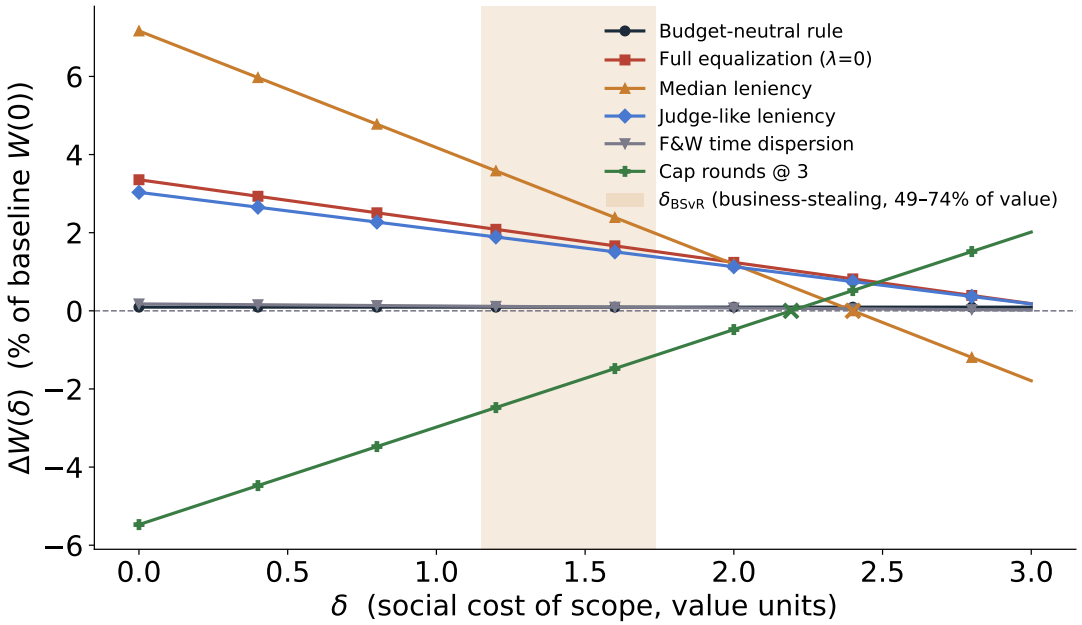


Figure 6: *Break-even social cost of scope and the business-stealing anchor*

Notes: Net welfare change $\Delta W(\delta) = \Delta W(0) - \delta\bar{v}\Delta\hat{S}$ relative to the status quo, as a percentage of baseline portfolio value, plotted against the social cost of one unit of granted scope δ (in mean-patent-value units), at $\omega = 1$ under the $(q, \hat{s})$ pricing. Markers ($\times$) locate each reform's break-even δ^*. The shaded band is the Bloom, Schankerman, and Van Reenen (2013) business-stealing anchor. Its left edge, $\delta_{\text{BSvR}} = 1.12$, is the anchor value implied by the 49% rivalry share of the marginal private return to R&D under their baseline closeness measure; the band extends to 1.69, the value implied by the largest rivalry share their alternative estimates support (74%, instrumented). Both are the limiting case $\tau = 1$, in which the entire rivalry wedge is counted as deadweight rather than transfer, so the band bounds the *mapped rivalry component* of the social cost of scope; it does not bound blocking, litigation, or follow-on costs falling outside that wedge. Every equalization reform breaks even beyond the band, at $\tau^* = \delta^*/\delta_{\text{BSvR}} \geq 1.42$ even at the band's adverse edge, so each remains welfare-improving for all $\tau \in [0, 1]$ under this pricing.

rises in δ while equalization's falls, the two cross at $\delta \approx 2.3$, and above that the cap is preferred, though only because it contracts the portfolio. Between 2.2 and 2.4, every reform improves on the status quo.

An external anchor for the social cost of scope. The break-evens above locate the welfare verdict on a δ axis but cannot speak to the true social cost of scope. Bloom et al. (2013) decompose the marginal private return to R&D and find that 49% of it is business stealing: of each dollar an innovation returns to its owner, 49 cents is market value taken from product-market rivals rather than created. I import that share at the level of the average patent. A grant privately worth $\bar{v}$ then imposes a mirrored cost of $0.49\,\bar{v}$ on the owner's rivals, and the instrument that inflicts it is the grant's exclusive scope, so I spread the cost over the average grant's $\bar{\hat{s}}$ units of scope: the cost of one unit of scope is $\delta_{\text{BSvR}} = 0.49/\bar{\hat{s}} = 1.12$, in mean-patent-value units, the denomination δ already carries in Equation (21) – about 1% of the average patent's value per percentile point of scope. In Figure 6, I show that the deadweight share at which a reform's verdict flips, $\tau^* = \delta^*/\delta_{\text{BSvR}}$, exceeds one for every reform under the mean pricing: 2.14 for median equalization at the anchor, and 1.42 in the most adverse rivalry share their estimates support. Equalizing examiner leniency therefore passes a cost-benefit test at every admissible deadweight share,

including the limiting case $\tau = 1$ in which business stealing is counted as pure social waste and no transfer at all. That pass covers rivalry alone: pushing τ to one exhausts the rivalry wedge but says nothing about the other costs of scope – litigation, blocking of follow-on work, licensing frictions – and the break-evens are the numbers to price those against. Appendix M.1 derives the anchor, states what the import does and does not assume, and reports the wider range their alternative estimates support.

Budget-neutral reallocation: the horizontal-equity view. The experiments above change the size of the granted portfolio, so their value content is entangled with a level effect this partial-equilibrium model does not price. To isolate the allocative content, I hold the total number of grants and the total granted scope fixed at their status-quo values and let a single consistent examiner (leniency set to the mean) redistribute them under the by-the-book rule.

The two budget constraints require two instruments. I solve for the grant threshold μ^* at which the equalized corps issues the status-quo number of grants, and then, holding μ^* fixed, rescale the base offer ratio γ_s so that the corps issues the status-quo total granted scope. The second step does not disturb the first: Equation 16 clarifies that the offer ratio applies only after a grant has been decided, so γ_s cannot move the acceptance set, and the calibration is triangular for that reason.[70]

This budget-neutral reform is free of the social cost of scope, because aggregate scope is fixed by construction, so the scope-cost term drops out entirely, and it leaves portfolio value essentially unchanged. Instead, the allocation moves and 11.7% of the granted-scope budget is reassigned. Of those, 11.5 percentage points are churn within quality cells, scope moving to a different applicant at the same invention quality, and only 0.19 points are a systematic shift across quality, while 14.3% of applications change acceptance status. That neutrality is an outcome, not an accounting identity: the equalized corps re-ranks every application under a common threshold and offer rather than undoing each draw one by one, so the by-the-book rule was free to concentrate its cuts in either tail of the quality distribution. It does not, because the acceptance margin reads filed scope, not quality. Examiner inconsistency thus operates almost entirely as a lottery over which applications of a given quality receive scope rather than as a distortion of the quality gradient of grants.

Who wins and who loses. The direction of the incidence follows from the reform itself: granted scope moves from applicants who drew lenient examiners to those who drew strict ones. Appendix Figure M.1 reports the magnitudes. The top leniency decile loses 2.9 percentage points of the scope budget, the bottom decile gains 5.8, and the profile across quality deciles is flat, with at most 0.08 points reassigned per decile. The reform therefore corrects a disadvantage that is uncorrelated with the invention itself and thus invisible to any quality-based instrument. Since the bundle also resets the common threshold and offer, the incidence is a net transfer toward the strictly examined rather than an exact undoing of each draw.

[70] Step (i) is solved numerically to a tolerance of 0.2% on the grant count; step (ii) is solved analytically on the exact piecewise offer map and binds to machine precision. The residual in the grant count of Table 2 is that numerical tolerance, not a slack constraint, which is why the scope column reads −0.00 and the grant column does not. Reported portfolio value moves by the same amount: with private value flat in granted scope, value per patent is constant across scenarios, so the two columns carry the same information in every row.

Robustness of the misallocation magnitudes. The portfolio-value level ΔW depends on calibrated inputs I can perturb without re-estimating the model, because the leniency-scope offer gradient γ_l enters only the post-acceptance pass-through from filed to granted scope and so re-prices analytically. In Appendix M.2, I show that consistency dominates prosecution length at every value in the calibration band, so the policy ranking does not depend on the gradient's calibrated value.

The composition weight ω, which prices marginal grants relative to the average granted patent, is set to one throughout, and that is an estimate rather than a convention. Modeling the value equation jointly with the grant-selection index in the manner of Avivi (2025) and estimating the correlation between them on the initial-allowance margin gives a value-selection correlation $\rho_v = 0.072$ (Appendix N; the subscript keeps it apart from the representation ratio $\rho(x)$). The estimate puts the implied ω inside $[0.94, 1.02]$ across every reform, so the composition channel is worth at most two percent and its direction is not identified. Repricing inside that band moves median equalization between +6.4% and +7.4% against +7.16% at $\omega = 1$, the value being exactly linear in the weight. Appendix M reports both sweeps in full and states what they do and do not measure. The δ-free reallocation share is untouched by either, aggregate scope being fixed by construction.

These magnitudes carry three caveats from the estimated model. The first caveat concerns the grant condition, which is somewhat too steep in scope at strict examiners, so the model likely overstates how sharply they suppress broad claims and hence the aggregate scope that equalization restores. Second, the $\Delta\hat{S} > 0$ magnitudes are best read as an upper bound on the scope channel. And the grant rule carries no quality-leniency interaction, so the consistency counterfactual abstracts from any quality screen specific to strict examiners.

The fixed filing population is probed rather than modeled. In Appendix M.3, I show that imposing an entry elasticity and removing the entry-marginal filers symmetrically from reform and baseline barely moves the level of ΔW, and the trimmed filers are representative in value rather than the marginal-value mass a fixed-population critique presumes; the exercise is a trimming sensitivity, not a bound on reform-induced entry.

Four limitations qualify the results above. First, scope is measured as a within-cell percentile rank. Results that depend only on the ordering are invariant to this normalization, but quantities that sum scope across applications are not. Second, private value is observed only for granted patents, so the value of marginal scope is not identified and is set by assumption. Third, the estimated market-value price is nearly flat in quality and granted scope, so the counterfactuals measure changes in the number and breadth of granted rights, not changes in their value. Fourth, individual-level monotonicity, the compositional interpretation of the filed-scope gradient, and the fixed applicant population are assumptions the data cannot test. The budget-neutral reallocation share depends least on these limitations, which is why the policy analysis reports it first. Appendix N discusses each limitation in detail.

8 Conclusion

This paper shows that the patent lottery depends on how much scope an applicant requests, not on the quality of their invention. Across 5.4 million applications, roughly one granted patent in twenty-one owes its existence to the draw of a lenient examiner. That share rises steeply with the scope of the application's claims while staying nearly flat in the measured quality of the invention. The applications the lottery decides are broad filings, and their inventions span the same quality range as those the office grants anyway. Their breadth was allowed because of who examined the file rather than because of the invention beneath it. These are bad patents in the second of the literature's two senses (U.S. Federal Trade Commission, 2003): not weak inventions, but claims broader than the invention supports, with re-examination as the test.

A single selection margin explains the whole composition. I model examination as a threshold decision. An application is granted when the scope it claims, judged against the examiner's assessment of its quality, clears the assigned examiner's personal threshold. Among granted patents, the share granted due to the examiner draw is highest where a grant was least likely. Scope is the applicant's choice. Broader claims lower the odds of approval while raising the value of success, so the least likely grants are, by the applicant's own decision, the most ambitious ones. Quality affects the acceptance chance positively, as higher quality improves the examiner's assessment directly, but also negatively, as stronger inventions also file broader claims. This is why the examiner lottery can affect an invention with any quality. The acceptance threshold is a modeling choice, but I find no evidence that the data reject it. The threshold ties all three findings to a single number per application, the grant probability: it determines the increasing share of patents caused by the lottery across scope, the flatness of the share across quality, and the nearly uniform exposure of applications to the draw (the flip rate). The threshold also implies that a lenient examiner grants more at every quality level, rather than judging strong applications differently, a pattern which the data reproduce on a surface the estimation never targeted.

For patent policy, the important margin is examiner consistency, not prosecution length. Any increase in consistency adjusts scope almost entirely within quality cells, reallocating 11.7% of the scope budget among inventions of comparable quality. Higher consistency increases the number of grants, and with it the value of the granted patent portfolio; the additional granted scope it also creates is what the break-evens price. The value increase is more than twice as large as the social costs implied by the external rivalry benchmark at its baseline value. Capping prosecution instead destroys the value of late-completing grants without improving the allocation. For the leniency literature, the composition means existing lottery-based estimates price patent protection for unusually ambitious applications.

The findings add an empirical point to the general concern that patent breadth blunts the cumulative innovation on which long-run growth depends (Galasso and Schankerman, 2014): examiner inconsistency routes breadth by the size of the request rather than by the invention, for roughly one granted patent in twenty-one. I do not convert the mechanism into an aggregate productivity figure, because the value of the marginal broad grant is not identified here. What the paper shows is the allocation rule itself: the patent office's inconsistency hands out exclusion rights, and it hands them out by the requested scope.

Bibliography

Acemoglu, D. and U. Akcigit (2012, 02). Intellectual property rights policy, competition and innovation. *Journal of the European Economic Association 10*(1), 1–42.

Agan, A., J. L. Doleac, and A. Harvey (2023, 01). Misdemeanor Prosecution*. *The Quarterly Journal of Economics 138*(3), 1453–1505.

Andrews, I., M. Gentzkow, and J. M. Shapiro (2017). Measuring the sensitivity of parameter estimates to estimation moments. *The Quarterly Journal of Economics 132*(4), 1553–1592.

Angrist, J. D., G. W. Imbens, and A. B. Krueger (1999). Jackknife instrumental variables estimation. *Journal of Applied Econometrics 14*(1), 57–67.

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.

Avivi, H. (2025). Are patent examiners gender neutral? Working Paper, University College London.

Bloom, N., M. Schankerman, and J. Van Reenen (2013). Identifying technology spillovers and product market rivalry. *Econometrica 81*(4), 1347–1393.

Boldrin, M. and D. K. Levine (2013, February). The case against patents. *Journal of Economic Perspectives 27*(1), 3–22.

Borusyak, K., P. Hull, and X. Jaravel (2021, 06). Quasi-Experimental Shift-Share Research Designs. *The Review of Economic Studies 89*(1), 181–213.

Borusyak, K., P. Hull, and X. Jaravel (2023, June). Design-based identification with formula instruments: A review. Working Paper 31393, National Bureau of Economic Research.

Bound, J., D. A. Jaeger, and R. M. Baker (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association 90*(430), 443–450.

Bryan, K. A. and H. L. Williams (2021). Chapter 13 - innovation: market failures and public policies. In K. Ho, A. Hortaçsu, and A. Lizzeri (Eds.), *Handbook of Industrial Organization, Volume 5*, Volume 5 of *Handbook of Industrial Organization*, pp. 281–388. Elsevier.

Carley, M., D. Hegde, and A. Marco (2015). What is the probability of receiving a u.s. patent?

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and Y. Feng (2024). Binscatter regressions. *Stata Journal*. Forthcoming.

Cockburn, I. M., S. Kortum, and S. Stern (2002, "June"). "are all patent examiners equal? the impact of examiner characteristics". "Working Paper" "8980", "National Bureau of Economic Research".

Dahl, G. B., A. R. Kostøl, and M. Mogstad (2014, 08). Family Welfare Cultures *. *The Quarterly Journal of Economics 129*(4), 1711–1752.

de Rassenfosse, G., W. E. Griffiths, A. B. Jaffe, and E. Webster (2020, 12). Low-Quality Patents in the Eye of the Beholder: Evidence from Multiple Examiners. *The Journal of Law, Economics, and Organization 37*(3), 607–636.

deGrazia, C., J. Frumkin, and N. Pairolero (2018, February). Embracing invention similarity for the measurement of vertically overlapping claims. *Economics of Innovation and New Technology 29*(2), 113–146. Available at SSRN: https://ssrn.com/abstract=3166042.

Dobbie, W., J. Goldin, and C. S. Yang (2018, February). The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review 108*(2), 201–40.

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

Feng, J. and X. Jaravel (2020, January). Crafting intellectual property rights: Implications for patent assertion entities, litigation, and innovation. *American Economic Journal: Applied Economics 12*(1), 140–81.

Frakes, M. D. and M. F. Wasserman (2017, 07). Is the Time Allocated to Review Patent Applications Inducing Examiners to Grant Invalid Patents? Evidence from Microlevel Application Data. *The Review of Economics and Statistics 99*(3), 550–563.

Frakes, M. D. and M. F. Wasserman (2020). Procrastination at the patent office? *Journal of Public Economics 183*, 104140.

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.

Frakes, M. D. and M. F. Wasserman (2024, January). Deadlines versus continuous incentives: Evidence from the patent office. Working Paper 32066, National Bureau of Economic Research.

Frandsen, B., L. Lefgren, and E. Leslie (2023, January). Judging judge fixed effects. *American Economic Review 113*(1), 253–77.

Frankel, A. and N. Kartik (2019). Muddled information. *Journal of Political Economy 127*(4), 1739–1776.

Galasso, A. and M. Schankerman (2014, 11). Patents and Cumulative Innovation: Causal Evidence from the Courts *. *The Quarterly Journal of Economics 130*(1), 317–369.

Gao, T., X. Yao, and D. Chen (2022). Simcse: Simple contrastive learning of sentence embeddings.

Gaulé, 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.

Gouriéroux, C., A. Monfort, and E. Renault (1993). Indirect inference. *Journal of Applied Econometrics 8*(S1), S85–S118.

Hansen, N. (2006). The CMA evolution strategy: a comparing review. In *Towards a New Evolutionary Computation: Advances in the Estimation of Distribution Algorithms*, Volume 192 of *Studies in Fuzziness and Soft Computing*, pp. 75–102. Springer.

Heckman, J. J. and E. J. Vytlacil (2007). Chapter 71 econometric evaluation of social programs, part ii: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. Volume 6 of *Handbook of Econometrics*, pp. 4875–5143. Elsevier.

Higham, K., G. de Rassenfosse, and A. B. Jaffe (2021). Patent quality: Towards a systematic framework for analysis and measurement. *Research Policy 50*(4), 104215.

Hortaçsu, A., F. Luco, S. L. Puller, and D. Zhu (2019). Does strategic ability affect efficiency? Evidence from electricity markets. *American Economic Review 109*(12), 4302–4342.

Kelly, B., D. Papanikolaou, A. Seru, and M. Taddy (2021, September). Measuring technological innovation over the long run. *American Economic Review: Insights 3*(3), 303–20.

Kline, P., R. Saggio, and M. Sølvsten (2020). Leave-out estimation of variance components. *Econometrica 88*(5), 1859–1898.

Kline, P. and C. R. Walters (2019). On Heckits, LATE, and numerical equivalence. *Econometrica 87*(2), 677–696.

Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman (2017, 03). Technological Innovation, Resource Allocation, and Growth*. *The Quarterly Journal of Economics 132*(2), 665–712.

Kolesár, M. (2013). Estimation in an instrumental variables model with treatment effect heterogeneity. Working paper, Cowles Foundation, Yale University, Version 1.3.1, November 5, 2013.

Kuhn, J. M. and N. C. Thompson (2019). How to measure and draw causal inferences with patent scope. *International Journal of the Economics of Business 26*(1), 5–38.

Lemley, M. A. and B. Sampat (2012, 08). Examiner Characteristics and Patent Office Outcomes. *The Review of Economics and Statistics 94*(3), 817–827.

Lemley, M. A. and C. Shapiro (2005). Probabilistic patents. *Journal of Economic Perspectives 19*(2), 75–98.

Lu, Q., A. Myers, and S. Beliveau (2017, November). Uspto patent prosecution research data: Unlocking office action traits. *USPTO Economic Working Paper No. 10*. Available at SSRN: https://ssrn.com/abstract=3024621 or http://dx.doi.org/10.2139/ssrn.3024621.

Marco, A. C., J. D. Sarnoff, and C. A. deGrazia (2019). Patent claims and patent scope. *Research Policy 48*(9), 103790.

Matcham, W. and M. Schankerman (2026). Screening property rights for innovation. *Econometrica*. Forthcoming (accepted May 2026). Available at SSRN: https://ssrn.com/abstract=4519999 or http://dx.doi.org/10.2139/ssrn.4519999.

Melero, E., N. Palomeras, and D. Wehrheim (2020). The effect of patent protection on inventor mobility. *Management Science 66*(12), 5485–5504.

Ni, J., G. H. Ábrego, N. Constant, J. Ma, K. B. Hall, D. Cer, and Y. Yang (2021). Sentence-t5: Scalable sentence encoders from pre-trained text-to-text models.

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

Pennington, J., R. Socher, and C. Manning (2014, October). GloVe: Global vectors for word representation. In A. Moschitti, B. Pang, and W. Daelemans (Eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, pp. 1532–1543. Association for Computational Linguistics.

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

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica 56*(4), 931–954.

Romano, J. P. and M. Wolf (2005). Stepwise multiple testing as formalized data snooping. *Econometrica 73*(4), 1237–1282.

Romano, J. P. and M. Wolf (2016). Efficient computation of adjusted p-values for resampling-based stepdown multiple testing. *Statistics & Probability Letters 113*, 38–40.

Sampat, B. and H. L. Williams (2019, January). How do patents affect follow-on innovation? evidence from the human genome. *American Economic Review 109*(1), 203–36.

Schankerman, M. and F. Schuett (2021, 10). Patent Screening, Innovation, and Welfare. *The Review of Economic Studies 89*(4), 2101–2148.

Scotchmer, S. (1991, March). Standing on the shoulders of giants: Cumulative research and the patent law. *Journal of Economic Perspectives 5*(1), 29–41.

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

Smith, A. A. (1993). Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics 8*(S1), S63–S84.

Trajtenberg, M. (1990). A penny for your quotes: Patent citations and the value of innovations. *The RAND Journal of Economics 21*(1), 172–187.

U.S. Federal Trade Commission (2003, October). To promote innovation: The proper balance of competition and patent law and policy. Report, Federal Trade Commission, Washington, DC.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin (2017). Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, Volume 30. Curran Associates, Inc.

Zhou, X. and Y. Xie (2019). Marginal treatment effects from a propensity score perspective. *Journal of Political Economy 127*(6), 3070–3084.

Appendix

Table of Contents

A Summary statistics

Table A.1 reports application-level summary statistics for the estimation window of Section 3.1. The leniency rows show the raw material of the design: the leave-one-out examiner grant rate disperses widely (SD 0.185) around the same mean as its art-unit expectation, and their difference, the recentered deviation Δl, is centered on zero with a standard deviation of 0.129. The scope rows are reported in words per claim, the raw inverse of scope, before the within-year percentile transformation of Section 3.2; granted claims carry more words than filed claims on average; the granted mean runs over the granted subsample while the filed mean runs over all applications, so the contrast mixes the narrowing prosecution produces with selection into grant, and the paired within-application evidence is in Section 3.2. Forward citations to the publication are defined for granted and rejected applications alike, which is the coverage property the validation exercises of Section 3.4 use.

Table A.1: *Summary statistics, estimation sample*

	N	Mean	SD	P10	Median	P90
Granted	5,367,642	0.711	0.453	0.000	1.000	1.000
Examiner leniency (leave-one-out)	5,367,642	0.661	0.185	0.400	0.692	0.878
Expected leniency (art-unit pool)	5,367,642	0.661	0.133	0.485	0.679	0.813
Leniency deviation Δl	5,367,642	-0.000	0.129	-0.170	0.014	0.149
Invention quality (within-cell percentile)	5,364,971	0.506	0.289	0.110	0.510	0.910
Filed scope (words per claim)	5,367,642	55.7	40.8	27.8	47.3	89.3
Granted scope (words per claim)	3,772,516	69.8	49.8	33.5	58.5	113.8
Revision rounds (non-final rejections)	5,363,782	1.27	0.89	0.00	1.00	2.00
Forward citations (to the publication)	5,367,642	3.13	11.01	0.00	1.00	7.00
Realized renewals	3,814,653	-0.15	0.49	-1.00	0.00	0.00

Notes: Application-level summary statistics for the estimation window, publication years 2005–2020 (5,367,642 applications; the examiner-dummy estimators drop a further three observations, see Section 3.1). *Granted* is the grant indicator. *Examiner leniency* is the leave-one-out mean acceptance rate over the assigned examiner's other decisions; *expected leniency* averages the leniency of examiners deciding in the same art unit within an eighteen-month window on either side of the decision date; the *deviation* is their difference, the recentered variation the design uses. *Invention quality* is the text-based index of Section 3.3, expressed as a within-art-unit × publication-year percentile rank on the unit interval. *Filed* and *granted scope* are the average number of words per claim in the application and the issued patent respectively; more words delineate a narrower claim, so word count is the inverse of scope (Section 3.2), and the granted row conditions on grant. *Revision rounds* counts non-final rejections. *Forward citations* count citations to the application publication. *Realized renewals* are observed for granted patents only, computed as actual minus potential renewals, so zero means renewed at every opportunity.

B Proof of Proposition 1

Write the standardized acceptance index as $t(q,s) = (a(q,s)+l)/\sigma$, so that the grant probability is $P = \Phi(t)$ and the first-stage sensitivity to leniency is $\psi = \partial P/\partial l = \sigma^{-1}\varphi(t)$. Equation (12) writes the share of compliers among granted patents as $\phi_c^{D=1} = \psi\,\bar{d}\,/\,P$, with $\bar{d} = \mathbb{E}[(\Delta l)_+]$ a constant that does not depend on (q,s) under quasi-random assignment. Substituting,

$$\phi_c^{D=1}(q,s) \;=\; \frac{\bar{d}}{\sigma}\,\frac{\varphi\big(t(q,s)\big)}{\Phi\big(t(q,s)\big)} \;=\; \frac{\bar{d}}{\sigma}\,\lambda\big(t(q,s)\big),$$

where $\lambda(t) = \varphi(t)/\Phi(t)$ is the inverse Mills ratio.

Part (i). The inverse Mills ratio is strictly decreasing on $\mathbb{R}$: $\lambda'(t) = -\lambda(t)\big(t+\lambda(t)\big) < 0$, since $\lambda(t) > \max\{0,-t\}$ for all t (a standard bound on the hazard of the normal). Hence $\phi_c^{D=1}$ is a strictly decreasing function of t, and therefore of the acceptance index $a(q,s)$, holding l fixed. For any characteristic $x \in \{q,s\}$, $\partial\phi_c^{D=1}/\partial x = (\bar{d}/\sigma)\,\lambda'(t)\,\sigma^{-1}\,\partial a/\partial x$, which has the sign of $-\partial a/\partial x$ and is zero exactly when $\partial a/\partial x = 0$. Thus the complier share strictly increases in any characteristic that lowers the acceptance index and is invariant to any characteristic that leaves it unchanged. □

Part (ii). The treated-complier mass, the probability that an application is granted *because* the realized draw exceeded the expected one, is $\psi\,\bar{d} = (\bar{d}/\sigma)\,\varphi\big(t(q,s)\big)$, with $\bar{d} = \mathbb{E}[(\Delta l)_+]$ as in (12). This is the one-sided object; the two-sided flip rate ϕ_c of (11) replaces $\bar{d}$ with $\mathbb{E}[|\Delta l|]$ and is twice as large under mean-zero recentering. Both are proportional to $\varphi\big(t(q,s)\big)$, so the shape argument is common to them and only the level differs. The standard normal density φ is strictly single-peaked at 0, so each is single-peaked in the acceptance index and attains its maximum where $t(q,s) = 0$, i.e. where $P(q,s,l) = \Phi(0) = \frac{1}{2}$. □

B.1 The mechanism in four canonical threshold models

Proposition 1 is a statement about threshold-crossing models generally, and each structural literature carries a canonical member. The mappings below are exact on the point that matters, which margin the threshold prices, and each supplies the corresponding answer to why optimizing applicants do not equalize expected values across quality.

Menu costs. Leniency dispersion plays the role of a small aggregate shock and the grant rule that of an Ss band: the applications that respond are those nearest the trigger, selected on their gap, not a random draw. The gap here is claimed scope. Quality is a firm characteristic orthogonal to the gap: knowing productivity does not locate a firm within its band, and knowing quality does not locate an application relative to the threshold, because the index prices scope and admits quality only through a noisy reading of the file. The complier-share profile of Section 6.3 is the density at the adjustment margin, the same object that governs the selection effect in state-dependent pricing.

Trade. A tariff change moves the export cutoff and selects the firms nearest it; the cutoff prices productivity, so marginal exporters are the least productive exporters. The grant cutoff prices the size of the request, so marginal applications are broad rather than bad. On equalization: utility equalizes across locations only through the mobility margin. Quality is an immobile fundamental, land rather than labor, and its rents persist; the indifference condition holds on the mobile margin, scope, the margin on which the applicant's first-order condition of Section 7.3 operates.

Auctions and screening. Filed scope is a bid read in reverse: claiming more raises the surplus conditional on a grant and lowers its probability. Expected surplus constant in type would violate the envelope condition, under which interim utility is the integral of the allocation probability and therefore monotone in type: rents are information rents, and a screening authority cannot extract them. The bid would still reveal the type under single crossing, if higher quality raised the marginal value of scope. It does not here: the value of an additional claim lies in excluding neighboring variants, roughly independent of the quality of the core, the empirical content of the within-field value–quality correlation near zero behind $\omega = 1$ in Section 7.7, and of the two filing regimes of opposite sign in Section 7.3. Without single crossing the request pools the types, and screening on the request is not screening on the type.

Search. Filed scope is a reservation wage: refusing cheap certainty for a better outcome at the price of a longer spell and exit risk. Grant is job finding, revision rounds are duration, abandonment is exit, and the concession decay γ_{A} is the declining reservation wage of a finite horizon. Compliers are whoever stands at the margin the instrument moves, and the margin reads the request, which the office observes, not the type, which it does not. The value of search rises with the offer distribution, so the reservation-wage trade-off equalizes nothing across workers.

The common structure: distance to the margin is set by a choice the gatekeeper reads, the request, not by the fundamental it cannot, the type; and level equalization across types requires an arbitrage margin, entry, mobility, or full extraction, that priority removes by design.

Remark. Part (i) is stated holding the examiner draw l fixed at a reference value (the mean leniency at which the profiles are evaluated); the monotonicity is in the systematic index $a(q, s)$. Additive separability of l is what makes ψ depend on (q, s) only through a; a leniency-characteristic interaction would add a term to $\partial a/\partial x$ and is the empirically testable deviation discussed in the text. Neither part is new as mathematics. Part (i) is the monotonicity of the normal hazard and part (ii) the unimodality of the normal density; what is specific to this setting is that the argument of both is an acceptance index in which the applicant's own filed scope enters, so the composition is partly a statement about applicant behavior and not only about the screening technology.

B.2 Assessment noise and the "hard to judge" reading

The Remark above holds the assessment scale σ fixed and names a leniency-characteristic interaction as the testable deviation. Table B.1 measures both. A probit of the grant outcome on the recentered leniency deviation, run within a bin of quality or of filed scope, has slope β/σ and cannot on its own say

which of the two moves. The art unit $\times$ year pool effect enters the acceptance index additively rather than through β, so its slope in the same probit is $\bar{\sigma}/\sigma$, which varies with σ alone. The ratio of the two slopes is $\beta/\bar{\sigma}$, so both are recovered up to one common constant and their variation across bins can be read separately. The separation is not a refinement. Imposing $\beta \equiv 1$ returns a σ that rises by seven percent across quality, and freeing β reverses that to a fall of fifteen percent, with the pass-through falling by twenty.

Neither margin supports the "hard to judge" reading. Noise falls across quality rather than rising, so the applications assessed least precisely are the weakest ones. Weighting applications by the complier density puts the compliers' noise 0.32 percent above the population mean, against a spread of half across cells. The reading fails on the mechanism before it fails on the magnitude. Since $\psi = \varphi(t)/\sigma$, the derivative $\partial\psi/\partial\sigma$ carries the sign of $t^2 - 1$. A noisier assessment therefore adds to the complier margin only more than one standard deviation from the threshold, which means a grant probability outside $[0.16, 0.84]$, and no cell of the surface is there. Over the range the data cover, noise thins the margin rather than supplying it.

Two further senses of "hard to judge" concern the quality measure rather than the examiner, and Table H.4 in Appendix H reports both. The first is imprecision, the disagreement between scores built on two independent halves of the text corpus at fixed construction, which loads on compliers negatively or not at all. The second is reference dependence, the disagreement between the art-unit and the coarser technology-center comparison pool, on which compliers rank 0.57 percentage points higher once ranked within art unit $\times$ year. Both are properties of the measure rather than of the examiner's assessment, and both are an order of magnitude below the scope margin that carries the composition.

Table B.1: *Assessment noise, leniency pass-through, and the complier margin*

	β/σ	$\bar{\sigma}/\sigma$	σ	β
Assessment noise across quality, filed scope pooled				
lowest quality bin	2.484	0.523	1.000	1.000
2nd quality bin	2.401	0.555	0.942	0.910
3rd quality bin	2.365	0.582	0.898	0.854
4th quality bin	2.309	0.605	0.863	0.802
5th quality bin	2.323	0.602	0.869	0.812
6th quality bin	2.310	0.610	0.856	0.796
7th quality bin	2.327	0.611	0.855	0.801
highest quality bin	2.331	0.616	0.848	0.796
Assessment noise across filed scope, quality pooled				
narrowest scope bin	2.326	0.646	1.000	1.000
2nd scope bin	2.396	0.592	1.093	1.125
3rd scope bin	2.422	0.591	1.094	1.139
4th scope bin	2.406	0.573	1.129	1.168
5th scope bin	2.389	0.564	1.147	1.177
6th scope bin	2.366	0.563	1.147	1.167
7th scope bin	2.372	0.557	1.161	1.184
broadest scope bin	2.262	0.552	1.171	1.139
Compliers less population, weights from Figure D.5				
quality (percentile points)	−0.68			
filed scope (percentile points)	+0.81			
assessment noise σ (percent)	+0.317			
examiner bootstrap s.e.	(0.036)			
95% interval	[+0.25, +0.39]			
Dispersion of σ, and the sign of $\partial\psi/\partial\sigma$				
across cells, max/min	1.533			
across cells, s.d. (percent of mean)	8.6			
cells where noise raises the margin	0 of 64			

Notes: The Table reports the assessment-shock scale σ of Equation (5) and the leniency pass-through β, estimated on the estimation sample ($n = 5,353,106$ applications, 12,698 examiners, art unit × publication year pools of at least 30). Within each bin, a probit of the grant outcome on the recentered leniency deviation has slope β/σ, which on its own cannot say which of the two varies. The art unit × year pool effect enters the acceptance index additively rather than through β, so its slope in the same probit is $\bar{\sigma}/\sigma$, which varies with σ alone, and the ratio of the two slopes is $\beta/\bar{\sigma}$. The pool effect is measured as the leave-one-out pool grant rate on the probit scale, so an application never enters its own regressor. Both implied columns are relative to the first row of their block and are identified up to one common constant, so levels carry no units and only ratios across bins are interpretable. Complier means are the modified-outcome form $\mathbb{E}[X \mid \text{complier}] = \sum_i \psi_i X_i / \sum_i \psi_i$ with ψ the surface of Figure D.5, matched to applications on quality and filed scope. The same weights return the complier quality gap of Section 6.3 to within a tenth of a percentile point, which is the check that they are being applied correctly. Standard errors resample examiners with replacement, $B = 300$. They hold ψ and the pool effect at their point values, so the sampling error of the complier surface itself is not propagated. The last block reports the dispersion against which the complier displacement should be read, and the count of cells in which a noisier assessment would raise rather than thin the complier margin. That count follows from $\partial\psi/\partial\sigma$ carrying the sign of $z^2 - 1$ at $z = \Phi^{-1}(P)$, so noise adds to the margin only where the cell grant rate falls outside $[0.16, 0.84]$. Quality is a percentile within art unit × publication year, so every pool enters every quality bin equally and the quality block is balanced on pool composition by construction. Filed scope is ranked within publication year, so pools enter its bins unequally and the scope block should be read as directional.

C Proof of Proposition 2

Setup and solution concept. There are two parties. An applicant A holds a joint prior $w^0 \in \Delta(Q \times L)$ over own quality $q \in Q = [0,1]$ and examiner leniency $l \in L = [0,1]$. An examiner E holds a prior $\phi^0 \in \Delta(Q)$ over application quality and applies a threshold grant rule g_E together with the offer (16), both taken as structurally given: the examiner has no objective function and does not optimize. The solution concept is correspondingly one-sided, and deliberately so. A solution consists of (i) an applicant policy σ_A^* mapping each history h_r to an action in $S_A(h_r)$; (ii) a belief system μ_A^* assigning posterior $w_r^A \in \Delta(Q \times L)$ at each h_r via Bayes' rule; and (iii) an examiner belief system μ_E^* assigning $\phi_r^E \in \Delta(Q)$ at each h_r via Bayes' rule from signals $\xi_r = q + \eta_r$. Optimality requires σ_A^* to maximize the applicant's expected payoff at every information set given μ_A^* and g_E. This is a single-agent stochastic control problem with an exogenous, empirically disciplined environment, not an equilibrium of a game; no mutual best-response condition is imposed or claimed.

