NBER WORKING PAPER SERIES

THE EFFECT OF APPLICATION FEES ON ENTRY INTO PATENTING

Gaétan de Rassenfosse
Adam B. Jaffe

Working Paper 33492
http://www.nber.org/papers/w33492

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2025

This study received funding from the USPTO and was independently prepared as background material for the USPTO's report to the Committee on the Judiciary of the Senate and the Committee on the Judiciary of the House of Representatives, as required by the Unleashing American Innovators Act of 2022. The authors thank Lauren Ailes and Brett Lockard for their helpful comments and discussions. The paper also benefited from comments from USPTO employees and participants at the NBER Productivity Seminar (Boston) and the European Commission's Joint Research Center (Seville). The views expressed herein are solely those of the authors. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Effect of Application Fees on Entry into Patenting
Gaétan de Rassenfosse and Adam B. Jaffe
NBER Working Paper No. 33492
February 2025
JEL No. O30, O31

ABSTRACT

Ensuring broad access to the patent system is crucial for fostering innovation and promoting economic growth. To support this goal, the U.S. Patent and Trademark Office offers reduced fees for small and micro entities. This paper investigates whether fee rates affect the filing of applications by small and micro entities. Exploiting recent fee reforms, the study evaluates the relationship between fee changes and the number of new entrants, controlling for potential confounding factors such as legislative changes. The findings suggest that fee reductions alone are insufficient to significantly increase participation in the patent system among small and micro entities.

Gaétan de Rassenfosse
Ecole polytechnique fédérale de Lausanne
College of Management of Technology
Odyssea Station 5
CH-1015 Lausanne
Switzerland
gaetan.derassenfosse@epfl.ch

Adam B. Jaffe
188 Brookline Avenue
Apartment 26A
Boston, MA 02215
and NBER
adam.jaffe@motu.org.nz

1. INTRODUCTION

A necessary, though not sufficient, condition for the patent system to drive innovation is the active participation of innovative companies in utilizing it. Ensuring widespread engagement with the institution, therefore, becomes a crucial policy goal (Acemoglu and Robinson 2013). U.S. patent law includes provisions aimed at enhancing accessibility to the patent system, particularly through its fee structure. The Patent Law Amendments Act of 1982 introduced a 50-percent fee reduction for independent inventors, nonprofit organizations, and small businesses.[1] The Leahy-Smith America Invents Act (AIA) of 2011 expanded accessibility by introducing further fee reductions for micro entities, thus targeting individual inventors, micro-businesses, and certain nonprofit organizations. The fee reductions for small and micro entities reflect a clear legislative intent to ensure access to the patent system.[2] This policy is grounded in the overarching principle of fostering innovation by making the patent system more accessible to a diverse range of applicants, regardless of their size or financial capacity (House Report 112-98, Doody 2012, *inter alia*). The Unleashing American Innovators Act (UAIA) of 2022 further reinforced this approach by increasing fee reductions for small and micro entities, currently providing a 60 percent discount for small entities and an 80 percent discount for micro entities. Other patent offices are moving in the same direction, with the European Patent Office having recently introduced a 30-percent fee discount to support small applicants.[3]

This policy objective is well-founded, for recent empirical evidence suggests that patenting is a privilege of the few. Using a representative sample of U.S. firms over the period 2008–2015, Mezzanotti and Simcoe (2023) report that most U.S. firms that perform R&D do not patent. They also find that patenting is disproportionately concentrated among the largest firms (as measured by worldwide sales) and exhibits strong variations across industries. Furthermore, the means to achieve this policy objective is sensible. Numerous empirical studies have documented that fees affect the behavior of applicants, including the number of applications (de Rassenfosse and van Pottelsberghe 2007, 2012), the length of the renewal period (Baudry and Dumont 2006, Thompson 2017), the quality of applications (de Rassenfosse and Jaffe 2018), the composition of patent documents (van Zeebroeck et al. 2008), and the validation behavior in regional patent systems (Harhoff et al. 2009). The U.S. Patent and Trademark Office (USPTO) regularly estimates fee elasticities on its own, finding similar effects (USPTO 2013, 2017, 2020).

While the policy of fee reductions to enhance access for small and micro entities is laudable, its actual efficacy remains to be scrutinized. The patent filing process can be particularly daunting for new applicants, with non-fee-related barriers potentially serving as significant deterrents (Masurel 2002, Sichelman and Graham 2010, Athreye et al. 2021). Many small and micro entities may lack awareness that their inventions may be patentable, or even if they are aware, they may be dissuaded by misunderstandings or fears regarding the patent system—such as concerns about disclosure or about their likelihood of success in obtaining a

[1] The objective was to keep fees for those groups low when undiscounted USPTO fees were increased due to Congress changing the overall cost recovery goal for the USPTO to be fully fee funded. See 35 U.S.C. § 41(h) and "Patent and Trademark Office Appropriation Authorization," Public Law 97-247, 96 Stat. 317 (Aug. 27, 1982).

