## NBER WORKING PAPER SERIES

# DEADLINES VERSUS CONTINUOUS INCENTIVES: EVIDENCE FROM THE PATENT OFFICE

Michael D. Frakes Melissa F. Wasserman

Working Paper 32066 http://www.nber.org/papers/w32066

NATIONAL BUREAU OF ECONOMIC RESEARCH 1050 Massachusetts Avenue Cambridge, MA 02138 January 2024

We are grateful to numerous workshop participants and colleagues for helpful comments. We acknowledge funding from NIA grant 1R01AG076586-01A1. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2024 by Michael D. Frakes and Melissa F. Wasserman. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Deadlines Versus Continuous Incentives: Evidence from the Patent Office Michael D. Frakes and Melissa F. Wasserman NBER Working Paper No. 32066 January 2024 JEL No. D03,J33,K0,O34

## ABSTRACT

A quota system with an associated deadline may retain the possibility of worker procrastination and related deadline behaviors. A performance appraisal system based on continuous temporal incentives, on the other hand, has the potential to alleviate deadline effects but may lose some of the quality-related benefits associated with the flexibility of a quota/deadline system. We explore these tradeoffs by observing patent examiner behavior and examination quality outcomes surrounding a 2011 reform at the U.S. Patent and Trademark Office that built on its bi-weekly quota system by adding a set of bonuses tied to daily examination-pendency measures. We find a substantial reduction in deadline effects and near complete temporal smoothing in examiner behavior in connection with the reform, leading to large reductions in average examination pendency while resulting in no corresponding reductions in the accuracy of examinations.

Michael D. Frakes Duke University School of Law 210 Science Drive Box 90362 Durham, NC 27708 and NBER Michael.frakes@law.duke.edu

Melissa F. Wasserman University of Texas College of Law 727 E Dean Keeton Street Austin, TX 78705 mwasserman@law.utexas.edu

An online appendix is available at http://www.nber.org/data-appendix/w32066

### I. INTRODUCTION

A foundational question within agency theory is the degree of deference to which the principal should grant the agent in executing the task(s) at issue (Holmstrom 1977). Deference to agents may be beneficial in harnessing certain informational and other advantages that the agent has over the principal; however, it may naturally come at the expense of preserving space for the agent's actions to depart from the principal's interests. These fundamental tradeoffs arise with respect to multiple dimensions of task performance. This paper will explore one such dimension: the timeliness of task completion.

A classic tool to incentivize task completion is the imposition of a quota with a deadline. That is, an organization may base financial rewards on the achievement of designated production levels by specified dates. Quotas of this nature do embrace a degree of deference, extending flexibility to agents in determining how to time and sequence their efforts over the quota period. Discrete temporal incentives of this nature—and the flexibility they afford—may nonetheless leave a residual role for procrastination or related behaviors whereby employees exhibit a disproportionate share of their effort near the quota deadline.<sup>1</sup>

To avoid concerns over residual procrastination, principals may elect to exercise greater control over the temporal aspects of the agents' assignments. For instance, they could increase the frequency by which quotas are set. Alternatively, they could avoid a discrete quota-based structure altogether and impose a higher-powered incentive system based on a continuous temporal function of task completion (Holmstrom and Milgrom 1987; Heath, Larrick & Wu 1999). For instance,

<sup>&</sup>lt;sup>1</sup> We define "procrastination" as a situation in which economic actors counterproductively delay tasks that they had previously intended to undertake—or at least that others, such as supervisors, expected them to undertake—by certain times. Lapses in self-control that may lead to procrastination are often thought to stem from models of individual decisionmaking that involve either salience costs—e.g., differential discount rates applied to costs and benefits of tasks—or present-biased preferences—e.g., higher discount rates applied to the short term than to the long term (Akerlof 1991; O'Donoghue and Rabin 1999).

imagine rewards based on the average time-to-completion in days of tasks performed. Of course, bear in mind the tradeoff identified above. That is, while the constant pressure associated with a continuous temporal incentive structure may stem procrastination tendencies, the diminished flexibility and discretion associated with this continuous approach may induce performance-reducing anxiety and stress (Olsen and Torsvik 2000; Chung et al. 2020) and may discourage task-sorting and sequencing strategies that otherwise produce efficiency gains.

In this paper, we explore these tradeoffs empirically and investigate the relative effects of a discrete quota structure versus a continuous temporal incentive system. In particular, we explore whether the imposition of a continuous system leads to a reduction in delayed work efforts while also testing for offsetting changes in task quality arising from the diminished temporal discretion. For these purposes, we evaluate a 2011 reform to the performance appraisal plan faced by patent examiners at the U.S. Patent and Trademark Office (Patent Office). With this reform, the Patent Office expanded upon its biweekly (two-week) quota system to add new incentives tied to average review-completion durations calculated at the daily level. Following the reform, scores based on such durations form the basis for various promotion outcomes and bonuses. The stated intent of the 2011 reform was indeed to reduce bunching of task completions at the end of quota periods.<sup>2</sup>

Facilitating our ability to explore whether this reform reduced worker delays as the Patent Office intended are certain institutional features of the patent context that allow us to empirically identify delayed work efforts in the first place. For instance, the substantial backlog that examiners face allows us to avoid some of the challenges that researchers face in interpreting deadline effects in the sales context, where deadline bunching may either reflect delayed salesperson efforts or

<sup>&</sup>lt;sup>2</sup> According to a Patent Office representative immediately prior to the roll-out of this reform, this new "docket management element" has "incentives built in for examiners to move their cases quicker than on a biweekly cycle" (Patent Public Advisory Committee Meeting, April 2011). E-mail from Robert L. Stoll, Former Commissioner for Patents, Patent & Trademark Office, to Melissa Wasserman (Sept. 7, 2023 12:35 CST) (on file with author) (noting that the Patent and Trademark Office implemented the continuous incentive system because "[w]e wanted to move things more quickly and we provided incentives to do so.").

salespersons responding to weak consumer demand in the present period by "pulling in" sales that would otherwise come in subsequent periods (Oyer 1998). Importantly, our analysis also builds on our prior work (Frakes and Wasserman 2020, hereinafter "FW"), which took advantage of the iterative, multi-round nature of the patent examination process to construct nuanced empirical markers reflective of worker delays / procrastination behavior.

In this prior work—which predominantly focused on the pre-reform period—we began by documenting substantial evidence of first-round review bunching at biweekly quota deadlines. We then set forth evidence consistent with predictions that procrastinating examiners faced with a rush to complete delayed first-round reviews just prior to deadlines would issue ill-conceived and easily-overcome rejections—termed "shotgun" rejections—at such moments. This strategy would afford examiners the ability to perform the proper substantive evaluation in later rounds of review, unlike an ill-conceived acceptance strategy, which would be final and non-correctable in nature.<sup>3</sup> Indeed, we found evidence that examiners did correct for deficient first-round reviews with accurate decisions in later rounds, with the ultimate consequence of this behavior being application "churn"—i.e., a prolonged examination process due to a wasted first round of review.

We begin the present analysis by exploring whether the documented deadline-bunching patterns and markers of compromised / rushed reviews at deadlines were reduced when the Patent Office implemented performance incentives at a continuous, daily level. Our findings are striking. End-loading of first-round reviews—completing reviews on the last day of the quota period promptly falls by roughly half following the onset of the Patent Office's reform. Moreover, the markers indicative of rushed shotgun rejections at deadlines that we consistently document in the

 $<sup>^{3}</sup>$  FW also acknowledge that certain types of task sorting may also explain some of the documented deadline bunching and shotgun rejections. More specifically, FW note that it is possible that patent examiners are reserving close call reviews to the end of the bi-week and then rejecting those close call applications in order to garner more time to review the application in the next round. This type of close call task sorting is also a type of time mismanagement, as patent examiners are spending more time on reviewing a patent application than the Patent Office allocates.

pre-reform period virtually disappear upon the reform. For instance, whereas prior to the reform end-loaded applications were allowed at a nearly 60-100% lower rate (relative to the mean) than non-end-loaded applications, this differential fell to essentially zero following the reform.

In fact, our results demonstrate a near total, day-by-day smoothing of examiner behavior following the reform. That is, we find a dramatic and complete flattening over the generalized biweek period in the affirmative decisions examiners make on the first-round review (e.g., rejection/allowance) and in the ultimate outcomes associated with those first-round reviews (e.g., total application duration, ultimate application allowance, etc.). Post reform, examination outcomes simply no longer depend on the day on which the initial review was completed.

To the extent the pre-reform shotgun-rejection markers are reflective of procrastination or related delay-based behaviors that harm examination quality, our findings imply that the addition of a meaningful though arguably modest set of financial rewards tied to a continuous incentive system may reduce or eliminate harmful delays in work effort that exist within temporally discrete monitoring structures. Taking matters in reverse, if one assumes that the default performance-monitoring approach is continuous in nature, our findings imply that the introduction of a quota/deadline system may induce procrastination or related practices in the first place.

Even though we document the near elimination of deadline effects upon the 2011 reform and with it, the likely reduction of some degree of procrastination and the application churn that it generates—we acknowledge the possibility that the reform may lead to other effects that compromise *net* examination quality. After all, the primary drawback of a continuous incentive system is the lack of flexibility it may leave workers, which could lead to lower quality work on any given application, no matter how it is smoothed over the bi-week.

5

To explore the overall effect of the 2011 reform on examination quality, we estimate difference-in-difference specifications that draw on the fact that GS-14 examiners can sign off on their own decisions while lower GS-level examiners require supervisory sign off on their decisions before receiving workload credit, which tends to come with a delay. This structure leaves the temporal incentives of lower GS-level examiners blunted—both pre- and post-reform—given their less-than-complete control over the timing of the receipt of credit for their efforts. As a result, while all examiners faced altered incentives due to this reform, examiners at GS-14 levels faced an arguably stronger change in incentives at that time.

Using this framework, we estimate the overall effect of the 2011 reform on two separate but related dimensions of patent examination quality: the timeliness of review and the accuracy of review. Beginning with timeliness, our findings suggest that the net effect of the reform was indeed beneficial—i.e., we document a net reduction in average examination pendency emerging subsequent to (and not prior to) the reform for the treatment relative to control examiners. That is, our results suggest that the pendency gains associated with the reduction in shotgun rejections (i.e., the reduction in wasted first-round reviews) implied by the examination-smoothing findings summarized above are not otherwise offset by deleterious pendency effects stemming from the reform that apply to applications across the board, regardless of where in the bi-week they fall.

This reduction in examination pendency may benefit welfare insofar as long patent review times have been found to interfere with the licensing of new technologies (Gans et al. 2008; Farre-Mensa et al., 2020), in addition to more broadly impeding innovation and job creation and frustrating the launch of new products (Hegde et al. 2020).<sup>4</sup> The other side of the examination-

<sup>&</sup>lt;sup>4</sup> For instance, examination delays may lead firms to rely increasingly on trade secrets instead of patents, potentially preventing public disclosure of information that future inventors may build upon. One study estimated that an additional year of delay in reviewing patent applications at the U.S. Patent Office, European Patent Office, and the Japanese Patent Office costs the global economy 9.2 billion dollars (London Economics 2010).

delay coin is production levels. The quicker that the Patent Office can dispose of applications, the more that it can work through its backlog of applications awaiting review. The Patent Office has frequently identified its backlog as among its biggest challenges (Patent Office 2008).<sup>5</sup> Even to this day, over one million patent applications are awaiting review and less than 40% of initial reviews conducted meet the agency's pendency target (U.S. Patent & Trademark Office 2021).