Decoupling. Because the examiner forms quality beliefs exclusively from the private signal $\xi_r = q + \eta_r$– and not from the applicant's scope choice s_r–the examiner's posterior ϕ_r^E evolves independently of σ_A. From the applicant's perspective, ϕ_r^E is therefore an exogenous stochastic process that can be forecast without reference to own strategy. This decouples the two belief systems: the applicant can treat the examiner's dynamic threshold as a feature of the environment rather than a strategic response. The applicant's problem reduces to a finite-horizon single-agent stochastic control problem, which is solved by backward induction on the round index $r \in \{0, \ldots, R-1\}$, the text's convention, with $R-1$ the terminal round.

State and action spaces. The applicant's state at round r is the pair $(w_r^A, \bar{s}_r)$: the posterior $w_r^A \in \mathcal{P} \equiv \Delta(Q \times L)$, endowed with the weak topology and compact and convex; and the *scope ceiling* $\bar{s}_r \in [0, \bar{s}_A]$. The action space at round r is

$$S_A(\bar{s}_r) \;=\; [0,\, \bar{s}_r], \qquad \bar{s}_0 = \bar{s}_A, \qquad \bar{s}_r = s_0 \text{ for all } r \geq 1,$$

so prosecution may narrow claims and re-broaden them up to, but never beyond, the original filing s_0: the new-matter bar caps every amendment at the initial disclosure, which is exactly the constraint the solved model imposes. The permanence of granted-scope concessions (Sections 2 and 5.1) operates through the offer (16), not through the action set. $\bar{s}_A$ is the upper bound on filed scope imposed by existing prior art, and 0 is the outside option of abandoning the application. Each $S_A(\bar{s}_r)$ is compact and convex, and the correspondence $\bar{s} \mapsto S_A(\bar{s})$ is continuous (indeed both upper and lower hemicontinuous, being an interval with continuous endpoints), which is what Berge's theorem requires below. The grant probability $d_r(s, w_r^A)$, the granted scope $\hat{s}(s)$, and the revision cost $C_A(s)$ are continuous in s by the smoothness of the normal CDF and the continuity of the cost structure; payoffs are bounded by construction.

Backward induction – base case ($r = R-1$). At the terminal round the game ends whatever happens, so both the continuation branch and the abandonment branch of (18) are worth zero and the applicant maximizes the terminal instance of that objective,

$$U_A^{R-1}(s \mid w_{R-1}^A) \;=\; d_{R-1}(s, w_{R-1}^A) \cdot \nu s \;-\; \kappa s^\chi,$$

with κs^χ the scope-acquisition cost. d_{R-1} is jointly continuous in (s, w_{R-1}^A): the grant probability integrates a bounded function of (q, l, s), continuous in s by the smoothness of the normal CDF, against w_{R-1}^A, and integration of a bounded continuous function is continuous in the measure under the weak topology. Hence U_A^{R-1} is jointly continuous in state and action, and continuous in s on the compact set $S_A(\bar{s}_{R-1})$. By the Extreme Value Theorem the maximum is attained, so the argmax correspondence $\sigma_{R-1}^*(w_{R-1}^A, \bar{s}_{R-1}) = \operatorname{argmax}_{s \in S_A(\bar{s}_{R-1})} U_A^{R-1}(s \mid w_{R-1}^A)$ is non-empty at every state. By Berge's Maximum Theorem, applicable because $S_A(\cdot)$ is a continuous, compact-valued correspondence and the objective is jointly continuous, the value function $V_{R-1}(w_{R-1}^A, \bar{s}_{R-1}) = \max_{s \in S_A(\bar{s}_{R-1})} U_A^{R-1}(s \mid w_{R-1}^A)$ is bounded and continuous in the state.

Backward induction – inductive step. Suppose at round $r+1$ there exists a non-empty argmax correspondence σ_{r+1}^* and a bounded, continuous value function $V_{r+1}(w_{r+1}^A, \bar{s}_{r+1})$. Upon a rejection at round r the applicant updates beliefs via Bayes' rule:

$$w_{r+1}^A \;=\; B^A(w_r^A,\, s,\, \text{reject}) \;\propto\; w_r^A(q, l) \cdot \Pr(\text{reject} \mid q, l, s),$$

where the rejection likelihood is the marginal over the examiner's signal history ϕ_r^E, which the applicant never observes: from the grant rule (15),

$$\Pr(\text{reject} \mid q, l, s) = \mathbb{E}\big[\Pr(\text{reject} \mid q, l, s, \phi_r^E) \,\big|\, q\big] = \Phi\!\left(\frac{\mu - \alpha_q\, \mathbb{E}[\mathbb{E}_r[q] \mid q] + \Pi(s) + b_r\, r + b_0\, \mathbf{1}[r=0] - l}{\sigma_{\text{tot}}(r)}\right),$$

the second equality exact under the Gaussian learning structure because the dispersion of the realized posterior mean around its signal-integrated value is precisely the $\sigma_e(r)$ component of $\sigma_{\text{tot}}^2(r)$ in (32), with Π the scope penalty. Rejection is the event the applicant conditions on; whether they then revise or abandon is their own decision and carries no further information about (q, l). The likelihood is bounded and continuous in (q, l, s) by the smoothness of Φ and of Π, so the grant probability $d_r(s, w_r^A) = 1 - \int \Pr(\text{reject} \mid q, l, s)\, \mathrm{d}w_r^A(q, l)$, now defined solely from the stated applicant state, is jointly continuous in (s, w_r^A) under the weak topology, and the Bayes map $(w_r^A, s) \mapsto B^A(w_r^A, s)$ is weakly continuous, its normalizer bounded away from zero because Φ is interior on the compact support. The round-r objective is

$$U_A^r(s \mid w_r^A, \bar{s}_r) \;=\; d_r(s, w_r^A) \cdot \nu s \;-\; \kappa s^\chi \;+\; (1 - d_r(s, w_r^A)) \max\Big\{0,\; -C_A(s, r) + \beta\, V_{r+1}\Big(B^A(w_r^A, s),\; \bar{s}_{r+1}\Big)\Big\},$$

the objective (18), with κs^χ the scope-acquisition cost, $C_A(s, r)$ the round-r revision cost, and the outer maximum the abandonment option. Since d_r, C_A, V_{r+1}, and B^A are continuous, so is their composition, and the pointwise maximum of two continuous functions is continuous; the ceiling is constant at s_0 from

round one on, so the constraint correspondence is trivially continuous. Hence U_A^r is jointly continuous in state and action, and continuous on the compact set $S_A(\bar{s}_r)$. By the Extreme Value Theorem the argmax $\sigma_r^*(w_r^A, \bar{s}_r)$ is non-empty at every state; by Berge's Maximum Theorem, whose hypotheses of a continuous compact-valued constraint correspondence and a jointly continuous objective are now verified, the resulting value function V_r is bounded and continuous, closing the induction.

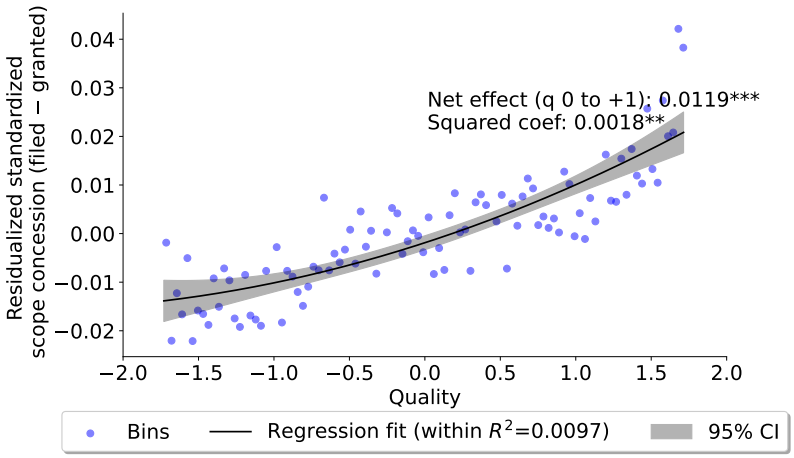
Remark on set-valued maximizers. Continuity and compactness deliver existence outright, with no appeal to randomization: quasiconcavity is not required by the Extreme Value Theorem and is not assumed here. Where U_A^r fails to be strictly quasiconcave in s the argmax may be set-valued, and any selection from it, or any distribution over it, is optimal. In the solved model the action set is a finite grid, so a measurable pure selection always exists and is what the numerical solution returns. Non-monotonicity of $\partial U_A^r / \partial s$ therefore at most generates multiple local optima, never a requirement to mix; the existence and pure-selection conclusions stand regardless.

Belief consistency. The applicant's belief system μ_A^* is constructed by applying B^A at every on-path history under σ_A^*; it is consistent with σ_A^* and g_E by construction. The examiner's belief system μ_E^* follows from the signal process independently of σ_A^*, and is similarly consistent by construction. Beliefs at histories reached with probability zero are unrestricted and play no role, since the examiner's rule g_E is exogenous and does not respond to them.

Conclusion. Backward induction delivers a non-empty optimal policy σ_r^* at every round r and state $(w_r^A, \bar{s}_r)$, together with a bounded, continuous value function. The triple $(\sigma_A^*, \mu_A^*, \mu_E^*)$ satisfies sequential optimality for the applicant and on-path Bayes consistency for both parties, which is Proposition 2. □

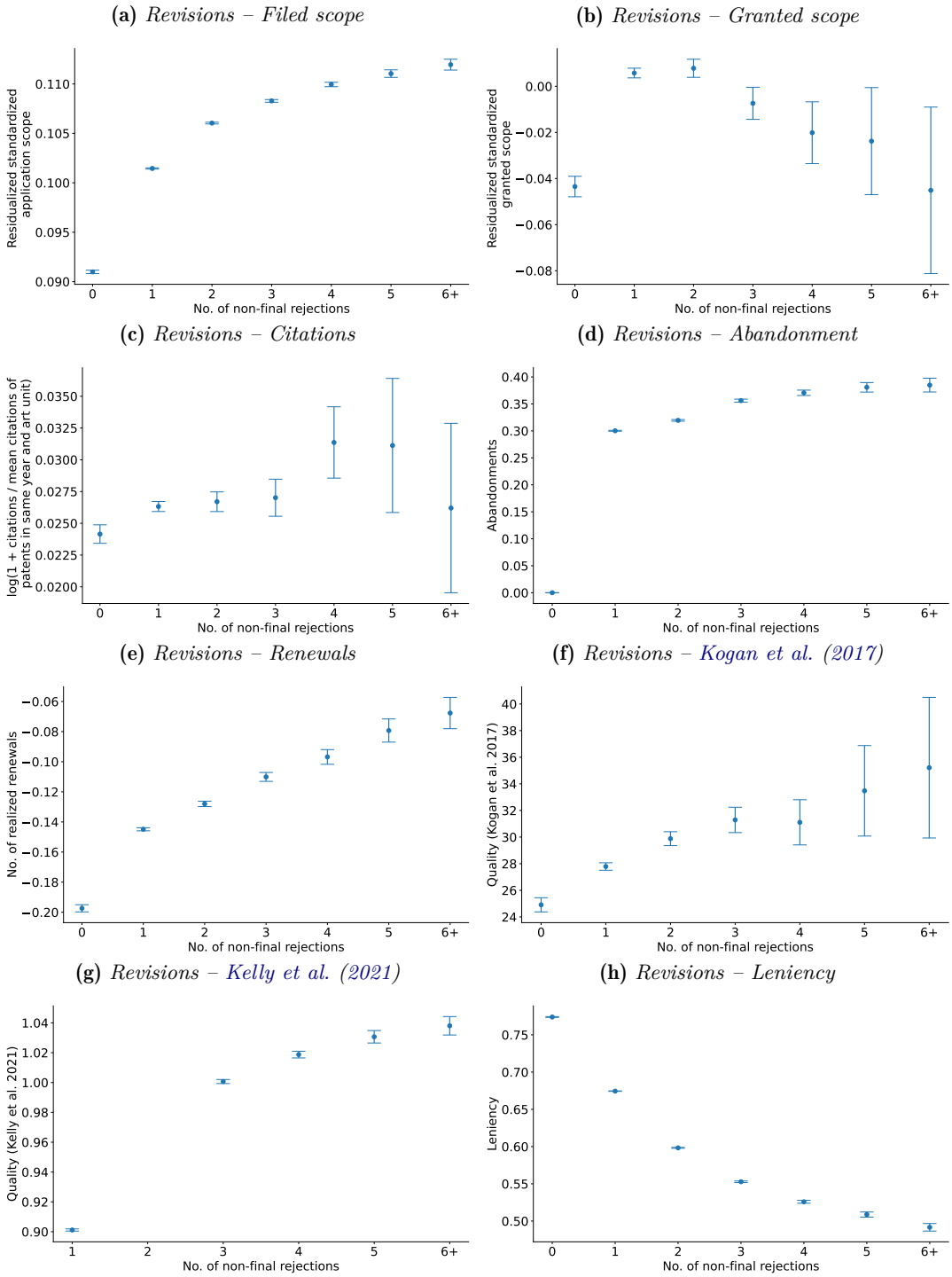
D Additional output

Figure D.1: *Conditional correlation: difference between filed and granted scope by quality*



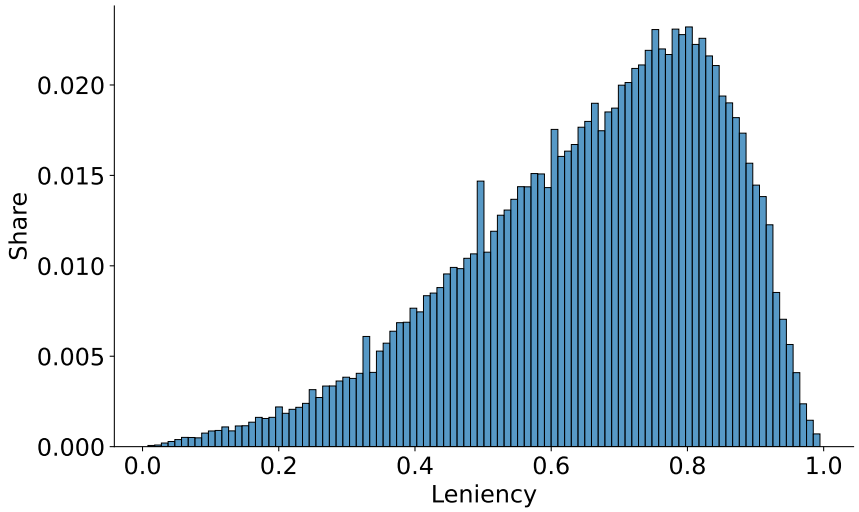
Notes: Binned scatter of the residualized difference between filed and granted scope against standardized invention quality, on the specification of Figure 2 (accepted applications, $n = 3{,}812{,}246$). Both scopes are within-year percentiles, so a positive value is a narrowing of the claims between filing and issue. The linear and squared terms and the net effect are reported in column (5) of Table K.2.

Figure D.2: *Additional patent and patent application attributes by the number of revisions*



Notes: The Figure shows the mean values of patent and patent application attributes by the number of review rounds the patent application had to go through before being accepted, abandoned, or receiving a final rejection. Because the number of patent applications going through multiple revision rounds decreases rapidly, I top-code the number of revisions at 6. The error bars represent the 95% confidence intervals based on the standard error of the means. Panels D.2a–D.2c complete the main-text Figure 3.

Figure D.3: *Distribution of examiner leniency*



Notes: The Figure plots the distribution of examiner leniency, measured as the leave-one-out acceptance rate among all previously examined applications by the same examiner. The sample is the 2005–2020 estimation sample described in Section 3.1; the leave-one-out measure is undefined for an examiner's first observed decision, and leniency values of exactly zero or one are excluded. The dashed vertical line marks the sample median. The interquartile range spans approximately 26.5 percentage points, against a mean acceptance rate of approximately 66%; this dispersion is the identifying variation for the complier shares of Section 6 and the compression experiments of Section 7.7.

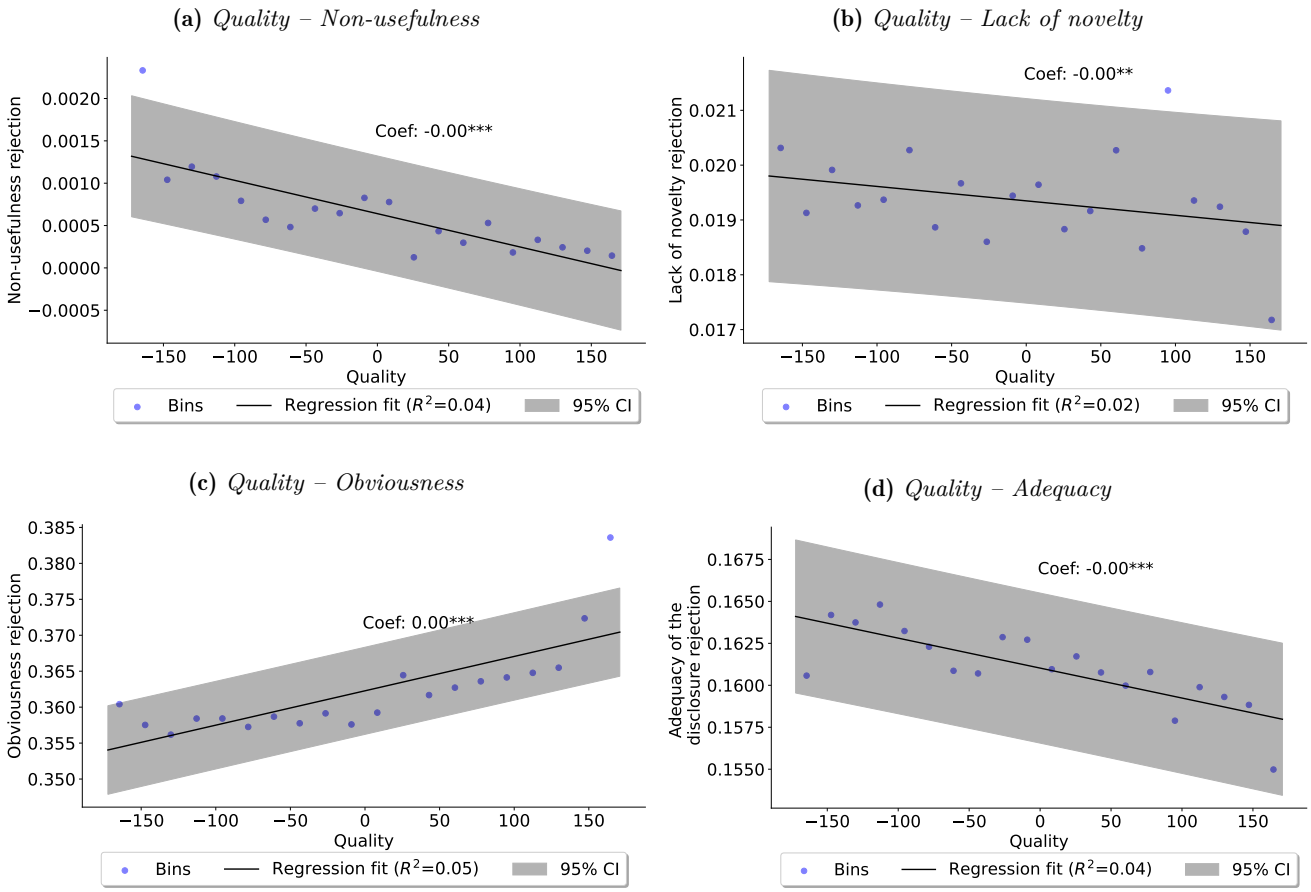
Table D.1: *Unconditional first stage*

	FE-OLS	UJIVE
Leniency pass-through $\bar{\psi}$	0.667	1.062
	(0.001)	(0.004)
Art unit × year FE	✓	✓
Controls	✓	✓
Examiner leave-out		✓
Observations	5,364,969	5,364,969

Notes: Unconditional first stage of the grant indicator on examiner leniency, on the 2005–2020 core sample (publication years; largest connected set of 12,705 examiners over 9,701 art unit × year cells). Column 1 is the fixed-effects OLS slope of grant on the recentered leniency deviation Δl_p, controls absorbed by Frisch–Waugh. Column 2 instruments that deviation with the examiner leave-out fit (UJIVE). The two constructions do different work: the leave-out fit, shared by both columns, removes the own-observation bias that inflates a raw leniency regression, while instrumenting the noisy leave-one-out index corrects its measurement-error attenuation, which is why Column 2 is larger and Column 1 is the conservative floor. Controls are small- and micro-entity indicators, continuation-type dummies, examiner experience, and a first-year-active indicator. Standard errors clustered on art unit × publication year (9,701 clusters). The two columns are the two pass-through conventions discussed in Appendix D; the main text uses the FE-OLS figure, the conservative of the two. Mean favourable deviation $\bar{d} = 0.0503$ and grant rate 0.7106.

The complier surface and rejection reasons.

Figure D.4: *Conditional correlations – Rejection subsample*



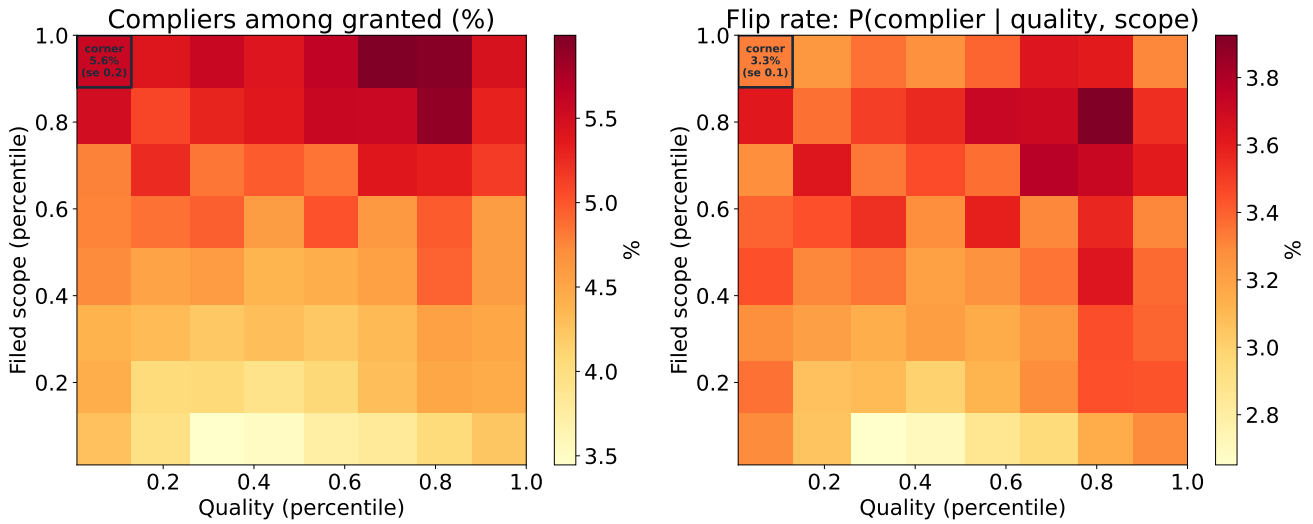
Notes: The Figure plots the conditional correlations between my quality measurement and the different rejection reasons. The sample size for these regressions is 2,088,978. As in Figure 2, the control variables include small and micro entity indicators, categorical dummies for the type of continuation of a patent application, and examiner experience dummies, with art unit × year fixed effects absorbed. No post-treatment variable enters the conditioning set; in particular the number of review rounds, which is an outcome of prosecution, is not controlled for. The binned scatter plots are produced using `binsreg` (Cattaneo et al., 2024) and covariates are included via auxiliary regressions and residualizing the outcome variable of interest.

Panel D.5a plots the two dimensional complier surface. The figure shows that the direction of the scope gradient is universal: in all quality bands the share is higher at the broad end of filed scope than at the narrow end, and in all one hundred scope bands it is higher at the bottom of the quality distribution than at the top, running from 4.9% to 6.5% across scope in the lowest quality band and from 4.9% to 4.1% across quality among the narrowest filings. The scope margin is the wider of the two, spanning between one and two percentage points within a quality band against between a third of a point and one point within a scope band, so the grants that examiner inconsistency creates are concentrated on quality-scope mismatches: ambitious claims attached to inventions whose quality does not stand out.

Panel D.5b plots the flip rate on the two dimensional space. The figure highlights that concentration is sharper among granted patents than among applications. The flip rate of Equation (11), the share of all applications whose grant status the draw decides, moves less over the plane than the share among granted patents. Between the two corner cells that bound the granted-share gradient it moves from 6.4%

Figure D.5: *Complier composition over the quality-scope plane*

(a) *Compliers among granted patents (%)* **(b)** *Flip rate: share of all applications (%)*



Notes: The Figure plots the complier distribution over the joint quality × filed scope plane, estimated by the examiner-dummy UJIVE modified outcome $\mathbf{1}[q \in B_i \wedge s \in B_j] \cdot D$ on an 8×8 grid of equal-mass marginal quantile bins (Section 5.3). Panel (a) reports the share of compliers among granted patents, Equation (12); Panel (b) the flip rate, the share of all applications in the cell whose grant status is decided by the examiner draw. Both axes are in-sample percentile ranks. The low-quality × broad-scope cell is outlined and annotated with its cluster-robust standard error. Cells with fewer than two hundred applications are masked; none are on this grid. Across the 100×100 evaluation grid underlying the surface, the flip rate falls monotonically in the cell grant rate (rank correlation -0.74) and peaks in the cells closest to a grant probability of one half, the single-peakedness that part (ii) of Proposition 1 predicts.

to 7.2%, a spread of less than one percentage point against the 2.4 points separating the same cells in the share among granted patents. The mass of applications sitting at the acceptance margin therefore varies less across the plane than the share it represents among granted patents, because the grant rate itself falls from 77% at the narrow extreme of the scope distribution to 58% at the broad extreme. Section 6.3 shows that this is not a coincidence but the signature of the threshold model. The numerator and the denominator of the inverse Mills ratio move differently, and over the range of acceptance indices the data cover, it is the denominator that moves more: along the scope margin, roughly three fifths of the log change in the share among granted patents is the grant rate and two fifths the density at the margin. The corner is accordingly a statement about the composition of the patents that exist and can be asserted, which is the population that matters for the welfare accounting of Section 7.7.

Figure D.4 shows that rejection reasons do not provide an alternative explanation for the complier composition. The two time-intensive rejection reasons, lack of novelty and obviousness, cannot by themselves generate the observed pattern. While Panel D.4c shows that obviousness rejections are positively correlated with application quality,[71] Panel D.4b shows that lack-of-novelty rejections are essentially uncorrelated with quality. Table K.3 quantifies both: a one-standard-deviation increase in quality raises the probability of an obviousness rejection by 0.48 percentage points, about a sixtieth of the 27.8% base

[71] Narrowing claims is a standard response to an obviousness rejection. USPTO MPEP §2143 sets out the rationales an examiner must establish for a prima facie case of obviousness; it is the grounds so established that a narrowing amendment is drafted to overcome.

rate, while its effect on lack-of-novelty rejections is −0.03 percentage points with a standard error of 0.01: precisely estimated, marginally distinguishable from zero, and an order of magnitude smaller than the obviousness response, on 2,088,978 applications.

Reading the rejection-reason coefficients. The scale of Table K.3 differs from that of Table 1, where quality enters as a raw percentile rank on the unit interval and the coefficients are therefore full-range effects. In the rejection-reason regressions the quality percentile is standardized to unit variance and demeaned within art unit × year cells before estimation, so the coefficient is the effect of a one-standard-deviation move within a cell.

These profiles narrow the rejection-reason alternative without closing it. What the two panels rule out is that the complier composition is an artifact of which reason examiners invoke varying systematically with quality: on the novelty margin it does not, and on the obviousness margin it runs in the direction that reinforces rather than substitutes for the mechanism of Section 6.3. What they do not rule out is heterogeneity in how each rejection reason responds to scope or to leniency. A rejection type could be flat in quality and still be the channel through which scope draws scrutiny, or through which lenient and strict examiners differ. Testing that requires interacting rejection reason with scope and with leniency, which these marginal profiles do not do (Appendix D). I therefore read them as evidence that rejection-type composition in quality is not driving the result, not as evidence that rejection reasons are irrelevant to it.

E Robustness: many-examiner instruments and the UJIVE estimator

The main-text complier shares of Section 6 use the examiner-dummy UJIVE modified-outcome estimator described in Section 5. This appendix first collects the implementation details of that estimator, the smoothing and inference choices and the derivation of the flip rate, and then presents two robustness exercises: a semiparametric partially linear estimator of the same first-stage sensitivity, and a many-instrument comparison confirming that a one-dimensional leniency index summarizes the examiner variation.

Sample form of the modified-outcome estimator. In sample, the modified-outcome coefficient of Section 5.3 is the just-identified instrumental-variables ratio

$$\hat{\theta}(B) \;=\; \frac{\sum_p U_p \,\mathbf{1}[X_p \in B]\, D_p}{\sum_p U_p D_p}, \tag{22}$$

whose numerator and denominator both scale with the complier share; the share therefore cancels, which is why $\theta(\cdot)$ in (8) is a proper distribution over any partition and requires no auxiliary weighting scheme.

Implementation: smoothing, resolution, and inference. The complier profiles of Section 5.3 are reported in two resolutions. The binned version uses twenty equal-mass bins of the percentile axis. The smoothed version replaces the bin indicator in (22) with a Gaussian kernel $K_h(X_p - x)$ of bandwidth $h = 0.07$ over the unit percentile interval, comparable to a natural cubic spline with seven degrees of freedom; the population density in the ratio's denominator is estimated as the mean kernel weight at x, so numerator and denominator share the kernel and the leading smoothing bias common to both cancels in the ratio $\rho(x)$. That cancellation does not make the ratio unbiased. A density ratio smoothed with a common kernel retains an $O(h^2)$ term in the *curvature* of the two densities, which survives precisely where numerator and denominator curve differently, so the reported bands cover the smoothed ratio rather than the underlying one and are correctly centered only to that order. Results are reported for $h \in \{0.05, 0.07, 0.10\}$ and are insensitive to the choice, which bounds the practical importance of the residual term without eliminating it; the outermost percentiles are kernel-edge regions and are trimmed. Standard errors come from the influence function of the instrumental-variables ratio and are clustered on the art unit $\times$ year cell, the level at which examiners are assigned. Because the profiles are read as curves rather than at single points, I also report *simultaneous* bands: a Gaussian multiplier (wild cluster) bootstrap with 2,000 draws delivers a studentized sup-t critical value (2.68 on the quality axis, 2.85 on the scope axis) and, from the same draws, a joint test of the null that the curve is flat.

Derivation of the complier shares. Under the linear probability model of Section 5.5, application p at characteristic value x is a complier in one of two events: $l_p > l_p^* > \tilde{l}_p$ (granted by an above-average examiner, would have been rejected by the expected one) or $l_p < l_p^* \leq \tilde{l}_p$ (rejected by a below-average examiner, would have been granted by the expected one). The probability of either event for a given deviation Δl_p is $\psi(x) \cdot |\Delta l_p|$, and taking expectations over the instrument distribution gives Equation (11) in the text. The second equality there uses two facts: that $\psi(x)$ factors into the level $\bar{\psi}$ times the shape

$\rho(x)$, and that Δl_p is independent of x under quasi-random assignment, so its distribution can be taken outside the conditioning.

The share among granted patents conditions on the favorable event alone: the probability of $l_p > l_p^* > \tilde{l}_p$ conditional on the realized draw is $\psi(x) \cdot (l_p - \tilde{l}_p)_+$, so integrating over the instrument distribution gives the treated-complier mass $\psi(x)\,\bar{d}$ with $\bar{d} = \mathbb{E}[(\Delta l_p)_+]$, and dividing by the conditional grant rate at x yields Equation (12) in the text. In estimation, $\bar{\psi}$ comes from the pooled first stage (7), $\rho(x)$ from the modified outcomes (9), and the conditional grant rate $P[D = 1 \mid x]$ from a Nadaraya-Watson average on the same kernel and grid. $\bar{\psi}$, $\bar{d}$, and the conditional grant rate are held fixed at their sample values, and standard errors and simultaneous bands are computed from $\rho(x)$ alone. For the two scalars this is exact bookkeeping: they enter multiplicatively, so their uncertainty moves the level of a profile and not its shape. The conditional grant rate is not shape-neutral in the same way, since it varies with x and enters the denominator pointwise; the bands therefore condition on its estimate rather than integrate over it, an omission that is negligible at this n rather than innocuous by construction – each kernel cell averages a binary outcome over hundreds of thousands of applications, putting the denominator's pointwise relative sampling error at the order of a tenth of one percent, well inside the bands on $\rho(x)$. Because $\rho(\cdot)$ integrates to one against the population density, the grant-weighted average of (12) collapses to $\bar{\psi}\,\bar{d}\,/\,P(D = 1)$, the aggregate share of Section 6.1.

Semiparametric partially linear estimator. Following the residualization approach of Robinson (1988), I also estimate the conditional first-stage sensitivity $\psi(q)$ from the partially linear system

$$D_{p_{kt}} = \alpha_{kt} + \psi(q)\, l_{p_{kt}} + g(q_{p_{kt}}, s_{p_{kt}}, X_{p_{kt}}) + \upsilon_{p_{kt}} \tag{23}$$

$$l_{p_{kt}} = \gamma_{kt} + m(q_{p_{kt}}, s_{p_{kt}}, X_{p_{kt}}) + \eta_{p_{kt}}, \tag{24}$$

where α and γ are art unit $\times$ year fixed effects and the nuisance functions $g(\cdot)$ and $m(\cdot)$ absorb the nonlinear effects of quality, filed scope, and covariates on both grant and leniency. I estimate the nuisances by random forests with cross-fitting (double machine learning), and identify the varying coefficient $\psi(q)$ as the best linear projection of the cross-fitted signal on a tensor of natural cubic splines in the quality and scope percentiles. This estimator collapses the examiner dummies into the single recentered index $\Delta l_p = l_p - \tilde{l}_p$ and relaxes the linear-control structure of the main estimator.

On the **scope** margin the two estimators agree: the complier share rises steeply across the filed-scope distribution under both. On the **quality** margin they do not. The modified-outcome estimator is nearly flat, moving within half a percentage point of its mean with an interior lift; the partially linear estimator declines. Because this disagreement bears directly on the paper's second headline fact, I traced it. It is not binning (the kernel-smoothed modified outcome, matched to the same effective degrees of freedom, reproduces the modified-outcome shape), not the instrument set (the scalar recentered index run through the modified outcome agrees with the examiner-dummy version), not the control set (re-residualizing the same instrument on the double machine learning conditioning set leaves the divergence intact), and not the flexibility of the learner (replacing the random forest with linear nuisances leaves the partially linear profile essentially unchanged). What remains is the estimand itself. The partially linear conditional effect

is a variance-weighted projection, proportional to $\mathrm{Cov}(\Delta l, D \mid q)/\mathrm{Var}(\Delta l \mid q)$, whereas the complier mass is the covariance alone; where the residual instrument variance varies with quality the two curves separate. The modified-outcome estimator is the object that answers the question posed in Section 5.3, and it is the one reported in the main text. Nor does the choice between a common and a conditional dose reopen the question: reweighting the profile point by point by the conditional favorable dose of Appendix F moves it by at most three percent anywhere, so the dose-weighted profile shares the near-flatness of the reported one.

Many-instrument robustness. Collapsing the roughly twelve thousand examiners into the single leave-out index Δl_p is, in effect, a one-dimensional summary of the examiner fixed effects in the grant equation. Because the expected leniency $\tilde{l}_p$ is constructed by leaving out the application's own examiner draw, this index is already insulated from the own-observation bias that afflicts the classical many-examiner instrument set (Bound et al., 1995; Angrist, Imbens, and Krueger, 1999). I verify this directly: I re-estimate the first stage with the full set of examiner dummies as instruments and compare estimators that differ only in how they handle many-instrument and own-observation bias.

Estimators. Let D_p be the grant indicator, Z the matrix of examiner dummies, and W the controls (art unit $\times$ year fixed effects and the covariates defined in Section 5.3). Two-stage least squares (2SLS) projects D on $[Z, W]$ in sample; with many instruments its fitted first stage is overfit and biased toward the OLS association (Bound et al., 1995). The jackknife instrumental variables estimator (JIVE; Angrist et al., 1999) replaces each fitted value with its leave-one-out analogue, removing the own-observation component. The unbiased jackknife estimator (UJIVE; Kolesár, 2013) additionally leaves the application out of the *control* projection, making it consistent under many-instrument and many-control asymptotics, the relevant regime here, where the art-unit $\times$ year fixed effects are themselves numerous. The leave-out fitted values require the leverages of the joint examiner-by-cell design, which is a bipartite two-way fixed-effects problem with no closed form; I compute them by the Johnson-Lindenstrauss random-projection estimator used by Kline, Saggio, and Sølvsten (2020), restricting the sample to the largest leave-out-connected component of the examiner-cell graph (which retains all but three observations).

Result. On the 2005–2020 core sample (n = 5,367,639; 10,270 art-unit $\times$ year cells; 12,706 examiners; 22,986 design parameters), the examiner dummies are jointly strong (joint first-stage $F \approx 29.8$). The mean leverage is only $K/n \approx 0.004$, so the own-observation and many-instrument distortions are structurally small. Table E.2 reports the result: the honest leave-out (UJIVE) first stage recovers 96.6% of the in-sample 2SLS explained variation in grant, and JIVE 93.7%.[72] The implied overfit of the examiner-dummy first-stage projection is thus at most three to four percent, so the design is not materially contaminated by the many-instrument problem at this sample size. The ratio holds the instrument set fixed and compares in-sample with leave-out fits, so what it diagnoses is that projection, not the separately constructed scalar index; that the single leave-out index Δl_p adequately summarizes the examiner variation in grant

[72] JIVE lies below UJIVE because, unlike UJIVE, it does not leave the application out of the high-dimensional control projection and therefore over-corrects; UJIVE is the appropriate benchmark with many controls (Kolesár, 2013).