[2] See the legislative history of PL 97-247. For example, representative McClory said: "There are those who maintain that proposed fee increases will discourage individual inventors and small businesses from using the patent system. H.R. 6260 would clearly alleviate that concern in that it provides a 50 percent reduction in all patent fees for independent inventors, small businesses, and nonprofit organizations." Congressional Record (June 8, 1982) p. H12915.

[3] OJ EPO 2024, A3, available at <https://www.epo.org/en/legal/official-journal/2024/01/a3.html> (last accessed January 31st, 2025).

patent. Furthermore, the substantial costs associated with patent attorneys, ranging from several thousand to tens of thousands of dollars, coupled with non-monetary costs that first-time patenting entails, may deter even the most optimistic potential applicants from engaging in the process.

As far as we can ascertain, no study has explicitly addressed the question of whether fee reductions are effective at encouraging *entry* into the U.S. patent system. Nicholas (2011) studies the impact of the 1883 Patents Act in Britain, which reduced filing fees by 84 percent—from £25 ($4900 in today's currency) to £4. He reports that patenting increased 2.5-fold after the reform, and the response was similar across corporate and independent-owned patents. Furthermore, the sectoral distribution of patenting did not change significantly. He concludes that the "inventive activity moved inside the British patent system" as a result of the reform. More recently, Li (2012) investigates the factors behind the explosion of patent applications in China. The author concludes that regional patent subsidy programs, covering deductions and reimbursements of application fees, have been a major driving force of this growth. He observes the surge in all types of organizations, including firms, universities, research institutes, and individuals. While these results suggest that financial incentives have been used successfully to broaden access, it remains to be seen whether these findings, covering different periods and institutional contexts, can be transposed to the modern U.S. patent system.

The paper proposes an in-depth econometric assessment of three major fee reforms that occurred in 2004, 2013, and 2022. As a preliminary step to the statistical analysis, we first needed to harmonize applicant names in order to identify new applicants. We define new applicants as U.S. or foreign applicants who have not filed a utility patent application in the last three or five years. (However, we find that the results are very robust to the time window used.) Equipped with these data, the econometric analysis seeks to quantify the changes in the number of new entrants following fee changes. Because every fee reform has its specificities, we adopt case-specific regression models for each reform.

Descriptive statistics suggest that the rate of new entrants is below ten percent for applicants claiming undiscounted fees, meaning that every month, less than one in ten 'large' applicants did not file at least one patent application in the last few years. This rate of new entrants reaches an average of 30–50 percent over the study period for small entities and 90 percent for micro entities.[4] The rate of new entrants among entities that pay discounted fees has been steadily declining, which is a cause for concern. Regarding the impact of the fee reforms, all findings point in the same direction—despite the different approaches adopted. Fee reforms have had limited to no impact in stimulating applications from new entrants. The major shifts in the number of new entrants over the study period are associated with non-fee-related events, such as the transition from a first-to-invent to a first-inventor-to-file system and the COVID-19 pandemic.

The rest of the paper is organized as follows. Section 2 provides background information on the fee changes and the identification of new applicants. Section 3 presents descriptive statistics of relevant dimensions of the data. We introduce the general econometric approach in Section 4 and present the econometric results in Section 5. Section 6 discusses the case of domestic patent applications. The last section offers some concluding thoughts.

[4] The uncertainty surrounding the small entity entry rate stems from the treatment of applications arising from 'individuals,' as Section 3 explains (see, in particular, the discussion accompanying Figure 6).

2. BACKGROUND

The objective of the empirical analysis is to study the extent to which fee changes affect access to the patent system, as captured by the number of new applicants. At a high level, studying this facet of the fee policy requires identifying the relevant fee changes, identifying new applicants, and estimating appropriate regression models. The present section discusses the first two points.

2.1 Identifying relevant fee changes

The USPTO fee schedule evolves regularly, with fee levels changing, fees being introduced or discontinued, and, more rarely, new categories of fee payers being created. The schedule of fees may appear complex at first, with different fee codes and categories of fee payers. However, all applicants must ultimately pay filing, search, and examination fees at the time of filing a non-provisional patent application. The USPTO also requires the payment of additional fees, such as excess claim fees or excess page fees. However, to the extent that these additional fees cover characteristics of patent documents and reflect handling expenses, they are more likely to affect the 'shape' of patent documents (van Zeebroeck et al. 2008) rather than the participation rate of applicants.[5]

Consequently, we measure the fees paid by summing up filing, search, and examination fees. The USPTO kindly provided these data for the period ranging from October 1st, 2000, to January 1st, 2023. Table 1 provides an overview of these fees for the three categories of fee payers, namely 'Undiscounted,' 'Small,' and 'Micro.' Applicants must pay the undiscounted fees, except if they qualify for the small entity status or certify that they are entitled to micro entity status with respect to a particular application.[6] The small entity status is available if the applicant and all entities with rights in the invention are individuals; small businesses whose number of employees, including affiliates, does not exceed 500 persons; or non-profit organizations (including universities). The micro entity status is available if the inventor, applicant, and all entities with rights in the invention are small entities that meet additional criteria such as income and patent portfolio size.[7]