While delaying patent determinations may be costly, getting the decision wrong is costly as well. For instance, issuing invalid patents—such as patents on novel or obvious innovations may leave markets with diminished access to the patented products (in light of patents' exclusionary effects), while ostensibly failing to provide commensurate innovative benefits. Accordingly, we also evaluate the net effect of the 2011 reform on examination accuracy. To flag for review errors, we take advantage of the fact that many U.S. patent applicants also seek protection on the underlying innovation from the European Patent Office (EPO), an office with similar patentability standards but that invests considerably more resources per application than the U.S. Accordingly, we form examination accuracy flags by observing whether the EPO decisions on the underlying innovations align with that of the U.S. examiners. Ultimately, our difference-in-difference findings suggest no offsetting deterioration in examination review accuracy—i.e., no effects on either granting invalid patents or rejecting valid patents.

Altogether, our analysis of the Patent Office's 2011 reform suggests that the adoption of high-powered—albeit less discretionary—incentives to perform timely reviews led to a reduction in costly examination delays without sacrificing examination accuracy.

<sup>&</sup>lt;sup>5</sup> The addition of continuous incentives to an existing quota system may increase examination production levels through two channels. First, in a direct sense, continuous incentives may cause examiners to exceed the quota. Second is an indirect channel. As above, continuous incentives may discourage examination delays that otherwise lead to a greater number of unnecessary later rounds of review. By reducing overall examination pendencies, a continuous incentive system may allow examiners to pick up more applications in the first place. As discussed below, certain institutional features limit our ability to test the direct channel; however, our analysis does suggest productivity gains through the second channel.

The paper proceeds as follows. In Section II, we situate our analysis within various literatures, including the literature on the determinants of patent examination quality and the personnel and organizational economics literatures discussing performance appraisal structures. In Section II, we provide background on the patent examination process and the Patent Office's 2011 reform to its performance-appraisal system. In Section III, we describe the data, and in Section IV, we conduct our analysis of the 2011 performance-appraisal reform on patent examiner behavior. In Section V we explore the implications of the examination smoothing associated with the 2011 reform for the overall quality of patent examination. Section VI concludes.

### II. BACKGROUND

### A. Literature Review

Our analysis contributes to several literatures, beginning with the nascent literature on patent examination quality. Motivating these inquiries are foundational studies documenting rampant heterogeneity in patent examiner decisions (Cockburn et al. 2003). Studies illuminating the determinants of this heterogeneity include those exploring the effects of examiner experience and examiner resource allotments on granting proclivities and on the accuracy of reviews (Lemley and Sampat 2012; Frakes and Wasserman 2017; Lei and Wright 2017; Frakes and Wasserman 2023). Related studies shed light on administrative-agency-level incentives to shape examiner granting practices to address funding shortages and backlogs (Frakes and Wasserman 2013, Frakes and Wasserman 2015).<sup>6</sup> Of closest relevance to the present study is Frakes and Wasserman (2020),

<sup>&</sup>lt;sup>6</sup> Further related are studies that draw on examiner heterogeneity in practices to construct instrumental variables designs to study the effects of patenting on various outcomes, such as the effects of patents on cumulative innovation (Sampat and Williams 2019) or the effects of patents on venture capital funding (Farre-Mensa et al. 2020). Relatedly, Aneja et al. (forthcoming) explore the effects of examination scrutiny in early rounds of review on the inclinations of female inventors to continue with the application process.

which is summarized in the Introduction and discussed in greater detail below, along with related studies on deadline effects among patent-prosecuting attorneys (Balasubramaian et al. 2017).

Patent-examination investigations of this nature have not been confined just to the academic literature, with concerns over examination quality leading to a series of congressional hearings and investigations by the Office of the Inspector General of the Department of Commerce and by the Government Accountability Office (OIG 2014; OIG 2015; GAO 2016).

Our study further contributes to the growing literature within administrative law that recognizes the impact that internal structure of agencies—over which administrators have significant control—has on the dynamics of an agency's decision-making. More specifically, this literature identifies that seemingly ministerial tweaks to internal agency structures can have wide-ranging public policy impacts (Nou 2015 and Carpenter 2015). Within this field, even fewer studies have focused on the impact that the structure of performance appraisal systems or quality assurance programs have on agency outcomes (Ames et al. 2020).

Our analysis likewise contributes to the personnel and behavioral economics literature studying quotas. Much of the existing literature here focuses on the comparison between quotas and linear commission plans in the sales context (Basu et al. 1985; Holmstrom and Milgrom 1987; Darmon 1997; Kim 1997; Oyer 2000; Chung et al. 2014), with a small subset of this literature exploring the effects of varying the temporal frequency by which quotas are set (Chung et al. 2020). These sales-focused studies, however, are centered more on the implications of compensation structures for motivations to hit certain production levels—e.g., addressing concerns over diminished incentives to produce when workers are far out of the range of hitting their quotas. These studies have largely not focused on the implications of deadlines for delayed work effort.

Certain other studies, however, have addressed the self-control (and related) implications of workplace-based deadlines (Kremer et al. 2015; Cadena et al. 2011; Duflo et al. 2011; Asch 1990).

We are not aware of any studies that have empirically explored the procrastination implications of quotas/deadlines—a discrete versus continuous inquiry in the spirit of the quotas versus commissions debate—as we attempt to do in this paper. Moreover, further building on the personnel and organizational economics literatures, our analysis explores dimensions of quality beyond productivity and timeliness of review. The unidimensional nature of the outcome of interest in the sales context—whether a sale occurred—limits the ability to study other dimensions of quality. Moreover, among the empirical literature specifically focused on procrastination behaviors, only a few studies have explored the impact of decreasing procrastination on the quality of tasks performed (Cadena et al. 2011; Kremer et al. 2015; Benzarti 2017).

Furthermore, as discussed in the Introduction, various institutional features of the patent examination context offer certain advantages to studying the impacts of discrete versus continuous incentive systems on worker delays / procrastination. Mainly, such features facilitate the ability to empirically identify worker delays in the first place. As we will develop in detail in our empirical analysis below, the iterative nature of the patent examination process—combined with theoretical predictions as to how examiners navigate this process—allows us to develop affirmative markers suggestive of procrastination.

Further aiding the ability to empirically identify instances of delayed work effort, the scope of possible non-procrastination explanations behind deadline bunching itself is arguably narrower in the patent context. For instance, unlike with sales, there is little concern that deadline bunching by patent examiners reflects the need to pull in work from future periods to meet insufficient opportunities in the present period. With substantial, ever-present backlogs facing examiners, such opportunities abound. Relatedly, in some workplace contexts, deadline effects may stem not from delayed work but from unexpectedly difficult tasks conducted earlier in the observation period. Easing such concerns, patent examiners facing a complex review are legally instructed to allow the application should they fail to find a basis to reject within a designated number of hours.<sup>7</sup>

A further, more general advantage of turning to the patent examination context to study questions of this nature is that the entire examiner corps is essentially tasked with the same standardized, easily-codified job—i.e., to review claims, apply patentability standards, and render patentability decisions (Hegde et al. 2022). This facilitates our ability to explore the determinants of examiner decisionmaking while aggregating the decisions of thousands of examiners.

Finally, building on the above quality-versus-production point, our study offers an empirical contribution to the multi-tasking literature (Holmstrom and Milgrom 1991). However, instead of broadly exploring the quality and productivity implications of fixed wages / salaries versus incentive pay, our study focuses more specifically on the temporal monitoring structure of an incentive-pay system.

### B. Background on Patent Examination

Incoming patent applications are routed to an Art Unit, a group of 15 to 30 examiners who review applications in the same technological field. Upon arrival, the Supervisory Patent Examiner (SPE) of that Art Unit assigns the application to a specific examiner. Though not always purely random—insofar as there is some evidence of sub-specialization within Art Units—this assignment process is nonetheless tangential to the patent worthiness of the application (Righi and

<sup>&</sup>lt;sup>7</sup> Examiners are instructed as such due to the otherwise potentially never-ending nature of their jobs, which is essentially to prove a negative—e.g., that applications are not new.

Simcoe 2017) and thus effectively random for the purposes of the key examiner decisions explored below, mainly whether examiners allow applications in the first round.<sup>8</sup>

Upon assignment, the examiner conducts a prior art search—that is, a search of previous patents, applications, or publications that are material to the patentability of the focal invention. Upon completion of this search, the examiner assesses the invention's patentability based on the criteria outlined in the Patent Act—e.g., whether the invention is novel and nonobvious.

After this assessment, an examiner composes what we term a "first office action on the merits" (FOAM or first-round review), which either allows the patent to issue or outlines the reasons for why the invention fails to meet the patentability standards.<sup>9</sup> In the latter case—i.e., in the case of a FOAM rejection—the decision is non-final. An applicant responds to a non-final rejection by amending the claims<sup>10</sup> or disputing the rejection. Upon receipt of the applicant's response, the examiner will issue a second office action that will either: (1) allow the patent to issue, (2) finally reject the application, or (3) again non-finally reject the application. If the examiner sets forth all of the reasons for rejection in the FOAM and believes the invention still fails to meet the standards, she will finally reject the application. If the examiner issued an incomplete first-round review—i.e., failed to include all bases for rejection—she may issue a second non-final rejection, which includes new grounds for rejection.<sup>11</sup> In the face of a final rejection, the aggrieved applicant must abandon the application altogether, appeal the denial to the Patent Trial and Appeal Board, or continue the examination process by filing a repeat application.<sup>12</sup>

<sup>&</sup>lt;sup>8</sup> Lemley and Sampat (2012) and Frakes and Wasserman (2017) interviewed examiners and supervisors and confirmed that SPEs employ various randomization approaches to assigning applications.

<sup>&</sup>lt;sup>9</sup> By FOAM, we refer to the first office action associated with a given application. Technically, the first office action after the filing of a Request for Continuing Examination may be viewed as a FOAM by the Patent Office. Since that action would be continuing the same application file, however, we view it as different from the actual initial office action, which is where our focus lies.

<sup>&</sup>lt;sup>10</sup> An applicant may add more limitations to the claim and narrow the scope of the invention in order to overcome a rejection that the invention is not novel or obvious in light of the person of ordinary skill in the art.

<sup>&</sup>lt;sup>11</sup> An examiner may not render a final rejection that contains a new ground of rejection unless the new ground is necessitated by the applicant's claim amendments, or, in certain cases, if the new ground is based on information submitted in the applicant's information disclosure statement. MPEP 706.07(A).

<sup>&</sup>lt;sup>12</sup> Repeat applications generally fall into one of two categories: continuation applications and Requests for Continued Examination (RCE).

There are two main types of examiners working within each Art Unit. Examiners at pay grades GS-13 and below on the General Schedule pay scale are known as known as "assistant" examiners. After completing a special evaluation program, examiners may become "primary" examiners, generally reaching GS-14 at this time. Primary examiners attain full authority to sign off on all aspects of their reviews without the need for supervisory review. While assistant examiners independently review, and complete the bulk of the work associated with, the applications assigned to them, they must have their reviews and decisions approved by a supervisor—either by a primary examiner or by their SPE.<sup>13</sup>

## C. <u>Quotas</u>

Examiners are expected to attain a certain number of work credits, known as "counts," on a bi-weekly (every two weeks) and quarterly period, where such expectations are a function of the complexity of the field in which the examiner is working and of her position in the GS pay scale.<sup>14</sup> Counts have historically been earned upon the issuance of a FOAM or a disposal action.<sup>15</sup> In order to be promoted and be eligible for bonuses, an examiner typically need not only meet her workload goals but surpass them. Failure to earn the target amount of work credits can ultimately result in termination of employment. The interviews that we have conducted with examiners and with former SPEs confirm that the implications of failing to hit these targets create strong incentives for examiners to maintain examination throughput. The strong deadline effects documented below (pre-reform) further support the contention that examiners are highly motivated to hit their quotas.