Table E.2: *Many-instrument robustness of the first stage*

	2SLS	JIVE	UJIVE
First-stage signal (relative to 2SLS)	1.000	0.937	0.965

Notes: The table reports the first-stage signal $\sum_i \hat{D}_{-i} D_p$, the (leave-out) covariance of the fitted treatment with the realized grant, for the examiner-dummy instrument set, normalized to the in-sample 2SLS value. 2SLS uses in-sample fitted values; JIVE (Angrist et al., 1999) and UJIVE (Kolesár, 2013) use leave-one-out fitted values, with UJIVE additionally jackknifing the control projection. Leverages are computed by Johnson-Lindenstrauss random projection (Kline et al., 2020). Sample: publication years 2005–2020; $n = 5,367,639$; 12,706 examiner instruments; 10,270 art-unit × year cells; mean leverage $K/n = 0.004$; joint first-stage $F = 29.8$. Absolute first-stage signal: 2SLS 59,107, JIVE 55,373, UJIVE 57,066. Controls W as defined in Section 5.3.

rests on the agreement between the scalar-index and examiner-dummy modified-outcome profiles reported elsewhere in this appendix, not on this ratio.

The level of the complier share: a conservative floor. The diagnostics above concern the *composition* of the complier pool: the 96.6% bounds the many-instrument overfit of the examiner-dummy projection behind it, and the profile-agreement checks support summarizing that variation with the one-dimensional index. The *level* of the aggregate complier share is a separate question, because the leave-out index Δl_p is a noisy proxy for an examiner's true leniency: it averages a finite number of prior decisions, so classical measurement error attenuates the first-stage slope, and instrumenting the index with the examiner dummies (UJIVE) removes that attenuation ($\hat{\psi}$ rises from 0.67 to 1.06). The aggregate share decomposes as

$$\text{complier share} \;=\; \frac{\psi \cdot \bar{d}}{\Pr(D_p = 1)}, \tag{25}$$

the product of the first-stage pass-through ψ and the average positive leniency dose $\bar{d} = \mathbb{E}[(\Delta l_p)_+]$, divided by the grant rate ($\Pr(D_p{=}1) = 0.71$). Whether each input is measured with, or corrected for, noise brackets the share:

- **Floor,** 4.7% – pass-through and dose both in measured units ($\psi = 0.67$, $\bar{d} = 0.050$). The attenuated, correction-free reading used as the headline throughout the paper.
- **Ceiling,** 7.5% – the pass-through deattenuated by the examiner-dummy instrument (UJIVE, $\psi = 1.06$), with the dose left in measured units. The largest internally defensible value.
- **Midpoint,** 6.0% – both objects in true-leniency units: the deattenuated pass-through with a dose shrunk to $\bar{d} = 0.050 \cdot \sqrt{\lambda} = 0.040$, where $\lambda = \hat{\psi}_{\text{OLS}}/\hat{\psi}_{\text{UJIVE}} = 0.63$ is the estimated reliability. The estimand-consistent point estimate.

The square root in the midpoint is where the two corrections differ. Reliability is a *variance* ratio, $\lambda = \text{Var}(l^{\text{true}})/\text{Var}(l^{\text{obs}})$, and it is λ itself that attenuates a regression coefficient, which is why the pass-through is deattenuated by dividing by it. The dose is not a coefficient but a *scale*: for a mean-zero

deviation, $\bar{d} = \mathbb{E}[(\Delta l)_+] = \mathbb{E}\,|\Delta l|/2$, and under normality this equals $\sigma_{\Delta l}/\sqrt{2\pi}$, proportional to a standard deviation, so classical measurement error within a scale family inflates it by $\sqrt{1/\lambda}$ and the correction is $\sqrt{\lambda}$, not λ. Applying λ to both would double-count the correction on the dose. The midpoint is therefore the estimand-consistent point under normality of the recentered deviation and a classical, scale-family error; the floor requires neither assumption. The composition results of Section 6 are invariant to this choice, because the pass-through scales every point of the quality and scope profiles equally; only the level moves. I adopt the 4.7% floor as the headline so that no conclusion rests on a measurement-error correction, and note that under the stated assumptions the correction can only raise the estimate. Even the floor is close to four times the 1.2% benchmark of de Rassenfosse et al. (2020) ($4.7/1.2 = 3.9$), and the corrected readings are five to six times it. Both numbers measure inconsistency in the application of one office's own standard, but they are not the same estimand, and the gap has a construction-based reading rather than a substantive one. Theirs is a rate of *detected* disagreement. A granted US patent enters their numerator only if a twin application was refused at another office, so the rate is bounded above by the share of families with mixed grant–refusal outcomes, 25.2% of US grants in their sample, and a marginal US grant whose twins were all granted contributes nothing to it by construction. My estimate counts grants that flip with the identity of the examiner drawn inside an art unit, which does not require the outcome to be observed a second time. I therefore read their figure as a floor on the same phenomenon rather than as a competing estimate of it. I do not rest this on sample selection: their twin applications are of higher value than the average filing, but the authors' own comparison of the estimated inventive-merit distributions across offices finds selection to be immaterial, and they argue that population rates are unlikely to be much higher than their sample estimates.

F Instrument validity: monotonicity and complier balance

The examiner-leniency design requires that assignment be as-good-as-random within art unit × year (addressed in Section 5), that leniency affect outcomes only through the grant decision (exclusion), and that no application is granted by a stricter examiner but rejected by a more lenient one (monotonicity, or "no defiers"). The joint test of exclusion and monotonicity of Frandsen et al. (2023) requires a second-stage outcome and is therefore reported for the Sampat and Williams (2019) and Farre-Mensa et al. (2020) subsamples in Appendix J (p-values 1.000 and 0.570). This appendix reports the diagnostics available in the first-stage-only universe of applications: a direct monotonicity check and a balance test on predetermined covariates.

Monotonicity. Under monotonicity, the first-stage effect of leniency on grant must be non-negative in every subpopulation; a negative first stage in a well-populated subgroup would signal defiers. I provide three pieces of evidence. First, re-estimating the first stage within subgroups, the sensitivity of grant to recentered leniency is positive across all sixteen publication years in the sample (range 0.86–1.02) and across all one hundred quality percentiles (range 0.78–0.93); it is also non-negative in 97.9% of the 657 art units, with fourteen negative point estimates (the fifth percentile across art units is 0.00). Table F.1 summarizes. The minimum in that table, −55.75, should be read as precision rather than as evidence of defiers: the estimate is indistinguishable from zero and from most positive values, and it is the only one of its size, since the next most negative estimate is −2.60. Of the fourteen negative point estimates, three are individually significant at the five percent level. The joint assessment is formal, and its two halves point in different directions. Art units are disjoint subsamples, so the 632 studentized slopes with non-degenerate standard errors are independent; the count of significant negatives, three, is then significantly fewer than the sixteen that chance would produce were every true slope exactly zero ($p = 10^{-4}$), so negative first stages are rare rather than systematic. The minimum studentized slope, however, is −5.3, against an expected minimum of −3.2 across 632 independent tests of an exact zero, so non-negativity in every unit is rejected ($p < 10^{-4}$): in at least one art unit, and at most a handful, the local first stage genuinely runs the wrong way, and not all of the negative estimates are imprecise. What the design requires is the average-monotonicity condition, whose test follows, not unit-level exactness. The scalar-type diagnostic below carries the same reading: monotonicity is an approximation holding almost everywhere, not an exact restriction. Art units enter this table only if the subgroup estimator returns a non-degenerate fit, which is why 657 of the 755 units in the sample appear. Second, I implement a studentized sup-test of the non-negativity of the binned first stage (Frandsen et al., 2023): applications are sorted into ten bins and the UJIVE first-stage slope is tested for non-negativity in every bin. The test does not reject monotonicity whether the bins are formed on my quality index (minimum studentized slope 29.6 against a critical value of 2.8; $p = 1.00$) or on the Kelly et al. (2021) score ($p = 1.00$). Third, I test the scalar-type assumption directly. A two-dimensional examiner, lenient on quality but strict on scope, would generate defiers on the scope margin even where the average condition holds within coarse subgroups. Within each art unit × period cell, I split every examiner's caseload at the cell median of filed scope (average words per claim; broad = fewer words), compute the examiner's grant rate separately on the two halves, and compare the

within-cell ranking of examiners implied by the broad half against that implied by the narrow half. The benchmark is a placebo that randomly re-halves each examiner's caseload at exactly the scope split's half sizes, so the sampling-noise structure of the two grant-rate estimates is identical by construction and the placebo correlation is the agreement that noise alone permits. Table F.2 reports both correlations under four-year and two-year period definitions. The rankings agree at 96 percent of the noise ceiling, and the gap of 0.03 does not move when the period is halved, which points away from within-cell examiner drift as its dominant source without proving the residual is zero. The difference is precisely estimated ($z = -12.8$, clustered on art unit), so an exactly scalar type is rejected at this sample size. Reading the magnitude as an upper bound on any scope-specific dimension rests on a sign assumption: residual within-cell drift, and any other component shared across the split, loads on the scope split and not on the placebo, inflating the gap rather than masking it. Under that assumption the 0.03 is a ceiling; without it, an estimate. What the design requires is that defiers on the scope margin be rare, and a ranking displacement of three to four percent of the noise ceiling is the scalar approximation holding to within a few percent.

Table F.1: *Monotonicity: first-stage sensitivity across subgroups*

Subgroup partition	# subgroups	Min	Median	Max	Share ≥ 0
Publication year	16	0.86	0.90	1.02	100%
Quality percentile	100	0.78	0.86	0.93	100%
Art unit	657	−55.75	0.70	2.36	97.9%

Notes: First-stage sensitivity of grant to recentered leniency Δl_p, re-estimated within each subgroup. Under monotonicity the coefficient is non-negative in every subgroup. Partitions that condition on the instrument itself (e.g. leniency percentile) leave no residual variation and are uninformative for this check.

Table F.2: *Scalar leniency: examiner rankings across the scope margin*

	4-year cells	2-year cells
Rank correlation, scope split	0.756	0.680
Rank correlation, placebo split (noise ceiling)	0.782	0.710
Agreement ratio	0.966	0.959
Difference (s.e.)	−0.027 (0.002)	−0.029 (0.002)
z	−12.8	−12.8
Cells	1,861	2,864
Art units	577	557
Examiner-cells	30,893	41,398

Notes: Within each art unit × period cell, every examiner's decided applications are split at the cell median of filed scope (average words per claim; broad = fewer words), and the examiner's grant rate is computed separately on each half. Reported are the caseload-weighted means of the within-cell Spearman correlations between the examiner rankings implied by the two halves. The placebo split randomly re-halves each examiner's caseload at the same half sizes (twenty draws), so its correlation is the ranking agreement that sampling noise alone permits. Examiners qualify with at least twenty decisions per half and a non-degenerate overall grant rate; cells qualify with at least eight qualifying examiners. Weights are qualifying examiners per cell; the standard error of the difference is clustered on the art unit. Sample: publication years 2005–2020, 5,367,642 applications.

Complier characteristics. Random assignment of applications to examiners within art unit × year, the balance of the instrument itself against predetermined covariates, is established in Section 5. Here I characterize the compliers directly using the modified-outcome estimator of Borusyak et al. (2023), which recovers the mean of any characteristic among compliers, $\mathbb{E}[X \mid \text{complier}]$. On the two central margins the complier mean confirms the main-text profile: compliers are of essentially average invention quality ($\mathbb{E}[q \mid \text{complier}] = 0.50$ against a population mean of 0.51) but of distinctly broader filed scope (0.53 against 0.51). On a predetermined firm characteristic, whether the applicant is publicly listed, compliers differ significantly from the granted population: 34.5% of compliers are listed firms versus 28.7% of all granted applications, a gap of 5.8 percentage points ($z = 17.7$). Table F.3 collects these complier means.

Table F.3: *Complier characteristics: means versus the granted population*

Characteristic	Reference	Compliers	Gap (z)
Invention quality (mean)	0.51 (pop.)	0.50	−4.5
Filed scope (mean)	0.51 (pop.)	0.53	16.1
Publicly listed (share)	0.287 (granted)	0.345	17.7

Notes: Complier means via the UJIVE modified-outcome estimator of Borusyak et al. (2023). Each mean is the efficient pooling of the treated- and untreated-arm estimates of the same complier mean; z is the studentized pooled-mean-minus-reference gap, with standard errors clustered on art unit × publication year. The reference column is the population mean for quality and scope and the granted-sample mean for public listing. Quality and scope are the paper's central margins (Section 6); public listing is a predetermined firm characteristic, defined only among granted applications, so its row is the treated-only coverage estimand on the same UJIVE variation; no untreated arm exists to pool for it. The quality gap is precisely estimated and small: 0.6 percentile points, the object Section 6.2 reports.

Distributional balance and caseload composition in measured leniency. Table 1 tests balance through a single linear projection of leniency on quality and filed scope. The complier estimator, however, is identified from the *bin-conditional* mean of the instrument, so it is worth testing that null directly. Let B index a bin of the filed-scope percentile and let $g_B = \mathbf{1}[s \in B]$ be the corresponding modified outcome. The treated and untreated arms of the estimator are

$$\hat{\theta}_T(B) = \frac{\overline{\Delta l \, g_B D}}{\overline{\Delta l \, D}}, \qquad \hat{\theta}_U(B) = \frac{\overline{\Delta l \, g_B (1-D)}}{\overline{\Delta l \, (1-D)}}, \tag{26}$$

both of which identify the complier density $f_c(B)$ when assignment is as-good-as-random, so their difference is a per-bin balance test. Because Δl is mean zero by construction, that difference has the closed form

$$\hat{\theta}_T(B) - \hat{\theta}_U(B) \;=\; \frac{\Pr(s \in B)\, \mathbb{E}[\Delta l \mid s \in B]}{\overline{\Delta l \, D}}, \tag{27}$$

so the statistic is the bin-conditional instrument mean, rescaled by the first stage. Two features of Equation 27 matter for reading it. The rescaling is by $\overline{\Delta l \, D} \approx 0.0095$, which magnifies the underlying imbalance roughly a hundredfold; and since $\sum_B \Pr(B)\, \mathbb{E}[\Delta l \mid B] = \mathbb{E}[\Delta l] = 0$, the per-bin deviations sum to zero and are therefore not independent tests.

This diagnostic rejects on the scope axis and not on the quality axis, and the contrast survives the

clustering used in Table 1: clustering two-way on customer number × art unit, eleven of twenty scope bins reject (maximum $|z| = 4.1$) against one of twenty quality bins (maximum $|z| = 2.0$). Clustering instead on art unit × publication year gives thirteen and six. Panel A of Table F.5 reports both. The asymmetry has a mechanical explanation. Measured leniency is examiner j's realized grant rate, $l_j = n_j^{-1} \sum_{i \in j} D_i$, and the grant rate falls steeply in filed scope (0.769 to 0.578 from the narrowest to the broadest bin, against 0.701 to 0.714 across quality). An examiner whose caseload happens to be scope-heavy therefore records

$$\mathbb{E}[l_j \mid \bar{s}_j] \;=\; \lambda_j \;-\; b\,(\bar{s}_j - \mathbb{E}[\bar{s}]), \qquad b > 0, \tag{28}$$

where λ_j is the examiner's true strictness and $\bar{s}_j$ the mean filed scope of the caseload: they are *measured* as less lenient without being stricter. Caseloads are persistent in scope (cross-half correlation of the examiner's mean filed scope within a cell, 0.35), so an application's own scope is positively correlated with its examiner's caseload mean, and $\mathbb{E}[\Delta l \mid s]$ inherits a negative slope in s. Empirically the bin-level imbalance correlates 0.97 with the bin grant rate. A characteristic thus breaks distributional balance only if it is *both* caseload-persistent and grant-shifting. Filed scope is both. Quality is if anything more caseload-persistent (0.84) but nearly flat in grants, rising by a point where scope falls by nineteen. I conclude that sorting alone does not generate the asymmetry.

Sign of the bias. The consequence for the paper's central object follows from decomposing the representation ratio. With a mean-zero instrument,

$$\rho(B) \;=\; \underbrace{\frac{\mathrm{Cov}(\Delta l,\, D \mid B)}{\overline{\Delta l}\;\overline{D}}}_{\text{complier mass}} \;+\; \underbrace{\frac{\mathbb{E}[\Delta l \mid B]\;\Pr(D = 1 \mid B)}{\overline{\Delta l}\;\overline{D}}}_{\text{contamination}}. \tag{29}$$

The contamination term carries the sign of $\mathbb{E}[\Delta l \mid B]$, which is negative in broad-scope bins and positive in narrow ones. Estimated complier mass is therefore displaced down exactly where the complier share is highest, and

$$\operatorname{plim} \partial_s \hat{\rho}(s) \;<\; \partial_s \rho(s)\,: \tag{30}$$

the estimated scope gradient is attenuated. Contamination of this form cannot manufacture the rising profile documented in Section 6.2; it can only flatten it.

Verification. I confirm the sign two independent ways. Netting the contamination term of Equation 29 out algebraically turns the profile measured with the paper's instrument, which runs from 0.983 to 0.911 over the $p2.5$–$p97.5$ range of filed scope (a *falling* 0.93×), into one running from 0.838 to 1.087 (a rising 1.30×). Alternatively, rebuilding leniency from scratch as a leave-one-out examiner mean of $D_i - \hat{g}(s_i)$, grants purged of the caseload's scope composition, collapses the bin-level imbalance correlation from 0.86 to 0.40 while leaving quality unchanged (0.76 to 0.77), and steepens the profile from 1.26× to 1.50×. Every figure in this comparison is measured against the same unpurged leave-one-out benchmark, so the starting correlation is its 0.86 rather than the 0.97 of the paper's own instrument reported above. Compliers among

granted widen accordingly, from 3.87–6.50% to 3.48–6.97% across the scope range. Panel B reports three levels, and only the first is the paper's. Evaluated at Equation 25, my own instrument gives $\bar{\psi} = 0.666$ and an aggregate complier share of 4.7%, the headline floor of Section 6.1, recovered here as a check rather than reasserted. The two leave-one-out rows sit at $\bar{\psi} \approx 0.82$ not because the attenuation discussed in Section 5.3 has been undone, but because they are a different instrument on a different scale; their levels are therefore not comparable to the paper's, and what the pair identifies is the difference the purge makes. That difference is negligible: $\bar{\psi}$ moves from 0.8227 to 0.8217 and the implied share from 5.27% to 5.26%. The quality profile is untouched throughout, declining by a factor of 0.96 both before and after the purge. Contamination of measured leniency is thus a statement about complier *composition*, not about the level, and it runs against the paper's finding rather than toward it. Panel B of Table F.5 collects these comparisons.

One limit confines the purged numbers to evidence on the *direction* of a bias. The exercise is run on a scalar recentered-leniency index Δl, and that is not the estimator behind the paper's headline profile: Figure 4 is estimated from the full set of examiner dummies in UJIVE form (Section 5.3), which uses the examiner variation jointly rather than collapsing it into one index. The two therefore weight examiners differently, and no equivalence result links the weights the purged scalar index implies to those UJIVE delivers. That gap closes by computation rather than by a theorem: rebuilding the UJIVE instrument itself from the purged grants, on the same design matrices, the same leave-out construction, and the same Johnson–Lindenstrauss draws, so that the two instruments differ by the purge and by nothing else, steepens the examiner-dummy scope profile from 1.13× to 1.30× over the same p2.5–p97.5 range and moves the quality profile only from 1.00× to 1.01×. The conservative direction of the bias is therefore a property of the headline design, not of the scalar convention. The purge still does not produce a replacement value for the gradient, and the levels reported in either column are the purged instrument's own rather than the paper's.

Finally, the magnitude should be read in instrument units rather than in the units of Equation 27. The bin-conditional imbalance $\mathbb{E}[\Delta l \mid B]$ reaches at most 0.0028 against a standard deviation of Δl of 0.119, or 2.3 percent of a standard deviation. At $n = 5.4$ million applications it is precisely estimated rather than large.

The conditional leniency dose. Section 5.3 factors the conditional complier share into a scalar dose and a shape, $\psi(x)\,\bar{d} = \bar{\psi}\,\rho(x)\,\bar{d}$, and this is legitimate only if the dose does not itself vary with x. What licenses the factorization is the scalar-index structure of Section 5.2: with leniency entering the grant rule additively and separably, ψ depends on (q, s) only through the acceptance index, and the dose is a property of the assignment process rather than of the application. That is not an innocuous assumption. Were the dose to vary with x, $\rho(x)$ would not be carrying all the conditional variation and the product would misstate the profile's *shape*, not merely its level. The condition is testable and I test it here. I sort applications into twenty equal-mass bins on each conditioning axis, the construction the complier profiles use, and compute the bin-conditional dose, clustering standard errors and the joint test on art unit × year. Two dose statistics are reported because two are in circulation: the object actually pulled out of Equation 12 is the *one-sided* favorable deviation $\mathbb{E}[(\Delta l)_+ \mid X = x]$, while the two-sided $\mathbb{E}[|\Delta l| \mid X = x]$ is

the natural flip-rate analogue and is twice as large under mean-zero recentering. The error the factorization commits at x is exactly $\bar{d}(x)/\bar{d} - 1$, so the table reports that ratio in percent.

The construction reproduces the pipeline it is meant to audit: the unconditional $\mathbb{E}[(\Delta l)_+] = 0.05028$ against the 0.0503 used throughout, and $\mathbb{E}[\Delta l] = -6 \times 10^{-6}$ confirms the recentering.

Table F.4: *Conditional leniency dose: deviation from the pooled dose*

		Deviation from pooled dose (%)			
Axis	Dose statistic	Max abs.	Max − min	SD across bins	Wald $\chi^2(19)$
Quality	$(\Delta l)_+$	3.03	4.39	1.14	54.0
Quality	$\lvert\Delta l\rvert$	2.12	2.79	0.70	149.8
Filed scope	$(\Delta l)_+$	2.02	3.50	1.08	116.5
Filed scope	$\lvert\Delta l\rvert$	3.18	5.95	1.73	199.8

Notes: Twenty equal-mass bins per axis on the estimation sample (n = 5,367,639 applications in 10,270 art unit × year clusters). Entries are the bin-conditional dose expressed as a percentage deviation from its pooled value, which is exactly the error the factorization $\psi(x)\,\bar{d}$ commits at that bin. The Wald statistic tests equality of the dose across all twenty bins, clustered on art unit × year; the five percent critical value is approximately 31.

Two things follow. First, *the joint test rejects on every row and that fact is not informative.* At five and a third million applications a Wald statistic is a statement about precision, not about magnitude, and the same trap recurs wherever this sample is tested bin by bin. What the table establishes is the size of the approximation: the conditional dose is flat to within roughly three percent of itself on either axis, and the spread across bins is around one percent. The scalar pull-out is an approximation whose error has now been measured rather than assumed away.

Second, the small residual variation runs in the direction of the caseload-composition bias discussed above. The dose *falls* in quality, from about +3.0% of the pooled value in the bottom decile to −1.4% in the top. Across filed scope it *rises* through the eighteen interior bins, from about −1.5% to +2.0%, with both endpoint bins reverting, so it should not be described as monotone. Because the complier profile also rises in filed scope, holding the dose fixed at its pooled value understates the profile's slope by roughly a tenth of that slope: the three-and-a-half-percent dose movement multiplies the profile's *level*, so its effect on the slope is the level-to-slope ratio times larger than the dose movement itself. As with the caseload composition analyzed above, the approximation attenuates the scope gradient rather than manufacturing it.

Table F.5: *Distributional balance and caseload composition in measured leniency*

	Filed scope	Invention quality
Panel A: preconditions and the balance test		
Caseload persistence (split-half r)	0.35	0.84
Grant rate, lowest → highest bin	0.769 → 0.578	0.701 → 0.714
Bins rejecting, cluster art unit × year	13/20	6/20
max $\lvert z \rvert$	9.1	4.1
Bins rejecting, two-way cluster	11/20	1/20
max $\lvert z \rvert$	4.1	2.0
Max $\lvert \mathbb{E}[\Delta l \mid B] \rvert$, in sd of Δl	0.023	0.022
corr(imbalance, bin grant rate)	0.97	0.21
Panel B: consequence, by leniency measure		
	$\bar{\psi}$ / share (%)	scope ρ, p2.5 → p97.5
Paper instrument Δl	0.6665 / 4.72	0.983 → 0.912 (0.93×)
Leave-one-out grant rate	0.8227 / 5.27	0.794 → 1.004 (1.26×)
Leave-one-out, scope-purged	0.8217 / 5.26	0.716 → 1.077 (1.50×)
Algebraic netting-out	—	0.838 → 1.087 (1.30×)
Quality profile ρ, same three measures		
Paper instrument Δl	—	1.113 → 0.855 (0.77×)
Leave-one-out grant rate	—	0.996 → 0.953 (0.96×)
Leave-one-out, scope-purged	—	0.989 → 0.952 (0.96×)

Notes: Distributional balance of the examiner instrument on predetermined characteristics, and the caseload-composition mechanism behind the filed-scope rejection. Sample: 5,364,971 applications, publication years 2005–2020. 20 equal-mass bins per axis. *Panel A*. Caseload persistence is the cross-half correlation of the examiner mean characteristic within an art unit × year cell (66,568 examiner × cell units with at least 15 applications per half); it measures whether examiner identity predicts the caseload. The balance statistic is the treated-minus-untreated contrast of Equation 26, which equals the bin-conditional instrument mean rescaled by the first stage (Equation 27); rejection is at the five percent level. Two-way clustering is on `customer_number` × `examiner_art_unit`, matching Table 1. The imbalance is also reported in standard deviations of Δl (sd = 0.119), since the rescaling in Equation 27 magnifies it by roughly a hundredfold. A characteristic breaks balance only if it is both caseload-persistent and grant-shifting: filed scope is both, quality is more persistent but grant-flat. *Panel B*. $\bar{\psi}$ is the pooled first stage, estimated after partialling out the art unit × year fixed effects and covariates, and the share is $\bar{\psi}\,\mathbb{E}[(\Delta l)_+]/P(D=1)$ (Equation 25). Following that convention the dose $\mathbb{E}[(\Delta l)_+]$ is taken on the recentered but unresidualized instrument; the share is invariant to rescaling the instrument, but not to residualizing it, so the convention has to be stated. On this convention the first row reproduces the paperś headline floor. ρ is the representation ratio $f_{\text{complier}}/f_{\text{pop}}$ over the p2.5–p97.5 range. The scope-purged instrument is a leave-one-out examiner mean of $D_i - \hat{g}(s_i)$; the algebraic netting-out instead removes the contamination term of Equation 29 from the paper instrument. Both routes steepen the scope profile and leave the level and the quality profile unchanged.

A pre-treatment anchor for the expected-leniency benchmark. The recentred instrument Δl_p subtracts from the assigned examiner's leniency the mean leniency of the pool the application could have drawn. In the shipped construction that pool is assembled around the application's *decision* date, within a symmetric ±18-month window. Two features of this construction matter. The decision date is itself moved by leniency, since a more permissive examiner disposes of a case sooner, so the benchmark is not strictly predetermined; and because the window is symmetric, part of the pool is drawn from examiners

deciding up to eighteen months *after* the anchor, which no applicant could have observed ex ante.

I therefore rebuild the benchmark on an anchor that the examiner draw cannot move. The application's publication date is scheduled by statute at roughly eighteen months from the priority date, so it is fixed independently of who examines the case. Anchoring on it, restricting the window to the eighteen months *preceding* publication, pooling within the art unit that the design randomizes inside of, and leaving out the applicant's own examiner yields a pre-treatment instrument defined for 5,333,672 applications, 99.4% of the core sample, with a median pool of 909 decisions. The two instruments correlate at 0.93. Table F.6 reports both.

The first stage strengthens under the re-anchoring: the grant response per standard deviation rises from +0.0996 to +0.1148, a factor of 1.15. The pre-treatment instrument is therefore usable, and the objection is a robustness question rather than a threat to identification. It does not, however, explain the residual association between measured leniency and the text components documented above. That association is slightly larger on the pre-treatment instrument, not smaller, so re-anchoring can be ruled out as its source; the caseload composition analyzed earlier in this appendix remains the operative explanation.

The composite-quality row is the one a reader will hold against the balance claims of Sections 3.4 and 5.4, and the reconciliation is a decomposition rather than a correction. The −0.0380 is estimated on an object and a conditioning set the design never uses: the raw level of the composite index, with the fixed effects alone. Re-expressing the outcome as the within-cell percentile rank, the object every design exhibit conditions on, takes the association to −0.0106 (0.0026), so nearly three quarters of it is variation in the index's level that the rank never transmits; adding the control set of the first stage takes the rank to +0.0008 (0.0027). On the pre-treatment instrument the same two steps run −0.0522 to −0.0139 (0.0028) to −0.0024 (0.0030). Under the design's own object and conditioning there is no association left to explain, on either anchor (Table F.6, bottom panel). What the conditioning does not remove is the component loadings: impact and novelty stay individually associated with the instrument under the full control set (+0.0251 and −0.0303 per standard deviation on the shipped instrument, with the same signs and slightly larger magnitudes on the pre-treatment one), which is the caseload-composition channel above, and the reason the balance claims of the main text are scoped to the rank rather than to the index's parts.

Three properties of the rebuilt instrument should be stated by anyone using it. First, a backward-looking window under-estimates contemporaneous leniency because grant rates drift upward over the sample, so Δl_p built this way has mean +0.012 rather than zero; it is absorbed by the art unit × year effects but it is not recentred in the sense the main text uses. Second, measured leniency is not an examiner-level constant. Only 1.6% of the 12,709 examiners appearing in the leniency data carry a single value (12,706 of them enter the connected estimation core of Section 5), so any pool average is a weighted average over examiner-periods rather than over examiners. Third, the publication anchor is not universally pre-decision: for 3.6% of applications the decision precedes publication, and for those the anchor is predetermined by statute but not by timing.

Table F.6: *Expected-leniency benchmark: decision-date versus pre-treatment anchor*

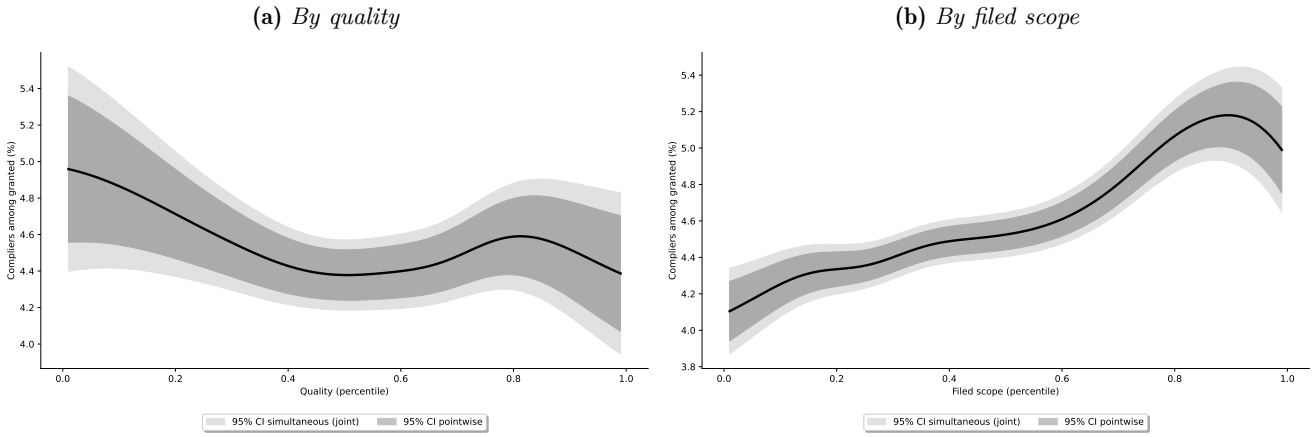
	Shipped Δl_p (decision anchor)	Pre-treatment Δl_p (publication anchor)	Ratio
Standard deviation	0.129	0.144	
Grant *(first stage)*	+0.0996	+0.1148	1.15
	(0.0006)	(0.0004)	
Impact component	+0.0145	+0.0149	1.03
	(0.0017)	(0.0018)	
Novelty component	−0.0231	−0.0269	1.16
	(0.0017)	(0.0018)	
Composite quality	−0.0380	−0.0522	1.37
	(0.0044)	(0.0055)	
Composite quality, reconciled to the design conditioning			
as within-cell rank (design object)	−0.0106	−0.0139	1.31
	(0.0026)	(0.0028)	
with the first-stage control set	−0.0242	−0.0395	1.63
	(0.0046)	(0.0058)	
rank and control set (design conditioning)	+0.0008	−0.0024	—
	(0.0027)	(0.0030)	

Notes: Each cell is the coefficient from a separate regression of the outcome on the recentered instrument, with art unit × publication year fixed effects absorbed and standard errors clustered on the same cell, expressed per standard deviation of the instrument in use. All outcomes other than the grant indicator are standardized. The shipped instrument centres the expected-leniency pool on the applications decision date within a symmetric ±18-month window. The pre-treatment instrument centres it on the statutory publication date, uses only the preceding 18 months, pools within art unit, and leaves out the applicants own examiner; it is defined for 5,333,672 applications, 99.4% of the core sample, with a median pool of 909 decisions. Estimation sample 5,333,618 applications after singleton fixed effects are dropped. The two instruments correlate at 0.93. Ratio is the pre-treatment coefficient over the shipped one. The backward-looking window under-estimates contemporaneous leniency because grant rates drift upward over the sample, so the pre-treatment instrument has mean +0.012 rather than zero; it is absorbed by the fixed effects but is not recentered in the sense used in the main text. The final three rows re-express the composite-quality association on the designs own terms: as the within-cell percentile rank that every design exhibit conditions on, with the control set of the first stage (entity status, continuation type, examiner experience, first-year indicator) added, and with both. They are estimated on the same sample with the same fixed effects and clustering (singleton cells retained, 5,333,672 applications); a ratio of two near-zero coefficients is noise and is not reported.

G Robustness: sample restrictions

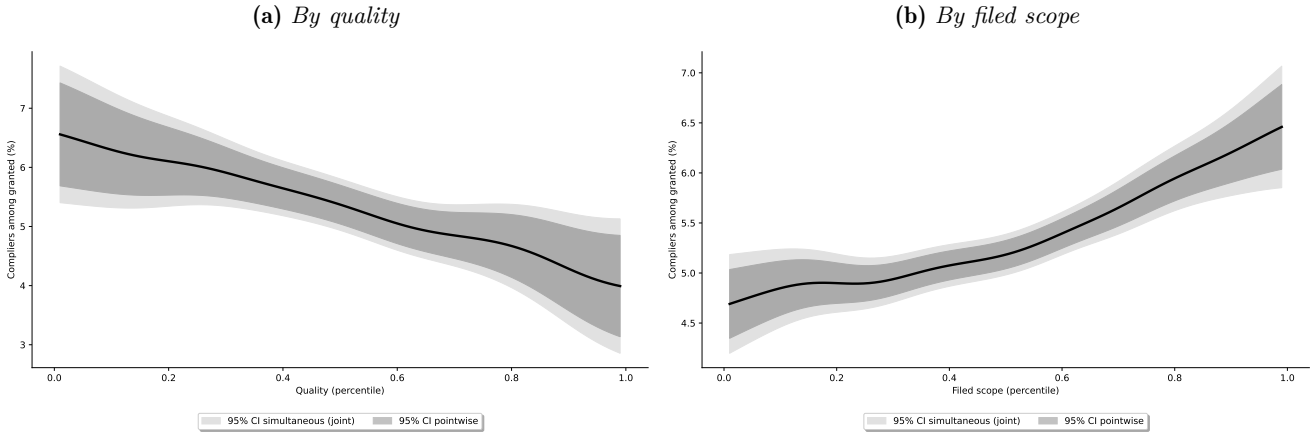
This appendix collects robustness checks that re-estimate the complier composition of Section 6 on alternative subsamples. Each exhibit replicates Figure 4, the conditional share of compliers among granted patents by quality and by filed scope, under a different sample restriction, holding fixed the examiner-dummy UJIVE modified-outcome estimator (Section 5.3), the standard errors, and the confidence bands. Across every restriction the active margin of Section 6 is preserved: the share rises in filed scope. The quality profile is flat excluding continuations (joint flatness $p = 0.103$) and declines in the random-digit cells, so on no restriction do marginal grants concentrate in high quality; no conclusion about where the lottery binds depends on the baseline sample.

Figure G.1: *Robustness: complier shares excluding continuations*



Notes: Replication of Figure 4 on the subsample of non-continuation applications (2,145,177 observations). Continuation applications, which inherit prosecution history from a parent filing and are frequently assigned to the same examiner, are excluded. Estimation follows the complier pipeline of Figure 4; shaded areas show 95% pointwise and 95% simultaneous confidence bands.

Figure G.2: *Robustness: complier shares in random-digit art units*



Notes: Replication of Figure 4 restricted to random-digit art unit × year cells (1,199,154 observations): the art unit and docket-year pairs in which supervisors allocated applications to examiners by the terminal digit of the application serial number, which is assigned sequentially and is therefore unrelated to the content of the application (Sampat and Williams, 2019). The cells are identified from the list used in that literature; an application is assigned to a cell by its publication year, offset by the statutory eighteen-month publication lag (publication year equal to the docket year plus one or two). Within these cells examiner assignment is random by construction, which is why they are commonly used as a robustness benchmark in the examiner leniency literature. Estimation, standard errors, and confidence intervals as in Figure 4.

The random-digit cells are not a random subset of the estimation window, and part of the difference between their quality profile and the baseline's is the difference their composition predicts: the matched-twin exercise below traces the elevated bottom edge to the era and technology mix, while the upper-half decline is specific to the units themselves. The cells concentrate in electrical and mechanical technology groups, art units in the 2800, 3600, and 3700 series make up 85% of the subsample against 51% of the window, and in earlier cohorts, with a median publication year of 2010 against 2013. Within them, the two descriptive gradients that generate the baseline's quality profile behave differently. The acceptance-quality profile peaks in the middle of the distribution instead of rising throughout, and high-quality applicants do

not file broader claims there, so the scope channel that lifts the baseline's total profile over its upper half (Section 6.3) has nothing to work with. Proposition 1 operates at fixed (q, s), so it reaches the total profile only through the scope composition: with that channel absent, the decline the direct quality premium imposes at fixed scope is no longer offset, consistent with the lower, declining profile the figure shows. The reading is consistency with the mechanism rather than a derivation from it – the cells' mid-peaked acceptance profile would not by itself deliver a monotone decline pointwise, so the marginal shape also carries how filings distribute across scope within these units. The instrument is not what differs, as the dispersion and dose of the leniency draw are nearly identical inside and outside the cells (a mean absolute deviation of 0.103 against 0.100). A composition-matched twin separates the two candidate explanations: re-estimating the identical design on non-random-digit applications drawn to the same technology-group $\times$ publication-year composition (0.9 million applications; the match binds where the random-digit cells dominate a stratum) reproduces the elevated bottom edge, 6.4% against the random-digit cells' 6.8%, but not the decline: over the upper half the twin rises to 6.0% in the top bin where the random-digit cells fall to 3.9%. The match holds technology group $\times$ publication year and nothing finer, so unit-level differences inside those strata, the art-unit mix among them, remain unbalanced: the twin is evidence consistent with the scope-channel reading, not an isolation of that channel from every remaining alternative. Table G.1 reports both twenty-bin profiles. The declining direction itself only reinforces the conclusion the profile serves: on no restriction do marginal grants concentrate among high-quality inventions.

Table G.1: *Random-digit cells against their composition-matched twin*

Quality bin	Random-digit cells		Composition-matched twin	
	share (%)	s.e.	share (%)	s.e.
1	6.77	0.52	6.43	0.51
2	6.44	0.44	5.44	0.40
3	6.08	0.40	4.73	0.33
4	6.10	0.35	4.70	0.30
5	6.25	0.32	4.50	0.28
6	5.82	0.27	4.29	0.25
7	5.90	0.25	4.61	0.24
8	5.60	0.25	4.30	0.24
9	5.62	0.23	4.44	0.24
10	5.40	0.22	4.52	0.21
11	5.42	0.23	5.06	0.21
12	5.02	0.23	4.81	0.22
13	5.02	0.24	4.76	0.24
14	4.68	0.24	4.93	0.25
15	4.92	0.26	5.14	0.27
16	4.66	0.30	5.41	0.28
17	4.71	0.33	5.40	0.31
18	4.49	0.38	5.52	0.32
19	3.92	0.45	5.79	0.37
20	3.88	0.52	6.01	0.42

Notes: The Table reports the share of compliers among granted patents by twenty equal-mass quality bins, from the examiner-dummy UJIVE modified-outcome estimator of Section 5.3, in the random-digit art unit × year cells and in the composition-matched twin: non-random-digit applications drawn to the random-digit cells' technology-group × publication-year strata and size. Bins in which the conditional grant share falls below five percent are masked, as in the corresponding figures. The twin reproduces the elevated level of the random-digit cells at the bottom of the quality distribution but not their decline over the upper half, which is therefore specific to the random-digit units themselves rather than to their technology-and-era composition.

H Quality measure: construction, validation, and decomposition

This appendix supplements Section 3 with the full discussion of the quality index's three design choices, summarized there in a single paragraph, and with three robustness exercises for the pre-grant quality index: sensitivity to construction choices, convergent validity against external measures, and a decomposition into its novelty and impact components.

Design choice 1: titles and abstracts, not claims. Since claims are the legally binding part of a patent and are the main focus of negotiation during the prosecution process, applicants tend to narrow their claims in response to examiner objections. Therefore, the text of the claims at any stage in the prosecution process

shows the examiner's decisions as much as the underlying invention. Titles and abstracts, however, are fixed in the filing text, reproduced in the first ("A1") publication at the statutory eighteen months, and remain mostly unchanged during the prosecution process. Using abstracts and titles therefore avoids one specific channel of contamination, as the text Q_p is built from cannot have been rewritten in response to the examiner's objections.

Design choice 2: within-cell percentiles. The percentile transformation $Q_p = P(FS_p/BS_p)$ is calculated separately for each art unit × publication year cell, employing the same partition that is used as the fixed effects in the first stage (Section 5). The resulting measure ranks each application relative to the cohort which it was evaluated alongside, taking into account both field-level differences in the average rate of development and the density of prior art, as well as year-to-year changes in either of these. An application with $Q_p = 90$ is in the top 10% of its art unit × year cell as the combination of novelty and impact, not in the top 10% of all USPTO applications. The cost of this within-field construction is in interpretation. Q_p measures quality relative to a technology field rather than on an absolute scale.

Because the index enters every specification through this rank rather than through the level of FS_p/BS_p, any distortion that applies a common strictly monotone transformation to the *ratio* within an art unit × year cell leaves Q_p *exactly* unchanged. Distortions outside this class are the subject of the construction-robustness table below and of the balance tests (Section 5.4 and Appendix F).[73]

Design choice 3: orthogonality to the examiner draw. Decisive for the causal design in Section 5, Q_p must be orthogonal to the examiner draw. This requirement is distinct from the timing of the text; the two are routinely run together, and only the first is load-bearing. Validity as a control does not require that Q_p be knowable at the time of the decision. It requires that the leniency of the examiner assigned not move it. A measure built entirely from information realized after the decision is a valid control if the draw does not shift it, and a measure built entirely from information available before the decision is not, if the draw does. The forward-similarity component FS_p is the part of the construction that could fail this: because it is built from patents filed in the five years after the focal application, the focal grant might itself shape subsequent filings in the art unit, making $\bar{v}_p^+$ endogenous to the examiner decision.

Two tests bear on it, at different levels of aggregation. At the level of the art unit, the exclusion-relevant test is whether leniency moves a unit's forward technological trajectory: regressing the art-unit × year mean forward similarity on the art-unit × year mean leniency with art-unit and year fixed effects, the coefficient is −0.05 standard deviations (s.e. 0.04, 9,896 unit-years; Table H.6). The raw cross-section of art-unit levels is positive, but that is technology composition, fast-moving fields also differing in grant rates, not a trajectory response. And because the leniency measure predicts the art-unit grant rate almost one-for-one, a poorly measured regressor cannot explain the within-unit null. Examiner leniency does not move the forward technological trajectory of an art unit.

[73] Cell-common rescalings of the similarity scale are within the class. Distortions that hit FS_p and BS_p separately need not be, since even a common additive shift to both can reorder a ratio, and truncation of the forward window for the most recent cohorts, bursts in filing volume, and drift in the density of the embedding space are cell-common only approximately. Appendix N draws the same boundary for the rank-defined scope results.

At the level of the individual application, the direct test is whether measured quality is balanced against the variation the design actually uses, and it is: testing the bin-conditional mean of the recentered leniency deviation separately in twenty equal-mass bins rejects in one of the twenty quality bins, the chance level for twenty tests, while the identical procedure rejects in eleven of twenty filed-scope bins, the detected imbalance that makes the quality null informative (Section 5.4 and Appendix F). The balanced object is Q_p itself, the within-cell rank every estimate conditions on; the raw level of the composite index does carry a small, precisely estimated association with the instrument, which Appendix F decomposes into the level-versus-rank gap and the control set of the first stage. Together with the art-unit null, this supports reading Q_p as predetermined with respect to the draw, which is what the design requires and all that it requires.[74]

Construction choices. The index is built from title-and-abstract text rather than from claims, for the fixed-at-filing reason of Section 3, embedded with sentence-T5 (Ni et al., 2021), and expressed as art unit × publication year percentiles (Section 3.4). Table H.1 reports the Spearman rank correlation of the baseline index with the un-percentiled measure: the cell-wise normalization reorders the index only modestly (Spearman 0.85 with the raw measure on 200,000 patents), so the ranking reflects genuine within-field variation rather than an artifact of cross-field level differences. Two further families of construction choices are varied, both of which leave the underlying embeddings intact and change only what each application is compared against: the comparison pool's reference class and anchor (coarsening the focal art unit to the technology center or the workgroup, anchoring on the publication date in place of the filing date, or drawing the pool as a fixed count of nearest-in-time documents rather than a fixed-duration window), and the comparison window's length and placement (six months to ten years, with and without a gap at the anchor). The rank correlations span 0.14 to 0.94, and the range tracks what a variant changes: anchoring, placement, and the normalization leave the ranking nearly intact (0.89–0.94), while it moves with how the comparison pool itself is drawn, its reference class, its membership, and how short its window is – that is, with *who* an application is compared against. That dimension is an object of study rather than a nuisance: reference-class dependence is measured directly, its complier loading the +0.57 percent residual of Table H.4, and under every variant in the family, the grant-conditioned pool excluded as post-treatment, the pooled complier mean on the variant's own rank stays within one percentile point of the population mean, computed in the same sweep behind that table.

[74] Neither test reaches a channel operating within the application rather than across the art unit. Continuation and divisional filings enter the forward window carrying text nearly identical to their parent, so a grant that generates them could raise FS_p for that application alone, which an art-unit average cannot detect. Section 6.2 re-estimates the composition excluding continuations; Appendix N collects the design's other limits.

Table H.1: *Quality measure: sensitivity to construction choices*

Variant	Alters	Rank corr. with baseline
Raw (un-percentiled)	Within-cell normalization	0.886
Publication-date anchor	Comparison-pool anchor date	0.939
Technology-center pool	Comparison-pool reference class	0.334
Workgroup pool	Comparison-pool reference class	0.504
Six-month windows	Comparison-window length	0.243
One-year windows	Comparison-window length	0.443
Three-year windows	Comparison-window length	0.837
Ten-year windows	Comparison-window length	0.817
180-day gap at the anchor	Comparison-window placement	0.936
365-day gap at the anchor	Comparison-window placement	0.926
Nearest-in-time pool (k=200)	Pool construction (fixed count)	0.140

Notes: Spearman rank correlation of each construction variant with the baseline index (filing-date-anchored comparison pool within the focal art unit, expressed as the within-cell percentile), on the estimation window 2005–2020, pairwise-complete per row (N between 5,352,310 and 5,364,971). Pool and window variants leave the underlying embeddings intact and change only what each application is compared against. Each variant is a one-dimension change from the publication-anchored construction, whose own row is the anchor contrast; rows below it therefore differ from the filing-anchored baseline in their stated dimension and the anchor date jointly. The nearest-in-time pool draws the 200 documents closest in time on each side of the anchor in place of the fixed-duration window. The raw row removes the within-cell percentile normalization. A variant whose comparison pool conditions on the grant outcome is excluded as post-treatment.

Convergent validity. The design measure is built to identify the acceptance margin, not to maximize the bulk correlation with realized value. Consistent with this, its *linear* conditional correlation with ex-post value proxies is positive and precisely estimated but small (standardized β between 0.001 and 0.013 for forward citations, renewals, and Kogan et al. (2017) market value). The measure's value content is instead concentrated in the *upper tail*, which is the economically relevant form of convergent validity for a screening measure. Table H.2 reports the value premium by quality bin, the mean of normalized forward citations and Kogan et al. (2017) market value in the top decile, top 5%, and top 1% of the quality distribution relative to the median bin. The premium rises steeply into the tail on citations, from +38% to +75%, and modestly on market value, from +8% to +13%; the steep half of the tail premium is a citation fact, not a price fact, which is what the nearly flat price map of Appendix N reflects. Because these proxies are observed only for granted patents, they condition on the grant selection the index is designed to sidestep; the direction of that selection cannot be signed (Section 3), so the premium is read as a within-granted fact rather than a bound in either direction. The by-revision-round patterns of the same proxies appear in Figure D.2.

Table H.2: *Convergent validity: invention-quality value premium by tail bin*

Invention quality	n	Norm. fwd. citations	Market value (Kogan et al.)
Median (p40–60)	762,543	0.024	27.8
Top 10%	386,801	0.033 (+38%)	29.9 (+8%)
Top 5%	195,282	0.036 (+49%)	30.8 (+11%)
Top 1%	41,942	0.042 (+75%)	31.5 (+13%)

Notes: Granted applications, publication years 2005–2020. Invention quality is the within-art-unit×year percentile rank of the text-based measure. Each cell is the mean over non-missing observations in the quality bin (n is the bin size); parentheses give the percent difference from the median (p40–60) bin. Forward citations are the level ratio of citations to the publication-year mean, with granted applications recording no citations counted as zero; the ratio is sparse (97.5% of granted applications are zeros), so bin means are tail-driven by construction. Winsorising the ratio at its 99th percentile retains premia of +24%, +29%, +42% across the three tail bins. Market value is the Kogan et al. (2017) estimate ($m), available for the matched subset.

Novelty-impact decomposition. The composite index combines two components: *novelty* (backward textual dissimilarity from prior art) and *impact* (forward textual similarity to subsequent work). These load with *opposite* sign and similar magnitude on the acceptance margin: broad, distinctive (novel) claims draw examiner scrutiny and scope concessions, whereas the impact component raises acceptance. The composite is therefore an order of magnitude flatter at the margin than either component, though the offset is not exact: the impact loading dominates slightly, so the composite carries a mild positive grant gradient. This is visible in the complier profile. Re-estimating the conditional complier share with the within-cell rank of the *novelty* component alone on the quality axis, rather than the composite, flips the profile from declining to rising: the share of lottery-decided grants *increases* by 6.8% from the bottom to the top of the novelty distribution, mirroring (at smaller magnitude) the 40% increase across filed scope, whereas across the composite index it *declines* by 15.5% (Table H.3). The novelty component is thus the behaviourally active half of the measure; the composite understates the lottery's sensitivity to it because the impact component pulls in the opposite direction.

Table H.3: *Decomposition: complier share by composite quality, novelty, and scope*

Axis of the complier profile	Bottom pctile	Top pctile	Change
Composite quality index	5.37%	4.54%	−15.5%
Novelty component only	4.79%	5.12%	+6.8%
Filed scope *(reference)*	4.03%	5.63%	+39.7%

Notes: Conditional share of lottery-decided grants (compliers among granted patents, in percent) at the bottom and top of each axis, on the untrimmed evaluation grid. The composite nets a grant-*positive* impact component against a grant-*negative* novelty component of similar size, and is therefore an order of magnitude flatter on the acceptance margin than either half: regressing grant on within-cell percentile ranks gives +0.061 on impact and −0.066 on novelty, against +0.013 on the composite itself. The component loadings do not sum to the composite's, because the composite ranks their ratio rather than adding them; its net loading is positive even though the novelty coefficient is the larger in magnitude. Because the complier share is proportional to a strictly *decreasing* inverse Mills ratio of the acceptance index (Section 6.3), a grant-negative component must show a *rising* complier profile: this is why the novelty component alone rises, like filed scope, while the composite, whose net grant loading is positive, declines.

Ambiguity of the quality assessment. Two senses of "hard to judge" concern the quality measure rather than the examiner, whose assessment noise Appendix B.2 takes up separately. Table H.4 reports both indices in all four conditioning forms. The measurement-noise index loads on compliers negatively or not at all in every form, so marginal grants are not applications whose quality the measure pins down poorly. That reading carries a caveat the table itself reports: the noise index fails the treated-versus-untreated balance check in every conditioning form, including the within-cell one, so its loading is not as clean a complier characteristic as the other rows; the direction of the conclusion would survive an imbalance that moves the loading's size, but the table does not establish the imbalance immaterial. The reference-class index is positive in every form and halves once ranked within art unit × year, so roughly half of the raw association is between-cell composition; what survives is the small application-level residual the text quotes, and its profile shape is flat.

Table H.4: *Complier loadings on the ambiguity indices*

	Complier − population (%)	z	Balance z	Flat p
Measurement noise: |half 1 − half 2|, fixed construction				
raw	−0.31	−0.56	+2.54	0.422
within quality deciles	+0.09	+0.16	+3.14	0.160
within filed-scope deciles	−0.48	−0.86	+2.94	0.506
within art unit × year	−0.54	−3.29	+2.96	0.088
Reference-class dependence: |art-unit − technology-center pool|				
raw	+1.26	+3.47	+0.03	0.013
within quality deciles	+1.22	+3.51	+0.13	0.049
within filed-scope deciles	+1.17	+3.22	+0.17	0.028
within art unit × year	+0.57	+3.46	−0.15	0.127

Notes: The Table reports, for two ambiguity indices in four forms each, the gap between the complier mean and the population mean of the index rank, in percent of the population mean, from the examiner-dummy UJIVE modified-outcome estimator of Section 5.3 on the estimation sample ($n = 5,367,639$; standard errors clustered on art unit × publication year). The measurement-noise index scores every application on two independent halves of the text corpus at fixed construction and takes the absolute disagreement; the reference-class index holds the corpus fixed and takes the disagreement between the quality rank built on the art-unit comparison pool and the same rank rebuilt on the coarser technology-center pool (comparison pools selected on grant outcomes are excluded as post-treatment). Each index is percentile-ranked and, in the conditional forms, re-ranked within deciles of quality, within deciles of filed scope, or within art unit × publication year cells; the last form separates an application-level property from art-unit-level composition and is the form quoted in the text. Balance z is the treated-versus-untreated arm difference of the modified-outcome construction; Flat p is the simultaneous sup-t test that the complier representation profile over the index is flat.

Reliability of the quality rank. Table H.5 bounds what measurement noise could do to the complier-quality profile. The split-half correlation scores every application on two independent halves of the text corpus at fixed construction; at 0.88, the Spearman-Brown reliability of the full-corpus rank is 0.94, so classical noise attenuates a binned profile over the rank by at most seven percent of its amplitude. The dispersion across construction variants is a separate object, a choice rather than noise, and is priced by the reference-class index above.

Table H.5: *Reliability of the quality rank*

Split-half correlation of the quality rank (Pearson)	0.880
Split-half correlation of the quality rank (Spearman)	0.880
Full-corpus reliability (Spearman-Brown)	0.936
Implied ceiling on attenuation of a profile amplitude	$\times$ 1.068
Pairwise correlation across construction variants (min)	0.076
Pairwise correlation across construction variants (median)	0.504
Pairwise correlation across construction variants (max)	0.988

Notes: The Table reports the measurement reliability of the quality rank on the estimation sample (5,344,793 applications complete on all rank columns). The split-half correlation scores every application on two independent halves of the text corpus at fixed construction and correlates the two within-cell percentile ranks; the Spearman-Brown adjustment $2r/(1+r)$ converts it to the reliability of the full-corpus rank. Under classical measurement error a binned profile over the rank is attenuated by the reliability, so the true amplitude of the complier-quality profile exceeds the observed one by at most the stated factor. The variant correlations are pairwise Pearson correlations among the ten retained construction variants of the index; their dispersion is a construction choice rather than noise, and is the object priced by the reference-class index of Table H.4.

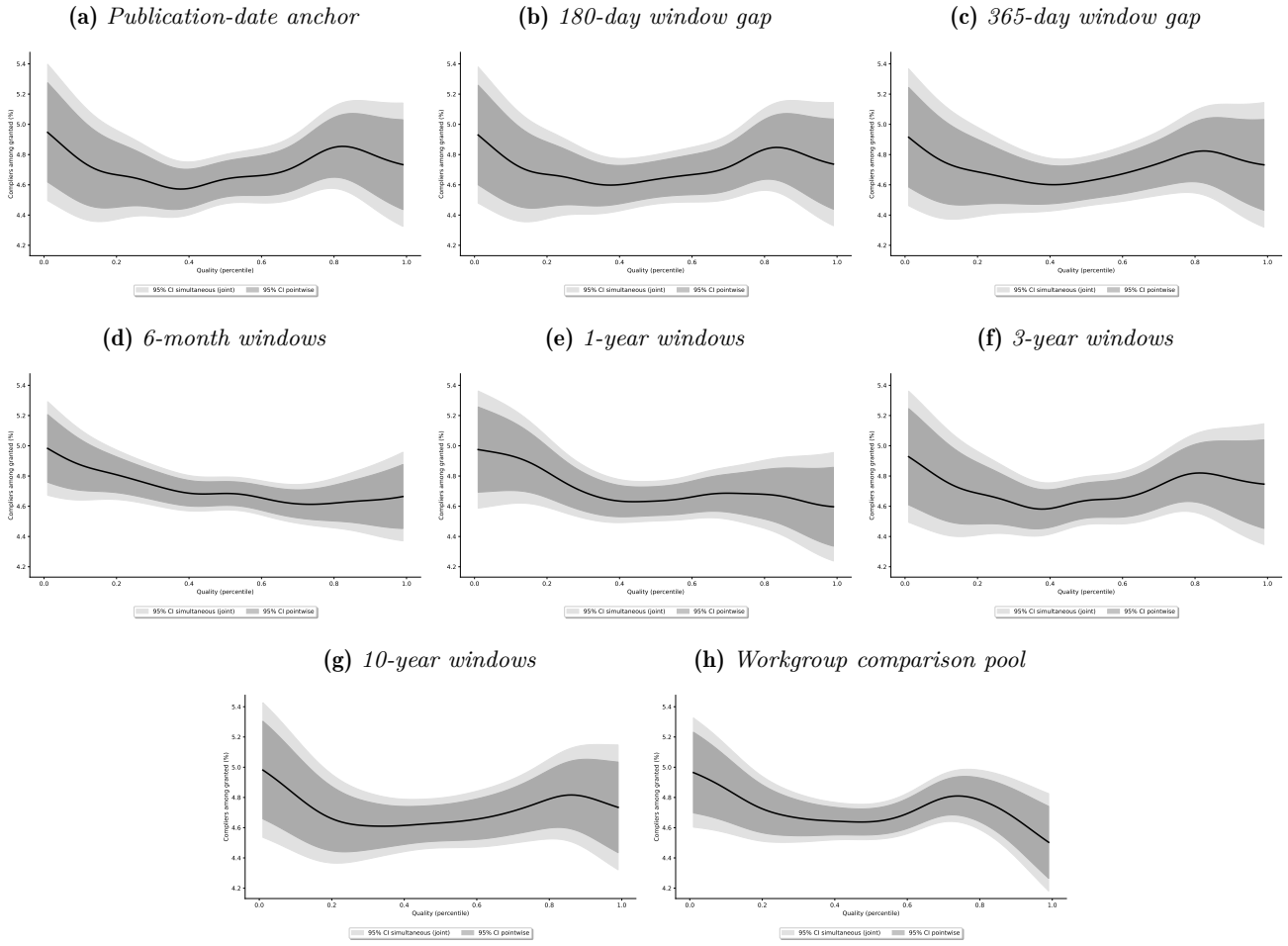
Table H.6: *Forward similarity and examiner leniency at the art-unit level*

	β	s.e.	p	N
Cross-section of art-unit means (raw units)	+0.038	0.004	0.000	701
Cross-section of art-unit means (SD units)	+0.321	0.031	0.000	701
Within unit: art unit $\times$ year, unit and year FE (SD units)	−0.054	0.037	0.153	9,896

Notes: The Table reports regressions of mean forward similarity (the impact component of the quality index) on mean examiner leniency. The first two rows collapse to art-unit means over the 2005–2020 estimation window and regress the level of one on the other, weighted by art-unit size, with heteroskedasticity-robust standard errors; the positive association reflects technology composition across fields. The third row is the exclusion-relevant design quoted in the text: art unit $\times$ publication year means with art-unit and year fixed effects, weighted by cell size, standard errors clustered on art unit, so the coefficient measures whether a unit's forward trajectory moves with the leniency of the examiners deciding in it. SD units standardize both variables.

Construction robustness of the complier profile. The near-flatness of the complier-quality profile is not specific to how the index is built. Figure H.1 re-estimates the headline profile of Figure 4 with the quality index rebuilt under eight alternative construction choices, each changing exactly one dimension of the construction: the anchor date, the length of the text-comparison windows, the gap between the backward and forward windows, and the level of the comparison pool. The estimator, instrument, leave-out design, and sample are identical across panels, so the panels differ only in the ranking the quality measure induces. Every variant moves within the same narrow band around the overall mean, and the simultaneous test of a flat profile rejects in none of them (p between 0.09 and 0.37).

Figure H.1: *Construction robustness: compliers among granted patents by quality, across index constructions*



Notes: Each panel plots the kernel-smoothed UJIVE share of compliers among granted patents by quality percentile, as in Figure 4, with the quality index rebuilt under one alternative construction choice per panel. The design, leave-out connected set, and instrument are built once from the grant outcome and shared across panels, so nothing but the quality ranking varies. Shaded areas are 95% simultaneous (light) and pointwise (dark) confidence bands; all panels share a common vertical scale. The simultaneous test of a flat profile does not reject in any panel.

I Scope measure: construction robustness and normalization dependence

Filed scope is measured in one specific way: the average number of words per claim (Section 3.2), expressed as a within-year percentile rank. This appendix asks two questions about that choice. The first is construction: whether the results depend on which part of the claims text the measure reads. The second is normalization: whether quantities that sum scope across applications depend on the cardinal scale behind the rank.

Alternative claim measures. Two in-family alternatives are available for every application: the length of the first claim, the object that Kuhn and Thompson (2019) measure, and the total length of the claims. They correlate with the headline measure at Spearman 0.63 and 0.47, so they are different orderings of the same concept rather than relabelings, and where the samples overlap the corpus first-claim length

agrees with the independently parsed first-claim word counts of Kuhn and Thompson (2019) at a rank correlation of 0.997. Table I.1 rebuilds the broadness axis from each alternative on the estimation sample. The acceptance decline from the narrowest to the broadest ventile is 19–22 percentage points under all three measures, with near-identical profiles (Figure I.1).

Figure I.2 and Table I.2 carry the test to the paper's headline object. The complier-scope profile is re-estimated with the identical estimator, instrument, leave-out design, and sample, so the rows differ only in the ranking the scope measure induces. The share of grants owed to the draw rises from the narrowest to the broadest filings under every measure: by 1.8 percentage points under the headline axis run through this pipeline, 2.3 under first-claim length, and 1.4 under total claim length, with the low-quality, broad-scope corner between 5.4 and 5.8 percent throughout. No alternative profile departs from the headline profile by more than a third of a percentage point at any point of the distribution – at this sample size that is outside the headline curve's own sampling band, as orderings correlating at 0.5–0.6 must be, but it is a tenth of the gradient the measures agree on. Claim count is deliberately not in this family: applications with more claims write shorter ones, acceptance is nearly flat in the count, and it enters the decomposition regressions as a control rather than as a scope measure.

Table I.1: *Scope measure: alternative claim measures, agreement and acceptance*

	Spearman vs. headline	Acceptance, narrowest 5%	Acceptance, broadest 5%	Decline (pp)	N
Average claim length (headline)	1.00	76.9	57.7	19.2	5,363,832
same, on the first-claim sample	1.00	76.7	56.9	19.9	4,880,897
First-claim length	0.63	79.0	57.4	21.6	4,880,897
Total claim length	0.47	78.8	56.9	21.9	5,363,832

Notes: Estimation sample (design-complete rows, publication years 2005–2020), $N = 5,363,832$. Each row measures filed scope from a different part of the claims text and rebuilds the broadness axis with the estimation convention: the within-publication-year percentile of the negated measure, so higher means broader. Acceptance columns are mean grant rates in the bottom and top broadness ventiles. The first-claim measure excludes the 9.0% of applications whose published first claim holds at most 10 words, dominated by the "(canceled)" notation of continuations whose original first claim was withdrawn during prosecution; the headline measure re-estimated on those same rows is the second row. Claim count is deliberately not a member of this family: applications with more claims write shorter ones (Spearman +0.26 between claim count and word-per-claim broadness), yet acceptance is nearly flat in claim count (Spearman +0.04) – it counts how many boundaries an application draws rather than how much language each one stakes out, and it enters the decomposition regressions as a control instead. At-issuance measures are excluded as post-decision text. Where the samples overlap ($n = 671,268$ granted, patent-matched), the corpus first-claim length agrees with the independent Kuhn–Thompson first-claim word counts at filing at a rank correlation of 1.00.

Figure I.1: *Acceptance in filed scope under alternative claim measures*

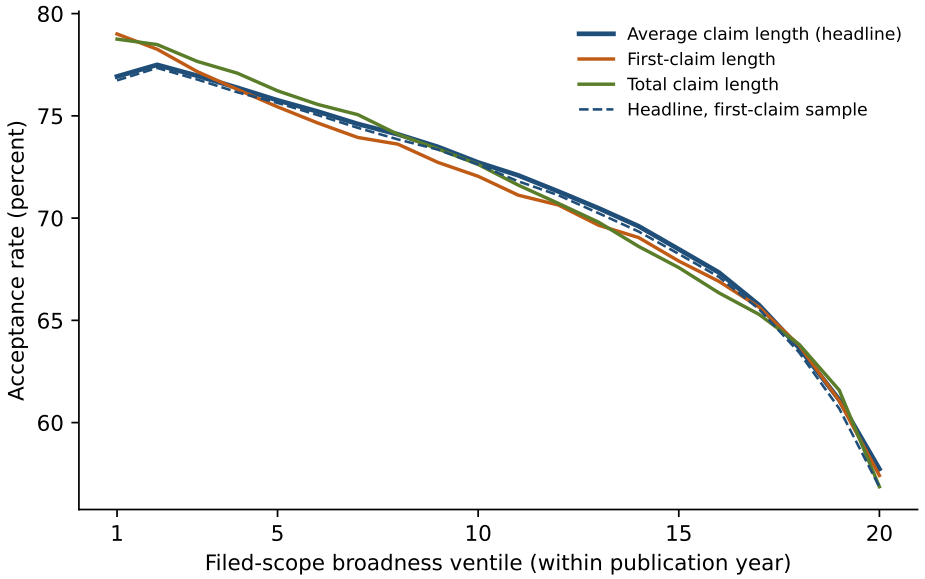


Figure I.2: *Compliers among granted patents by filed scope, under alternative claim measures*

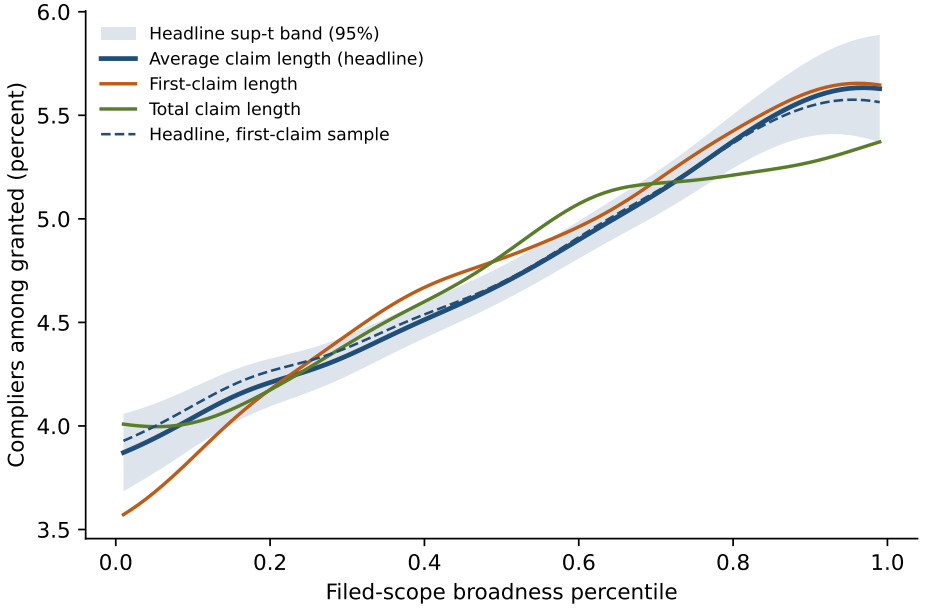


Table I.2: *Complier-scope gradient under alternative claim measures*

	Compliers among granted (%)				
	Narrowest ventile	Broadest ventile	Rise (pp)	Low-quality, broad corner	N
Average claim length (headline)	3.8	5.6	+1.8	5.6	5,364,969
same, on the first-claim sample	3.8	5.5	+1.7	5.4	4,880,896
First-claim length	3.3	5.6	+2.3	5.8	4,880,896
Total claim length	4.1	5.5	+1.4	5.7	5,363,956

Notes: Share of granted patents owed to the examiner draw, in the bottom and top ventile of each broadness axis and in the lowest-quality × broadest-scope corner of the 8×8 quality-scope surface. The design, leave-out connected set, and instrument are built once and shared by every row, so rows differ only in the ranking the scope measure induces; the first-stage scalar is the same one behind the paper's overall complier share, which this run reproduces at 4.7%. The headline row re-estimates the paper's own axis through the identical pipeline, so every other row is a like-for-like contrast. The first-claim rows exclude the 9.0% of applications whose published first claim holds at most ten words, dominated by the "(canceled)" notation of continuations; the second row shows the exclusion alone is innocuous. A simultaneous test of a flat profile rejects at $p < 0.001$ for every row – the object is a gradient under every measure – and no profile departs from the headline profile by more than a third of a percentage point at any point of the distribution.

Normalization dependence. Filed and granted scope enter as within-year percentile ranks of average claim length. Appendix N notes that quantities which *sum* scope across applications inherit a normalization that the ordering itself does not pin down. The rest of this appendix separates the part of the paper that is immune from the part that is exposed, and measures the exposure.

The immune part is larger than the objection suggests. A rank is unchanged by any strictly monotone transformation of claim length, so any result that uses only the ordering is exactly invariant to the choice of scale. That covers which applications fall in which equal-mass scope bin, the complier share estimated at each bin and therefore the shape of the profile, the ordinal reclassification counts, and any ratio taken across bins. It is not a robustness finding but an algebraic one, and it requires no re-estimation. The exposed part is the class of quantities that aggregate scope in levels: the granted-scope budget, $\Delta\hat{S}$, the reallocation share, and the break-evens built on δ. A sum of ranks carries no cardinal content, so these are conditional on the scale in a way the composition results are not.

Table I.3 quantifies that conditionality. It re-expresses the same scope ordering on a family of positive scales, $1/w^{\gamma}$ for $\gamma \in \{\frac{1}{2}, 1, 2\}$ in the average claim length w, every member of which agrees exactly on which patent is broader and differs only in by how much, and reports the share of the granted-scope budget carried by each rank decile. Deciles are formed on the rank, so the groups are held fixed across columns and only the weighting moves, which isolates normalization from composition.

The shares move substantially. The broadest decile carries between 14.5% and 34.9% of the budget depending only on the scale, the largest swing in any single decile's share is 20.4 percentage points, and the implied concentration, the broadest decile's share relative to the narrowest, ranges from 2.4 to 32.7. The percentile rank is not an extreme member of the family: its decile profile lies closest to $1/w$, with a maximum deviation of 3.3 percentage points, so the reported magnitudes do not appear to be inflated by the normalization. They are nonetheless conditional on it, and any share-of-budget statistic in Section 7.7 should be read that way.

The swings do not, however, propagate to the statistic the policy argument rests on. Table I.4 recomputes the budget-neutral reallocation share Z on each of the same scales, holding the counterfactual allocation fixed and changing only the scale it is measured on. Z moves from 11.7% on the rank to 10.6%, 11.1% and 12.3% on the three inverse-power scales: a total range of under two percentage points, a factor of 1.2 from smallest to largest, against a factor of 2.4 in the broadest decile's budget share. The reason is in the second row. Most of the reallocation runs through *which* applications are granted rather than through how much scope a given grant receives, and an application that changes grant status contributes its whole scope under every scale, which is why that part of Z moves little across the scales tried; the stability is an empirical property of the tested family, not an algebraic one, since the switchers' scope distribution differs from the baseline budget's. The component measured among applications granted in both states is the scale-sensitive one, ranging from 0.4% to 1.6%, but it is a small part of the total. The δ-free reallocation result is therefore considerably more robust to the cardinalization than the budget shares alone would suggest.

Two limits of these exercises. Both hold the allocation fixed and vary only the scale, so they bound the accounting component of normalization dependence and not the behavioural one: filed scope responds

to a revision cost defined on the rank scale, and a complete answer would re-estimate the model under the alternative scale. And the rank embeds a second normalization that the raw scales do not, since it is taken within publication year, so two patents with identical claim length in different years receive different ranks. That choice is held fixed throughout the table, and what varies is the cardinal spacing alone.

Table I.3: *Normalization dependence of the granted-scope budget*

Rank decile (1 = narrowest)	rank	$1/w^{1/2}$	$1/w$	$1/w^2$
	share of the granted-scope budget			
1	0.9%	6.0%	3.5%	1.1%
2	2.7%	7.6%	5.5%	2.4%
3	4.5%	8.4%	6.7%	3.6%
4	6.4%	9.0%	7.7%	4.8%
5	8.4%	9.6%	8.7%	6.1%
6	10.5%	10.2%	9.8%	7.7%
7	12.8%	10.8%	11.0%	9.7%
8	15.2%	11.5%	12.4%	12.5%
9	17.8%	12.4%	14.5%	17.1%
10	20.7%	14.5%	20.2%	34.9%
Broadest/narrowest	22.1	2.4	5.7	32.7

Notes: Each column re-expresses the same scope ordering on a different positive scale and reports what share of the total granted-scope budget each decile then carries. w is the average claim length in words. The three inverse-power scales are strictly decreasing in w and agree exactly on which patent is broader; the rank is computed within publication year, so it agrees with them within a year but can reorder cross-year pairs, and the four columns share an ordering up to that cross-year component rather than exactly. Deciles are formed on the rank, so the groups are held fixed across columns and only the weighting moves, which isolates normalization from composition up to the same cross-year reordering. Budget shares are invariant to multiplying a scale by a positive constant, so the inverse-power columns need no further normalization. Sample: $3,771,826$ granted patents in the core sample. The table bounds the accounting component of normalization dependence, holding the allocation fixed; it does not bound the behavioural component, since filed scope responds to a cost defined on the rank scale.