<sup>&</sup>lt;sup>13</sup> One group of assistant examiners—GS-13 examiners with partial signatory authority (Frakes and Wasserman 2017)—have the ability to sign off on FOAMs without supervisory review. The data used in the present study, however, does not distinguish among the different types of GS-13 examiners.

<sup>&</sup>lt;sup>14</sup> FW provide more detail on the quota process. Bi-weekly periods generally end on Mondays.

<sup>&</sup>lt;sup>15</sup> A disposal action can include an allowance by the examiner, abandonment by the applicant (typically after receiving a final rejection) or the filing of an RCE. Notably, it does not include a second office-action non-final rejection.

### D. Scope for Procrastination Within Quota System: Predictions and Prior Literature

Frakes and Wasserman (2020) (hereinafter FW) finds that nearly half of first-round reviews were completed on the last day of the quota period. FW seeks to investigate whether such deadline effects were reflective of examiner time mismanagement, particularly procrastination. Ultimately, FW sets forth a series of findings suggestive of (though not necessarily determinative of) some degree of procrastination in FOAMs.

To begin, FW theorizes that if examiners did delay/procrastinate on initial reviews but were nonetheless motivated to hit their quotas, they would find themselves in a time crunch near deadlines to complete the needed reviews.<sup>16</sup> Examiners might respond to these time pressures by issuing "shotgun" rejections—i.e., weakly-conceived rejections that may be easily overcome by applicants. For example, an examiner may reject an application as lacking novelty by citing a scientific article that, upon inspection, clearly does not disclose the invention. As a result, such a rejection could be easily overcome in the next round of review by noting that the article does not actually disclose the invention. These shotgun rejections could be quickly drafted in the shadow of the imminent deadline. Moreover, importantly, by issuing such rejections, examiners could afford themselves the opportunity to correct a mistaken decision during a later round of review. In this sense, examiners may strongly prefer a shotgun rejection to the alternative of issuing ill-conceived, quick allowances. Acceptances, after all, would be final and non-correctable in nature.

Focusing predominantly on the period prior to the 2011 docket-management reform, FW finds remarkable evidence consistent with this prediction of procrastination-induced shotgun rejections at deadlines. In particular, FW finds that FOAM decisions rendered on the last day of the quota period were roughly 90% (relative to the mean) more likely to be rejected than FOAM

<sup>&</sup>lt;sup>16</sup> FW further predicts such delays in the first place by presenting a model of examiner decisionmaking with quasi-hyperbolic discounting.

decisions rendered on days within the quota period. Further, FW finds evidence that this substantial rejection disparity in first-round reviews between applications that were end-loaded and not end-loaded was ultimately corrected in later rounds as predicted. That is, applications with end-loaded and non-end-loaded FOAMs were ultimately allowed at nearly equal rates.<sup>17</sup>

FW further bolsters a procrastination explanation for these empirical observations by showing that both clustering of reviews at deadlines and shotgun rejection patterns in the shadow of deadlines intensify once examiners begin to work-from-home, consistent with a prediction that time-inconsistent behaviors would only intensify upon the reduction in daily supervision.<sup>18</sup> Despite this additional evidence of procrastination, FW acknowledges the possibility that some portion of the deadline effects they observe—e.g., the clustering of reviews at deadlines and the rejection patterns at deadlines—could be explained by models of task sorting. More specifically, FW addresses a story in which examiners sort patent applications by leaving close-call applications until the end of the bi-week and beginning the quota period with more straightforward applications.

To be sure, though not fully acknowledged in this light by FW, an application-complexity sorting story may itself be consistent with models of procrastination.<sup>19</sup> An actor may simply be procrastinating on the completion of the more unpleasant tasks (such as close-call applications), as distinct from procrastinating across the board (on the review of all applications). This slight modification to the procrastination story may nonetheless explain the same pattern of results—i.e., bunching at deadlines and shotgun-rejection behavior—with the added nuance that the delayed reviews tend to be focused on the close-call end of the application-complexity spectrum.

<sup>&</sup>lt;sup>17</sup> FW finds evidence that end-loaded first-round decisions relative to non-end-loaded FOAMs were more likely to be associated with non-final rejections in the second round—i.e., with decisions that are effectively examiner admissions of insufficient first-round reviews.

<sup>&</sup>lt;sup>18</sup> There is no apparent reason to think that task sorting would increase upon the commencement of working from home.

<sup>&</sup>lt;sup>19</sup> After all, instead of envisioning an actor performing a task in some unitary sense, imagine an actor facing tasks of different types, where the actor applies different time discounts across task types according to their degree of unpleasantness / salience costs (Akerlof 1991).

That being said, FW acknowledges that examiners may sort reviews within the bi-week for reasons unrelated to time-inconsistent behaviors. For instance, examiners may hold on to close-call reviews—both until the end of the bi-week and into future rounds—to buy time to discover relevant prior art and to resolve the uncertainty over close-calls, even if it ultimately means spending longer on the overall review than expected by the Patent Office.

Importantly, the significance of the present investigation into the effects of deadlines versus continuous review systems does not hinge on which version of the above stories explains the observed deadline effects. After all, in each case, we observe a prolongment of examiner reviews, not just until the end of bi-weekly quotas but also conscious extensions of reviews until subsequent rounds, which may not take place until months later. Accordingly, each case can be seen as a form of time mismanagement that frustrates the Patent Office's goals of timely reviews and limited backlogs.<sup>20</sup> No matter which of these explanations lies behind the deadline bunching, the deadline rejection patterns, and the ultimate application delays, the empirical question remains as to whether these patterns of behavior may reverse upon the Patent Office's 2011 reform, which uses higher-powered incentives to encourage timelier reviews.

## E. 2011 Performance Appraisal Reform: Prediction

In 2011, the Patent Office executed the most substantial reform to its Performance Appraisal Plan in nearly twenty-five years. A central component to this reform was the creation of a "Docket Management Element" to its appraisal criteria. Under this system, a new component was added to examiners' overall performance scores (representing 20% of the overall score) where this component is a function of the average number of days that examiners spend processing office

<sup>&</sup>lt;sup>20</sup> Moreover, this prioritizing of easy tasks and delays on harder tasks may stem from a set of preferences that places different time discounts on easy versus hard tasks and thus can be seen as a behavioral practice highly related to classic models of procrastination based on present-biased preferences.

actions (Marco et al. 2017). In addition to its implications for the examiner's overall performance appraisal score, better performance on this docket management element has implications for the receipt of a special quarterly bonus known as the Pendency Award. Examiners may earn bonuses at levels equal to 0.25 - 0.75% of their salary if their docket management scores hit a series of designated targets. Further, examiners may be eligible for a supplemental award of an additional 0.5 to 1% of their salary if such targets are consistently met over four quarters (Marco et al. 2017).

The addition of these docket-management scores and associated awards creates incentives to avoid delaying office action reviews until the end of the quota periods. Consider an examiner contemplating completing a FOAM review on April 1<sup>st</sup>. Assume their next quota period end is April 12<sup>th</sup>. The examiner will face no quota-related consequences for completing the review on the 12<sup>th</sup>. However, this new docket management component provides an incentive to complete the review sooner than the 12<sup>th</sup>.<sup>21</sup> As a result, work effort may be smoothed throughout the quota period rather than heavily clustered at the end.

This smoothing of work effort not only will result in accelerating the office action decision but may also lead to a decrease in the incidence of low-quality shotgun rejections occurring at the end of the bi-weekly quota. After all, the theory set forth above underlying the issuance of shotgun rejections stemmed from the time pressures and heavy remaining workloads in the shadow of deadlines that were created by previous delays, pressures that one might expect to diminish with smoothed work effort. In other words, we predict that the imposition of the 2011 reform will lead to a reduction in the various time mismanagement behaviors discussed above in the presence of the Patent Office's bi-weekly quota system.<sup>22</sup>

<sup>&</sup>lt;sup>21</sup> This new component does not remove all temporal discretion, however, and provides some flexibility to patent examiners. After all, it provides for a running average number of days that an examiner spends processing FOAM.

<sup>&</sup>lt;sup>22</sup> We are not aware of other aspects of the 2011 performance appraisal reform or other contemporaneous policy changes that might be expected to lead to similar effects. We do note that beginning in October, 2010, the Patent Office reformed the way in which its Office of Patent Quality Assurance (OPQA) performs quality reviews on a random selection of each examiner's applications annually. Prior to this reform, the OPQA

In the empirical analysis below, we test these various predictions. That is, we first test whether examiners respond to the 2011 reform by spreading their examination reviews more smoothly throughout the bi-week period—i.e., less clustering around deadlines. Next, we test whether the increase in uniformity of work effort with respect to initial reviews also results in a smoothing of examination outcomes (e.g., rejection patterns) over the bi-week.

### III. DATA

To explore these questions, we collect information on over 7 million individual patent applications from the Patent Office's Patent Application Information Retrieval ("PAIR") database,<sup>23</sup> drawing on records from patent applications filed on or after March 2001 and through June 30, 2022. The PAIR database provides daily information on a number of characteristics of, and events associated with, each application, including the timing and disposition of the first and subsequent office actions and the name of the associated patent examiner. For each such examiner, we obtained information from a Freedom-of-Information-Act (FOIA) request regarding the precise timing of each GS-level promotion that the relevant examiner received over her career.

Our analysis will predominantly focus on first-round review decisions reached by the end of 2015. This still provides a number of post-treatment years from which to evaluate the effects of the 2011 reform, while providing enough time following the relevant first-round decisions to

stated that part of the quality assessment score it forms as a result of these reviews was based on the "quality of the actions taken during the course of the examination." Following the October 2010 OPQA reform, the Patent Office clarified this criterion to indicate that OPQA quality assurance calculations will specifically consider the quality of initial searches and the quality of first-office action reviews. We acknowledge the possibility that this modification to its quality assurance criteria might also encourage examiners to not compromise quality on first round reviews by delaying until quota deadlines. Ultimately, we think that the results we present below are more consistent with the docket-management continuous-time reform for several reasons. First and perhaps most importantly, we note that the OPQA process is not designed to have implications for individual performance evaluations and, in fact, is prohibited from consideration in individual performance reviews. The associated OPQA reviews—which are substantially fewer in number than the reviews associated with an individual examiner's formal performance appraisal—are designed to form office-wide quality assurance metrics aimed at tracking office-wide quality. Second, the timing of the findings documented below is far more consistent with the initiation of the docket management system—i.e., third quarter of 2011—than the timing of the effective implementation of the new OPQA system. Third, the reduced end-loading and improved examination quality that we document following 2011 is stronger in the case of GS-14 examiners—relative to a control group of GS-11 examiners—i.e., stronger in the case of examiners that have greater control over the temporal implications of their decisions. It is not clear why this heterogeneous effect would be expected to result from the OPQA reform. <sup>23</sup> More specifically, we collect data from the Patent Examination Research Dataset, which itself draws from the PAIR database and which is

<sup>&</sup>lt;sup>23</sup> More specifically, we collect data from the Patent Examination Research Dataset, which itself draws from the PAIR database and which is organized by the Office of the Chief Economist at the Patent Office.

track its long-term outcomes—e.g., ultimate application allowance, ultimate application duration, and outcomes at the European Patent Office—without concerns over right truncation of the data.