Table I.4: *Normalization dependence of the budget-neutral reallocation share*

Scale	rank	$1/w^{1/2}$	$1/w$	$1/w^2$
Reallocation share Z	11.7%	10.6%	11.1%	12.3%
of which among always-granted	1.1%	0.4%	0.8%	1.5%

Notes: The budget-neutral reallocation share $Z = \frac{1}{2}\sum_i |\hat{s}_i^{\text{cf}} - \hat{s}_i^{\text{base}}| / \sum_i \hat{s}_i^{\text{base}}$ recomputed with granted scope expressed on each of the scales of Table I.3, holding the counterfactual allocation fixed. Applications that are not granted contribute no scope under any scale. The second row restricts the numerator to applications granted in both the status quo and the counterfactual, so the difference between the rows is the part of the reallocation that runs through which applications are granted at all rather than through how much scope a given grant receives. Scales are mapped onto the rank by the within-rank-bin mean over the granted core sample. Estimated model, N = 7,000,000 simulated applications.

J What the examiner lottery identifies: marginal treatment effect curves

The composition of Section 6 and the mechanism of Section 6.3 are statements about which applications the examiner draws decide. This appendix draws out the consequence for the causal estimates the examiner lottery has produced: if the complier pool is not representative, those estimates differ systematically from the average effects.

In a judge fixed-effects IV design, the LATE is a weighted average of treatment effects over the complier subpopulation, so when that subpopulation tilts toward broad-claiming applications of unremarkable quality, the LATE weights treatment effects differently from the population average. Which way the ATE and LATE differ depends on how the treatment effect varies across the resistance distribution, and that is an object to be estimated, not a corollary of the composition. I estimate it with marginal treatment effect (MTE) curves (Heckman and Vytlacil, 2007), whose propensity score $P(X, Q, Z) = \Pr(D = 1 \mid X, Q, Z)$ incorporates quality, filed scope, and examiner leniency jointly, so the correction accounts for all three dimensions at once rather than one at a time. The estimating equation, the implementation, the sample matching, and the joint test of the identifying assumptions follow below.

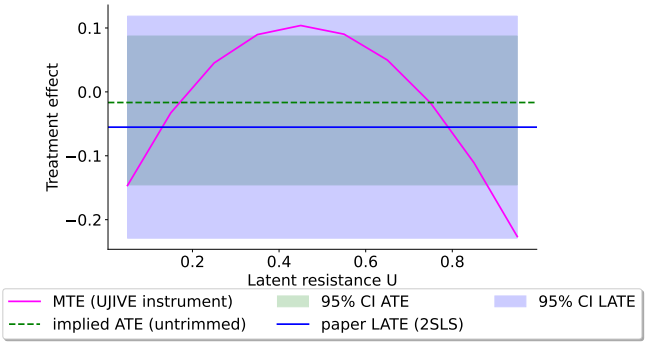
One caveat governs how far the results can be read. The design-based sections of the paper use the random assignment of examiners and little else; recovering a marginal treatment effect requires a selection model, whose substantive assumptions are stated below before the estimates are read. I therefore state the conclusion as a claim about *direction* and about consistency across settings, not about magnitude.

Figure J.3 reports the curves for an outcome of each paper, with the underlying regressions in Appendix K.3. In Sampat and Williams (2019), the curve is hump-shaped in resistance, while in Farre-Mensa et al. (2020) it rises monotonically, implying negative selection on potential gains. What the two share is the sign of the always-taker region. Applications granted under nearly any examiner carry negative estimated effects in both, so in each setting the LATE is identified on the part of the resistance distribution where a grant is worth most. In the Farre-Mensa et al. (2020) setting the integrated ATE lies below the LATE in every outcome and under both integration windows, a consequence of where the complier mass sits rather than of the slope of the curve; in the Sampat and Williams (2019) setting the two estimands are statistically indistinguishable, with small gaps of unstable sign, so the ordering claim rests on the first setting alone.

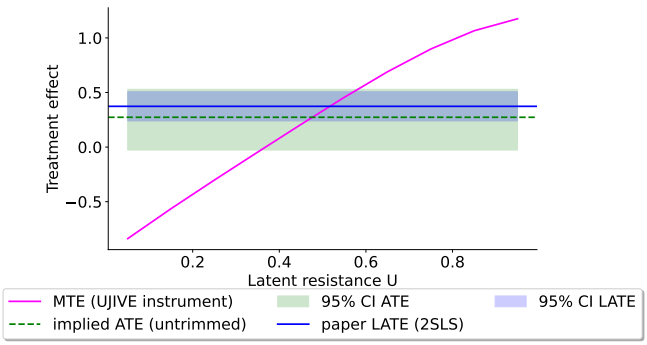
The strength of that pattern should not be overstated, and three ways of counting it give three different answers. Re-estimating both estimands on 5,000 common bootstrap resamples to test the difference between LATE and ATE, the gap is individually significant in two of the eight outcomes under bias-corrected and accelerated intervals, both of them citation outcomes of Farre-Mensa et al. (2020), and one survives (Romano and Wolf, 2005, 2016) familywise control. For both outcomes reported here, the difference is not statistically distinguishable from zero, so the ordering rests on point estimates alone. A uniform sign across one setting's outcomes is what a systematically non-representative complier pool predicts and what an unsystematic one does not; the Sampat and Williams (2019) gaps are small and of unstable sign, which is what a sample an order of magnitude smaller supports. The pattern suggests systematic overstatement where there is power to detect it, without establishing that any one published

Figure J.3: *Marginal treatment effect curves*

(a) *log(Scientific publications) – Sampat and Williams (2019)*



(b) *log(Follow-on patent applications) – Farre-Mensa et al. (2020)*



Notes: Each panel plots the MTE curve (Equation 31) for the primary outcome of the respective paper, estimated on the variation used by the examiner-dummy UJIVE estimator of Section 5: the leave-out deviation $\hat{D}^A_{-i} - \hat{D}^W_{-i}$ enters the selection probit as the instrument, and the sample is the connected set on which the leave-out fit is defined. For Sampat and Williams (2019) the examiner-leniency design is shown. The horizontal axis is the unobserved resistance term $U \in [0, 1]$; higher values correspond to applications that require a more lenient examiner to be accepted. The magenta curve is the MTE. The dashed green line is the ATE obtained by integrating that curve over the full support of U, with its 95% bootstrap confidence interval shaded in green. The solid blue line is the LATE delivered by each paper's own estimator, its published 2SLS specification re-estimated on exactly this sample so that the comparison is not a sample-composition artifact, with its 95% cluster-robust confidence interval shaded in blue. The two objects are therefore an average effect and the complier-weighted effect the original design reports, on common ground. The ATE integrated over the common-support window only is reported in the text rather than plotted. The two shaded bands are constructed by different routines and share no bootstrap draws, so their overlap is not a test that the two estimands are equal; the formal test, which re-estimates both on common resamples, is reported in Appendix Tables K.10 and K.11. Additional MTE outcomes are reported in Appendix Figure J.5.

LATE is significantly overstated. I therefore rest the argument on the pattern rather than on any single test. The remainder of this appendix describes the two curves outcome by outcome, reports the full test and the diagnostics behind it, and sets out how far the finding generalizes beyond these two papers.

Estimating treatment effects beyond the LATE requires a selection model of the patent review process. For the settings of Sampat and Williams (2019) and Farre-Mensa et al. (2020), I estimate marginal treatment effect curves via local instrumental variables, as implemented in the `localIV` package of Zhou and Xie (2019). Following the standard treatment of the generalized Roy model and instrumental variables (for more details see Heckman and Vytlacil (2007)), the marginal treatment effect (MTE) curve is defined, adjusted to my previously used notation, as:

$$MTE(x, u) = \frac{\partial \, \mathbb{E}[Y|X = x, Q = q, P(X, Q, Z) = p]}{\partial p}\bigg|_{p=u} \tag{31}$$

Three terms are new relative to the main text: the propensity score $P(X, Q, Z) = \Pr(D = 1 \mid X, Q, Z)$, which gives the probability of patent grant conditional on controls X, quality Q, and leniency Z; the latent resistance term u; and an outcome variable Y that is affected by the patent grant D. The MTE at u is the treatment effect for applications that are exactly indifferent at propensity score $p = u$: applications with low resistance are granted by nearly any examiner, while applications with high resistance are granted only under an unusually lenient draw.

Equation (31) carries assumptions that quasi-random assignment does not supply, and they should be stated first. Three are substantive. First, a scalar latent index: resistance is a single unobserved u that orders all applications, so that any two applications with the same propensity score are exchangeable in their selection behavior. The acceptance margin of Section 5.1 delivers exactly such an index, with $a(q, s) + l$ crossing a threshold, which is why the specification is a natural one here, but it is imposed rather than tested, and it rules out selection on dimensions orthogonal to that index. Second, a parametric propensity specification: the score is estimated from a probit in quality, controls, and the instrument, so the MTE inherits any misspecification of that first step. Third, and most consequentially, *support*. Equation (31) is defined only where the propensity score has mass, and the estimated score does not reach either extreme in these samples. The full-support integral therefore extrapolates the fitted curve into regions the design never populates, and no quantity of additional data of this kind repairs it, since the examiner draw simply does not move applications that far. I report the common-support integral alongside the full-support one throughout for this reason, and read the tails of every MTE figure as extrapolation.

This is the basis for the direction-only reading stated at the outset: the common-support comparison leans far less on the tails than the full-support level does. The direction is still an empirical regularity under the maintained specification rather than a support-robust implication, because both estimands weight the whole fitted curve and unsupported tails can in principle move the sign as well as the level.

One further feature of the specification makes the correction a joint one. The propensity score $P(X, Q, Z)$ incorporates quality, filed scope, and examiner leniency together, all three determinants of the grant decision identified in Section 5.1, so the observed selection those determinants carry is accounted for jointly; the resistance index u ranks applications by their unobserved resistance conditional on the score, not by a position in the (q, s, l) space itself. The complier composition over quality and filed scope implies non-representativeness along the observed dimensions, which the MTE weights jointly rather than one at a time.

The two settings in detail. Sampat and Williams (2019) study whether patents on human genes blocked follow-on research, finding no significant effect on subsequent scientific publications. Panel J.3a of Figure J.3 shows that the MTE curve for their primary outcome is hump-shaped in resistance: the applications granted under nearly any examiner and those granted only under an unusually lenient one both show negative estimated effects (-0.31 and -0.56 at the fifth and ninety-fifth percentiles of u), while the intermediate-resistance applications that carry the most weight in the LATE show positive ones, peaking at 0.10. Their own specification, re-estimated on this sample, returns a LATE of -0.055 (cluster-robust s.e. 0.089), a null consistent with the original finding. Integrating the MTE over the full support of u yields an ATE of -0.017, and over the common-support window $+0.015$; on 5,000 common bootstrap resamples the ATE-minus-LATE difference is $+0.039$ (BCa 95% CI $[-0.109, 0.320]$) over the full support and $+0.070$ ($[-0.051, 0.380]$) over the common-support window (Table K.10). This setting therefore delivers no ordering: the gap sits inside its interval under either integral and its sign is not stable across the setting's outcomes. What it contributes is the shape, the hump in resistance, together with the caveat that both tails of the fitted curve lie where the propensity score has little support, so I read neither the

full-support level nor its sign as robust.

Farre-Mensa et al. (2020) find that startups assigned a lenient examiner experience 55% larger employment and 80% higher sales growth. Panel J.3b shows that the MTE curve for follow-on patent applications rises monotonically in resistance, from −0.83 at the fifth percentile of u to 1.18 at the ninety-fifth: *negative selection on potential gains* in its textbook form, in which the applications granted most readily are those that benefit least from the grant. Their own specification, re-estimated on this sample, returns a LATE of 0.369 ([0.236, 0.502]), against an integrated ATE of 0.269 ([0.021, 0.524]), or 0.318 ([0.082, 0.542]) over the common-support window. Here the ordering does not follow from the slope of the curve, since an upward-sloping MTE is consistent with an ATE on either side of the LATE, but from where the complier mass sits relative to it: the large negative values belong to the always-taker region at low resistance, which the complier-weighted LATE does not reflect. The ordering conclusion therefore rests on the Farre-Mensa et al. (2020) setting; Sampat and Williams (2019) contributes the curve rather than the ordering.

The joint bootstrap. All p-values reported in this appendix are from the bias-corrected and accelerated (BCa) interval of the ATE-LATE difference, computed on 5,000 common resamples, with both estimands conditioned on the same covariates, the control set of the MTE outcome equation, so that their difference is a well-defined object. For Farre-Mensa et al. (2020) that matched conditioning shifts the LATE by at most 0.005 relative to the published specification the figures plot, and for Sampat and Williams (2019) the two coincide. For the two primary outcomes the difference is not statistically significant: −0.104 ($p = 0.521$) for Farre-Mensa et al. (2020) and +0.039 ($p = 0.568$) for Sampat and Williams (2019). It is significant, at the five percent level and one outcome at a time, in two of the eight, both citation outcomes of Farre-Mensa et al. (2020): total citations within five years (−0.344, $p = 0.029$) and mean citations per application (−0.235, $p = 0.003$, and −0.203, $p = 0.008$, after common-support trimming). Only one of the two, mean citations per application, survives the Romano-Wolf familywise adjustment across the outcomes of its own table ($p = 0.017$); the other sits just above the threshold once the family is in view ($p = 0.088$). Tables K.10 and K.11 report the test for every outcome together with the diagnostics behind it; those resamples are drawn over applications rather than over the art unit × year cells the standard errors elsewhere cluster on, and Table K.12 shows that resampling those cells instead reproduces the analytic cluster-robust standard error of the LATE and leaves every verdict unchanged. That check covers the Farre-Mensa et al. (2020) family, which is where every individually significant difference sits; no Sampat and Williams (2019) outcome is individually significant under either integral, so no verdict in that setting rests on the resampling unit.

How far this travels. The findings plausibly extend beyond the two papers studied here, though the evidence is not equally strong in both dimensions. The composition documented in Section 6 is a property of a threshold grant rule in which scope is penalized, so any examiner-leniency instrument applied to USPTO prosecution should inherit some version of it. The *scope* tilt is the well-supported half: it is the steeper gradient, it survives the caseload-composition purge, and it is what the mechanism of Section 6.3 predicts most sharply. The quality tilt is milder and its complier mean is close to the population's,

so I would not assert that every examiner-leniency LATE is materially mis-weighted on quality. This stands in contrast to judge fixed-effects designs in the economics of crime (e.g., Agan et al., 2023), where defendants do not make strategic pre-treatment choices that select a non-representative complier pool. In patent prosecution, applicants' strategic filing decisions generate the non-representative complier pool, and the MTE framework is the natural tool for recovering treatment effects informative about the average applicant rather than the marginal one.

Estimation samples. Implementation follows the original papers' specifications, augmented with my quality measure. Because the samples that reach estimation are materially smaller than the samples I match, I report the full ladder rather than its first rung. For Farre-Mensa et al. (2020), the replication file contains 34,435 rows, 34,418 of them distinct applications. Matching to my data without the publication-year restriction of Section 3.1 recovers 28,395; imposing that restriction, which the rest of the paper applies throughout, leaves 16,683 and drops 11,712 applications (41.2%), all of them published before 2005. The subsequent outcome and covariate requirements drop essentially nothing. The binding step is instead the estimator's: the leave-one-out examiner instrument is defined only on the largest connected component of the examiner-by-cell bipartite graph, which retains 10,289 applications (38.4% dropped) across 2,196 examiners and 873 art unit $\times$ year clusters. The leverage and scope filters remove no further observations. It is 10,289, not 28,395, that the estimates rest on; quoting the larger figure as the sample would start the ladder in one population and finish it in another.

For Sampat and Williams (2019) the ladder is shorter and cut in a different place: 1,545 distinct applications, 1,450 matched, and 1,196 reaching the examiner-design estimates (1,108 for the gene-level design), with the loss falling on the covariate requirement rather than on the connected set. This setting is not subject to the publication-year restriction, deliberately: its gene patents are filed between 2002 and 2006, and imposing the restriction would discard most of the sample. One consequence bears on how the two replications are compared: my quality measure is a percentile rank, and it is therefore computed over different populations in the two settings. Columns from the two designs are indexed by deciles of different underlying distributions and should not be read as though they shared a scale. For Sampat and Williams (2019), the primary outcome is follow-on innovation in 2011-2012, measured by the logarithm of scientific publications (Table 2, Panel A in the original paper); for Farre-Mensa et al. (2020), it is the logarithm of follow-on patent applications. In each setting, I re-estimate the original specification including my measure of quality as an additional control variable, estimate the propensity score from quality, controls, and examiner leniency, and trace out the MTE curve via Equation 31. The resulting curves for the primary outcomes are reported in Figure J.3.

The subsamples corroborate the *level* of the full-sample complier share but are too small to resolve its gradients. Figure J.4 reports the share of compliers among granted patents for both papers, estimated with the same examiner-dummy modified-outcome estimator used for the full sample in Figure 4a. In the Farre-Mensa et al. (2020) subsample the profile hovers near 5% along both the quality and the filed-scope axis, close to the 4.7% full-sample share, but is flat to within its confidence bands in each. In the Sampat and Williams (2019) subsample the quality profile rises from roughly 4.5% to an interior peak near 11% before settling near 7%, although the joint test of flatness does not reject it ($p = 0.071$); the scope profile

does reject flatness ($p = 0.000$), but its shape is dominated by a region of the scope distribution in which the gene-patent subsample contains almost no applications, visible as the break in the fitted curve. The bands are wide enough at these sample sizes that the point estimates should not be read at face value. What these subsamples establish is that the magnitude of the lottery share documented in Section 6 is not an artifact of the full-sample construction; they neither confirm nor contradict its quality and scope gradients, which are identified off the full sample.

The subsamples also permit a formal test of the identifying assumptions that is infeasible in my main first-stage-only analysis. Because both settings provide a second-stage outcome, I can apply the joint test for the exclusion restriction and monotonicity of Frandsen et al. (2023). The respective p-values of the test are 1.000 for the setting of Sampat and Williams (2019) and 0.570 for Farre-Mensa et al. (2020). In both settings, the tested implications of exclusion and monotonicity thus survive; the scalar latent-index structure, the parametric propensity specification, and the support conditions the MTE integration additionally requires are not what this test probes.

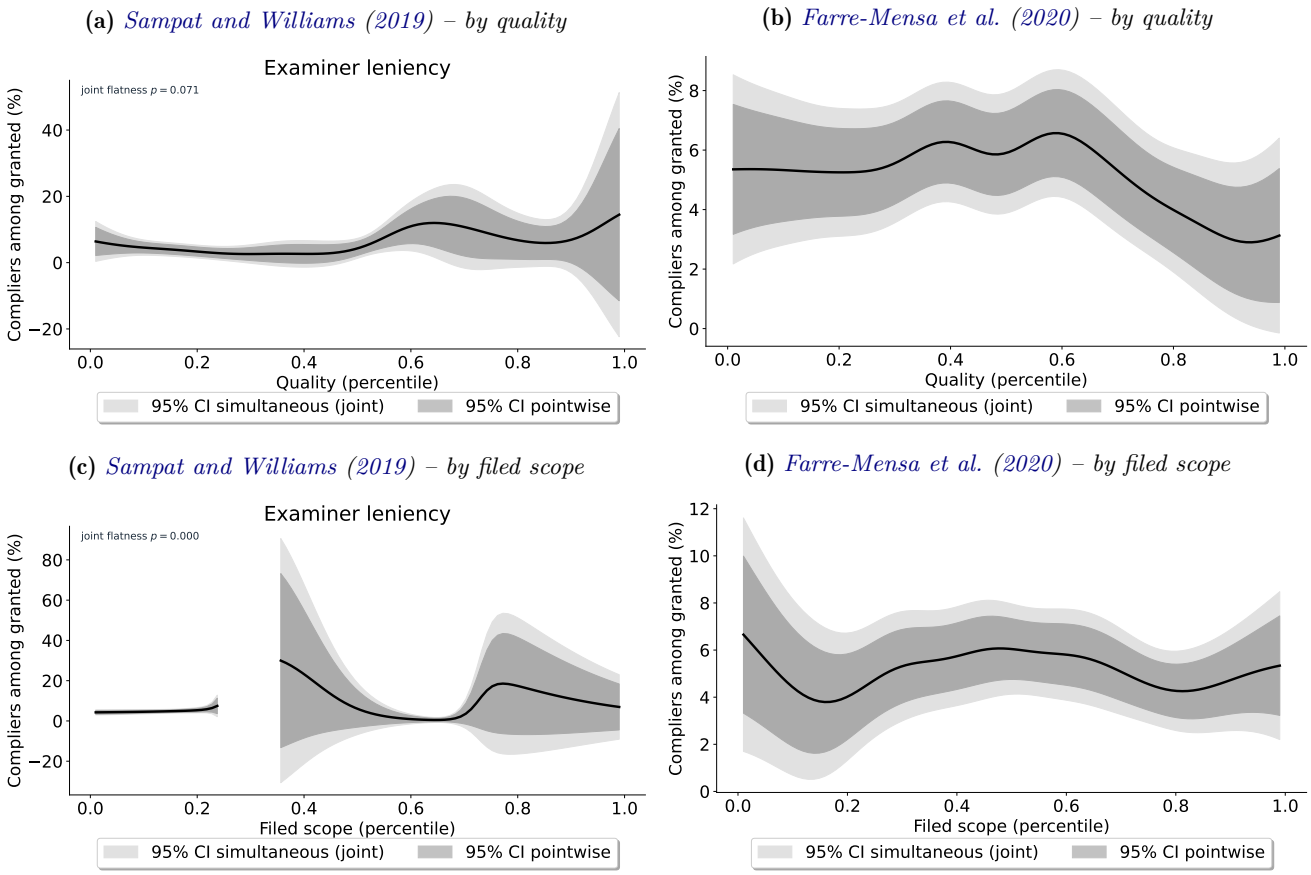
The MTE findings are robust to alternative outcome measures. Figure J.5 reports MTE curves for any follow-on scientific publication in the subsample of Sampat and Williams (2019), and for the logarithm of application and patent citations after five years, application citations after five years, and the logarithm of granted patents in the subsample of Farre-Mensa et al. (2020). Across these alternative outcomes, the qualitative pattern of the primary estimates carries over: the shapes of the curves, monotonically upward-sloping in the Farre-Mensa et al. (2020) setting and hump-shaped in Sampat and Williams (2019), mirror the primary outcomes in Figure J.3, and the integrated ATEs lie below the LATEs delivered by each paper's own published specification in every Farre-Mensa et al. (2020) outcome and for both the full-support and the common-support integral, while the Sampat and Williams (2019) gaps remain inside their intervals with unstable sign. Appendix Tables K.10 and K.11 test the comparison formally, outcome by outcome, by re-estimating the ATE and the LATE on common bootstrap resamples and under a common control set; the difference is individually significant for the two citation outcomes of Farre-Mensa et al. (2020) and nowhere else. The substantive interpretation of these results is developed at the start of this appendix.

Which Sampat and Williams estimand is on display. The LATE I report for that setting is their specification re-estimated *unweighted*; their published estimand is not. Their frame is gene-level – 293,652 gene rows over 1,545 applications, a mean of 190 genes per application and a maximum of 3,103, and the published estimates weight by that count. With dispersion of that size the weights bite. Re-estimating the identical specification with them moves every LATE by up to one and a half unweighted standard errors and carries four of the six across zero:

Design	Outcome	Unweighted	Gene-weighted	Shift (SE)
Examiner	log follow-on publications	−0.0552 (0.0887)	−0.0004 (0.0305)	+0.62
Examiner	Any follow-on publication	−0.0508 (0.0868)	+0.0131 (0.0224)	+0.74
Examiner	Any follow-on test	+0.0254 (0.0411)	−0.0164 (0.0342)	−1.02
Gene	log follow-on publications	+0.0243 (0.0252)	−0.0135 (0.0205)	−1.50
Gene	Any follow-on publication	+0.0263 (0.0225)	−0.0070 (0.0159)	−1.48
Gene	Any follow-on test	−0.0026 (0.0237)	−0.0159 (0.0217)	−0.56

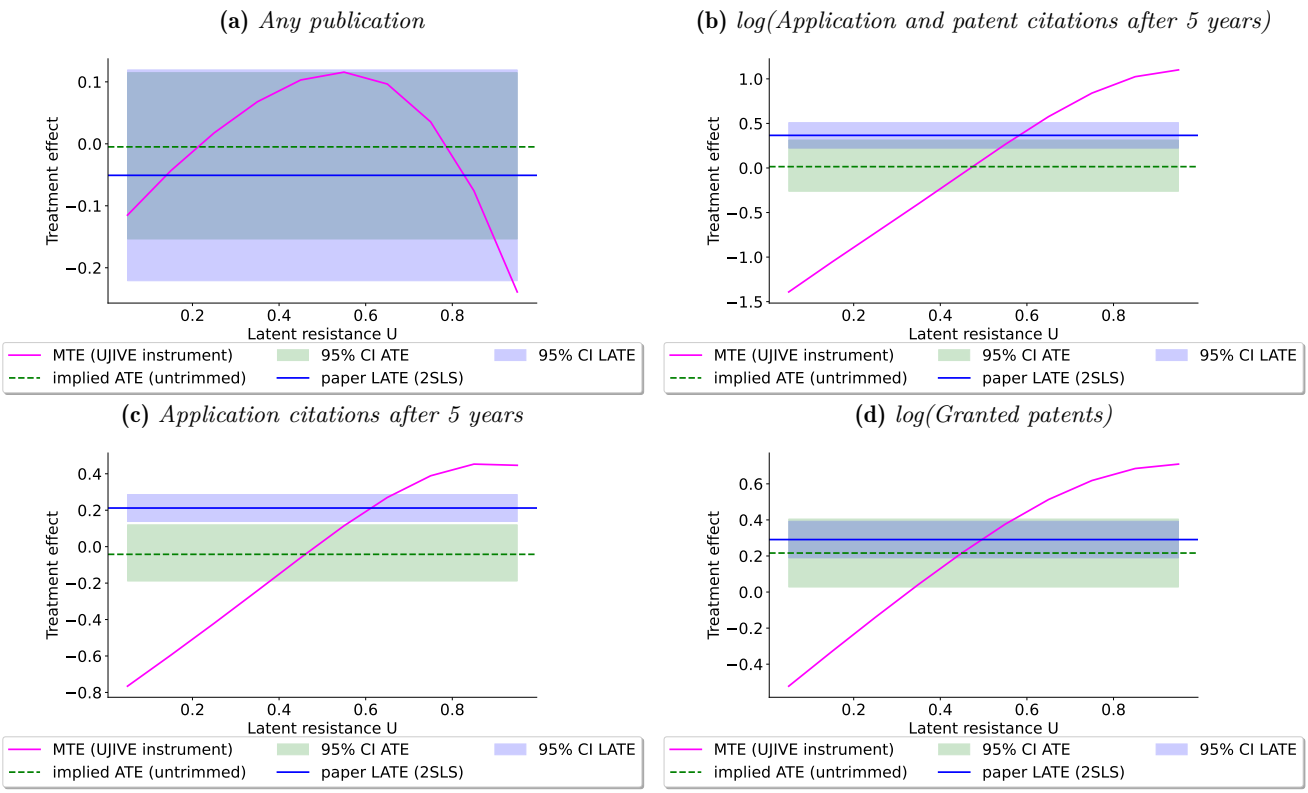
Every estimate, weighted or not, is statistically indistinguishable from zero, and the shifts run to between 0.6 and 1.5 unweighted standard errors. Two conclusions follow, and they point in opposite directions. The substantive reading is untouched: this setting delivers a null on follow-on innovation whichever estimand is used, which is what the comparison above rests on. But the *signs* of the point estimates are not robust to the weighting choice, so no argument here should lean on them, and the tables label which estimand they display rather than leaving it implicit.

Figure J.4: *Conditional share of patents accepted due to the examiner*



Notes: The Figure plots the share of compliers among granted patents, $\phi_c^{D=1}$, across the percentiles of quality (Panels J.4a and J.4b) and of filed scope (Panels J.4c and J.4d), for the sample of Sampat and Williams (2019) in the left column and for the sample of Farre-Mensa et al. (2020) in the right column. The estimator is the same examiner-dummy UJIVE modified-outcome estimator used for the full sample in Figure 4, applied here within each replication subsample; for Sampat and Williams (2019) the examiner-leniency design is shown. The shaded areas show 95% pointwise and 95% simultaneous (multiplier bootstrap) confidence bands. Where reported, the annotated p-value is that of the joint sup-test of the null that the curve is constant. Note the difference in vertical scale across panels: the replication subsamples are two to three orders of magnitude smaller than the full sample, so the bands are correspondingly wide and the point estimates should not be read at face value.

Figure J.5: *Additional MTE curves*



Notes: The Figure plots the results of estimating Equation 31 for any follow-on scientific publication in the subsample of Sampat and Williams (2019) in Panel J.5a and the results on follow-on innovation in Farre-Mensa et al. (2020) in Panel J.5b, J.5c, and J.5d. Estimation, sample, and the meaning of each plotted series are as in Figure J.3: the selection probit is instrumented with the leave-out examiner-dummy UJIVE deviation on the connected-set sample, the magenta curve is the MTE, the dashed green line the ATE integrated over the full support of U, and the solid blue line the LATE from each paper's own published 2SLS specification re-estimated on this same sample; shaded bands are the corresponding 95% confidence intervals. In every outcome shown, as in Figure J.3, the integrated ATE lies below that LATE. The bands are not comparably constructed and their overlap does not test equality of the two estimands; Appendix Tables K.10 and K.11 report the joint-bootstrap test of that difference for each outcome shown here.

K Regression tables underlying the figures

Several exhibits in the paper plot the output of a regression rather than a raw moment of the data. This appendix reports the underlying estimates in table form, so that every plotted curve can be read as coefficients with standard errors. The correspondence is one-to-one and stated in each table's note.

Two families of exhibit are deliberately absent. The by-revision panels (Figures 3 and D.2) and the leniency density (Figure D.3) plot sample means and a kernel density, not fitted coefficients. The structural fit and counterfactual exhibits compare simulated to empirical moments and are reported in Appendix L.

K.1 Conditional correlations

The binned scatter plots of Section 3.4 and Section 4 overlay a quadratic fit of the residualized outcome on the quality index. Tables K.1 and K.2 report that fit: the linear and squared quality terms separately, and the net predicted change across the plotted window, which is the statistic annotated on each panel.

Table K.1: *Conditional correlations: validation against external quality proxies*

	(1) Citations	(2) Renewals	(3) Kogan et al.	(4) Kelly et al.
Quality	0.0117***	0.0129***	0.0008	0.0267***
	(0.0015)	(0.0011)	(0.0014)	(0.0012)
Quality^2	0.0072***	−0.0030***	0.0050***	0.0049***
	(0.0010)	(0.0009)	(0.0013)	(0.0011)
Net effect ($q: 0 \to +1$)	0.0189***	0.0100***	0.0058***	0.0316***
	(0.0020)	(0.0014)	(0.0020)	(0.0016)
Controls	✓	✓	✓	✓
Art unit × year demeaning	✓	✓	✓	✓
Observations	3,812,246	3,812,246	1,094,509	1,945,744
R^2	0.002	0.004	0.013	0.024

Notes: Quadratic specification $y = \beta_1 q + \beta_2 q^2 + \gamma' X + \varepsilon$ estimated by OLS with CRV1 (ind_year_fe) standard errors in parentheses. *Net effect* is $\hat{\beta}_1(q_{hi} - q_{lo}) + \hat{\beta}_2(q_{hi}^2 - q_{lo}^2)$ over $q_{lo} = 0$ to $q_{hi} = 1$, with a delta-method standard error from the CRV1 (ind_year_fe) covariance of $(\hat{\beta}_1, \hat{\beta}_2)$. The upper evaluation point lies slightly beyond the standardized rank's support, whose upper endpoint is $\sqrt{3} \approx 1.73$ for a within-cell percentile, so the net effect is a fitted-quadratic contrast at two standard deviations rather than a sample-range change. Outcomes and the quality index are standardized and demeaned within art unit × year cells before estimation, so the cell fixed effects enter as a pre-transformation rather than as absorbed dummies and R^2 is a within-cell R^2. Controls are a new-examiner indicator, continuation-type dummies, dummies for the first five years of examiner experience, and small- and micro-entity indicators. Stars: $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Columns correspond to Panels 1a, 1b, 1c and 1d of Figure 1.

Table K.2: *Conditional correlations: acceptance and scope*

Regressor x	(1) Acceptance Quality	(2) Acceptance Filed scope	(3) Filed scope Quality	(4) Granted scope Quality	(5) Difference in scope Quality
x	0.0063***	−0.0386***	0.0001**	−0.0001	0.0101***
	(0.0006)	(0.0003)	(0.0000)	(0.0012)	(0.0009)
x^2	0.0003	−0.0189***	0.0001**	0.0010	0.0018**
	(0.0004)	(0.0004)	(0.0000)	(0.0009)	(0.0007)
Net effect ($x : 0 \to +1$)	0.0065***	−0.0575***	0.0002***	0.0010	0.0119***
	(0.0007)	(0.0006)	(0.0000)	(0.0015)	(0.0012)
Controls	✓	✓	✓	✓	✓
Art unit × year demeaning	✓	✓	✓	✓	✓
Observations	5,364,971	5,367,642	5,364,971	3,770,120	3,812,246
R^2	0.086	0.093	0.009	0.016	0.010

Notes: Quadratic specification $y = \beta_1 x + \beta_2 x^2 + \gamma' X + \varepsilon$ estimated by OLS with CRV1 (ind_year_fe) standard errors in parentheses. *Net effect* is $\hat{\beta}_1(x_{hi} - x_{lo}) + \hat{\beta}_2(x_{hi}^2 - x_{lo}^2)$ over $x_{lo} = 0$ to $x_{hi} = 1$, with a delta-method standard error from the CRV1 (ind_year_fe) covariance of $(\hat{\beta}_1, \hat{\beta}_2)$. The upper evaluation point lies slightly beyond the standardized rank's support, whose upper endpoint is $\sqrt{3} \approx 1.73$ for a within-cell percentile, so the net effect is a fitted-quadratic contrast at two standard deviations rather than a sample-range change. The regressor is the standardized within-cell-demeaned percentile rank of quality except in column (2), where it is the identically treated percentile rank of filed scope. Outcomes and the quality index are standardized and demeaned within art unit × year cells before estimation, so the cell fixed effects enter as a pre-transformation rather than as absorbed dummies and R^2 is a within-cell R^2. Controls are a new-examiner indicator, continuation-type dummies, dummies for the first five years of examiner experience, and small- and micro-entity indicators. Stars: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Columns (1)–(4) correspond to Panels 2a, 2b, 2c and 2d of Figure 2; column (5) to Appendix Figure D.1.

Table K.3: *Conditional correlations: rejection reasons*

	(1) Non-usefulness	(2) Lack of novelty	(3) Obviousness	(4) Adequacy
Quality	−0.039***	−0.026**	0.478***	−0.178***
	(0.006)	(0.013)	(0.030)	(0.025)
Controls	✓	✓	✓	✓
Art unit × year FE	✓	✓	✓	✓
Observations	2,088,978	2,088,978	2,088,978	2,088,978
R^2	0.042	0.020	0.052	0.042

Notes: Linear specification $y = \beta q + \gamma' X + \alpha_{kt} + \varepsilon$ estimated on the rejection-reason subsample with art unit × year fixed effects absorbed. The dependent variable is an indicator for the rejection reason, so coefficients and standard errors are reported in *percentage points* (×100). Standard errors (CRV1 (ind_year_fe)) in parentheses. Controls are small- and micro-entity indicators and continuation-type and examiner-experience dummies. Stars: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Columns correspond to the four panels of Figure D.4.

K.2 Complier shares

The complier-share figures plot a kernel-smoothed curve in the percentile of the conditioning covariate. Tables K.4 to K.7 report that curve at deciles, together with the objects that determine the plotted bands: the simultaneous-band multiplier, the bootstrap sup-test of flatness, and the first-stage share and dose that scale the estimand.

Table K.4: *Complier share among granted patents, by quality and filed scope*

Percentile	(1) Quality	(2) Filed scope
0.10	4.71 (0.14)	4.04 (0.05)
0.20	4.63 (0.11)	4.21 (0.04)
0.30	4.61 (0.08)	4.33 (0.04)
0.40	4.57 (0.07)	4.51 (0.03)
0.50	4.59 (0.06)	4.70 (0.03)
0.60	4.67 (0.07)	4.91 (0.03)
0.70	4.81 (0.08)	5.13 (0.04)
0.80	4.94 (0.10)	5.38 (0.05)
0.90	4.87 (0.13)	5.58 (0.07)
Joint flatness p	0.052	0.000
Sup-t critical value	2.65	2.80
First-stage $\bar{\psi}$	0.667	0.667
Mean favourable deviation $\bar{d}$	0.0503	0.0503
Bandwidth	0.07	0.07
Bootstrap draws	2,000	2,000

Notes: Share of compliers among granted patents, $\phi_c^{D=1}$, in percent, from the examiner-dummy UJIVE modified-outcome estimator, evaluated at deciles of the conditioning covariate. Cluster-robust standard errors in parentheses. *Joint flatness p* is the multiplier-bootstrap sup-test of the null that the curve is constant; the sup-t critical value is the simultaneous-band multiplier plotted in the figure. $\bar{\psi}$ is the pooled first-stage pass-through, the derivative of the grant probability with respect to leniency, and $\bar{d} = \mathbb{E}[(\Delta l)_+]$ the mean *favourable* deviation of the realised examiner from the expected pool. Neither is itself a complier share: the share among granted patents is $\bar{\psi}\,\bar{d}/P(\text{grant})$. Curves are kernel-smoothed at the reported bandwidth; grid points where the covariate carries less than 5% of the local mass are shown as "–". Estimation sample: 5,364,969 applications across 12,705 examiners. Columns correspond to Panels 4a and 4b of Figure 4.

Table K.5: *Decomposition of the complier share: representation ratio and grant rate*

	Quality			Filed scope		
Percentile	$\rho(x)$	$P(D{=}1 \mid x)$	$\phi_c^{D=1}$	$\rho(x)$	$P(D{=}1 \mid x)$	$\phi_c^{D=1}$
0.10	0.992	0.706	4.71	0.927	0.770	4.04
0.20	0.976	0.707	4.63	0.956	0.761	4.21
0.30	0.974	0.708	4.61	0.970	0.750	4.33
0.40	0.968	0.709	4.57	0.993	0.738	4.51
0.50	0.974	0.711	4.59	1.014	0.724	4.70
0.60	0.992	0.712	4.67	1.037	0.708	4.91
0.70	1.023	0.713	4.81	1.055	0.690	5.13
0.80	1.053	0.715	4.94	1.064	0.663	5.38
0.90	1.041	0.715	4.87	1.045	0.627	5.58
$p90/p10$	1.049	0.987	1.035	1.127	1.227	1.383
Share of log gradient (%)	139	−39	100	37	63	100

Notes: Decomposition of the complier share of Table K.4 into its two moving parts. The identity $\phi_c^{D=1}(x) = \bar{\psi}\,\rho(x)\,\bar{d}\,/\,P(D{=}1 \mid x)$ holds exactly, so the columns multiply to the reported share rather than approximating it. $\rho(x)$ is the complier representation ratio, above one where compliers are over-represented relative to the population; $P(D{=}1 \mid x)$ is the conditional grant rate, a within-bin sample mean and not an instrumented quantity. The final rows report the $p90/p10$ ratio of each factor, with the denominator entered as $P(D{=}1 \mid p10)/P(D{=}1 \mid p90)$ so that a factor ratio above one raises the share across the range, and the share of the log gradient in $\phi_c^{D=1}$ that each factor carries. On filed scope the two factors reinforce, and the grant rate carries the larger part (63% against 37%). On quality they oppose: the representation ratio carries more than the whole gradient (139%) while the mildly rising grant rate offsets part of it, its inverted ratio 0.987 below one and its contribution −39%. Estimator, bandwidth and sample are those of Table K.4.

Table K.6: *Complier share: sample-restriction robustness*

Panel A: by quality

Percentile	(1) Baseline	(2) Excl. continuations	(3) Random-digit cells
0.10	4.71 (0.14)	4.87 (0.16)	6.30 (0.38)
0.20	4.63 (0.11)	4.72 (0.13)	6.11 (0.29)
0.30	4.61 (0.08)	4.56 (0.10)	5.92 (0.22)
0.40	4.57 (0.07)	4.43 (0.08)	5.66 (0.18)
0.50	4.59 (0.06)	4.38 (0.07)	5.36 (0.17)
0.60	4.67 (0.07)	4.40 (0.07)	5.04 (0.17)
0.70	4.81 (0.08)	4.48 (0.08)	4.85 (0.20)
0.80	4.94 (0.10)	4.59 (0.11)	4.67 (0.27)
0.90	4.87 (0.13)	4.52 (0.13)	4.28 (0.36)
Joint flatness p	0.052	0.103	0.012
Sup-t critical value	2.65	2.74	2.60
First-stage $\bar{\psi}$	0.667	0.686	0.761
Mean favourable deviation $\bar{d}$	0.0503	0.0472	0.0497
Bandwidth	0.07	0.07	0.07
Bootstrap draws	2,000	2,000	2,000

Panel B: by filed scope

Percentile	(1) Baseline	(2) Excl. continuations	(3) Random-digit cells
0.10	4.04 (0.05)	4.25 (0.06)	4.85 (0.14)
0.20	4.21 (0.04)	4.33 (0.05)	4.90 (0.10)
0.30	4.33 (0.04)	4.40 (0.04)	4.93 (0.08)
0.40	4.51 (0.03)	4.49 (0.04)	5.07 (0.07)
0.50	4.70 (0.03)	4.53 (0.04)	5.19 (0.07)
0.60	4.91 (0.03)	4.62 (0.05)	5.40 (0.08)
0.70	5.13 (0.04)	4.81 (0.06)	5.66 (0.09)
0.80	5.38 (0.05)	5.07 (0.07)	5.95 (0.12)
0.90	5.58 (0.07)	5.18 (0.09)	6.21 (0.15)
Joint flatness p	0.000	0.000	0.000
Sup-t critical value	2.80	2.83	2.81
First-stage $\bar{\psi}$	0.667	0.686	0.761
Mean favourable deviation $\bar{d}$	0.0503	0.0472	0.0497
Bandwidth	0.07	0.07	0.07
Bootstrap draws	2,000	2,000	2,000

Notes: Each column re-estimates the complier share of Table K.4 on a different sample, holding the examiner-dummy UJIVE modified-outcome estimator, the standard errors and the bands fixed. Sample sizes: Baseline 5,364,969; Excl. continuations 2,145,177; Random-digit cells 1,199,154. Reading conventions are as in Table K.4. Columns two and three correspond to Figures G.1 and G.2.

Table K.7: *Complier share in the replication subsamples*

Panel A: by quality

Percentile	(1) Sampat-Williams	(2) Farre-Mensa et al.
0.10	4.55 (0.84)	5.33 (0.88)
0.20	3.37 (0.70)	5.25 (0.75)
0.30	2.60 (0.99)	5.53 (0.70)
0.40	2.65 (1.45)	6.27 (0.70)
0.50	4.55 (1.08)	5.93 (0.71)
0.60	11.16 (2.77)	6.54 (0.75)
0.70	10.64 (4.47)	5.26 (0.73)
0.80	6.71 (2.92)	3.97 (0.74)
0.90	6.94 (3.62)	3.02 (0.89)
Joint flatness p	0.071	0.109
Sup-t critical value	2.78	2.85
First-stage $\bar{\psi}$	0.327	0.835
Mean favourable deviation $\bar{d}$	0.0362	0.0337
Bandwidth	0.07	0.07
Bootstrap draws	2,000	2,000

Panel B: by filed scope

Percentile	(1) Sampat-Williams	(2) Farre-Mensa et al.
0.10	4.56 (0.38)	4.44 (1.29)
0.20	5.29 (0.46)	4.00 (0.93)
0.30	–	5.29 (0.80)
0.40	23.98 (16.79)	5.72 (0.72)
0.50	5.53 (4.33)	6.03 (0.65)
0.60	0.85 (0.82)	5.80 (0.67)
0.70	3.61 (2.68)	5.05 (0.62)
0.80	17.26 (12.25)	4.26 (0.58)
0.90	10.92 (8.64)	4.73 (0.74)
Joint flatness p	0.000	0.704
Sup-t critical value	2.75	2.91
First-stage $\bar{\psi}$	0.327	0.835
Mean favourable deviation $\bar{d}$	0.0362	0.0337
Bandwidth	0.07	0.07
Bootstrap draws	2,000	2,000

Notes: The same examiner-dummy UJIVE modified-outcome estimator as Table K.4, applied within each replication subsample; for Sampat and Williams (2019) the examiner-leniency design is shown. Sample sizes: Sampat and Williams (2019) 233,442 gene-level observations across 81 examiners; Farre-Mensa et al. (2020) 10,286 applications across 2,196 examiners. The Sampat and Williams (2019) frame is gene-level, so its count is observations rather than applications; the underlying application count is smaller by roughly two orders of magnitude and is given with the full sample ladder in Appendix J. These subsamples are two to three orders of magnitude smaller than the full sample, so the standard errors are correspondingly wide and the point estimates should not be read at face value. Columns correspond to the two columns of Figure J.4.

K.3 Marginal treatment effects

The MTE figures plot three objects at once: the curve, the ATE obtained by integrating it, and the LATE from each paper's own published specification re-estimated on the same sample. Tables K.8 and K.9 report all three, so the ATE-versus-LATE comparison developed in Appendix J can be read numerically. The LATE row is the paper's own estimator and is a different estimand from the LATE implied by the MTE machinery; it is the former that the figures plot.

Tables K.10 and K.11 then test whether the two differ. The figures invite the wrong inference. Their ATE and LATE confidence bands overlap, and overlapping intervals are not a test of equality, not even for independent estimates, and these two are neither independent nor comparably constructed. The plotted ATE interval is a percentile bootstrap of the integrated curve; the plotted LATE interval is an analytic cluster-robust interval from a two-stage least squares fit. They come from different routines and share no draws, so nothing in them identifies the covariance between the two estimands, which is what the variance of their difference depends on. The tables therefore re-estimate both objects on a common set of bootstrap resamples and report the confidence interval of the difference, which handles that covariance by construction.

One convention in these tables differs from the figures. A test of two estimands is only interpretable if both are conditioned on the same covariates, so the LATE in the test is estimated under the control set that the marginal treatment effect's own outcome equation uses, on the identical estimation frame. For Sampat and Williams (2019) that set is the published one and the two coincide exactly. For Farre-Mensa et al. (2020) it adds the filed-scope control that the MTE outcome equation carries, so the LATE in Table K.11 is not numerically the LATE plotted in Figure J.3: it is 0.373 against 0.369 for follow-on applications, and at most 0.005 apart in any outcome, against a bootstrap standard error of roughly 0.07. The figures and Table K.9 report the published specification exactly, because there the LATE is a description of what the original design delivers rather than one side of a hypothesis test. Nothing in the verdicts below turns on the choice; conditioning both estimands alike is what makes the difference a well-defined object.

Four refinements matter for how the resulting p-values should be read. First, the raw percentile interval is only first-order accurate, and the resample distribution here is neither symmetric nor median-unbiased, since the observed difference sits away from the median of its own resample distribution in every outcome. The headline interval is therefore bias-corrected and accelerated, which is second-order accurate. The other second-order-accurate option, a studentized interval, is not available: the integrated ATE has no closed-form variance estimator, so studentizing would require an inner bootstrap inside every outer replicate. Second, the outcomes within each setting are strongly dependent, being different follow-on measures constructed on the same applications, so the familywise adjustment is a Romano-Wolf stepdown (Romano and Wolf, 2005, 2016) computed from the joint resample distribution rather than a Bonferroni or Holm correction, which would treat that dependence as absent and be badly conservative. Third, the number of replicates is itself an inferential choice rather than a computational detail: at 500 draws the Monte Carlo standard error of a p-value near 0.05 is roughly 0.010, wide enough that the borderline outcomes could not be assigned a side of the threshold, so the tables are computed at 5,000. Fourth, the

resampling unit is itself a choice. The tables above resample applications independently, which matches the convention of the estimating cell but not the level the standard errors cluster on, examiner art unit $\times$ publication year. Table K.12 repeats the Farre-Mensa et al. (2020) test resampling those clusters instead, redrawing cluster labels on every replicate. The estimands are unchanged by construction, being full-sample quantities that do not depend on the bootstrap, and the clustered bootstrap reproduces the analytic cluster-robust standard error of the LATE marginally more closely than the iid one does, at a mean ratio of 0.99 against 0.98 across the outcomes, which is what one expects when the dependence the clustering absorbs is modest. The signs, the ordering of the outcomes, and which of them survives the familywise adjustment are all unchanged. That table is computed at $B = 1{,}000$, however, so by the Monte Carlo logic just stated its borderline entries cannot be assigned a side of the five percent threshold: it is evidence on the resampling scheme, not a replacement for the tables above.

The verdicts match Appendix J: the ATE lies below the LATE in every Farre-Mensa et al. (2020) outcome under both integration windows, the difference rejects individually only in its two citation outcomes, one of which survives the Romano-Wolf adjustment, and neither primary outcome rejects equality of the two estimands. One detail is specific to these tables: both citation rejections are BCa p-values, and in both the correction moves the verdict toward rejection relative to the percentile interval, which is why the familywise column is reported alongside.

One diagnostic in these tables is of independent interest. The correlation between the ATE and the LATE across resamples never exceeds 0.13, even though the two are computed on identical rows. The integrated ATE loads on the fitted curve across the whole support of U, whereas the two-stage least squares LATE loads on variation in the complier region; the two draw on nearly disjoint information. A consequence is that the naive procedure of adding the two variances, which the joint bootstrap was designed to avoid, would have been approximately right here.

Table K.8: *Marginal treatment effects: Sampat and Williams (2019)*

	(1) log(Publications)	(2) Any publication
ATE (full support)	−0.017 [−0.146, 0.087]	−0.005 [−0.154, 0.115]
ATE (common support)	0.015 [−0.081, 0.099]	0.028 [−0.060, 0.126]
LATE (paper 2SLS)	−0.055 (0.089)	−0.051 (0.087)
ATE − LATE (full support)	0.039	0.046
ATE − LATE (common support)	0.070	0.079
MTE at $U = 0.10$	−0.090	−0.079
MTE at $U = 0.50$	0.097	0.109
MTE at $U = 0.90$	−0.169	−0.157
Observations	1,196	1,196

Notes: Marginal treatment effect estimates for Sampat and Williams (2019), on the variation used by the examiner-dummy UJIVE estimator (arm A: the leave-out deviation enters the selection probit as the instrument). ATE integrates the MTE curve over the full support of U; the common-support ATE integrates over the region where the propensity score is identified. Bootstrap 95% confidence intervals in brackets. The LATE row is each paper's *own* published 2SLS specification re-estimated on exactly this sample, with its cluster-robust standard error in parentheses – it is the blue line in the corresponding figure and is a different object from the MTE-machinery LATE. *Caveat*: the LATE row is unweighted across applications, whereas the original estimand of Sampat and Williams (2019) is gene-weighted. MTE rows report the curve evaluated at the stated latent resistance, linearly interpolated on the estimation grid. Column (1) corresponds to Panel J.3a of Figure J.3 and column (2) to Panel J.5a of Figure J.5.

Table K.9: *Marginal treatment effects: Farre-Mensa et al. (2020)*

	(1) log(Follow-on appl.)	(2) log(All cites, 5y)	(3) log(Appl. cites, 5y)	(4) log(Granted patents)
ATE (full support)	0.273	0.015	−0.042	0.216
	[−0.027, 0.529]	[−0.262, 0.314]	[−0.189, 0.120]	[0.028, 0.405]
ATE (common support)	0.340	0.102	0.005	0.264
	[0.055, 0.559]	[−0.168, 0.365]	[−0.138, 0.147]	[0.070, 0.430]
LATE (paper 2SLS)	0.373	0.366	0.212	0.291
	(0.068)	(0.073)	(0.038)	(0.052)
ATE − LATE (full support)	−0.100	−0.351	−0.254	−0.075
ATE − LATE (common support)	−0.033	−0.264	−0.207	−0.027
MTE at $U = 0.10$	−0.701	−1.223	−0.681	−0.426
MTE at $U = 0.50$	0.331	0.099	0.027	0.297
MTE at $U = 0.90$	1.120	1.063	0.450	0.697
Observations	10,286	10,286	10,286	10,286

Notes: Marginal treatment effect estimates for Farre-Mensa et al. (2020), on the variation used by the examiner-dummy UJIVE estimator (arm A: the leave-out deviation enters the selection probit as the instrument). ATE integrates the MTE curve over the full support of U; the common-support ATE integrates over the region where the propensity score is identified. Bootstrap 95% confidence intervals in brackets. The LATE row is each paper's *own* published 2SLS specification re-estimated on exactly this sample, with its cluster-robust standard error in parentheses – it is the blue line in the corresponding figure and is a different object from the MTE-machinery LATE. MTE rows report the curve evaluated at the stated latent resistance, linearly interpolated on the estimation grid. Column (1) corresponds to Panel J.3b of Figure J.3; columns (2)–(4) to Panels J.5b, J.5c and J.5d of Figure J.5.

Table K.10: *Testing the average against the local average treatment effect: Sampat and Williams (2019)*

	(1) log(Publications)	(2) Any publication	(3) Genetic test
Estimands			
ATE (full support)	−0.017	−0.005	0.013
ATE (common support)	0.015	0.028	0.008
LATE (paper 2SLS)	−0.055	−0.051	0.025
Test of ATE = LATE, full-support integral			
Difference	0.039	0.046	−0.012
95% BCa CI	[−0.109, 0.320]	[−0.119, 0.321]	[−0.143, 0.183]
p (BCa)	0.568	0.524	0.910
p (percentile)	0.667	0.606	0.788
p (Romano-Wolf, FWER)	0.830	0.830	0.855
Test of ATE = LATE, common-support integral			
Difference	0.070	0.079	−0.018
95% BCa CI	[−0.051, 0.380]	[−0.039, 0.439]	[−0.133, 0.143]
p (BCa)	0.253	0.197	0.792
p (percentile)	0.342	0.318	0.740
p (Romano-Wolf, FWER)	0.492	0.492	0.749
Diagnostics			
Corr(ATE, LATE) across resamples (full support)	0.063	0.018	0.015
Bootstrap s.e. of LATE	0.225	0.233	0.214
Cluster-robust s.e. of LATE	0.089	0.087	0.041
BCa bias correction $\hat{z}_0$ (full support)	0.069	0.060	0.077
BCa acceleration $\hat{a}$ (full support)	0.0147	0.0032	0.0365
Usable bootstrap replicates	5,000	5,000	5,000
Jackknife groups	50	50	50
Observations	1,196	1,196	1,196

Notes: Test of the null that the integrated ATE equals the LATE, for Sampat and Williams (2019). Both estimands are re-estimated on *the same* 5,000 iid row resamples, so the interval for their difference accounts for the covariance between them – which neither band in the corresponding figure contains, so that overlap is not a test of equality. ATE (full support) integrates the MTE curve over all of U, ATE (common support) only where the propensity score has support, and LATE is each paper's *own* published 2SLS estimator, re-estimated on this sample under the control set of the marginal-treatment-effect outcome equation so that both estimands are conditioned alike. The BCa p inverts a bias-corrected and accelerated interval and is the single-hypothesis test, so the stars follow it; the percentile p is its first-order-accurate counterpart; the Romano-Wolf p is a stepdown adjustment controlling the familywise error rate across the outcomes of this table. The examiner-leniency design is shown, matching the figures; the alternative design is estimated on the same sample. Diagnostics: the resample correlation of the two estimands is the quantity the naive add-the-variances test assumes away; the bootstrap and cluster-robust standard errors of the LATE are shown side by side as a check on whether iid resampling understates within-cell dependence; $\hat{z}_0$ and $\hat{a}$ are the BCa median-bias and skewness constants, both zero when BCa and the percentile method coincide. The LATE row is Sampat and Williams's specification re-estimated *unweighted* on this sample. Their published estimand weights by the number of genes per application; re-estimating with those weights reverses the sign of every LATE reported here while leaving each one statistically indistinguishable from zero, so the null result is robust to the choice of estimand but the signs are not. The weighted estimates are reported in Appendix J. Sample: 1,196 applications. Column (1) corresponds to Panel J.3a of Figure J.3 and column (2) to Panel J.5a of Figure J.5; column (3) is not plotted. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table K.11: *Testing the average against the local average treatment effect: Farre-Mensa et al. (2020)*

	(1) log(Follow-on appl.)	(2) log(All cites, 5y)	(3) log(Appl. cites, 5y)	(4) log(Granted patents)	(5) Approval rate
Estimands					
ATE (full support)	0.269	0.018	−0.024	0.221	0.138
ATE (common support)	0.318	0.081	0.008	0.255	0.161
LATE (paper 2SLS)	0.373	0.362	0.211	0.291	0.200
Test of ATE = LATE, full-support integral					
Difference	−0.104	−0.344**	−0.235***	−0.070	−0.062
95% BCa CI	[−0.379, 0.209]	[−0.672, −0.041]	[−0.408, −0.083]	[−0.275, 0.163]	[−0.214, 0.097]
p (BCa)	0.521	0.029	0.003	0.607	0.454
p (percentile)	0.443	0.038	0.006	0.457	0.378
p (Romano-Wolf, FWER)	0.660	0.088	0.017	0.660	0.660
Test of ATE = LATE, common-support integral					
Difference	−0.056	−0.281*	−0.203***	−0.036	−0.039
95% BCa CI	[−0.302, 0.241]	[−0.567, 0.014]	[−0.356, −0.057]	[−0.219, 0.193]	[−0.175, 0.109]
p (BCa)	0.775	0.062	0.008	0.862	0.663
p (percentile)	0.584	0.056	0.010	0.590	0.512
p (Romano-Wolf, FWER)	0.834	0.152	0.028	0.834	0.834
Diagnostics					
Corr(ATE, LATE) across resamples (full support)	0.073	0.044	0.079	0.067	0.041
Bootstrap s.e. of LATE	0.068	0.072	0.039	0.050	0.036
Cluster-robust s.e. of LATE	0.069	0.074	0.038	0.052	0.038
BCa bias correction $\hat{z}_0$ (full support)	0.057	−0.095	−0.172	0.112	0.068
BCa acceleration $\hat{a}$ (full support)	0.0225	0.0190	0.0152	0.0134	−0.0029
Usable bootstrap replicates	5,000	5,000	5,000	5,000	5,000
Jackknife groups	50	50	50	50	50
Observations	10,289	10,289	10,289	10,289	10,289

Notes: Test of the null that the integrated ATE equals the LATE, for Farre-Mensa et al. (2020). Both estimands are re-estimated on *the same* 5,000 iid row resamples, so the interval for their difference accounts for the covariance between them – which neither band in the corresponding figure contains, so that overlap is not a test of equality. ATE (full support) integrates the MTE curve over all of U, ATE (common support) only where the propensity score has support, and LATE is each paper's *own* published 2SLS estimator, re-estimated on this sample under the control set of the marginal-treatment-effect outcome equation so that both estimands are conditioned alike. The BCa p inverts a bias-corrected and accelerated interval and is the single-hypothesis test, so the stars follow it; the percentile p is its first-order-accurate counterpart; the Romano-Wolf p is a stepdown adjustment controlling the familywise error rate across the outcomes of this table. Diagnostics: the resample correlation of the two estimands is the quantity the naive add-the-variances test assumes away; the bootstrap and cluster-robust standard errors of the LATE are shown side by side as a check on whether iid resampling understates within-cell dependence; $\hat{z}_0$ and $\hat{a}$ are the BCa median-bias and skewness constants, both zero when BCa and the percentile method coincide. That conditioning is why the LATE here differs from the one in Table K.9, by up to 0.004: that table and the figures report the published specification exactly, since there the LATE describes what the original design delivers rather than forming one side of a hypothesis test. Sample: 10,289 applications. Column (1) corresponds to Panel J.3b of Figure J.3; columns (2)–(4) to Panels J.5b, J.5c and J.5d of Figure J.5; column (5) is not plotted. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table K.12: *Testing the average against the local average treatment effect under a clustered bootstrap: Farre-Mensa et al. (2020)*

	(1) log(Follow-on appl.)	(2) log(All cites, 5y)	(3) log(Appl. cites, 5y)	(4) log(Granted patents)
Estimands				
ATE (full support)	0.269	0.018	−0.024	0.221
ATE (common support)	0.318	0.081	0.008	0.255
LATE (paper 2SLS)	0.373	0.362	0.211	0.291
Test of ATE = LATE, full-support integral				
Difference	−0.104	−0.344**	−0.235***	−0.070
95% BCa CI	[−0.380, 0.201]	[−0.680, −0.057]	[−0.418, −0.094]	[−0.282, 0.159]
p (BCa)	0.498	0.026	0.002	0.531
p (percentile)	0.438	0.062	0.010	0.508
p (Romano-Wolf, FWER)	0.687	0.084	0.021	0.687
Test of ATE = LATE, common-support integral				
Difference	−0.056	−0.281**	−0.203***	−0.036
95% BCa CI	[−0.299, 0.233]	[−0.592, −0.001]	[−0.366, −0.065]	[−0.229, 0.180]
p (BCa)	0.765	0.050	0.004	0.759
p (percentile)	0.586	0.080	0.012	0.644
p (Romano-Wolf, FWER)	0.853	0.145	0.031	0.853
Diagnostics				
Corr(ATE, LATE) across resamples	0.176	0.159	0.192	0.117
Bootstrap s.e. of LATE	0.069	0.072	0.037	0.053
Cluster-robust s.e. of LATE	0.069	0.074	0.038	0.052
BCa bias correction $\hat{z}_0$	0.050	−0.164	−0.264	0.018
BCa acceleration $\hat{a}$	−0.0044	−0.0063	0.0003	−0.0003
Usable bootstrap replicates	1,000	1,000	1,000	1,000
Jackknife groups	50	50	50	50
Observations	10,289	10,289	10,289	10,289

Notes: Companion to Table K.11, which reports the same test under an iid row resample. Here the bootstrap resamples *clusters* of examiner art unit × publication year, matching the unit the standard errors cluster on: 873 clusters, 1,000 usable replicates, with cluster labels redrawn on each replicate. The ATE and LATE rows are identical to Table K.11 by construction: they are full-sample quantities and do not depend on the resampling scheme. Only the intervals and p-values differ. *This table is not a corrected version of Table K.11*: it uses $B = 1{,}000$ replicates against 5,000 there, so the two differ in resampling scheme and in number of replicates simultaneously, and any difference between them cannot be attributed to the scheme alone. The Romano-Wolf stepdown controls the familywise error rate across the full 5-outcome bootstrap family, which includes 1 outcome not displayed here (the approval rate). Stars follow the BCa p-value. Columns correspond to those of Table K.11. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

L Structural model: specification, identification, and fit

This appendix collects the full specification of the dynamic prosecution model of Section 7, states which empirical moments identify each structural parameter, and reports the fit of the estimated model. The model is estimated by indirect inference (Gouriéroux et al., 1993) (Section 7.5): the parameter vector θ is chosen to minimize the weighted distance between a set of binned empirical moments and their simulated counterparts, the latter evaluated on a panel of seven million applications drawn from the estimated primitives and solved by value-function iteration at each candidate θ. Optimization proceeds in two stages: a low-discrepancy Sobol sweep of Θ screens the global surface for promising basins, and from the best Sobol points a covariance-matrix-adaptation evolution strategy with IPOP restarts (Hansen, 2006) and a local Nelder-Mead polish deliver the reported minimizer. Common random numbers are held fixed across parameter evaluations, so differences in the criterion reflect θ rather than the draw. The three subsections that follow list the parameters and their estimates (Appendix L.1), the identification argument and the inference procedure (Appendix L.2), and the fit (Appendix L.3); the last includes an untargeted comparison of the model's implied complier shares to the reduced-form estimates of Section 6.

L.1 Parameters

Table L.1 lists the parameters, their economic role, the admissible bounds imposed during estimation, and the point estimate at the converged run. Curvature χ, the examiner signal noise σ_η, and the offer's leniency gradient γ_l (held at its data-anchored value) are calibrated and held fixed; the remainder are estimated by indirect inference, with α_q reported as a pinned normalization (see the table note).

The scope penalty $\Pi(s) = \alpha_1 s + \alpha_2 s^2 + \alpha_3 s^3$ is a fitted polynomial rather than a shape imposed a priori, and it is not monotone on all of $[0,1]$: at the estimates the grant index is increasing in scope over a low-scope region and decreasing thereafter. This does not compromise the design-based results of Sections 5–6, which use only the sign of $\partial a/\partial s$ where the filings actually are. That sign is negative and is identified directly from the acceptance-by-scope profile; nothing in those sections requires Π to be globally increasing. The cubic term is what lets the acceptance-by-scope profile peak away from the interior midpoint, which the data require; the quadratic alone forces a symmetric peak.

Two naming points, since both have caused confusion. The applicant's belief dispersion q_{unc} and the examiner's prior standard deviation $\sigma_q = 1/\sqrt{12}$ are *different objects* and are not alternative values of one quantity: q_{unc} governs the subjective density an applicant holds over own quality (Section 7.1) and is estimated, while σ_q enters only the examiner's precision weight f_r in (14) and is calibrated to the spread of the quality rank. They are grouped separately below for that reason. The round drift of the grant index is b_r throughout; δ denotes the social cost of granted scope in Section 7.7 and nothing else.

The shock variance and the quality pivot. The grant-index shock of (15) combines two sources,

$$\sigma^2_{\text{tot}}(r) \;=\; \sigma_e(r)^2 + \sigma^2_{\text{grant}}, \qquad \sigma_e(r) \;=\; \alpha_q\, f_r\, \frac{\sigma_\eta}{\sqrt{r+1}}, \tag{32}$$

the first term being the dispersion of the examiner's posterior around its signal-integrated mean, equal to $\alpha_q \sigma_q \sqrt{f_r(1-f_r)}$ and therefore *rising* across rounds over the observed range, from $0.028\alpha_q$ at $r=0$ to $0.064\alpha_q$ at $r=5$; what falls is the posterior standard deviation itself, and only slightly ($0.287 \to 0.281$). The second, σ_{grant}, is an idiosyncratic case-specific component that does not vanish as $\alpha_q \to 0$ and so keeps the grant probability interior when quality carries little weight. The simulation draws the round-r shock independently across rounds, although the σ_e component is not in fact independent: successive posteriors share signals, so $\text{Corr}(r, r') = \sqrt{(r+1)/(r'+1)}$, rising from 0.71 to 0.91 over the observed range. The approximation is immaterial by construction rather than by calibration: $\sqrt{f_r(1-f_r)} \leq \frac{1}{2}$, so $\sigma_e \leq \alpha_q \sigma_q / 2$, which at the estimated α_q is 1.8% of σ_{grant} for *any* value of σ_η. The bound binds through α_q, not through the signal noise.

Quality enters the grant index only through the posterior mean, which is affine in true quality: substituting the signal-integrated mean of (14), $\alpha_q \, \mathbb{E}_r[q] = \alpha_q \bar{q} + \alpha_q f_r \, (q - \bar{q})$. The posterior-mean profile therefore pivots about $q = \bar{q}$ as rounds accumulate, steepening in f_r while its value at the prior mean is unchanged: applications above $\bar{q}$ are treated more generously the longer prosecution runs and those below it more harshly, and no amount of learning moves the profile uniformly, so a trend in generosity common to all applications has to come from b_r and b_0. That much is a property of the functional form and holds at any parameter values; at the calibrated f_r the slope $\alpha_q f_r$ is small enough that the pivot is invisible in the fitted profiles.

What the applicant does not internalize in the offer. The applicant's value function (18) is computed against the full filed claim rather than the offer-scaled granted scope. The restriction is weaker than it first appears: because $\hat{s}_r$ is proportional to s_r on the estimated support, the constant component of the concession enters the grant payoff only as a factor scaling ν, and is therefore absorbed by the estimated ν. What the assumption removes is the applicant conditioning on the round decay γ_{E}^r and the scope gradient $\gamma_{s\text{-sl}}$ of the offer ratio, which would otherwise feed back into the stopping and scope margins.

Table L.1: *Structural parameters, roles, and estimates*

Parameter	Role	Bounds	Estimate
Examiner grant rule			
α_q	Quality loading in grant index	[0, 0.06]	0.052¶
α_1	Linear scope penalty	[−0.5, 2.0]	−0.496†
α_2	Quadratic scope penalty	[0, 2.0]	0.607
α_3	Cubic scope penalty	[0, 0.02]	0.014
μ	Grant threshold	[0.1, 1.5]	0.676
σ_{grant}	Idiosyncratic grant noise	[0.25, 0.45]	0.407
b_r	Round drift of grant index	[0.01, 0.25]	0.039
b_0	First-round grant barrier	[0.50, 0.60]	0.592
Applicant beliefs			
q_{unc}	Dispersion of belief over own quality	[0, 1.0]	0.808
q_{err}	Bias of belief over own quality	calibrated	0.0
σ_{upd}	Bandwidth of the belief-updating kernel	[0, 0.5]	0.363
Applicant scope value and cost			
κ	Scope-cost scale	[0.05, 1.5]	0.255
χ	Scope-cost curvature	calibrated	2.0
ν	Filing-value multiplier (scope_mult)	[1.0, 4.0]	2.885
σ_s	Scope-policy dispersion (scope_e)	[0.2, 0.5]	0.294
c_0	Fixed revision cost	[0.009, 0.35]	0.243
c_s	Scope revision cost	[0, 2.0]	0.000†
c_r	Round revision cost	[0, 0.25]	0.245
c_q	Quality persistence in revision cost	[0, 0.30]	0.082
c_{q^2}	One-sided quality persistence (top half)	[0, 0.8]	0.227
σ_c	Revision-cost dispersion	[0.05, 0.65]	0.650†
Scope policy and examiner offer			
β_s	Filed-scope loading on leniency	[−0.05, 0.5]	−0.022
β_q	Filed-scope loading on quality	[−0.5, 0.5]	−0.004
γ_s	Base offer ratio (granted/filed scope)	[0.25, 0.95]	0.756
γ_l	Leniency gradient of the offer ratio	calibrated	0.141
$\gamma_{s\text{-sl}}$	Filed-scope gradient of the offer ratio	[0, 1.0]	0.000†
γ_q	Quality gradient of the offer ratio	[−0.05, 0.02]	−0.012
γ_{A}	Scope-concession decay	[0.50, 1.00]	0.723
γ_{E}	Per-round decay of the offer (scope_decay)	[0.70, 1.00]	0.999†
Calibrated (not estimated)			
σ_η	Examiner signal noise	fixed	3.0
$\bar{q}$	Examiner prior mean of quality	fixed	0.5
σ_q	Examiner prior SD of quality	fixed	$1/\sqrt{12}$

Notes: Point estimates from the converged indirect-inference run (a warm-started CMA-ES with a Nelder-Mead polish; full-sample weighted moment distance 827.7). This is the run used throughout Sections 7.5 and 7.7. The estimator minimizes the weighted distance between simulated and empirical moments (Section 7.5). Calibrated constants: curvature $\chi = 2.0$ (quadratic filing cost, breaking the scope-multiplier/κ confound), $\sigma_\eta = 3.0$, and the offer's leniency gradient $\gamma_l = 0.141$ (data-anchored to the measured round-zero conditional leniency slope). q_{err} is calibrated to zero, being collinear with q_{unc}. † Estimate on a bound, within 1% of the searched interval: an active constraint, for which the sandwich variance does not apply and no standard error is reported (Appendix L.2). Two further estimates sit near a bound without reaching it, c_r at 98.0% and b_0 at 92.0% of their intervals. ¶ Pinned as the level normalization of the grant index: because the examiner's posterior mean averages one half at every round, (α_q, μ) move the acceptance index along a single direction, so only one coordinate of the pair carries independent information and no standard error is reported for α_q (Appendix L.2).

Table L.2: *The filed-scope quality gradient by filing route*

Filing route		Obs.	Share	Grant rate	$\partial s/\partial q$	
Claims derived from a prior document						
Continuation	CON	371,235	9.7%	0.797	−0.0151	(0.0015)
Divisional	DIV	134,951	3.5%	0.787	−0.0144	(0.0026)
National-stage entry	NST	657,234	17.2%	0.693	−0.0143	(0.0012)
Regime total		1,163,420	30.5%		−0.0154	(0.0009)
Claims drafted without a prior US claim set						
No continuity record	–	1,529,525	40.1%	0.713	+0.0068	(0.0007)
Continuation-in-part	CIP	105,273	2.8%	0.699	+0.0013	(0.0028)
Provisional-based	PRO	1,013,976	26.6%	0.684	−0.0011	(0.0009)
Regime total		2,648,774	69.5%		+0.0033	(0.0005)
Pooled (the estimator's target)		3,812,246	100.0%		−0.0027	(0.0013)

Notes: Slope of FE-residualised filed scope on invention quality, among accepted applications, by filing route. Built on the estimator's own basis: 2005–2020 core sample, quality as a within-sample percentile, filed scope as the pooled filed-plus-granted claim-length percentile within publication year and oriented so that higher values are broader, residualised on art unit × year, filing route, examiner experience, first-year-active, and entity size. The pooled row is the object the `scp` moment targets and reproduces its recorded value of −0.00271; the script aborts if it does not. Each regime-total row is a separate regression on the pooled regime subsample, not an average of the route rows above it, so a total can sit outside the range its routes span when their quality distributions differ. Routes are the USPTO continuity codes; *No continuity record* means the application does not appear in the continuity file and so is a residual category rather than a filing type. Reissues and substitutes (141 observations combined) are omitted. The three derived routes agree to within 0.0008 despite being distinct procedures, while the spread across all six routes is 0.0219, some 8.1 times the pooled slope, and the two regimes carry opposite signs. Standard errors are heteroskedasticity-robust. Absorbing applicant (`customer_number`) fixed effects moves the unresidualised gradient from −0.0004 to −0.0035, so the pattern is a property of applications rather than of who files them.