To provide a marker of the legal validity of the patent applications of interest, we draw on administrative data from the Organisation for Economic Co-operation and Development's (OECD) Triadic Patent Family Database (TPFD), which collects information on patent application outcomes associated with patents filed at various world patent offices including the U.S. Patent Office and the European Patent Office (EPO). Using these data, we follow various studies (Frakes and Wasserman 2023, Lei and Wright 2017, Frakes and Wasserman 2017, and Lemley and Sampat 2012) and rely on the fact that many U.S. applicants likewise file for patent protection on the underlying innovation with the EPO. The EPO has essentially similar patentability standards to the U.S. but invests substantially more examination resources per application (including examination time allotments and examiner compensation) and works in examination teams (Picard and van Pottelsberghe de la Potterie 2013; Chien 2018). Ideally, to test for errors in the decisions reached by U.S. patent examiners, external researchers would hire knowledgeable experts to spend considerable time and review the decisions reached by the examiners. While such an exercise would be prohibitively expensive and burdensome for us as researchers, we can take advantage of this institutional feature of the world patent system and use EPO examination teams as a proxy for these hypothetical external reviewers.

We use this approach to proxy for both Type I and Type II examination errors—i.e., erroneous allowances and erroneous rejections. In short, we flag Type I errors by observing whether a U.S.-issued patent's "twin" is rejected by the EPO and flag Type II errors by observing whether a rejected U.S. application's "twin" is allowed by the EPO. At the outset, we note that one limitation of this examination-accuracy analysis is that we cannot perform it on the full sample

of U.S. patent applications, only those that are also the subject of EPO applications. Across all technologies, this entails limiting our analysis to roughly 1/3 of the primary analytical sample.

	FOAMs Completed in 2005-2015 (N=1,627,471)			FOAMs Completed in Three Years Pre-Reform (2008-2010)			FOAMs Completed in Three Years Pre-Reform (2008-2010)		
	Overall	End-	Non-	Overall	End-	Non-	Overall	End-	Non-End-
		loaded	End-		loaded	End-		loaded	loaded
		FOAM	loaded		FOAM	loaded		FOAM	FOAM
			FOAM			FOAM			
End-loading	0.317	-	-	0.415	-	-	0.214	-	-
Rate	(0.465)			(0.492)			(0.410)		
First-Round	0.170	0.108	.198	0.158	0.050	0.234	0.182	0.181	0.182
Review	(0.375)	(0.310)	(0.398)	(0.365)	(0.219)	(0.423)	(0.386)	(0.384)	(0.386)
Allowance									
Rate									
Overall	0.791	0.771	.799	0.764	0.738	0.782	0.812	0.814	0.811
Allowance	(0.407)	(0.421)	(0.400)	(0.424)	(0.439)	(0.412)	(0.390)	(0.388)	(0.391)
Rate									
Application	269.67	312.41	249.82	285.72	333.05	252.15	245.01	262.18	240.33
Duration (Days	(313.35)	(338.79)	(298.72)	(329.71)	(343.38)	(315.37)	(288.23)	(308.74)	(282.20)
Between									
FOAM and									
Disposition)									
Incidence of	0.310	0.322	0.305	0.324	0.328	0.319	.288	0.286	0.288
Application	(0.462)	(0.467)	(0.460)	(0.467)	(0.469)	(0.466)	(0.453)	(0.451)	(0.453)
Review Error									
(Based on EPO									
Benchmarking)									

TABLE 1: SUMMARY STATISTICS FOR GS-14 EXAMINERS

Notes: PAIR data in Columns 1-3 are for first-round review decisions completed between 2005 and 2015 for GS-14 examiners (at time of firstround review), based on 1,627,471 first-round reviews. Columns 4-6 focus on a subset of these records in the three years prior to the 2011 reform and Columns 7-9 focus on a subset of these records in the three years following the 2011 reform.

Focusing on GS-14 examiners, Table 1 presents summary statistics for the resulting sample of first-round reviews, separately for end-loaded and non-end-loaded reviews—i.e., reviews completed on the final day of the quota period—as well as separately for the three years before and after the 2011 reform. We note two preliminary observations from Table 1. First, we document a substantial reduction in the incidence of an end-loaded first-round review following the 2011 reform that incentivized examiners to complete reviews earlier in the bi-week.

Second, Table 1 demonstrates a substantial differential in the pre-reform period in firstround allowance rates between end-loaded (5%) and non-end-loaded (23%) reviews, along with the near elimination of this differential in the post-reform period (18% in both cases). This is consistent with the predictions of Part II, whereby examiners who delay on their reviews during the quota period but who are motivated to hit their quota will issue shotgun rejections to buy time in future rounds to correct the low-quality initial reviews and whereby these behaviors will diminish once examiners begin to smooth their reviews over the bi-week period following the 2011 reform. In the analysis below, we will take a range of steps to verify that this smoothing of behavior is consistent with a causal response of the reform itself and will also further support the shotgun-rejection story with a consideration of other application outcomes.

### IV. EXAMINATION-SMOOTHING ANALYSIS

### A. Smoothing of Initial Reviews

We begin this investigation into the relative effects of discrete versus continuous incentives on workplace timeliness by considering an observation of deadline bunching. In particular, we explore the degree to which examiners complete FOAM reviews on the final day of the bi-weekly quota period and then evaluate the degree to which initial reviews are smoothed throughout the biweek upon the imposition of a continuous temporal incentive system.

In Figure 1, we demonstrate the evolution of the end-loading rate year-by-year and separately by examiner GS-level. A separate investigation by GS-level is particularly helpful insofar as it allows us to distinguish GS-14 examiners from examiners at lower grades, the latter of which need a supervisor (either a SPE or an assigned "primary" GS-14 examiner) to sign off on the review. In this sense, end-loading by GS-14 examiners reflects clear examiner-specific deadline effects, whereas end-loading at lower GS-levels may reflect a combination of deadline effects by the reviewing examiner and the supervisor signing off on the review.

FIGURE 1. END-LOADING RATES BY GS-LEVEL AND YEARS BEFORE AND AFTER 2011 PATENT OFFICE PERFORMANCE APPRAISAL REFORM



Notes: this figure plots the percentages of first-round reviews that are completed on the last day of the bi-weekly quota period, separately by year and by examiner GS-level.

As reflected in Figure 1, we document a precipitous drop in end-loading rates across examiners of all GS levels at the time of the 2011 performance appraisal reform. Consider, for instance, reviews by GS-14 examiners. In the period prior to the 2011 reform, examiners consistently end-loaded between 40 and 48 percent of their FOAM decisions. In the year of the reform, this rate falls to roughly 26 percent, and in the post-reform period, the rate by which GS-14 examiners end-load remains consistently in the low 20s. While this analysis reflects only time-series variation, the single-year drop in end-loading rates at the time of the reform is substantially larger than the relatively stable year-by-year changes in end-loading rates in both the pre- and post-reform years, bolstering a causal interpretation of these observations. Moreover, in Figure A1 of the Online Appendix, we estimate a regression version of the trends depicted in Figure 1 to

demonstrate that this precipitous drop cannot be explained by annual fluctuations in observable application characteristics (e.g., number and length of claims).

In Figure 2, we focus specifically on the 2011 reform year itself (fiscal) to demonstrate how and when end-loading began to decline so strongly. In particular, this Figure presents a histogram of the completion of FOAM reviews, broken down into daily frequency bins. To make sure that we focus exclusively on the reviewing examiner behavior—and not a combination of examiner and supervisor behavior—we show this figure for GS-14 examiners (though similar patterns are present for other grades, as expected from Figure 1). The daily frequency distribution depicted in Figure 2 demonstrates a striking degree of bunching in the completion of first-round reviews at the end of bi-weekly quota periods. Examiner reviews are lightest at the beginning of each bi-week period. As the bi-week nears its end, workload counts gradually begin to increase, with a spike on the last day (the lull in the preceding two days reflects the fact that these are weekends given that quota deadlines generally occur on Mondays). Consistent with FW, we also document spikes at the end of each quarter, further suggestive of deadline effects to the extent examiners are given the option to catch up on missed bi-weekly targets at the end of each quarter.

Importantly, the magnitude of the end-of-quota spikes falls substantially during the course of this reform year. In particular, this drop mostly occurs in the May / June period of 2011. This is consistent with the timing of the roll-out of the new docket-management system, which predominantly occurred in the third quarter of the 2011 fiscal year (i.e, April). Figure 2 thus suggests a near immediate examiner response to the new incentives imposed by the docket management system.<sup>24</sup>

<sup>&</sup>lt;sup>24</sup> The Docket Management element of the Patent Office's Performance Appraisal Plan—which created the daily pendency incentives—was established following an October 22, 2010 MOU between the Patent Office and examiner union (the Patent Office Professional Association or POPA). However, it was not until roughly the third quarter of the 2011 Fiscal Year before the new system was rolled out and, particularly, before the PALM database incorporated this new system into its coding (the PALM database keeps track of daily activities and events at the Patent Office).



FIGURE 2. FREQUENCY DISTRIBUTION OF FIRST-OFFICE-ACTION REVIEWS ACROSS INDIVIDUAL DAYS IN FISCAL YEAR 2011 (GS-14 EXAMINERS)

Notes: this figure depicts a histogram of the dates of completion of the first-round reviews during the 2011 fiscal year for GS-14 examiners (the corresponding figure using all examiners is virtually identical). The dashed reference line at April 1, 2011 indicates the beginning of the third quarter of the 2011 fiscal year. It was during this quarter that the new docket

management element came into effect.

Finally, in Figure 3, we show a generalized histogram over a 14-day bi-weekly period, effectively zooming in (and averaging over) on the patterns depicted in Figure 2. In Panel A, we do this focusing on a three-year period prior to the 2011 reform. In Panel B, we focus on a three-year period after 2011. Consistent with Figures 1 and 2, we document a large reduction in end-loading and a significant smoothing of examiner work efforts following 2011. While the review

For a discussion of these timing matters, see the meeting transcript and slides associated with the Patent Public Advisory Meeting at the Patent Office on April 11, 2011 (transcript: https://www.uspto.gov/sites/default/files/about/advisory/ppac\_transcript\_110414.pdf).

completions do not become uniform in the post-reform period, there is an evidently notable shift towards greater uniformity, with the spike being shifted somewhat towards the first week and the Friday, Saturday and Sunday before the quota-ending Mondays.

## FIGURE 3. FREQUENCY DISTRIBUTION OF FIRST-OFFICE-ACTION REVIEWS ACROSS GENERALIZED BI-WEEKLY PERIOD, PRE- AND POST-2011 REFORM (GS-14 EXAMINERS)



Notes: this figure depicts a histogram of the completion times of the first-round reviews for GS-14 examiners over a generalized bi-week period (the corresponding figure using all examiners is virtually identical). Day 14 corresponds to the deadline date.

#### B. Smoothing of Examination Outcomes: End-loaded versus Non-end-loaded

We next test whether this first-round review smoothing also results in a smoothing of examination outcomes. After all, as predicted in Part II, if examiners smooth their effort over the bi-week, they will experience less of a time-crunch to meet their quota at the deadline and thus less of a desire to issue shotgun rejections in the face of such deadline time pressures.