Horizon. The finite horizon $R = 12$ is a computational approximation rather than a statutory limit, so it has to be slack at the estimates for the round-cap counterfactuals of Section 7.7 to read as reductions in prosecution length rather than as artifacts of where the model was truncated. Re-solving the model at $\hat{\theta}$ over horizons $R \in \{3, 4, 6, \ldots, 16\}$ under common random numbers, the mass of applications still live at the terminal round is zero at $R = 12$ in a seven-million-application panel and 3×10^{-7} already at $R = 10$; the aggregate grant rate has converged by $R = 6$ and moves by at most 0.02% across every horizon at or above the baseline. Lengthening the horizon to sixteen rounds changes nothing, so the grant losses reported in Table 2 are caused by the caps being short, not by the baseline being close to an edge.

L.2 Identification and inference

Although all parameters are estimated jointly, each is identified primarily by a distinct feature of the data. Table L.7 states the leading source of identifying variation for each parameter, together with the weight the moment carries in the objective; Table L.8 groups the same mapping into parameter blocks and gives the economic reasoning behind each assignment. The argument is heuristic, in the spirit of the sensitivity measure of Andrews, Gentzkow, and Shapiro (2017): each parameter is assigned to the moment set to which the simulated moments $m^{\text{sim}}(\theta)$ are most responsive at $\hat{\theta}$. The acceptance-by-leniency profile carries

the dominant weight (200) because it pins the central mechanism–the slope of the grant probability in leniency.

The weighting matrix. Two choices in W depart from textbook practice and are made deliberately. First, each binned profile is standardized by the cross-bin standard deviation of its data reference curve rather than by per-bin sampling standard errors. With hundreds of thousands of applications per bin, per-bin standard errors shrink toward zero and a conventional efficient (inverse-sampling-variance) weighting would let residual bin-level noise dominate the criterion; standardizing by the amplitude of each profile across bins instead targets the *shape* of the conditional-mean curve, the slope and curvature that identify the scope penalties, the filing gradients, and the concession and offer parameters. Second, W is diagonal, combining that within-profile standardization with a small set of across-profile weights that equalize the influence of shape-carrying moments that would otherwise be swamped; the acceptance-by-leniency profile is the clearest case, since its amplitude is small but its slope identifies σ_{grant}. Both choices trade asymptotic efficiency for identification stability, which is why the across-profile weights are part of the estimand and not only of its efficiency (Section 7.5).

The calibrated leniency gradient. The offer's leniency gradient γ_l is held at a data-anchored value rather than estimated. At round zero (16) implies a granted-to-filed ratio of exactly $\pi(l)$, so the leniency slope of that round-zero ratio pins γ_l free of the acceptance selection that flattens the unconditional granted-scope profile. The round-zero ratio is not itself an auxiliary statistic in (20), so the anchoring adds no moment to the objective: granted scope by leniency remains a weighted target that the estimated parameters have to fit.

Two assignments are easily confused. The concession rate γ_{A} is identified from how granted scope tracks filed scope across the accepted, while the per-round decay γ_{E} is identified from the round-varying scope-leniency slope. Separately, the leniency loading of filed scope noted in Section 7.3 is identified from the scope-by-leniency profile as an equilibrium co-movement, and is disciplined separately from the offer's leniency gradient γ_l, which moves *granted* rather than filed scope.

Local identification. The assignment above is heuristic, so I check it against the object that actually governs local identification: the curvature of the objective, $\Lambda' W \Lambda$, with $\Lambda = \partial m^{\text{sim}}/\partial \theta'$ evaluated at $\hat{\theta}$ by central finite differences under common random numbers. Its columns are expressed in bound-range units, so a unit move is the same fraction of the searched interval for every parameter and the diagnostic is not an artifact of scaling. The stacked Jacobian is $1{,}126 \times 26$ over the 15 active moment sets.

The spectrum is full rank but sharply graded. Its three smallest eigenvalues, in bound-range units, load almost entirely on three parameters: the smallest, at 3.3×10^{-5}, is the offer's quality gradient γ_q nearly alone, and the next two are the belief pair, σ_{upd} against q_{unc}, in their two orthogonal combinations. These are the directions the objective measures worst, and they are the same three parameters whose sandwich errors in Table L.3 run several times their estimates; what the moment set cannot measure with any precision is how fast applicants learn.

One weakly identified direction does not appear in this spectrum, because bound-range units conceal

it: the searched interval for α_q is narrow, so its column is short in these units even though (α_q, μ) move the acceptance index along a single line, one unit of α_q against half a unit of μ. Re-scoring the estimated model directly shows that the compensated reassignment, α_q to zero with μ lowered by $\alpha_q/2$, reproduces the criterion to within simulation noise, while displacing μ alone by ± 0.005 costs several hundred weighted units. The pair spans one identified direction, and the inference below handles it by pinning α_q as the level normalization of the grant index.

A second near-flat pair sits in the offer block. The Jacobian columns of the base offer ratio γ_s and the offer decay γ_{E} are correlated at 0.99 at $\hat{\theta}$: because most acceptances occur early and the granted claim is a near-common multiple of the filed claim there, the level and the per-round decay of the offer move the moments along nearly the same direction, and the round-resolved endpoint variation separates them only weakly. This is consistent with γ_{E}'s estimate sitting against its bound, and the identification table's assignment of the decay to the offer's fall across rounds should be read with that near-collinearity in mind. The applicant-side decay γ_{A} is not part of the pair; its column correlates with either at roughly 0.45.

A third flat direction is economic rather than numerical, and a probe prices it directly. The applicant's problem is homogeneous of degree one in the payoff block, the value multiplier ν, the scope cost κ, and the revision-cost coefficients: scaling all seven jointly leaves every policy and every simulated moment unchanged, and no auxiliary statistic observes a payoff level. Re-scoring the estimated model with the block scaled by ± 2 percent moves the criterion by at most a few weighted units, against several hundred for a comparable proportional move in μ alone and a simulation-noise floor of roughly twenty, and leaves every fit diagnostic unchanged to the reported digit. The payoff scale is therefore a normalization, not an estimate: I read ν as the numeraire, the remaining six cost coefficients as relative to it, and the block's standard errors as conditional on that normalization.

The reason is the one given in Section 7.2. With the signal noise σ_η calibrated rather than estimated and large enough to leave the examiner's posterior near its prior over the observed round range, belief updating barely moves any simulated moment, so the profiles that were supposed to identify the updating rate–the level and dispersion of revision rounds by quality–carry little about it. The corresponding row of Table L.7 should be read as the moment set that would identify σ_{upd} in a specification where learning had bite, not as one that does. Its estimate is accordingly reported with a standard error several times its size, and nothing in Section 7.7 depends on it: the counterfactuals move the grant rule and the offer, and the learning block enters only through prosecution length, which the model matches through the cost parameters.

The per-round scope decay γ_{E} sits at its upper bound, and a weakly measured margin and a corner are the same finding seen twice: the objective gives the estimator little reason to hold γ_{E} away from unity, so it does not. The estimate $\hat{\gamma}_{\mathrm{E}} \approx 1$ means the offer's scope concession does not decay across rounds, and Section 7.7's round-cap counterfactual should be read with that in mind.

Inference. Under standard indirect-inference regularity conditions (Gouriéroux et al., 1993), $\hat{\theta}$ is consistent and asymptotically normal with sandwich variance

$$\widehat{\mathrm{Var}}(\hat{\theta}) \;=\; \left(1+\tfrac{1}{S}\right)\left(\Lambda' W \Lambda\right)^{-1} \Lambda' W\, \widehat{\Omega}\, W \Lambda \left(\Lambda' W \Lambda\right)^{-1}, \tag{33}$$

where Λ is the binding-function Jacobian, $\widehat{\Omega}$ the asymptotic variance of the empirical auxiliary statistics, W the weighting matrix of Section 7.5, and $S = N_{\mathrm{sim}}/N$ the simulation ratio, whose finite value inflates the variance by $1 + 1/S$. The bread is the Gauss–Newton form: under misspecification the population criterion's Hessian carries an additional residual-weighted curvature term of the binding function, which (33) omits, so the variance is the standard local approximation at the pseudo-true parameter rather than a misspecification-robust one.

Two features of $\hat{\theta}$ govern what (33) can deliver. The first is the (α_q, μ) direction established above, which the Jacobian measures poorly precisely because it is flat: a sandwich error for either coordinate alone would describe a near-singular inversion rather than the sample. I therefore pin α_q as the level normalization of the grant index, at no cost to the fit, since the compensated reassignment is exact, and invert the remaining 25×25 block, which has full rank, exactly rather than taking a pseudo-inverse of the whole. Every column of Λ is non-degenerate at $\hat{\theta}$, and the flat directions above are joint combinations rather than columns, so nothing else is removed from the inversion; the payoff-block standard errors it produces are read as conditional on the scale normalization, as stated above.

Standard errors for the twenty-five follow from (33) with $\widehat{\Omega}$ estimated by a block bootstrap over examiners, the level at which the instrument varies, on 500 resamples. Twenty of the twenty-five are more than two standard errors from zero. The five that are not are informative about where the model's information runs out rather than about noise: c_s and $\gamma_{s\text{-sl}}$ sit at their lower bounds within one standard error of zero, and the belief pair q_{unc} and σ_{upd} and the offer gradient γ_q are the genuinely imprecise interior estimates, the same parameters the flattest directions of the spectrum load on. For the coordinates on bounds the sandwich is at best descriptive: a constrained estimator's limiting distribution is a projection onto the feasible cone rather than the interior normal, which is why the parameter table declines to print their standard errors; where this section states such a coordinate in standard-error units, the unit is the unrestricted sandwich scale, and no comparison with zero for those coordinates is a test.

The second requirement is that $\hat{\theta}$ lie in the interior, and five of the twenty-six do not: α_1, c_s and $\gamma_{s\text{-sl}}$ at their lower bounds, and σ_c and γ_{E} at their upper bounds. At a bound the asymptotic distribution is not normal and a symmetric interval misstates it in a known direction, so the reported errors for those five should be read as describing curvature rather than as confidence intervals. The case worth stating explicitly is γ_{E}, whose estimate of 0.9994 lies 0.03 standard errors from its upper bound of one: the data are consistent with no per-round decay in the offer at all, which is the reading Section 7.7 attaches to the round-cap counterfactual. Table L.6 reports the position of every estimate within its box, together with the condition number.

An over-identification test rejects, and the useful thing is to say so and explain what it does and does not establish. The full-dimensional statistic is not available: forming it requires inverting $\widehat{\Omega}$ across all 1,126 bins, and a bootstrap covariance built from 500 resamples has rank at most 499, so any such

quantity would live on a subspace and its degrees of freedom would not be what they appear. Averaging each moment set to five bins reduces the problem to 74 dimensions, where the bootstrap comfortably supports an inverse; the statistic is then 1.4 million against 49 nominal degrees of freedom, and at ten bins per set 2.9 million against 109. I read these as descriptive magnitudes rather than calibrated tests: the estimates minimize a differently weighted criterion over the original bins, so the aggregated statistic does not inherit a χ^2 reference distribution. Both reject at any conventional level.

That is in part what an over-identification test does to a structural model at $n = 5.4$ million. The auxiliary statistics are estimated so precisely that a discrepancy far too small to matter economically is still many standard errors wide, and a consistent test rejects any fixed approximation error once the sample is large enough. But that argument is available to every structural paper and establishes only the unsurprising part, that the model is not literally true. The informative question is where the statistic comes from and whether it reaches the estimates, and both can be answered directly.

Table L.4 answers the first. It reports each profile's residual in its own units alongside its share of J, the two being the same misfit priced on different scales. The distribution is concentrated: the abandonment hazard by round supplies 84 percent of the statistic on its own, and it is the one profile the model genuinely fails, placing the hazard at 0.53 by the fifth round against 0.29 in the data. The model abandons too readily and increasingly so as prosecution lengthens. Of the remainder, acceptance by filed scope contributes 7.6 percent, acceptance by leniency 6.3, and granted scope by filed scope 3.0; no other profile reaches one percent, and outside the abandonment hazard the largest discrepancy anywhere in the objective is 0.12 on the scope percentile scale and the typical one is under 0.03. The statistic is not diagnosing a model that fits nothing precisely. It is diagnosing one profile that fits badly, against fourteen that fit closely.

Table L.5 answers the second, and is the reason the first answer matters. It reports the sensitivity of the estimates to the auxiliary statistics in the sense of Andrews et al. (2017), as the movement in each parameter induced by shifting an entire profile by one bootstrap standard deviation, expressed in that parameter's own standard errors. The abandonment hazard is the *least* influential of the fifteen sets by this measure, by an order of magnitude: were it wrong by a full bootstrap standard deviation at every round, no parameter would move by more than 0.011 standard errors, against movements of up to two-thirds of a standard error for the profiles that do the identifying work. What the two results establish jointly is leverage, not immunity. The observed hazard residual is not one bootstrap standard deviation but several hundred of them, the residual table's own scale column putting the largest bin at $934\,\hat{\sigma}$, so multiplying the local sensitivity through the observed misfit does not bound the displacement of $\hat{\theta}$, and I do not claim that it does. What the sensitivity shows is that, per unit of misfit, the estimates respond to this set an order of magnitude less than to the sets that identify them: the rejection is driven almost entirely by a moment with little influence on the estimates. The exposure that remains is concentrated in the profile the model fails: the round-cap counterfactuals operate directly on this exit margin, and their magnitudes are the most specification-conditional numbers the policy table reports. The standard errors are unaffected for a separate reason, requiring $\widehat{\Omega}$ only through a 25×25 projection that sits well inside the bootstrap's rank.

Two further points follow from the same table. First, the abandonment hazard was added to the objective to identify σ_{upd}, and σ_{upd} remains among the weakest-identified parameters, its standard error several times its estimate. The moment intended to pin it down barely moves it, and the same calibration of σ_η explains both facts and the misfit itself: a model in which learning does almost nothing has little to govern the timing of exit, so it neither fits the exit profile nor is disciplined by it. Second, the sensitivity ranking does not follow the weights. Acceptance by leniency carries the largest weight in W and is indeed the most influential profile, but granted scope by leniency is a close second while carrying unit weight. Influence on the estimates is a property of the Jacobian and the covariance together, not of the weight alone, which is why the assignment in Table L.7 is stated as a heuristic and this table as the computed version of it.

Sampling uncertainty in the counterfactuals of Section 7.7 is a separate matter and is not propagated from $\hat{\theta}$; what is reported there is sensitivity to calibrated inputs.

Table L.3: *Structural estimates and sandwich standard errors*

Parameter	Estimate	(SE)	$\lvert t\rvert$
c_0	0.2432	(0.0041)	59.4
q_{unc}	0.8082	(1.8896)	0.4
α_q ¶	0.0518	–	–
α_1 ‡	−0.4956	(0.0108)	45.8
α_2	0.6070	(0.0119)	51.0
b_r	0.0389	(0.0018)	22.0
κ	0.2549	(0.0054)	47.3
c_s ‡	0.0000	(0.0051)	0.0
c_r	0.2450	(0.0077)	32.0
σ_c ‡	0.6499	(0.0077)	84.6
σ_{grant}	0.4066	(0.0016)	262.1
σ_{upd}	0.3629	(1.1517)	0.3
μ	0.6764	(0.0044)	154.2
ν	2.8853	(0.0389)	74.1
β_s	−0.0223	(0.0042)	5.3
β_q	−0.0040	(0.0020)	2.0
b_0	0.5920	(0.0009)	689.8
σ_s	0.2944	(0.0006)	453.4
γ_{A}	0.7233	(0.0351)	20.6
γ_s	0.7562	(0.0155)	48.9
γ_{E} ‡	0.9994	(0.0179)	55.8
c_q	0.0819	(0.0079)	10.3
c_{q^2}	0.2271	(0.0312)	7.3
α_3	0.0141	(0.0005)	25.9
$\gamma_{s\text{-sl}}$ ‡	0.0001	(0.0121)	0.0
γ_q	−0.0116	(0.0532)	0.2

Notes: Indirect-inference sandwich standard errors, equation (33), with $\widehat{\Omega}$ from a block bootstrap over examiners on 500 resamples and the finite-simulation factor $(1 + 1/S)$ at $S = N_{\text{sim}}/n = 1.30$. $|t|$ is against zero. ¶ α_q is pinned as the level normalization of the grant index rather than reported with a standard error: because the examiner's posterior mean averages one half at every round, (α_q, μ) move the acceptance index along a single direction, and re-scoring the estimated model under the compensated reassignment reproduces the criterion to well inside simulation noise. A sandwich error for either coordinate alone would describe the near-singular inversion, not the sample. Every retained column is non-degenerate and the 25×25 block is inverted exactly. ‡ Estimate lies on a bound, where the asymptotic distribution is not normal: the reported error describes curvature and should not be read as a symmetric confidence interval. 20 of 25 identified parameters exceed two standard errors from zero. The over-identification statistic and the decomposition of the fit that underlies it are reported in Appendix L.2.

Table L.4: *Moment fit in natural units, and the decomposition of J*

			Residual, own units			
Moment set	Unit	Bins	RMS	Max	Max/$\hat{\sigma}$	% of J
Abandonment hazard by round	prob.	5	0.1359	0.2426	934	84.1
Granted scope by filed scope (resid.)	scope pctile	98/100	0.0453	0.1204	74	3.0
Grants by revision round	share of apps	10/12	0.0034	0.0092	49	0.6
Filed scope by round (resid.)	scope pctile	5	0.0236	0.0489	49	0.3
Acceptance by filed scope (partialled)	prob.	98/100	0.0276	0.0687	29	7.6
Acceptance by leniency	prob.	99/100	0.0235	0.1071	22	6.3
Rounds by quality (resid., demeaned)	rounds	100	0.0168	0.1115	21	0.0
Var(acceptance) by leniency	prob.^{2}	99/100	0.0182	0.0855	21	−1.8
Scope-leniency slope by round	pctile/quintile	4	0.0051	0.0082	20	−0.1
Var(filed scope) by quality	pctile^2	100	0.0012	0.0046	11	0.1
Granted scope by quality (resid.)	scope pctile	100	0.0032	0.0161	9	0.0
Filed scope by quality (resid.)	scope pctile	100	0.0026	0.0158	8	−0.0
Granted scope by leniency (resid.)	scope pctile	99/100	0.0197	0.0624	7	−0.0
Acceptance by quality (resid.)	prob.	100	0.0038	0.0134	6	−0.1
Filed scope by leniency (resid.)	scope pctile	99/100	0.0044	0.0192	4	−0.1

Notes: Residuals are $m^{\text{sim}}(\hat{\theta}) - m^{\text{data}}$, in each profile's own units; profiles marked *resid.* are residualised on the fixed-effect set before binning, so their level is a deviation. A bin count of the form a/b means $b - a$ bins are empty on one side and are excluded: those enter both moment vectors as an exact zero by construction and measure the support difference between a 5.4M-row sample and a 7.0M-row simulation rather than fit. $\hat{\sigma}$ is the bin's block-bootstrap standard deviation over 500 examiner resamples, which at $n = 5.4$M is very small: the Max/$\hat{\sigma}$ column is the quantity J aggregates, and the two residual columns are the same misfit priced in economic units. The final column apportions the coarsened $J = 1,364,175$ of Appendix L.2 across sets as $\sum_{i \in b} g_i(\widehat{\Omega}^{-1} g)_i / J$, which sums to 100 by construction and admits negative entries where one set's residual is partly accounted for by another's. Rows are ordered by Max/$\hat{\sigma}$.

Table L.5: *Sensitivity of the estimates to the auxiliary statistics*

	Leading		Second	
Parameter	Moment set	$\Delta\hat{\theta}/\mathrm{SE}$	Moment set	$\Delta\hat{\theta}/\mathrm{SE}$
c_0	Acceptance by leniency	+0.16	Acceptance by quality (resid.)	−0.09
q_{unc}	Acceptance by quality (resid.)	−0.04	Acceptance by leniency	+0.03
α_q	Grants by revision round	+0.11	Filed scope by quality (resid.)	+0.07
α_1	Acceptance by leniency	+0.08	Acceptance by filed scope (partialled)	−0.06
α_2	Filed scope by leniency (resid.)	+0.17	Filed scope by quality (resid.)	−0.17
b_r	Acceptance by leniency	−0.63	Grants by revision round	+0.43
κ	Acceptance by quality (resid.)	+0.28	Grants by revision round	+0.27
c_s	Acceptance by leniency	+0.26	Scope-leniency slope by round	−0.18
c_r	Acceptance by leniency	+0.15	Acceptance by quality (resid.)	−0.11
σ_c	Acceptance by leniency	+0.24	Filed scope by quality (resid.)	+0.09
σ_{grant}	Acceptance by leniency	+0.67	Acceptance by quality (resid.)	−0.28
σ_{upd}	Acceptance by leniency	+0.08	Acceptance by quality (resid.)	−0.04
μ	Grants by revision round	−0.43	Filed scope by leniency (resid.)	−0.13
ν	Acceptance by leniency	−0.26	Acceptance by quality (resid.)	+0.14
β_s	Filed scope by leniency (resid.)	−0.68	Acceptance by leniency	+0.58
β_q	Grants by revision round	−0.12	Acceptance by leniency	−0.07
b_0	Acceptance by leniency	+0.43	Grants by revision round	−0.13
σ_s	Var(filed scope) by quality	+1.77	Acceptance by quality (resid.)	−0.14
γ_{A}	Acceptance by leniency	+0.16	Acceptance by quality (resid.)	−0.09
γ_s	Granted scope by leniency (resid.)	−2.03	Granted scope by quality (resid.)	+0.56
γ_{E}	Granted scope by leniency (resid.)	+2.14	Granted scope by quality (resid.)	−0.36
c_q	Acceptance by quality (resid.)	−0.13	Grants by revision round	+0.09
c_{q^2}	Acceptance by quality (resid.)	+0.17	Acceptance by leniency	−0.09
α_3	Acceptance by leniency	−0.27	Acceptance by quality (resid.)	+0.21
$\gamma_{s\text{-sl}}$	Granted scope by leniency (resid.)	−1.64	Granted scope by filed scope (resid.)	+0.32
γ_q	Granted scope by quality (resid.)	−0.05	Granted scope by leniency (resid.)	−0.02

Notes: Sensitivity of the estimates to the auxiliary statistics, following Andrews et al. (2017). Entries are $\Sigma = (G'G)^{-1}G'D$ evaluated at $\hat{\theta}$, with G the binding-function Jacobian expressed in the criterion's own coordinates, each bin scaled by the square root of its weight in the objective, so the map is that of the weighted estimator actually used, equivalent to $(\Lambda' W \Lambda)^{-1}\Lambda' W$ in natural units. The perturbation shifts every bin of one profile by one block-bootstrap standard deviation, and entries are expressed in units of that parameter's own standard error. An entry of 1.0 therefore means: were the model wrong about that entire profile by one bootstrap standard deviation, the estimate would move by one standard error. This is the inverse of the map in Figure L.1, which reports how strongly each moment *responds* to a parameter; a moment can respond sharply and still contribute little to identification if another parameter moves it more. Every parameter has a row, including the pinned and bounded coordinates; rows for coordinates on bounds or inside a flat direction are computed from the same unrestricted map and carry the descriptive reading of Appendix L.1, not a test.

Table L.6: *Local identification diagnostics at the converged estimate*

Parameter	$\hat{\theta}$	Position in bounds	Inference
c_0	0.2432	68.7%	interior
q_{unc}	0.8082	80.8%	interior
α_q	0.0518	86.3%	interior
α_1	−0.4956	0.2%	at bound: sandwich N/A
α_2	0.607	30.4%	interior
b_r	0.03894	12.1%	interior
κ	0.2549	14.1%	interior
c_s	9.976×10^{-7}	0.0%	at bound: sandwich N/A
c_r	0.245	98.0%	interior
σ_c	0.6499	100.0%	at bound: sandwich N/A
σ_{grant}	0.4066	78.3%	interior
σ_{upd}	0.3629	72.6%	interior
μ	0.6764	41.2%	interior
ν	2.885	62.8%	interior
β_s	−0.02228	5.0%	interior
β_q	−0.004015	49.6%	interior
b_0	0.592	92.0%	interior
σ_s	0.2944	31.5%	interior
γ_{A}	0.7233	44.7%	interior
γ_s	0.7562	72.3%	interior
γ_{E}	0.9994	99.8%	at bound: sandwich N/A
c_q	0.08187	27.3%	interior
c_{q^2}	0.2271	45.4%	interior
α_3	0.01409	70.5%	interior
$\gamma_{s\text{-sl}}$	5.558×10^{-5}	0.0%	at bound: sandwich N/A
γ_q	−0.01163	54.8%	interior
Condition number of $\Lambda' W \Lambda$: 4.11×10^9			

Notes: Local identification diagnostics at the converged estimate. $\Lambda = \partial m_{\text{sim}}/\partial \theta'$ is the binding-function Jacobian, computed by central finite differences with common random numbers so that differenced simulation noise cancels; W is the estimation weight matrix. Columns of Λ are expressed in bound-range units, so a unit move is the same fraction of the searched interval for every parameter and the condition number is not an artefact of scaling. The condition number describes the local curvature of this objective at this estimate under these weights; it is not a rank test and does not establish global identification. "Position in bounds" is $(\hat{\theta} - \underline{\theta})/(\overline{\theta} - \underline{\theta})$. For parameters at a bound the usual sandwich variance does not apply, because the estimator is not in the interior of the parameter space; no symmetric standard error is reported for them.

Table L.7: *Identification: leading moment for each parameter*

Parameter	Primarily identified by	Moment weight
σ_{grant}, μ	Acceptance-rate gradient in leniency (the S-curve)	200
α_q	Pinned as the grant-index normalization (Appendix L.1); the near-flat quality profile follows from the calibrated signal noise	1
$\alpha_1, \alpha_2, \alpha_3$	Acceptance rate by filed-scope percentile	1
b_r, b_0, c_r	Grant rate by prosecution round (rounds distribution)	10
$q_{\text{unc}}, \sigma_{\text{upd}}$	Mean and variance of revision rounds by quality‡	10
c_q, c_{q^2}	Variance of revision rounds by quality percentile	10
c_0, c_s, σ_c	Level and dispersion of the rounds distribution	10
γ_s	Granted scope by filed-scope percentile	4
$\gamma_{s\text{-sl}}$	Slope of granted on filed scope	4
$\gamma_{\text{A}}, \gamma_{\text{E}}$	Scope concession across revision rounds	15
β_s	Filed scope by leniency	1
$\beta_q, \kappa, \nu, \sigma_s$	Filed scope by quality percentile (level and dispersion)	1
γ_l	Calibrated to the granted-scope leniency gradient (not estimated)	–
$\chi, \sigma_\eta, \bar{q}, \sigma_q, q_{\text{err}}$	Calibrated (not estimated)	–

Notes: Leading identifying moment per parameter; identification is joint. ‡This assignment is *not* borne out by the Jacobian: with σ_η calibrated, the two belief parameters are the two flattest directions of $\Lambda' W \Lambda$ and the moment set does not identify them. The row states the moment that would identify them in a specification where learning had bite. See the local-identification diagnostics above. "Moment weight" is the weight the moment set receives in the weighted indirect-inference objective. The offer's leniency gradient γ_l is calibrated to the measured round-zero conditional leniency slope rather than estimated, so it has no weight of its own. Diagnostic moment sets that are computed but excluded from the objective (weight 0) are omitted.

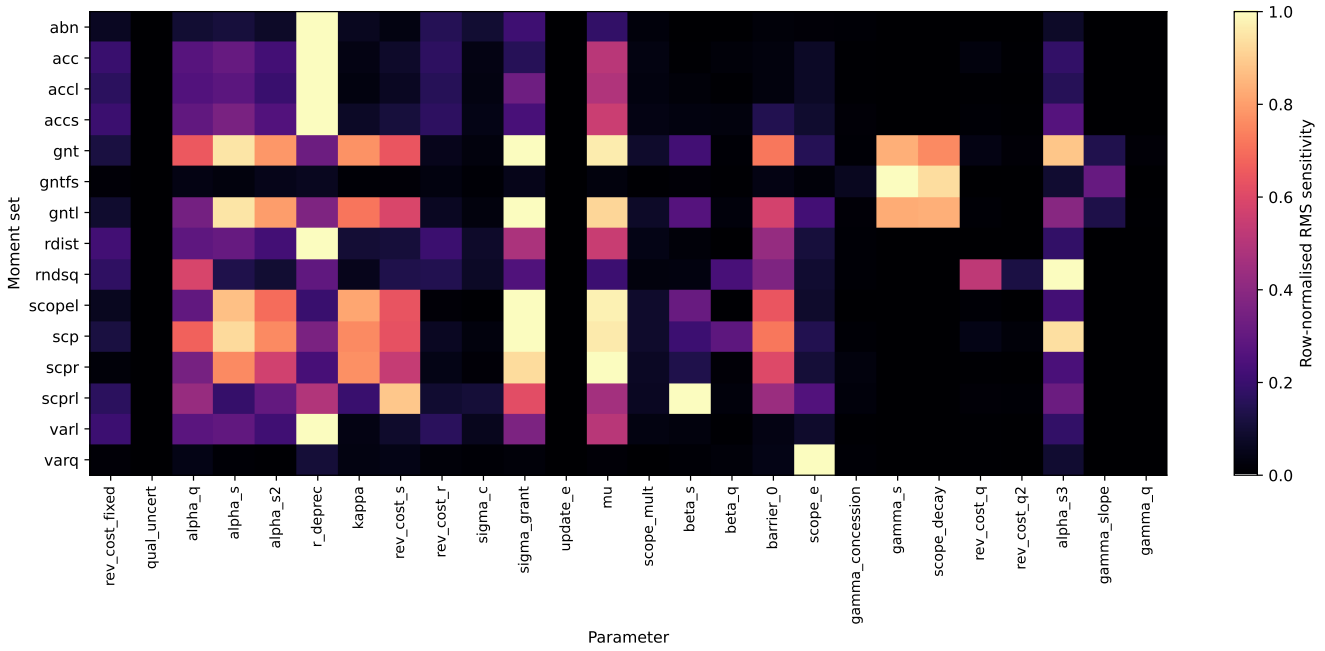
Table L.8: *Identification: parameter blocks and their informative moments*

Parameter block	Identifying moments	Source of identification
Grant condition $\alpha_q, \alpha_1, \alpha_2, \alpha_3, \mu, b_r, b_0, \sigma_{grant}$	Acceptance by quality; residualized acceptance by filed scope; acceptance by leniency; acceptance-by-round distribution; variance of acceptance by leniency	Scope penalties from the acceptance-scope slope net of quality and leniency, with the cubic term taking the location of the peak; μ from the acceptance level; b_r, b_0 from the round profile; σ_{grant} from the acceptance variance.
Beliefs and learning $q_{\text{unc}}, \sigma_{\text{upd}}$	Mean and variance of revision rounds by quality	Faster belief convergence shortens prosecution, so the level and dispersion of rounds by quality would pin down the applicant's belief dispersion and the rate at which rejections revise it. The two separate only weakly. Both are estimated, and both imprecisely: with the examiner's signal noise σ_η calibrated large, learning is close to vacuous, the simulated moments respond faintly to either parameter, and their standard errors run several times their estimates (Table L.3).
Applicant costs $\kappa, c_0, c_s, c_r, c_q, c_{q^2}, \sigma_c$	Acceptance-by-round distribution; variance of rounds by quality; filed-scope level	Revision costs govern how far applicants pursue prosecution before abandoning; the round distribution and its dispersion identify the cost components, and the two quality-persistence terms the differential speed of prosecution across the quality distribution.
Applicant scope policy $\beta_s, \beta_q, \nu, \sigma_s, \gamma_{\text{A}}$	Mean filed scope by leniency and by quality; variance of scope by quality; granted scope by filed scope; filed-scope-leniency slope by round	Filing gradients from scope-by-leniency and scope-by-quality; the level and dispersion of filed scope from the scope-by-quality profile; γ_{A}, the concession decay, from granted-vs-filed scope across rounds.
Examiner scope offer $\gamma_s, \gamma_{s\text{-sl}}, \gamma_{\text{E}}$	Level of granted scope conditional on acceptance; granted scope by filed scope; granted scope by leniency	Base offer ratio from the granted-scope level; the filed-scope gradient from the slope of granted on filed scope, which the level alone cannot match; the per-round decay from how the offer falls across rounds. The leniency gradient γ_l is calibrated rather than estimated.

Untargeted validation moments: the conditional complier shares $\phi_c(q)$ and $\phi_c(s)$, estimated independently of the model by the design-based estimator of Section 5. They carry no weight in the objective; the model must reproduce them through the mechanism above without having been fitted to them.

Figure L.1 displays the full empirical moment-sensitivity matrix underlying this assignment, evaluated at the converged estimate by central finite differences with common random numbers.

Figure L.1: *Moment sensitivity to parameters (empirical Jacobian)*



Notes: Row-normalized root-mean-square sensitivity of each targeted moment set (rows) to each estimated parameter (columns), $\partial m^{\text{sim}}/\partial\theta'$ evaluated at $\hat{\theta}$ by central finite differences with common random numbers. Brighter cells mark the parameters to which a moment set's simulated counterpart responds most strongly; this is the empirical analogue of the leading-moment assignment in Table L.7.

L.3 Fit

The fit is reported in two forms. Table L.9 gives the level and the slope of every binned profile that is conditioned on quality or on filed scope, data against model. Those profiles are flat or near-flat, so two numbers exhaust them and a panel would show sampling noise around a level. Figures L.2 and L.3 then plot the profiles whose *shape* carries information: the three that describe the acceptance margin and the four that describe scope and prosecution length. The remaining components of the objective are slopes computed across revision rounds, or hazards, which do not read as profiles at all.

Start with what the model gets right. It reproduces the level moments the offer mechanism targets to within a tenth of a percentile of scope: mean granted scope (0.426 in both series), mean filed scope among accepted applications (0.545 in both), and the scope concession (0.119 in both; Table L.9). Every profile that is flat in quality in the data is flat in quality in the model, and the acceptance rate itself matches to three decimals (0.711 in both). It reproduces the sign and shape of the two scope-margin profiles the design-based half of the paper rests on: acceptance is hump-shaped in filed scope, rising over the narrow claims and falling thereafter (Figure L.2b), and granted scope rises steeply in filed scope (Figure L.3a). The acceptance-by-leniency profile, which carries the dominant weight, is matched almost exactly in both level and slope, and filed scope falls gently across the leniency range in both series (slopes -0.016 in the data and -0.018 in the model, Figure L.3c).

Five residuals remain, and they are not of equal importance. The largest by a wide margin is the

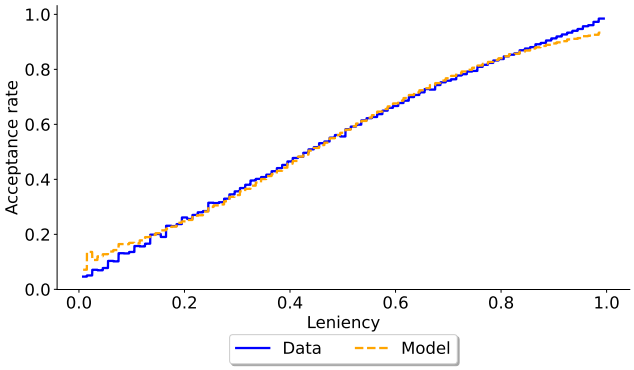
abandonment hazard by round: late in prosecution the model abandons just over half of the applications still live, against under three in ten in the data (0.53 against 0.29 in the final round bin, a root-mean-square gap of 0.136 across bins, three times the next largest). It accounts for 84 percent of the over-identification statistic and for almost none of the identifying content, which is why it is reported here rather than read as a constraint on the estimates (Section 7.5). Second, the slope of granted scope in filed scope is overstated (0.78 against 0.63): with the offer ratio's own scope gradient free to move, the granted claim tracks the filed claim more closely at the top of the distribution than the data do. Third, the acceptance hump in filed scope peaks too far to the right: the model's maximum sits near the 44th percentile, against a broad data plateau over the 14th to 23rd, and its decline over broad claims is four-fifths as steep (-0.083 against -0.100). These two are the same limitation seen from opposite sides: the examiner's scope penalty is a smooth low-order polynomial in scope, which cannot produce a sharp early peak and a steep tail at once. Fourth, the rise in revision rounds with quality is under a third of the data's (0.023 against 0.073), concentrated in the top quality quintile, which the model does not generate at all. Fifth, granted scope rises in examiner leniency about twice as steeply in the model as in the data (0.106 against 0.052) and sits too low over strict examiners (Figure L.3b).

Table L.9: *Moment fit: level and slope of the profiles conditioned on quality and on filed scope*

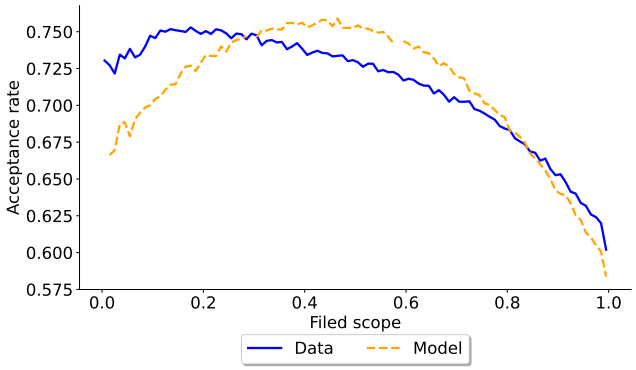
			Level		Slope	
Moment	Binned by	w	Data	Model	Data	Model
Acceptance rate	Quality	1	0.711	0.711	0.022	0.026
Mean filed scope	Quality	1	0.545	0.545	−0.003	−0.002
Variance of filed scope	Quality	1	0.068	0.073	−0.003	0.000
Mean revision rounds†	Quality	2	1.271	1.299	0.118	0.068
Granted scope, residualized	Quality	1	0.426	0.426	−0.008	−0.008
Granted scope	Quality	–	0.426	0.426	−0.008	−0.008
Scope concession (filed − granted)	Quality	–	0.119	0.119	0.005	0.006
Acceptance rate†	Filed scope	1	0.716	0.712	−0.100	−0.083
Granted scope†	Filed scope	4	0.406	0.395	0.629	0.778

Notes: Model column computed from the simulated panel at the converged estimate (the same run used for the policy counterfactuals of Section 7.7). Data column reports the empirical targets under the v2 sample conventions (core sample 2005–2020, within-year pooled scope percentiles, application and examiner controls partialled out). Scope is the within-year percentile of average claim length. "Level" is the mean across the 100 percentile bins; "slope" is the coefficient of a linear fit across those bins, with the conditioning variable on the 0–1 percentile scale. w is the weight the moment set carries in the indirect-inference objective; $w = 0$ marks a set that is computed but excluded, and "–" one that is not a moment set at all. All rows except those two condition on acceptance where the outcome is a granted quantity. Rows marked † also appear as figures, because their shape is not summarized by a slope; the profiles conditioned on examiner leniency are reported only as figures, for the same reason.

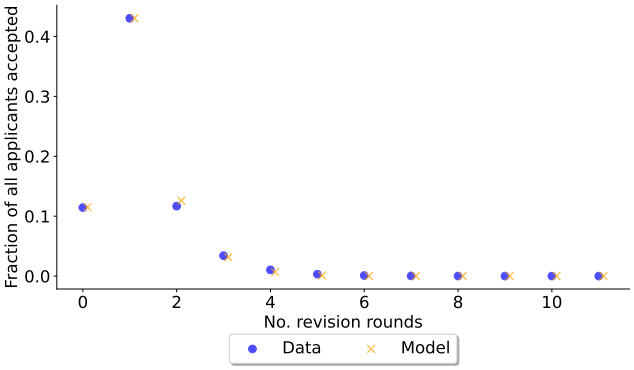
Figure L.2: *Targeted moment fit: the acceptance margin*



(a) By examiner leniency



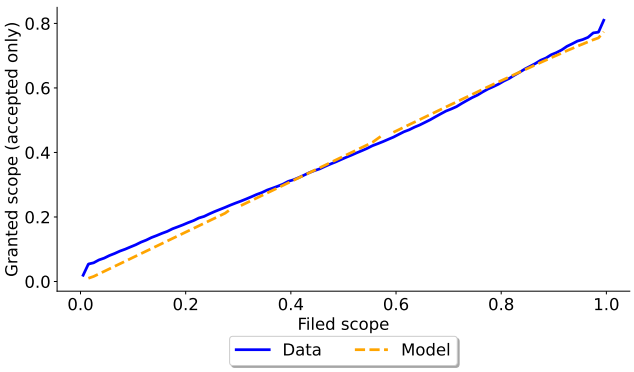
(b) By filed scope



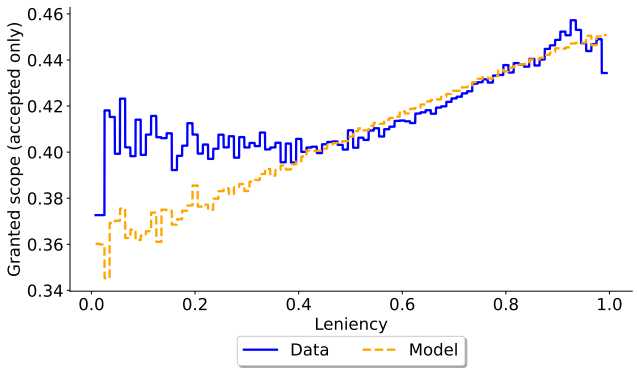
(c) By prosecution round

Notes: Solid line, data; dashed line, model at the converged estimate. All three profiles enter the estimation objective. Acceptance by *quality* rises only mildly in both data and model, about two percentage points end to end with the model tracking the rise, and is reported in Table L.9 instead; it can also be read off Figure L.4, which plots it separately within each leniency quintile. Filed scope is a within-year percentile, and the data series in Panel b is residualized on the fixed-effect set of Section 5 and on quality and leniency, so that the profile isolates the examiner's scope penalty. Panel a carries the dominant weight in the objective (Table L.7). Panel c plots the fraction of *all* applications granted at each round, not the round-conditional hazard.

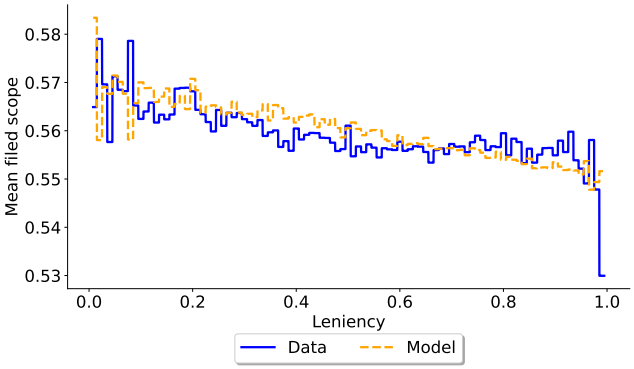
Figure L.3: *Targeted moment fit: scope and prosecution length*



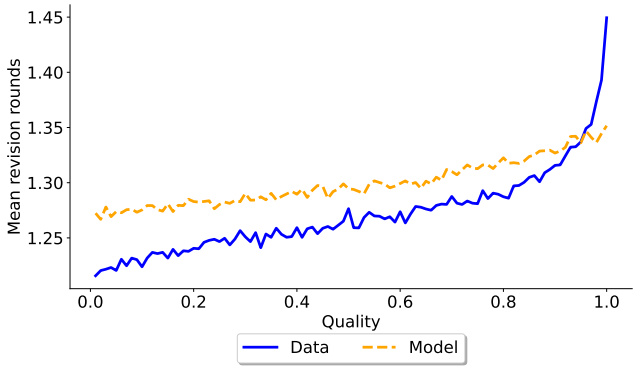
(a) Granted scope by filed scope



(b) Granted scope by examiner leniency



(c) Mean filed scope by examiner leniency



(d) Mean revision rounds by quality

Notes: Solid line, data; dashed line, model at the converged estimate. All four profiles enter the estimation objective. Scope is the within-year percentile of average claim length, pooled over filed and granted claims, and the data series are residualized on the fixed-effect set. Panels a and b condition on acceptance, as granted scope is only defined there; Panels c and d are over all applications, matching their estimation targets. The two profiles of filed scope conditioned on *quality*–its mean and its variance–are near-flat in both data and model and are reported in Table L.9 instead. Panel a carries the largest residual among the profiles plotted here: the model overstates how steeply granted scope rises with filed scope at the top of the distribution. Panel b is its counterpart in leniency, where the model's rise is about twice the data's.

Moments outside the objective. Because indirect inference fits what it is shown, the informative diagnostics are the profiles the estimator was not asked to match. Two appear in the bottom block of Table L.9: granted scope by quality without the leniency residualization, and the scope concession by quality. The residualized profile is not among them, because it is a weighted target of the objective (Section 7.5), where the offer's quality gradient γ_q reproduces both its level and its mild decline almost exactly. The two untargeted companions read the same offer: the concession's level and its rise in quality are matched jointly rather than traded against one another.

Why a loaded schedule rather than a belief-path offer. The record contains filed scope at entry and granted scope at issue, and nothing between (Section 7.4), so any model of the intra-prosecution concession path is identified only through its endpoint implications. That bounds what the offer's quality gradient can be asked to represent, and it supplies a test. The natural alternative to the reduced-form gradient γ_q is an offer conditioned on the examiner's round-r posterior. On endpoint data that family has exactly one distinguishing implication: the grant-round profile of the within-round quality gradient of the pass-through ratio, b_r. Under Bayesian updating the posterior's loading on true quality is $f_r = (r+1)\kappa/(1+(r+1)\kappa)$ with $\kappa = \sigma_q^2/\sigma_\eta^2$, so $f_0/f_1 = (1+2\kappa)/(2(1+\kappa)) > 1/2$ at every signal noise, and an offer decay $\gamma_{\mathrm{E}} \leq 1$ only shrinks later rounds relative to earlier ones. Every member of the family therefore predicts $|b_0| \geq |b_1|/2$: the first-round gradient cannot be less than half the second's. Table L.10 reports the estimated profile on the estimator's own construction. The bound is violated: $|b_1|/2 - |b_0| = 0.0044$, two and a half times its standard error of 0.0017. The family is therefore rejected at every $(\sigma_\eta, \gamma_{\mathrm{E}})$, and the maintained calibration, under which the posterior is nearly uninformative and every b_r should be near zero, is rejected at rounds one and two. The rejection is exactly as wide as the family: the bound assumes signals of equal precision across rounds, and it can fail only if the first signal is nearly uninformative while a later one is substantially sharper, a configuration the endpoint record can neither support nor rule out. Two qualifications apply. The constant-gradient form the model carries fails the same profile from the other side, predicting a gradient largest at $r = 0$ and declining where the data are flat at $r = 0$ with a hump at rounds one and two. Neither family fits the profile, so replacing the reduced form with a belief-path offer would add no identifiable content on endpoint data while losing the pooled-slope fit; the pooled slope is the only object both forms identify, it is what the calibration targets, and the round profile is reported as a limitation of both. And b_r conditions on a grant at round r, so the profile is a grant-selected equilibrium object; the belief-path prediction concerns the same selected object, so the comparison is like for like. The repair is data rather than modeling: the intermediate amendments exist, unparsed, in the application file wrappers.

The device's reach is bounded by the model's own structure. γ_q enters neither the grant rule (15) nor the applicant's problem, whose value function is computed against the full filed claim (Section 7.3), so no counterfactual can move the concession's quality gradient, and none of the exercises in Section 7.7 needs it to. The reform studied there operates through leniency, and its incidence loads on leniency rather than quality. An audit repricing the simulated panel across the gradient's plausible range confirms the separation: the welfare level is essentially unchanged, and the incidence exhibits that invert and reprice the offer are exactly invariant, because the repricing ratio cancels the gradient term by term. The

Table L.10: *The concession's quality gradient by grant round: testing the belief-path alternative*

Grant round r	b_r	s.e.	N	clusters
0	−0.0002	0.0016	574,297	8,785
1	−0.0092	0.0011	2,238,709	9,524
2	−0.0085	0.0018	615,460	9,321
3	−0.0062	0.0032	180,796	8,978
4	−0.0018	0.0056	55,114	7,874
5	−0.0067	0.0098	16,963	5,561
Belief-path bound $\lvert b_0\rvert \geq \lvert b_1\rvert/2$: violation $0.0044 > 2 \times$ s.e. $= 0.0035$				

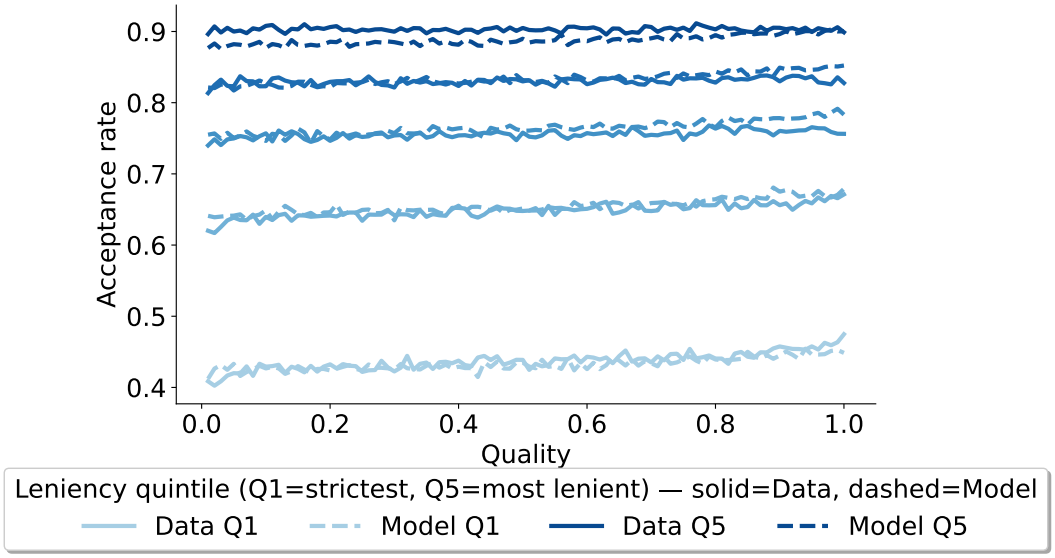
Notes: The Table reports the within-round quality gradient of the concession: for each grant round r (transaction counts at issue), the slope b_r from regressing the pass-through ratio, granted over filed scope clipped to $[0, 2]$ on the accepted sample with filed scope ≥ 0.05, on the quality rank centered at one half. Scope percentiles use the estimator's own construction: filed and granted claim lengths ranked against a pooled within-publication-year base and flipped so that higher values denote broader scope; the sample is the 2005–2020 estimation frame. Standard errors are clustered on art unit × publication year. The bound in the bottom row is parameter-free: under Bayesian updating the posterior's loading on true quality satisfies $f_0/f_1 = (1 + 2\kappa)/(2(1 + \kappa)) > 1/2$ for every signal noise, and an offer decay $\gamma_{\mathrm{E}} \leq 1$ only shrinks later rounds, so every belief-path offer predicts $\lvert b_0\rvert \geq \lvert b_1\rvert/2$. The observed violation exceeds twice its standard error, rejecting the family at every $(\sigma_\eta, \gamma_{\mathrm{E}})$; the maintained calibration, under which every b_r is near zero, is rejected at rounds one and two ($\lvert b_r\rvert$/s.e. = 8.1 and 4.6).

gradient's role is to make the simulated concession schedule match the observed one; no policy conclusion routes through it.

The third, and the substantive one, requires a figure. Figure L.4 plots the acceptance rate by quality *separately within each leniency quintile*: the joint surface, of which only the two marginals enter the objective. What is at stake here is a specification assumption rather than a hard prediction. The grant rule (15) is additively separable–leniency shifts the threshold and carries no interaction with quality–and Proposition 1 does not go through without that separability. Separability is *imposed*, not tested: the model has no quality-by-leniency parameter, so it cannot fit such an interaction and cannot report that it found none. What this panel can do is weaker but not empty. If a substantial interaction were present in the data, the fitted model would be unable to reproduce the joint surface, and the failure would show up as within-quintile tilt against the model's within-quintile profile. The panel shows a mild version of exactly that signature: the data's within-quintile rise in quality flattens as leniency grows, from a slope of 0.037 in the strictest quintile to essentially zero in the most lenient, faster than the probit's own curvature delivers, while the model's slope declines only mildly (0.024 to 0.020). The interaction this pattern suggests, quality mattering less to the most lenient examiners, is modest in level terms, with mean within-quintile gaps below 1.4 percentage points everywhere. I stop short of claiming the discrepancy would appear "here and nowhere else in the fit". An omitted interaction distorts the estimated parameters, and those parameters generate the targeted marginals too, so a misspecification of this kind could be partly absorbed elsewhere rather than surfacing cleanly in this one panel. The evidence is that separability holds in levels and bends in slopes: the five levels line up from 0.43 in the strictest quintile to 0.90 in the most lenient, the largest pointwise gap is 4.0 percentage points, in the middle quintile, and the one signed discrepancy is the slope flattening just described. An untargeted check of an imposed assumption can deliver no more;

the quintile levels themselves are close to implied by the targeted acceptance-by-leniency profile, so the panel's informative content is the within-quintile comparison, not the spacing.

Figure L.4: *An untargeted joint moment: acceptance by quality within leniency quintile*



(a)

Notes: Solid lines, data; dashed lines, model at the converged estimate. The joint surface does not enter the estimation objective; only its two marginals do. Applications are split into quintiles of examiner leniency (Q1 strictest, Q5 most lenient) and the acceptance rate is plotted against the quality percentile within each; only the extreme quintiles are labeled, and the shading runs light to dark from Q1 to Q5. The model contains no quality-by-leniency interaction, so the within-quintile agreement is not imposed; the spacing of the five levels, by contrast, is close to implied by the targeted acceptance-by-leniency profile of Figure L.2a. The complier share, the one untargeted moment that speaks directly to the leniency channel the counterfactuals exploit, is reported in the main text (Section 7.6, Figure 5).

Bandwidth robustness of the model's scope profile. The interior minimum in the model's complier share by filed scope (Section 7.6) is not an artifact of the smoothing bandwidth. Re-running the estimator on the simulated panel over Gaussian kernel bandwidths $h \in \{0.05, 0.06, \ldots, 0.15\}$, holding the evaluation grid and the trimmed range fixed, leaves an interior minimum at all eleven, located between the thirty-eighth and forty-first percentile (Table L.11). Its depth is not similarly stable: the fall from the narrow end to the trough more than halves across the range, the boundary sensitivity any kernel estimate carries at the edge of its support. The left arm is therefore present in sign and imprecise in size.

Table L.11: *Bandwidth sensitivity of the model's complier profile in filed scope*

	Interior minimum		Endpoints		
Bandwidth h	Location	Share	Narrow	Broad	Fall to trough
0.05	0.376	4.42	5.44	5.86	1.01
0.06	0.386	4.43	5.37	5.82	0.94
0.07	0.386	4.44	5.32	5.78	0.88
0.08	0.396	4.45	5.27	5.75	0.82
0.09	0.396	4.46	5.22	5.71	0.76
0.10	0.406	4.47	5.18	5.67	0.71
0.11	0.406	4.48	5.14	5.64	0.66
0.12	0.406	4.49	5.10	5.60	0.61
0.13	0.406	4.50	5.07	5.57	0.57
0.14	0.406	4.52	5.04	5.53	0.52
0.15	0.406	4.53	5.01	5.50	0.48

Notes: The model-implied share of compliers among granted patents, by filed scope, re-estimated on the simulated panel at each Gaussian kernel bandwidth h; the evaluation grid and the trimmed range are held fixed at the published values, and $h = 0.07$ is the bandwidth of the figure in Section 7.6. *Location* is the percentile of filed scope at which the curve attains its minimum over the plotted range; *Share* and the two *Endpoints* columns are in percent; *Fall to trough* is the narrow endpoint minus the trough, in percentage points. The minimum is interior at all 11 bandwidths and moves by three percentiles across the whole range, so its location is not a smoothing artifact; its depth is not similarly stable, which is the boundary sensitivity any kernel estimate carries at the edge of its support. The same sweep on the quality axis yields no interior minimum at any of the 11 bandwidths, the fitted minimum sitting on the upper boundary, so the non-monotonicity is specific to the scope axis. The estimator is the one of Section 5.3 applied to the simulated panel ($n = 7{,}000{,}000$, 99 leniency draws, overall share 4.90%), and it reproduces the published curves exactly at the figure's bandwidth.

M Counterfactuals: robustness of the welfare magnitudes

This appendix first derives the external anchor for the social cost of scope that Section 7.7 uses to locate its break-evens, and then reports in full three robustness exercises. The first two perturb the calibrated inputs the headline magnitudes depend on, the offer's leniency gradient and the composition weight behind the welfare level ΔW, and the third bounds the consequence of the model's fixed filing population. None requires re-estimating the model: the leniency-scope offer gradient γ_l enters only the post-acceptance offer ratio mapping filed to granted scope, so it re-prices analytically.

M.1 The external anchor for the social cost of scope

Bloom et al. (2013) decompose the marginal private return to R&D into a productivity component and a product-market-rivalry (business-stealing) component. Evaluated at their median output-to-R&D-stock ratio, the marginal private return splits into 10.7 points of own-productivity effect against 10.4 points of business stealing, so rivalry accounts for 49% of it. That share, rather than any figure they report as such, is what Section 7.7 imports, and two qualifications attach to the import. It is a magnitude bridge and not an identity: their ratio is defined per unit of R&D stock, for large listed US firms between 1981 and 2001, whereas mine is per unit of granted scope, and nothing guarantees the two margins scale alike. And it inherits their split parameter $\sigma = \frac{1}{2}$, the fraction of a rival's market-value loss that is redistribution rather than induced price decline.

Only the rivalry half of their decomposition maps to scope: broad claims exclude, they do not transmit knowledge, so the positive technology-spillover half has no counterpart in a scope cost. Business stealing is, however, largely a transfer from rivals rather than a social loss, so only some fraction $\tau \in [0, 1]$ of that wedge is a genuine deadweight cost. Writing the blocking cost as $\gamma_{\text{block}}(\tau) = \tau\, \gamma_{\text{BSvR}}$ and mapping it to the scope axis via $\delta = \gamma_{\text{block}}/\bar{\hat{s}}$ gives $\delta(\tau) = \tau\, \delta_{\text{BSvR}}$, with $\gamma_{\text{BSvR}} = 0.49$ and $\delta_{\text{BSvR}} = 1.12$ at the anchor. The social cost of scope is thus linear in the deadweight share, the mean granted scope entering both $\delta(\tau)$ and the break-even δ^* through the same normalization so that it drops out of the ratio τ^*, and the whole question reduces to one unknown scalar. That linearity is what makes the anchor decisive rather than merely suggestive.

Because a larger rivalry share is more adverse to these reforms, it matters that the anchor is the smallest value Bloom et al.'s estimates support. The 49% figure follows from the Jaffe closeness measure that serves as their baseline. Applying the same decomposition to their Mahalanobis measure, which they argue on axiomatic grounds dominates the Jaffe index, raises the rivalry share to 61%, and to their instrumented estimates, 74%, carrying δ_{BSvR} to 1.41 and 1.69 respectively. Both cut against the reforms. Even at 74%, every break-even deadweight share under the mean pricing remains above one: $\tau^* = 1.42$ for median equalization, 1.89 for judge-dispersion compression, and 1.88 for full equalization, against 2.14, 2.84 and 2.83 at the anchor. The binding case is median equalization at 1.42. The mean-pricing verdict therefore survives the full range of rivalry shares their estimates admit, which is a stronger statement than the anchor alone delivers.

One adjustment that is sometimes confused with a repricing of the value side should be distinguished from it. The composition weight ω of Appendix M.2 reprices *marginal* grants relative to their cell mean. The raw complier-value ratio is 1.151, but the observability decomposition below attributes the excess over one to selection into the granted-only value proxy rather than to the marginal grant's value, so the main text's $\omega = 1$ is the estimate rather than merely a conservative choice, and the reform-specific weights the decomposition implies lie inside [0.94, 1.02]. Wherever 1.151 appears below, it is this raw, observability-contaminated ratio, retained because the Monte Carlo perturbs the raw object.

Read on this scale, the model's own scope-cost constant is an equivalence rather than a second piece of evidence: $\gamma_{\text{block}} = 0.20$, or $\delta \approx 0.46$, is exactly the value implied by $\tau = 0.41$ at the anchor, or $\tau = 0.27$ at the most adverse rivalry share, that is, by assuming that somewhere between a quarter and a little under half of the rivalry wedge is deadweight. I report it to make that implicit assumption explicit, not as independent corroboration of the external anchor. It is a calibration input, and nothing in the estimation identifies it.

M.2 Sensitivity to the offer gradient and the calibration band

Table M.1 varies γ_l across $[0.09, 0.19]$. This band is the calibrated anchor $\gamma_l = 0.1407$ plus and minus an assumed spread of 0.05. It is not an estimated confidence interval: γ_l is calibrated to the offer anchor rather than estimated, so it has no sampling distribution, and the 0.05 is a judgment about how far the anchor might plausibly be displaced. The dispersion share moves only within $[1.1, 2.0]$% and the median-leniency break-even δ^* within $[2.32, 2.49]$; consistency dominates intensity at every point of the band,

so Section 7.7's policy ranking does not depend on the gradient's calibrated value. Table M.2 reports a 2,000-draw Monte Carlo that jointly perturbs γ_l and the complier-value composition weight ω. The two inputs are not on the same footing and the table should not be read as a single sampling band: ω is drawn from $\mathcal{N}(1.151, 0.024^2)$, whose spread is a delta-method standard error estimated from the complier-value decomposition, whereas γ_l is drawn from $\mathcal{N}(0.1407, 0.05^2)$ censored at the edges of $[0.04, 0.24]$, whose spread is the same assumed displacement used for the sensitivity grid. Roughly a third of an uncensored normal would fall outside the $[0.09, 0.19]$ grid, which is what one standard deviation means and not a discrepancy between the two exercises; the censoring at two standard deviations places about 2% of the draws at each endpoint. For median equalization it returns a mean level of +9.1% with a 5–95 range of $[+8.6\%, +9.6\%]$, and for the judge-dispersion reform a mean of +4.3% with range $[+4.0\%, +4.6\%]$. Both distributions lie above the corresponding $\omega = 1$ point estimates (+7.16% and +3.03%) reported in Section 7.7, because averaging over the sampled ω raises the level. The break-even means from the same draws ($\delta^* = 3.05$ and 4.52) are likewise ω-averaged and are not the $\omega = 1$ break-evens quoted in the text.

Those draws center ω on 1.151, and that center should not be read as the best estimate of the composition weight. It is estimated on the market-value measure, which exists only for patents assigned to publicly listed firms, and the complier pool is observed at a rate of 0.345 against 0.287 among the treated. Part of the wedge is therefore differential *observability* rather than differential value. The same decomposition run on forward citations, which are recorded for every granted patent and so carry no such selection, returns $\omega = 1.010$. Estimating the correlation between the value equation and the grant-selection index directly (Appendix N) puts the implied weight inside $[0.94, 1.02]$. Two routes that share neither an outcome nor an estimator therefore agree that the weight is one to within a few percent, and the reported $\omega = 1$ magnitudes are the estimates. The Monte Carlo band is retained as a sensitivity exercise on a deliberately generous ω, not as a confidence interval around the headline.

One caution about the same table, because the sign is easy to invert. The wedge for realized renewals is reported as 1.466, but renewals enter as the difference between actual and potential and are therefore negative: the complier mean of -0.217 against a treated mean of -0.148 says that marginal grants are renewed *less* often, so the ratio above one corresponds to lower value rather than higher. The δ-free reallocation share is untouched by either exercise: aggregate scope is fixed by construction and the value scale $\bar{v}$ cancels from every ratio, so the band is driven by composition and the offer gradient, not by value-level noise. Both exercises hold the estimated parameters at their point estimates, so these ranges describe sensitivity to calibrated inputs and do not carry sampling uncertainty from the indirect-inference step.

Table M.1: *Sensitivity of the misallocation magnitudes to the offer gradient γ_l*

ψ_l	L_1 (realised)	L_1 (equalised)	X	$\bar{\hat{s}}$	δ^*_{med}	δ^*_{jdg}
0.0900	0.1652	0.1634	1.1%	0.421	2.32	2.91
0.1407†	0.1677	0.1650	1.6%	0.426	2.40	3.19
0.1900	0.1701	0.1667	2.0%	0.432	2.49	3.50

Notes: Sensitivity of the misallocation magnitudes to the leniency-scope offer gradient ψ_l, recomputed analytically without re-estimating the model (the gradient enters only the post-acceptance filed-to-granted pass-through, so acceptance, rounds and filed scope are invariant to it). L_1 (realised) and L_1 (equalised) are the granted-scope L_1 gaps under the realised leniency distribution and under full equalisation, in which every examiner is placed at the pool mean; λ is reserved elsewhere for the inverse Mills ratio and for the per-round offer decay, so the intervention is named rather than symbolised. X is the dispersion-attributable share removable by examiner consistency; $\bar{\hat{s}}$ is mean granted scope; δ^*_{med} and δ^*_{jdg} are the median- and judge-equalisation break-evens. † marks the calibrated anchor $\psi_l = 0.1407$.

Table M.2: *Welfare-level sampling band from the (γ_l, ω) bootstrap*

	$\Delta W(0)$ (%)			δ^*	
Reform	Mean	SE	[5, 95] pctile	Mean	SE
Median-leniency equalisation	+9.10	0.31	[+8.59, +9.61]	3.05	0.15
Judge-dispersion compression	+4.28	0.20	[+3.95, +4.61]	4.52	0.46

Notes: Sampling band on the welfare level from a 2,000-draw Monte Carlo that jointly resamples the offer gradient $\psi_l \sim \mathcal{N}(0.1407, 0.05^2)$ and the complier-value composition weight $\omega \sim \mathcal{N}(1.151, 0.024^2)$, re-pricing each draw analytically. Reported are the mean, standard error and [5, 95] percentile interval of $\Delta W(0)$ and the break-even δ^*. The value scale $\bar{v}$ cancels from every ratio, so the band reflects composition and the offer gradient, not value-level noise.

M.3 The entry-elasticity bound

Table M.3 reports the trimming sensitivity of the level welfare change described in Section 7.7: an entry elasticity e is imposed and the entry-marginal filers, proxied by the lowest-quality applications, which rank the first-round continuation value up to idiosyncratic cost dispersion, are removed at rate e times the extensive-margin move, symmetrically from reform and baseline. Across $e \in \{0, 0.25, 0.5, 1\}$ the level of ΔW moves by at most six hundredths of a percentage point.

The table is built to show that this invariance is not an artifact of the removed filers being marginal in value. Its *Filers cut* and *Value cut* columns are near-equal at e=1: the dropped applications carry a share of accepted value proportional to their number and are accepted at close to the pool rate (60–76%), so they are welfare-active rather than deadweight the model discards. The flat ΔW column is instead a consequence of the reform's near-orthogonality to the quality axis. The dropped low-quality applications hold almost exactly the same share of accepted value under the reform as under the status quo, the counterpart of the 0.19-point across-quality reallocation reported in Section 7.7, so trimming them rescales the numerator and denominator of ΔW alike. The δ-free core, being budget-fixed, is unaffected by construction.

Table M.3: *Entry-elasticity bound on the level welfare change*

	ΔW (%) at entry elasticity e				At e=1	
Reform	0	0.25	0.5	1	Filers cut	Value cut
Median-leniency equalisation	+7.16	+7.16	+7.16	+7.16	7.5%	7.3%
Judge-dispersion compression	+3.03	+3.03	+3.03	+3.04	3.5%	3.4%
Lower-median re-anchoring (F&W)	−14.94	−14.92	−14.91	−14.88	15.5%	15.1%

Notes: Level welfare change $\Delta W(0)$ (portfolio value before scope-cost netting) under a reduced-form entry margin: the entry-marginal filers, proxied by the lowest-quality applications, are removed at rate e times the reform's own extensive-margin move, symmetrically from reform and baseline (identical draws). e=0 reproduces the fixed-population level; e=1 is a unit elasticity. *Filers cut* and *Value cut* report, at e=1, the share of applications removed and the share of accepted portfolio value they carry; the two are near-equal and near-identical across reform and baseline, so the removed filers are welfare-active – accepted at 59–75% – not marginal in value. ΔW is invariant to the entry margin because the reform is near-orthogonal to the quality axis (the counterpart of the 0.08% across-quality budget-neutral reallocation), not because the removed mass is valueless. Estimated model, $N = 7,000,000$ simulated applications.

M.4 Incidence of the budget-neutral reallocation

Figure M.1 reports the incidence of the budget-neutral reform of Section 7.7, with applicants sorted by the leniency of their status-quo examiner. The direction is mechanical; the content of the figure is the magnitudes and the near-exact flatness of the same decomposition run across quality deciles.

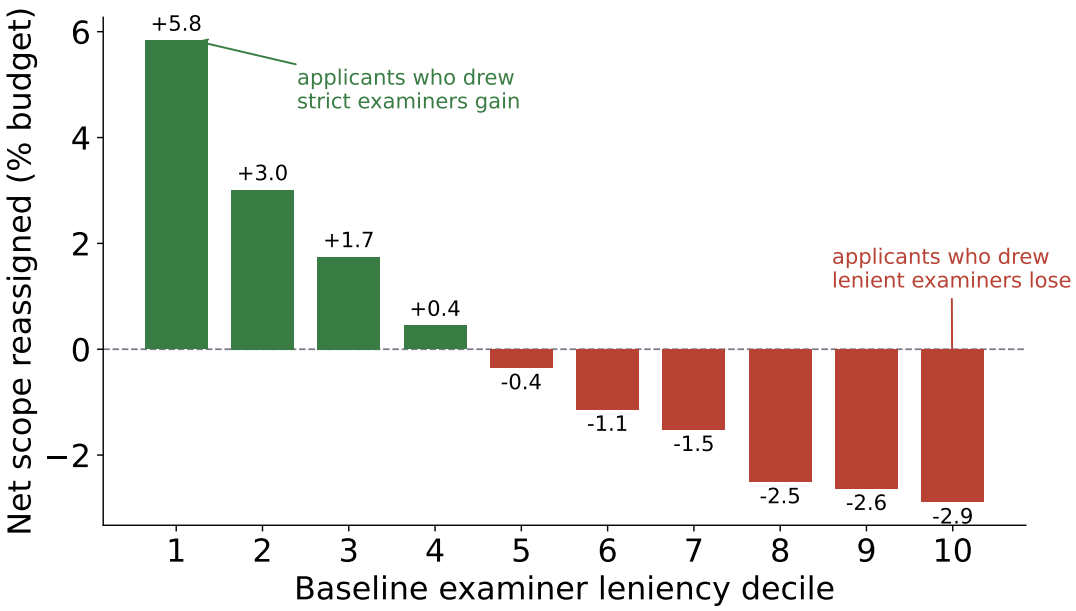


Figure M.1: *Incidence of the budget-neutral reallocation by examiner leniency*

Notes: Net granted scope reassigned by the budget-neutral consistency rule, as a percentage of the granted-scope budget, with applicants sorted into deciles by the leniency of their status-quo examiner (decile 1 strictest, decile 10 most lenient). Positive bars (green) are net winners, negative bars (red) net losers. The reform transfers scope from applicants who drew lenient examiners to those who drew strict ones; the analogous decomposition by invention quality is flat (net reassignment at most 0.08 points per decile). Estimated model, $N = 7{,}000{,}000$ simulated applications.

N Limitations of the design

The qualifications attached to individual results in the main text share a few common causes, collected here. What follows is not a list of things left for future work with this data. Each is a limit of the design itself, and none is repaired by more observations of the same kind.

Ordinal scope measure. Scope is a within-cell percentile rank. Every result defined through the ordering alone, the scope bins, the complier profiles over them, the ordinal reclassification, is exactly invariant to any monotone rescaling of claim length; every *quantity* that sums scope across applications, including the granted-scope budget, is normalization-dependent, and Appendix I finds that exposure material, which is why Section 7.7 leads with the δ-free and ordinal results. No cardinal, legally validated measure of claim scope exists, and constructing one is not a data problem. The mapping from scope to social cost is maintained rather than identified, and the $\tau = 1$ benchmark bounds only the mapped rivalry wedge: litigation, blocking, and follow-on costs outside it are not bounded by it (Appendix M.1).

Value observed only on grants. The private value of marginal scope is not identified from a sample selected on grant; the quality-only pricing imposes a zero rather than testing one, and the value-maximizing quality-scope schedule is not identified either. This is why the paper does not claim that scope is handed to inventions whose quality does not warrant it, only that lottery grants disproportionately involve broad filings of unremarkable measured quality and that consistency would reallocate scope within quality cells.

The validation correlations of Section 3.4 are likewise computed within the granted sample, and the direction of the resulting selection bias cannot be signed, since the external benchmark of Kogan et al. (2017) is itself measured around an examiner-influenced grant.

Nearly flat price. The market-value price the counterfactuals use is nearly constant over the $(q, \hat{s})$ plane, rising by only eight percent from the median bin to the top quality decile and thirteen percent to the top percentile, and flat in granted scope (Table H.2). The counterfactuals therefore price how many rights are issued and how broad they are, not how much any one of them is worth. The one channel that could repair this is testable rather than assumed: reproducing the joint value-allowance design of Avivi (2025) on the initial-allowance margin, where the examiner can have moved value only through the grant itself, gives a value-selection correlation of $\rho_v = 0.072$ (95% interval $[0.038, 0.107]$) once art unit $\times$ year is absorbed, implying a composition weight inside $[0.94, 1.02]$, which is why the reforms are priced at a weight of one.[75] The selection channel exists and is small; it is not a route to pricing which grants are worth more.

Monotonicity, scope choice, and entry. *Individual* monotonicity is untestable; the average monotonicity the data can test permits offsetting compliers and defiers, and the additive-leniency threshold rule of Section 5.1 is maintained as the structural assumption that delivers it. Filed scope is chosen rather than randomized and cannot be separated from unobserved applicant type, so the scope gradient of Section 6.2 is a composition fact and carries no causal reading. And entry is not modeled: no filing-rate moment identifies the outside option, so the applicant population is fixed by assumption, with the symmetric-trimming exercise bounding the consequence of that assumption rather than testing entry.

The reallocation share. The budget-neutral reallocation share is the result least exposed to these limits. It needs no price and no value-maximizing schedule, because grants and aggregate scope are fixed by construction, and it is computed inside the model, so it inherits the design's identification rather than the granted sample's selection; computed inside the model also means conditional on the estimated grant rule, the filing and offer equations, and fixed entry. Its one exposure is the cardinalization, and Appendix I measures it directly: the share moves only between 10.6% and 12.3% across the alternative scales, because applications that change grant status contribute their whole scope under every scale, so the range brackets the normalization rather than the model.

[75] Without the absorption the correlation is 0.40, so the apparent association is composition across technology fields rather than selection within them. The within-cell estimate is statistically indistinguishable from the 0.105 Avivi (2025) reports on a different sample.