We begin this outcome-smoothing inquiry in a binary fashion and compare first-round review outcomes as a function of the first-round review being end-loaded (completed on the last day of the quota period) versus non-end-loaded. Further, we begin by considering first-roundreview allowance decisions. In particular, for each calendar year leading up to and following the 2011 reform, we estimate the following specification:

$$FOAM\_Allow_{ait} = \alpha + \gamma_{i} + \beta_{1}End - loaded_{ait} + \beta_{2}X_{ait} + \varepsilon_{ait}$$
(1)

where *a* indexes the individual application, *i* indexes the examiner, *t* indexes the year in which the first office action is completed. *End-loaded*<sub>ait</sub> indicates whether the FOAM was completed on the last day of the quota period. **X**<sub>ait</sub> captures certain application characteristics.<sup>25</sup> *FOAM\_Allow*<sub>ait</sub> indicates whether the application was allowed on the FOAM. Examiner fixed effects,  $\gamma_i$ , alleviate concerns that examiners with high propensities to end-load differ in their allowance propensities. In this sense, we compare allowance outcomes within given examiners and across applications that they review at the end of quota periods and those that they review within the quota period.  $\beta_1$  captures this coefficient of interest. As with Figures 2-3, we limit this analysis to those reviews conducted by GS-14 examiners.<sup>26</sup>

In Figure 4 Panel A, we plot the  $\beta_1$  coefficients (and 95% confidence intervals) along the y-axis and the corresponding first-round-review year along the x-axis. As demonstrated, in the years leading up to the 2011 reform, we consistently document a large negative coefficient—ranging from -0.1 to -0.2—suggesting that end-loaded applications are rejected at a substantially higher rate relative to non-end-loaded applications, consistent with shotgun rejection predictions and consistent with the mean differences in Table 1. This magnitude is staggering at roughly 60-100% of the mean first-round-review allowance rate. In 2011—during the middle of which the

<sup>&</sup>lt;sup>25</sup> In our primary specifications, X includes an indicator for the incidence of a small entity applicant. In the Online Appendix, we demonstrate that the temporal smoothing analysis of the 2011 reform is virtually identical with the inclusion of other claims-related characteristics of the applications. We leave those as a robustness check as we ideally want application characteristics at the time of filing (to avoid endogeneity concerns) and claims characteristics are only available at the time of application publication, which generally occurs 18 months later. In estimating specification (1), standard errors are clustered to correct for autocorrelation within given examiners over time.

<sup>&</sup>lt;sup>26</sup> That being said, we do document similar pattern of results at lower-grades, albeit at smaller magnitudes as one would expect given this attenuation point—we return to this point below.

Patent Office implemented the relevant reform—this relationship falls by over half, and in the years following the reform, the  $\beta_1$  coefficient consistently hovers near zero.

FIGURE 4: ESTIMATED RELATIONSHIP BETWEEN THE INDICATED OUTCOME ASSOCIATED WITH A FIRST-ROUND-REVIEW DECISION AND THE INCIDENCE OF THAT INITIAL REVIEW OCCURRING ON THE LAST DAY OF THE QUOTA PERIOD, BY YEAR (GS-14 EXAMINERS)



Notes: each point estimate (and associated 95% confidence interval indicated in the vertical bars) across each figure is from a separate regression of the indicated outcome variable on the incidence of the first-round review being end-loaded, for the indicated year. Each specification includes examiner fixed effects and controls for the incidence of a small-entity applicants, with standard errors clustered at the examiner level. Sample sizes range from 107,887 in 2005 to 226,293 in 2015 for Panels A-C and 31,364 in 2005 to 49,905 in 2015 for Panel D.

This pattern suggests a striking effect of the 2011 reform on the degree to which examiners issue rejections for first-round reviews completed at the end of quota periods. Consistent with a causal response to the reform, we observe no trend towards smoothing of first-round-review allowances prior to the reform. If anything, the trend prior to the reform had been in the direction of stronger and stronger differentials in first-round-review allowances between end-loaded and

non-end-loaded reviews. Further consistent with a causal interpretation of these findings, the change in magnitude of the  $\beta_1$  coefficient in the year surrounding the reform is substantially higher than the average year-by-year swings in this measure before and after the reform.<sup>27</sup> Moreover, we are aware of no other development contemporaneous with the 2011 performance-appraisal reform that would be expected to have such a large impact on first-round-review allowance decisions.<sup>28</sup>

In the remaining panels of Figure 4, we conduct identical exercises but now focusing on the additional markers of shotgun rejections at deadlines. We begin with the incidence of the application ultimately being allowed over the examination period. Recall from Part II that under the shotgun-rejection theory, while examiners in the pre-reform period would be more likely to reject first-round reviews for applications end-loaded in the initial round relative to those not endloaded in the initial round, those applications would all look roughly similar—or at least far more similar—when viewed on the basis of their final dispositions.

We confirm those predictions in Panel B. As demonstrated, in the years prior to the 2011 reform, applications whose initial reviews are end-loaded are ultimately allowed only 1.2 to 3.2 percentage points less frequently than applications whose initial reviews are not end-loaded, representing a 1.7-4.6% differential relative to the mean. This is tiny in comparison with the 60-100% relative differential in first-round-review allowance rates. It is not surprising to observe some minor differential in overall allowances rates in the pre-reform period even if examiners endeavor to review the application with sufficient rigor in later rounds. After all, a shotgun-

<sup>&</sup>lt;sup>27</sup> This is similar to the point we raised in connection with Figure 1, though we note that our analysis here relies on more than just time series variation, as we are also drawing on cross-sectionalvariation in the incidence of end-loading.

<sup>&</sup>lt;sup>28</sup> We estimate separate annual specifications in Figure 4 so as to demonstrate in an absolute sense how end-loaded and non-end-loaded FOAMs differ in their dispositions in the post-reform period—i.e., to demonstrate that it falls to near zero in the post-period. The results from Figure 4 are nearly identical when we estimate a single specification that includes all years in Figure 4 and that includes year dummies and their interaction with the end-loading incidence and thereafter plot the coefficients of the interactions.

rejection practice prolongs the prosecution period due to a wasted first round and increases the chances that an applicant will abandon their application (which is coded as a non-allowance).

Following the 2011 reform, we would continue to expect no meaningful overall-allowance differential between applications with end-loaded and non-end-loaded first-round reviews. If anything, we would simply expect that the minor differential we do observe prior to the reform to go away upon the adoption of the reform. We indeed find this is the case in Figure 4 Panel B.

The real consequence of the shotgun-rejection behavior is examination pendency, since such rejections on the first-round reviews were meant to give examiners time to correct the matter on subsequent reviews. Accordingly, in Figure 4 Panel C, we build on Panels A and B by likewise demonstrating the year-by-year progression in the relationship between end-loaded versus non-end-loaded first-round decisions and the overall duration of the focal application. The overall duration of the focal application is defined as the time from the day the FOAM is completed until the ultimate disposition of the patent application.<sup>29</sup> This duration metric enables us to better isolate the effect that end-loading an initial review has on subsequent examination pendency.<sup>30</sup>

In the pre-reform period, we indeed find a strong positive relationship between first-round end-loading and examination duration, consistent with the shotgun-rejection theory. Following the 2011 performanace appraisal reform, we find a drastic diminishment in the link between examination-duration and first-round end-loading, with little such relationship left after the reform.

Again, the shotgun-rejection story entails examiners responding to deadline-looming time crunches by issuing weak rejections in order to offer more accurate patentability determinations in later rounds. Accordingly, both before and after the reform, we would predict that the examination

<sup>&</sup>lt;sup>29</sup> We actually take the natural log of the application duration to address outliers (while adding 1 to those applications allowed in the first-round review to avoid taking the log of 0). In the Online Appendix, we show that the implications of this analysis are identical when treating duration in levels.

<sup>&</sup>lt;sup>30</sup> There is no appreciable difference in the pendency of a patent application whose FOAM is endloaded versus non-endloaded when pendency is defined as the period from when the patent application is docketed to a patent examiner until the FOAM is completed.

accuracy associated wth the application's final disposition would not differ between those endloaded and not end-loaded on the first round, despite the massive difference in review scrutiny applied in the first round itself. We test this using the EPO benchmarking approach discussed in Part III, flagging for the existence of either a Type I error (allowing an invalid patent) or a Type II error (rejecting a valid patent). In Panel D of Figure 4, we specifically demonstrate that the degree to which examiners commit examination errors does not differ between end-loaded and non-endloaded first-round reviews and that this holds both pre- and post-reform. In Figure A18 of the Online Appendix, we demonstrate that this is true when we likewise explore Type I and Type II errors separately.

Altogether, these findings suggest that when examiners face a bi-weekly quota that they are incentivized to hit, they cluster much of their reviews at the deadlines, and that the key harm associated with these time-mismanagement tendencies is an increase in examination duration, with no ultimate harm in the accuracy of patent review. Moreover, these findings demonstrate that adding a meaningful set of rewards to provide timely reviews at daily levels leads to a temporal smoothing of examiner behavior and the apparent elimination of these negative examination-duration outcomes, while continuing to show no harms in ultimate patent review accuracy.

#### C. <u>Smoothing of Examination Outcomes: Day-by-Day</u>

We next unpack examiner behavior on a day-by-day basis over the bi-week, rather than lumping together the first thirteen days as Figure 4 does. With continuous incentives in place, not only would we predict a smoothing in outcomes between end-loaded and non-end-loaded firstround reviews (in a binary sense), but we would also predict a smoothing in examination outcomes over the full course of the bi-week. Even in the pre-period, this analysis builds on that of FW, which limited its analysis to a comparison of end-loaded and non-end-loaded reviews. To parallel the above analysis, we begin this inquiry with a consideration of first-round allowance decisions. In Panel A of Figure 5, we plot the mean first-round-review allowance rate for each day over the generalized bi-week. We do so separately for the pre-reform period (2008-2010) and the post-reform period (2012-2014), focusing on a six-year window around the reform and excluding the reform-year itself given that the reform occurred in the middle of the year.

FIGURE 5: RELATIONSHIP BETWEEN THE INDICATED OUTCOMES ASSOCIATED WITH THE FIRST-ROUND DECISIONS AND THE INCIDENCE OF THAT FIRST-ROUND DECISION OCCURRING ACROSS THE GENERALIZED BI-WEEKLY PERIOD, BEFORE AND AFTER 2011 PATENT OFFICE PERFORMANCE APPRAISAL REFORM (GS-14 EXAMINERS)



Notes: each figure presents means of the indicated outcome across the relevant day of the bi-week (from 1 being the first day to 14 the last day) in which the first-round review is completed. These means are presented separately for the three-year period (2008-2010) prior to the reform and the three-year period (2012-2014) following the reform, excluding the year of the reform. Sample sizes for Panels A-C are 341,423 in the pre-period and 536,913 in the post-period. Sample sizes in Panel D are 101,895 in the pre-period and 130,983 in the post-period.

In the pre-reform period, we demonstrate substantial variation in first-round allowance rates over the bi-week. We observe a significant drop in first-round allowance rates on the last

day of the quota period, which is consistent with examiners issuing shotgun rejections with the deadline imminent. Interestingly, we also observe a sharp, nearly monotonic fall in first-round allowance rates over the first thirteen days. We then plot evidence of a remarkable degree of smoothing of these daily allowance differentials following the continuous-incentives reform. In the post-reform era, we demonstrate near total uniformity in the first-round allowance rate over the generalized bi-week. The mean first-round allowance rate varies by only 0.2 percentage points across the bi-week. Put simply, following the reform, the likelihood that a first-round review is allowed is no longer a function of the day on which it is completed.

In Figure 5 Panel B, we plot the mean overall application allowance rate by the day the application's first-round review was completed over the generalized bi-week. In the pre-reform period, we show a generally monotonically decreasing pattern with a discernable dip on the final day. However, the degree of this variation is far more muted relative to the corresponding figure for first-round allowance rates. Again, this is consistent with the shotgun-rejection theory, in which examiners will ultimately review applications in later rounds with the proper rigor despite any rushed first-round reviews. Finally, post-reform, we likewise see a near complete smoothing of overall allowance patterns over the days of the bi-week in which the first round was completed.

In Figure 5 Panel C, we perform a similar exercise looking at the log of application duration as the relevant outcome. Our conclusions are much the same. Pre-reform, we see a spike in duration on the final day of the bi-week, along with a generally monotonic rise over the preceding days. In Figure A14 of the Online Appendix, we show the corresponding figure in levels and not logs. In the pre-reform period, the application duration whose first-round review is completed on the last day of the bi-week is, on average, over one hundred and fifty days longer than the application duration whose first-round review. In both

levels and logs, we then see a near complete temporal smoothing effect in connection with the 2011 reform.

In Figure 5 Panel D, we complete our analysis, showing the corresponding figure for the likelihood that the examiner commits either a Type I or Type II error. In both the pre-reform and post-reform period, this validity marker does not vary by the day on which the first-round review is completed. This result is consistent with the above-stated suggestion that examiners' time-mismanagement tendencies did not affect the accuracy of patent review. In Figure A19 of the Online Appendix, we demonstrate that this likewise holds when evaluating Type I and II errors separately.

### D. Task Sorting versus Procrastination

Models of procrastination based on present-biased preferences often predict a ramping up in anticipation of a deadline rather than a clustering entirely at the cusp of the deadline (Fischer 2001). Accordingly, the pre-reform increase in reviews throughout the course of the bi-week rather than just the spike at the end of the bi-week (see Panel A of Figure 3)—may be consistent with procrastination behaviors. Moreover, the pattern of first-round review outcomes throughout the bi-week—e.g., the generally monotonic decline in first-round allowance rates over the course of the bi-week (Figure 5)—may be reflective of a corresponding gradual realization of the time crunch that comes with the approaching deadline and the delays in work efforts, rather than a singular time crunch on the last day of the bi-week.

However, some aspects of this monotonic pattern in review outcomes throughout the biweek may also represent task sorting. Consider for instance an examiner preference for picking up easy allowances earlier in the bi-week and preserving closer calls for later in the bi-week. To be sure, as we discussed in Part II, task-sorting and procrastination are not necessarily mutually exclusive, as this close-call task sorting behavior may be a reflection of procrastination on particular tasks—i.e., procrastination on the more unpleasant close-call tasks. After all, presentbiased-induced procrastination behaviors need not always entail delaying on *all* tasks, but rather sometimes delaying on tasks whose associated costs are discounted at a different rate than other tasks. That being said, examiners may preserve close calls for the end of the bi-week for reasons unrelated to procrastination altogether, including in connection with an "option value" story in which examiners purposely delay close calls to buy time to discover relevant prior art.

In the Online Appendix, we provide evidence suggesting that some portion of the prereform patterns in Figure 5 likely represent task sorting, though likely a very small part of those patterns. To begin, in Figure A11 of the Online Appendix, we plot day-by-day (over the generalized bi-week) means of several patent application characteristics that are proxies for patent scope: the number and length of independent claims. Applications with greater scope may prove to be more difficult to examine—i.e., a patent examiner may need to conduct a more exhaustive prior art search and craft more rejections for a patent application that covers a larger scope of invention. We do find some evidence that, pre-reform, examiners are more likely to perform firstround reviews on applications with a narrower scope towards the beginning of the bi-week and initial reviews on applications with a larger scope towards the end of the bi-week.

However, when we present a version of Figure 5 in the Online Appendix (Figures A7-A10) that controls for a slate of characteristics of those applications reviewed on the focal day of the biweek, the pattern depicted in this alternative approach is highly similar, if not nearly identical, to that depicted in Figure 5. This suggests that the degree of task sorting present throughout the biweek is modest relative to the other behavioral explanations behind Figure 5—e.g., shotgun rejections due to procrastination-induced time crunches in anticipation of deadlines. Further consistent with this interpretation, we also demonstrate that the degree to which the *actual* first-round allowance rates fall over the generalized bi-week is substantially greater than the degree to which the *predicted* first-round allowance rates fall over this period, where such predictions are based on the observable claim characteristics (Figure A12 of the Online Appendix).<sup>31</sup>

No matter what combination of procrastination and task sorting may explain the pre-reform patterns in Figure 5, it is remarkable how much the continuous-incentive reform of 2011 smooths all of the relevant outcomes nearly perfectly over the bi-week. That is, while deadline/quota systems may leave open the scope for delayed work effort and task sorting (and interactions between the two), the imposition of the continuous incentive system appears to have unwound all such behaviors and eliminated all observed deadline effects.

In the Online Appendix, we further subject the findings presented in Figure 5 to a range of additional specification checks. For instance, we conduct a randomization-inference exercise and demonstrate that the striking bi-week pattern of first-round allowance rates is substantially inconsistent with what one would expect from chance alone (Figures A3 and A4). Further, while Figure 5 lumps the pre- and post-reform groups into three-year respective groups, we also plot versions of Figure 5 that show the corresponding bi-weekly patterns on a year-by-year basis (Figure A6). Doing so, and consistent with Figure 4, we demonstrate that the dramatic smoothing depicted in Figure 5 did not pre-date the reform. Finally, we show results using yet another marker of shotgun-rejections: the incidence of a second office-action non-final rejection (Figures A15-A17). Recall that such rejections are suggestive of incomplete first-round reviews. We likewise find evidence that the incidence of this new marker spikes on the final day of the bi-week and

<sup>&</sup>lt;sup>31</sup> Additional analyses set forth in FW further support the contention that task-sorting cannot explain all of the pre-reform deadline effects. That is, we note that FW finds that initial review bunching and markers for shotgun rejections intensify upon telecommuting / work-from-home.

monotonically increases leading up to the deadline, along with evidence of a smoothing of this incidence over the bi-week following the reform.

### V. IMPLICATIONS OF EXAMINATION SMOOTHING

### A. Overview

The above results suggest that the implementation of the continuous incentive system encouraged examiners to exhibit more continuous work efforts resulting in a smoothing of examination outcomes, including the near complete elimination of shotgun rejections. To the extent the 2011 reform diminished examiner procrastination, one would expect to see an overall improvement in the quality of patent examination post-reform, at least along the timeliness dimension. After all, if procrastination-induced shotgun rejections are eliminated, the churning of patent applications should diminish, and patent pendency should decrease.

Nevertheless, it is possible that the reform resulted in some offsetting harms. As noted in the Introduction, one of the primary drawbacks of a continuous incentive system is the lack of temporal flexibility and discretion that comes with a pure quota system. The diminished flexibility associated with higher-powered continuous incentives may result in added stress and anxiety, which could, in turn, result in lower-quality work product across the board.<sup>32</sup> The ability to sort and sequence work efforts over a discrete time period in particular ways—e.g., working in parallel rather than in sequence—may also potentially lead to quality gains.

It is unclear, a priori, which of these mechanisms associated with the adoption of the continuous incentive system wins out—i.e., the benefits of smoothing and reduced time mismanagement / procrastination versus the potentially lower quality output produced on any

<sup>&</sup>lt;sup>32</sup> It also may be possible that some of the pre-reform task sorting we identified was beneficial and its post-reform elimination results in a decrease in the quality of patent examination.

given day due to the reduced flexibility of a continuous incentive system. In this Part, we examine whether the adoption of the 2011 reform increased the *net* quality of patent examination, probing two dimensions of patent examination quality: timeliness of review and accuracy of review.

Simply observing the change in examination quality metrics before and after the 2011 reform would be problematic as examination pendency and accuracy may change meaningfully over time for reasons unrelated to the adoption of a continuous incentive system. To surmount this problem and to shed light—even if only suggestive—on the welfare effects of the 2011 reform, we estimate a difference-in-difference specification that draws on a prediction that the 2011 reform would be more likely to affect the behavior of one group of examiners over another.

In particular, we assume that while all examiners faced altered incentives due to this reform, primary examiners—i.e., GS-14 examiners—faced an arguably stronger change in incentives at that time. While not the case for GS-14 examiners, lower GS-level examiners require supervisory sign-off on their decisions before receiving workload credit. As such, lower GS-level examiners—both pre- and post-reform—lack complete control of the timing of the completion of, and the receipt of credit for, their reviews, unlike GS-14 examiners who retain total control over such matters. Given this diminished control by lower-GS-level examiners and given that the 2011 reform is meant to leverage examiner control over the timing of their decisions, we predict a more muted response to the reform for lower-GS-level examiners relative to GS-14 examiners.<sup>33</sup>

<sup>&</sup>lt;sup>33</sup> For instance, suppose that examiners prior to the reform are disinclined to accelerate their work product within the bi-week because such efforts may be thwarted anyway by supervisory delays. Given that supervisor delays will continue after the reform (since supervisory timings were not the target of the reform), examiners may continue to be disinclined to accelerate their work after the reform. In other words, the reform would not be expected to greatly change their behavior. We acknowledge that examiners may also respond to supervisor delays with the opposite reaction. That is, even pre-reform, lower-GS-level examiners may try to complete their reforms as early as possible to make room for supervisor delays. Of course, even if this were the case, it would perhaps simply continue in a similar fashion post-reform, as supervisor delays would still prolong receipt of work credit. Again, the key point is that supervisory sign off—which exists pre- and post-reform—leaves lower-GS-level examiners in weaker control of the timing of receipt of work credit, which leaves lower-GS-level examiners less responsive to a reform that is designed to alter examiner incentives regarding the timing of their work.

Of course, since this difference-in-difference structure is based on a stronger response to the reform for the treatment group relative to the control group—as distinct from no response among the control group—some care must come in the interpretation of the results. That is, our findings cannot be interpreted as the absolute magnitude of the effects of the reform. Nonetheless, to the extent we estimate a differential response of the reform across these groups, this analysis will shed light on the existence and sign of a net response of the reform on examination quality.

To clarify this approach further, we specifically select GS-11-and-below examiners as the control group and exclude GS-12 and GS-13 examiners from this analysis. Both before and after the reform, GS-11-and-below examiners did not receive credit for their work until their supervisor signed off on their office action review. This was also true for GS-12 and 13 examiners pre-reform. However, following the reform, pursuant to a new "Auto Count" rule, GS-12 and GS-13 examiners would receive credit for the purposes of their document management scores and rewards as of the time they finished their own work.<sup>34</sup> This Auto Count component to the 2011 reform complicates the ability to use GS-12 and 13 examiners as a control set given the multitude of changes they experience at once in 2011.<sup>35</sup>

#### B. First-Stage Analysis

As above, our goal with this difference-in-difference analysis is to explore the net effect on examination quality resulting from two behavioral responses to the 2011 reform: the smoothing of work over the bi-week and the potentially lower quality work on any given day due to the

 $<sup>^{34}</sup>$  As a result of this auto-count rule and as a result of the linkage between the data system used in the docket-management system and the PAIR data underlying our analysis, FOAM completions by GS-12 and 13 examiners after the reform are recorded in the PAIR data as of the time the examiner completes their review, even if that is before the supervisors have signed off. Prior to the reform, for GS-12 and 13 examiners, FOAM decisions were recorded as of the date the supervisor signed off.

<sup>&</sup>lt;sup>35</sup> Finally, we note that the types of decisions and tasks that primary examiners (GS-14 examiners) make that are implicated by the docketmanagement incentives that they face under the 2011 reform do not include tasks associated with their efforts to sign-off on assistant examiner decisions. As such, while we predict that examiners will be incentivized via the reform to reduce delays in their own examinations, we do not predict that supervisors / primary examiners will be incentivized via the new docket-management system to avoid delays in these sign-off tasks. Accordingly, we do not predict any temporal acceleration in supervisory sign-off in connection with the reform. E-mail from Robert L. Stoll, Former Commissioner for Patents, Patent & Trademark Office, to Melissa Wasserman (Sept. 7, 2023 12:35 CST) (on file with author) (noting that the 2011 reform did not change incentives for the sign-off of assistant examiners work product).

reduced flexibility of continuous incentives. For the reasons just stated regarding examiner control over the timing of receipt of work credit, we theorize that GS-14 examiners will be more responsive to the reform with respect to both such behaviors. It is only possible, however, to explore the first-stage of this analysis—i.e., the differential effect of the reform for treatment relative to control examiners on examiner decisionmaking—with respect to the smoothing behavior of examiners. Nonetheless, we hope that showing the first-stage for the smoothing analysis will validate more generally the prediction that GS-14 examiners will be more responsive to the reform.





Notes: each figure presents means of the indicated outcome variable by the day of the bi-week (1 being the first day and 14 being the last day) on which the first-round review was completed, separately for each year surrounding the 2011 reform.

In Figure 6, we show results from this first-stage exercise and indeed demonstrate a stronger smoothing response for GS-14 examiners relative to GS-11-and-below examiners. For these purposes, we focus on the pattern of first-round allowance decisions by examiners. The stronger smoothing effect on such allowances by GS-14 examiners is a function of two things. First, pre-reform, the slope of the first-round allowance rates over the bi-week was substantially stronger for GS-14 examiners. This is perhaps due to the weaker control over the timing of the receipt of credit for task completions for the control examiners, leaving them marginally less likely to engage in deadline-related behaviors pre-reform. Second, while both treatment and control examiners experienced smoothing of first-round allowance rates over the bi-week post-reform, the GS-14 examiners appeared to do so more completely than GS-11-and-below examiners. Moreover, we note that the reduction in end-loading for GS-14 examiners was larger post-reform relative to control examiners.<sup>36</sup>

### C. Specification and Results

Next, in exploring whether the 2011 reform is associated with a divergence in the quality of examination for GS-14 relative to GS-11-and-below examiners, we estimate the following dynamic difference-in-difference specification:

<sup>&</sup>lt;sup>36</sup> Pre-reform, the end-loading rate for control examiners exceeded that of treatment examiners by 10 percentage points. Post-reform, that gap grew to 16, reflecting a larger reduction in end-loading following the reform for GS-14 examiners. We acknowledge that this GS-11-and-lower comparative analysis is subject to the assumption that the 2011 reform did not change the degree of delays in supervisory sign-offs. For instance, perhaps the larger drop in end-loading for GS-14 relative to GS-11 examiners could reflect not a differential in the strength of the examiner incentives across these grade levels but an increase in the period between when assistant examiners complete their review and their assigned primary examiners (or SPEs) sign off on such review (leaving some additional amount of reviews to be completed at the end of the quota). While there is no direct feature of the new docket management system that would change supervisor incentives as it relates to the timing of their sign-off responsibilities, we nonetheless acknowledge this possibility.

(2)  

$$Y_{ikte} = \alpha + \beta_{1} TREATMENT_{ie} + \beta_{2} \sum_{q=2008}^{2014} Year_{t}$$

$$+ \beta_{3} \sum_{q=2008}^{2014} Year_{t} X TREATMENT_{ie} + \beta_{4} Experience_{ie}$$

$$+ \beta_{5} \sum_{q=2008}^{2014} Year_{t} X Experience_{ie} + \delta_{k} + \beta_{6} Controls_{i} + \varepsilon_{ikte}$$

where *i* indicates the individual application, *e* the examiner, *t* the year of the first-round review and *k* the technological Art Unit associated with the application. As discussed further below, we will explore three outcomes, Y: the natural log of examination duration, the incidence of firstround-review allowance, and the incidence of an examination error (again using the EPO benchmarking approach). The coefficients of interest in this regression are those of the interactions between the indicators for the calendar years associated with the first-round review and the indicator for whether the review was completed by a GS-14 examiner (treatment examiner).

GS-levels are, of course, naturally linked to examiner experience in years, the latter of which may also independently impact examination outcomes. To flexibly separate GS-level effects from experience effects when estimating the relevant dynamics, we include in the regression dummy variables for different levels of examiner experience in two-year bins and the interactions between such dummies and the year dummies. We also include Examiner Art Unit fixed effects,  $\delta_k$ , to address concerns that GS-14 and GS-11-and-below examiners may tend to cluster in different Art Units.<sup>37</sup> Finally, we include certain application-characteristic controls.<sup>38</sup>

<sup>&</sup>lt;sup>37</sup> The results are virtually identical when we instead include Art-Unit-by-year fixed effects.

<sup>&</sup>lt;sup>38</sup> Here, we are consistent with Figure 4's regression approach above in that we control primarily for the incidence of a large-entity applicant. And, in a robustness check in Figure A21 of the Online Appendix, we demonstrate that the results are nearly identical when we include a rich set of controls related to claim scope associated with the applications at the time of application publication.



FIGURE 7: DYNAMIC EXAMINATION QUALITY RESULTS FOR GS-14 RELATIVE GS-11-AND-BELOW EXAMINERS

Notes: the presented point estimates and 95% confidence intervals in each panel are from separate regressions of the indicated outcomes on a series of dummy variables for each calendar year associated with the first-round review, on a dummy variable indicating a treatment examiner (GS-14), and on a series of interactions between the treatment dummy and the calendar year dummies. The presented coefficients are of the latter interaction terms, leaving out the 2010 interaction as the reference period. The specifications also include the other controls indicated in equation (2). Standard errors are clustered at the Art Unit level. N = 1,666,972 for Panels A and B and N = 407,144 for Panel C.

One of the key benefits of approaching this difference-in-difference design dynamically—i.e., tracing the relevant differentials year-by-year—is that it allows us to confirm that the treatment and control groups are not trending differentially prior to the reform. Any such pre-trends may call into question the key assumption that the treatment and control groups would have otherwise trended together but for the reform. Encouragingly, as can be observed in Figure 7, there appears to be no discernible trend in either direction in the pre-2011 period.

As further evidenced by Panel A of Figure 7, the divergent response in the differential rate of examination pendency emerges in the reform year of 2011 and continues thereafter. The negative point estimates post-reform indicate that the examination pendency of GS-14 examiners

decreases at a greater rate than GS-11-and-below examiners upon the adoption of the continuous incentive system, implying that the net effects of the 2011 reform were to cause a decrease in patent examination duration and thus an improvement in the timeliness of review. That is, while Figures 4 and 5 already suggested a reduction in examination delays due to a reduction in shotgun rejection behavior near deadlines, the results from Figure 7 suggest that the pendency gains from reduced shotgun rejections are not otherwise offset by lower quality reviews across the board—i.e., for those reviews completed on any given day, regardless of proximity to the deadline.

To clarify that the mechanism underlying the pendency reduction in Panel A of Figure 7 is a reduction in shotgun rejections, we plot a corresponding figure using the incidence of a firstround allowance as the dependent variable—see Panel B of Figure 7. Given the theorized reduction in shotgun rejections, we predict an increase in allowance rates for treatment relative to control examiners. We indeed find that this is the case.

Finally, to probe the accuracy dimension of patent examination quality we use the EPO benchmarking approach discussed in Part III which flags for the existence of Type I or II errors. As evidenced by Panel C of Figure 7, there appears to be no trend pre- or post-reform in the differential error rates between GS-14 and GS-11-and-below examiners. As a result, we do not find that that accuracy of patent examination was compromised by the reform.

In Figure A22 of the Online Appendix, we demonstrate that the patterns depicted in Figure 7 cannot be explained by changes in examination characteristics over time differentially by those applications reviewed by GS-14 and GS-11-and-below examiners. That is, when we form predicted rates of first-round allowance based on a range of application characteristics—e.g., claim counts and claim lengths—and when we plot a corresponding figure but using predicted allowance

43

rates rather than actual allowance rates as the dependent variable, we estimate a flat differential trend between treatment and control examiners over time.

Altogether Figure 7 suggests that the continuous-incentive reform resulted in an increase in the *net* quality of patent examination. More specifically, these figures suggest the 2011 reform resulted in a decrease in examination duration without sacrificing accuracy in patent review.

## D. Additional Mechanisms Analysis

In this sub-Part, we differentially explore the effects of the 2011 reform on examiners whom we predict are more (versus less) likely to procrastinate / end-load in the pre-reform period. If we find that the reductions in examination pendency following the 2011 reform are stronger for the predicted high end-loading group relative to those who are less likely to end-load / procrastinate in the first place, such a finding would reinforce that the mechanism underlying the improvements in pendency documented in Panel C of Figure 7 are emanating from the smoothing effects documented in Figures 4 and 5 above and from the resulting reduction in shotgun rejections.

We form these predictions by stratifying examiners according to the pre-reform endloading rates of the Art Units to which the examiners are assigned (though we find nearly identical results when we stratify by the focal examiner's own pre-reform end-loading rates). More specifically, we estimate a specification analogous to equation (2) but instead of interacting year dummies with an indicator for a GS-14 examiner, we interact year dummies with the associated Art Unit's pre-reform end-loading rates (doing so only on GS-14 examiners). In Figure 8, we show results of this exercise both using the examination duration and first-round allowance rates as the dependent variable. Bolstering our above mechanism conclusions, we indeed find that, at the moment of reform, the reduction in examination pendency is much stronger for those examiners whom we predict were more likely to be heavy end-loaders pre-reform. Further consistent with a shotgun-rejection interpretation of these pendency results, we also find that the reform is associated with a larger increase in first-round allowance rates for examiners predicted to be the biggest end-loaders pre-reform (and thus the biggest shotgun rejectors pre-reform).





Notes: the presented point estimates and 95% confidence intervals in each panel are from separate regressions of the indicated outcomes on a series of dummy variables for each calendar year associated with the first-round review, on a measure of the pre-reform end-loading rate of the examiner's Art Unit, and on a series of interactions between the Art-Unit's end-loading baseline and the calendar year dummies. The presented coefficients are of the latter interaction terms, leaving out the 2010 interaction as the reference period. The specifications also include controls for the incidence of a small-entity applicant. Standard errors are clustered at the Art Unit level. N = 1,666,972.

In Figure A23 of the Online Appendix, we further differentiate the specification underlying Figure 8 by GS-14 and GS-11-and-below examiners. That is, we estimate a triple-differences version of specification (2) that interacts the  $\sum_{q=2008}^{2014} Year_t X TREATMENT_{ie}$  terms with the prereform end-loading rate of the focal examiner's Art Unit (while including the double interaction terms). This takes the mechanism verification one step further, as we would predict that the stronger response to the reform for the high-end-loading (pre-reform) Art Units would be even stronger for those examiners predicted to be more responsive to the reform. As demonstrated by Figure A23, we find that this is the case.

#### E. Worker Retention

Even if the introduction of incentives at a daily level did not lead to decreased average examination quality—based on review accuracy metrics—and harms to average examination pendency, it is possible that the additional stress placed upon examiners may lead to other negative outcomes. In a final exercise, we consider one such outcome: retention. As demonstrated by Figure A24 of the Online Appendix, we do not find evidence that the number of active GS-14 examiners decreases following the 2011 reform relative to the number of active GS-11-and-below examiners, easing concerns that the continuous incentive system negatively impacted retention.

### VI. CONCLUSION

Quotas with associated deadlines have the potential to accelerate work effort that would have otherwise been completed beyond the deadline. Nonetheless, deadlines preserve the possibility for residual procrastination. At the same time, the flexibility associated with a quota system with a deadline can produce various quality gains, extending discretion to workers on how to time and sequence their work efforts. Higher powered temporal incentives structured on a more continuous level have the potential to reduce the time mismanagement possibilities of a deadline system, but nonetheless run the risk of interfering with the benefits of the flexibility afforded by a quota / deadline system. In this paper, we have explored these classic quantity-versus-quality tradeoffs.

Our results suggest that the Patent Office's implementation of a continuous incentive system encouraged patent examiners to exhibit more continuous work efforts. The result was a roughly fifty percent decrease in the end-loading of first-round reviews and the virtual disappearance of markers suggestive of "shotgun" rejections—that is, behaviors in which examiners facing procrastination-induced time crunches at deadlines issue ill-conceived rejections in order to buy time in future rounds of review. Our results suggest that following the reform, examination outcomes no longer depended upon the day on which the initial review was completed. Effectively viewing our analysis in reverse, one can also interpret these findings to suggest that imposing a quota / deadline system in the first place may create the opportunity for procrastination / time-inconsistent behaviors to materialize.

Further, our findings suggest that the examination smoothing benefits associated with the Patent Office's imposition of a set of continuous temporal incentives were not accompanied by corresponding harms to examination quality resulting from the diminished flexibility in the examiner's schedule. That is, while we document notable gains in examination pendency associated with the reduction in deadline behaviors, we find no change in review accuracy.

These findings have several implications. First, as noted above, the Patent Office has suggested that its backlog of patent applications is one of its greatest challenges. While a host of solutions have been put forth to decrease patent pendency, the leading proposal is to hire more patent examiners (Schultz and Madigan 2016), an intervention that is arguably more costly than implementing a continuous deadline incentive system among the existing workforce. Alternatively, while keeping the same number of examiners, the Patent Office can increase throughput by decreasing the time allocations extended to examiners; however, research has suggested that such a move would be accompanied by deteriorations in examination accuracy (Frakes and Wasserman 2023, Frakes and Wasserman 2017); notably we find no such harms in connection with the Patent Office's 2011 reform. Ultimately, our findings highlight the

underappreciated role that performance appraisal systems play in the quality of outcomes in the administrative state or workplaces more generally. A seemingly minor tweak to the docket management score of patent examiners had a substantial impact on examination outcomes.

Finally, our analysis sheds further light on the investigation set forth in Frakes and Wasserman's (2020) into the causes of the substantial end-loading of examination reviews documented in the pre-2011-reform period. We present novel evidence suggesting that some—but only a small portion—of the end-loading may emanate from application sorting by review complexity. Moreover, by demonstrating a substantial smoothing of review completion dates and the elimination of markers of shotgun rejections upon a reform that is designed to encourage more timely review completions, our findings lend further evidence to a delay / procrastination interpretation of the analysis in Frakes and Wasserman (2020) an historical exercise, as there remains little evidence following the continuous-incentives reform indicative of procrastination tendencies of examiners.

#### REFERENCES

- Akerlof, George. 1991. "Procrastination and Obedience," *American Economic Review* 81(2): 1–19.
- Ames, David, Cassandra Handan-Nader, Daniel E. Ho and David Marcus. 2020. "Due Process and Mass Adjudication" Crisi and Reform," *Stanford Law Review* 72: 1–78.
- Aneja, Abhay, Oren Reshef, and Gauri Subramani. "Try, try, try again? Persistence and the Gender Innovation Gap," Mimeo.
- Asch, Beth J. 1990. "Do Incentives Matter? The Case of Navy Recruiters," *Industrial and Labor Relations Review* 43: 89S–106S.
- Balasubramanian, N., Lee, J., Sivadasan, J. 2017. "Deadlines, work flows, task sorting, and work quality," *Management Science* 64 (4): 1804–1824.

Benzarti, Y., 2017. How taxing is tax filing? Using revealed

- Basu, Amiya K., Rajiv Lal, V. Srinivasan, and Richard Staelin. 1985. "Salesforce Compensation Plans: An Agency Theoretic Perspective." *Marketing Science* 4 (4): 267–91.
- Benzarti, Youssef. 2017. "How Taxing Is Tax Filing? Using Revealed Preferences to Estimate Compliance Costs," NBER Working Paper 23903.
- Cadena, Ximena, Antoinette Schoar, Alexandra Cristea, and Heber M. Delgado-Medrano. 2011. "Fighting Procrastination in the Workplace: An Experiment," NBER Working Paper No. 16944.
- Carpenter, Daniel. 2015. "Internal Governance of Agencies: The Sieve, the Shove, the Show," *Harvard Law Review Forum* 129: 189–196.
- Chien, Colleen. 2018. "Comparative Patent Quality," Arizona State Law Journal 50: 71-140.
- Chung, Doug J., Thomas Steenburgh and K. Sudhir. 2014. "Do Bonuses Enhance Sales Productivity? A Dynamic Structural Analysis of Bonus-Based Compensation Plans," *Marketing Science* 33(2): 165–187.
- Chung, Doug J., Das Narayandas and Dongkyu Chang. 2020. "The Effects of Quota Frequency: Sales Performance and Product Focus," *Management Science* 76(4): 2151–2170.
- Cockburn, I., S. Korum, and S. Stern. 2003. "Are All Patent Examiners Equal? Examiners, Patent Characteristics, and Litigation Outcomes," in *Patents in Knowledge-Based Economy*, W. M. Cohen and S.A. Merril (Eds.), (Washington, DC: National Academies Press).
- Darmon, René Y. 1997. "Selecting Appropriate Sales Quota Plan Structures and Quota-Setting Procedures," *The Journal of Personal Selling and Sales Management*, 17(1): 1–16.
- Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2011. "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya," *American Economic Review* 101(6): 2350–90.
- Farre-Mensa, Joan, Deepak Hedge, and Alexander Ljungqvist. 2020. "What is a Patent Worth? Evidence from the U.S. Patent 'Lottery," *Journal of Finance*, 75(2): 639–682.
- Frakes, Michael D., and Melissa F. Wasserman. 2013. "Does Agency Funding Affect Decisionmaking?: An Empirical Assessment of the PTO's Granting Patterns," *Vanderbilt Law Review* 66: 67-125.
- Frakes, Michael D., and Melissa F. Wasserman. 2015. "Does the U.S. Patent and Trademark Office Grant Too Many Bad Patents?: Evidence from a Quasi-Experiment," *Stanford Law Review* 67: 613-676.
- Frakes, Michael D., and Melissa F. Wasserman. 2017. "Is the Time Allocated to Review Patent Applications Inducing Examiners to Grant Invalid Patents?: Evidence from Micro-Level Application Data," *Review of Economics and Statistics*, 99(3): 550–563.
- Frakes, Michael D., and Melissa F. Wasserman. 2020. "Procrastination at the Patent Office," *Journal of Public Economics* 183: 104140.

- Frakes, Michael, and Melissa Wasserman. 2023. "Investing in Ex Ante Regulation: Evidence from Pharmaceutical Patent Examination," *American Economic Journal: Economic Policy* 15: 151-83.
- Gans, Joshua S., David H. Hsu, Scott Stern. 2008. "The Impact of Uncertain Intellectual Property Rights on the Market for Ideas: Evidence from Patent Grant Delays," *Management Science* 54(5).
- Government Accountability Office. 2016. Patent Office Should Define Quality, Reassess Incentives, and Improve Clarity, GAO-16-490.
- Hedge, Deepak, Kyle F. Herkenoff and Chenqi Zhu. 2022. "Patent Publication and Innovation." *NBER Working Paper* 29770.
- Hedge, Deepak, Alexander Ljungqvist and Manav Raj. 2020. "Quick or Broad Patents? Evidence from U.S. Startups." *Swedish House of Finance Research Paper* 20-11.
- Heath, Chip, Richard P. Larrick and George Wu. 1999. "Goals as Reference Points," *Cognitive Psychology* 38 (1): 79–109.
- Holmstrom, Bengt, 1977. "On Incentives and Control in Organizations." PhD dissertation, Stanford University.
- Holmstrom, Bengt, and Paul Milgrom. 1987. "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica* 55 (2): 303–328.
- Holmstrom, Bengt, and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, & Organization* 7: 24-52.
- Kim, Son Ku 1997. "Limited Liability and Bonus Contracts," *Journal Economics & Management Strategy* 6(4): 899–913.
- Kremer, Michael, Supreet Kaur, and Sendhil Mullainathan. 2015. "Self Control at Work." *Journal of Political Economy* 123 (6): 1227–77.
- Lei, Zhen, and Brian Wright. 2017. "Why weak patents? Testing the examiner ignorance hypothesis," *Journal of Public Economics* 148: 43–56.
- Lemley, Mark A. and Bhaven Sampat. 2012. "Examiner Characteristics and Patent Office Outcomes," *Review of Economics and Statistics* 94(3): 817–827.
- London Economics. 2010. *Economic Study on Patent Backlogs and a System of Mutual Recognition*, available at http://www.ipo.gov.uk/p-backlog-report.pdf.
- Marco, Alan, Andrew Toole, Richard Miller and Jesse Frumkin. 2017. "USPTO Patent Prosecution and Examiner Performance Appraisal," USPTO Economic Working Paper No. 2017–08.
- Nou, Jennifer. 2015. "Intra-agency Coordination," Harvard Law Review 129: 422-490.
- O'Donoghue, Ted, and Matthew Rabin. 1999. "Doing it Now or Later," *American Economic Review* 89(1): 103–124.
- Office of the Inspector General. 2014. Abuse of Telework Program at the USPTO, Inspector General Referral No. PPC-CI-12-1196-H.

- Office of the Inspector General. 2015. USPTO Needs to Strengthen Patent Quality Assurance Practices, OIG-15-026A.
- Olsen, Trond, and Gaute Torsvik. 2000. "Discretion and Incentives in Organizations." *Journal* of Labor Economics 18(3): 377-404.
- Oyer, Paul. 1998. "Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality," *Quarterly Journal of Economics* 113: 149–185.
- Oyer, Paul. 2000. "A Theory of Sales Quotas with Limited Liability and Rent Sharing," *Journal* of Labor Economics 18(3): 405-426.
- Patent Public Advisory Committee Meeting. April 14, 2011, available at https://www.uspto.gov/sites/default/files/about/advisory/ppac/ppac\_transcript\_110414.pd f.
- Picard, Pierre M., and Bruno van Pottelsberghe de la Potterie. 2013. "Patent Office Governance and Patent System Quality," *Journal of Public Economics* 104: 16–17.
- Righi, Cesare, and Timothy Simcoe. 2019. "Patent Examiner Specialization," *Research Policy* 48(1): 137–148.
- Sampat, Bhaven, and Heidi L. Williams. 2019. "How Do Patents Affect Follow-on Innovation: Evidence from the Human Genome," *American Economic Review* 109(1): 203-36.
- Schultz, Mark, and Kevin Madigan. 2016. "The Long Wait for Innovation: The Global Patent Pendency Problem," *Center for the Protection of Intellectual Property*.
- United States Patent and Trademark Office. 2008. Performance and Accountability Report for the Fiscal Year 2008. Available at https://www.uspto.gov/sites/default/files/about/stratplan/ar/USPTOFY2008PAR.pdf.
- United States Patent and Trademark Office. 2013. Performance and Accountability Report for the Fiscal Year 2013. Available at https://www.uspto.gov/sites/default/files/about/stratplan/ar/USPTOFY2013PAR.pdf.
- United States Patent and Trademark Office. 2021. Performance and Accountability Report for the Fiscal Year 2022. Available at https://www.uspto.gov/sites/default/files/documents/USPTOFY21PAR.pdf
- United States Patent and Trademark Office. "Patent Examination Research Dataset, Application Data File." https://www.uspto.gov/ip-policy/economic-research/research-datasets/patent-examination-research-dataset-public-pair. Accessed 05/14/2018
- United States Patent and Trademark Office. "Patent Examination Research Dataset, Transaction History File." https://www.uspto.gov/ip-policy/economic-research/researchdatasets/patent-examination-research-dataset-public-pair.