Effective November 29th, 2000, the American Inventors Protection Act (AIPA) provides for the publication of pending patent applications eighteen months from the earliest claimed priority date. Prior to AIPA, patent documents were published only at the time of grant; data on ungranted patent applications was, therefore, not available. Since the goal of the empirical analysis is to observe all patent applicants, regardless of whether their patent applications will eventually be granted, our study period starts after AIPA (see also Figure 2).

[5] In FY 2023, patent application fees accounted for 24 percent of the total fees collected by the USPTO. Issuance fees accounted for 10 percent and maintenance fees for 54 percent.
[6] An entity that falsely claims a fee reduction, unless acting in good faith, faces a fine of at least three times the unpaid amount, see 35 U.S.C. § 41(j).
[7] Concretely, inventor Inga may be an individual, but if she has promised to license her new invention to a non-small entity, she cannot obtain small entity discounts when filing her patent application. For more information about the micro entity status, see <https://www.uspto.gov/patents/laws/micro-entity-status> (last accessed January 31st, 2025).

Table 1. *Fee change dates and the combined fees for filing, search, and examination, by entity status*

Fee change	***Background for fee change***	***Dates***		***Length (days)***	***Fees***			***CPI***
		Start	**End**		**Undiscounted**	**Small**	**Micro**	
1	.	10.01.00	09.30.01	364	710 .	355 .	.	.
2	35 U.S.C. 41(f)	10.01.01	12.31.02	456	740 *(4.23)*	370 *(4.23)*	.	*(2.13)*
3	35 U.S.C. 41(f)	01.01.03	09.30.03	272	750 *(1.35)*	375 *(1.35)*	.	*(2.25)*
4	35 U.S.C. 41(f)	10.01.03	09.30.04	365	770 *(2.67)*	385 *(2.67)*	.	*(1.82)*
5	35 U.S.C. 41(f)	10.01.04	12.07.04	67	790 *(2.60)*	395 *(2.60)*	.	*(3.19)*
6	Congress. Adj.	12.08.04	09.29.07	1025	1000 *(26.58)*	425 *(7.59)*	.	*(-0.31)*
7	35 U.S.C. 41(f)	09.30.07	10.01.08	367	1030 *(3.00)*	435 *(2.35)*	.	*(9.79)*
8	35 U.S.C. 41(f)	10.02.08	09.25.11	1088	1090 *(5.83)*	462 *(6.21)*	.	*(3.66)*
9	Congress. Adj.	09.26.11	10.04.12	374	1250 *(14.68)*	530 *(14.72)*	.	*(4.55)*
10	35 U.S.C. 41(f)	10.05.12	03.18.13	164	1260 *(0.80)*	533 *(0.57)*	.	*(2.16)*
11	Fee setting	03.19.13	01.15.18	1763	1600 *(26.98)*	730 *(36.96)*	400 *(-24.95)*	*(0.52)*
12	Fee setting	01.16.18	10.01.20	989	1720 *(7.50)*	785 *(7.53)*	430 *(7.50)*	*(6.60)*
13	Fee setting	10.02.20	12.28.22	817	1820 *(5.81)*	830 *(5.73)*	455 *(5.81)*	*(5.05)*
14	Congress. Adj.	12.29.22	.	1098	1820 *(0.00)*	664 *(-20.00)*	364 *(-20.00)*	*(14.89)*

Notes: Fees for electronic filing reported. Fees reported in contemporaneous U.S. $. Numbers in parentheses next to the fees indicate percentage changes in fees. The micro entity status did not exist prior to March 2013, hence the empty values. Background for fee change: '35 U.S.C. 41(f)': inflationary adjustment under 35 U.S.C. § 41(f); 'Congress. Adj.': Congressional adjustment; 'Fee setting': AIA Section 10 fee setting authority. The column CPI reports the inflation over the relevant period, with data coming from <https://data.bls.gov/cgi-bin/cpicalc.pl> (last accessed January 31st, 2025).

Table 1 reads as follows. Consider the third fee change on row number three. It was effective for 272 days, from January 1st, 2003, to September 30th, 2003. The undiscounted fees for filing, search, and examination were set at $750, corresponding to a 1.35-percent increase from the previous fees. The small entity fees were set at $375 (also a 1.35% increase), and there was no micro entity status at the time. Micro entities, as we know them today, paid the small entity fees. The inflation rate over the (previous) period was 2.25 percent. The fee increase was, therefore, below the rise in the consumer price index.

As explained in the next section, our preferred definition of a new entrant is an entity that did not file a U.S. utility patent application in the previous five years. Since we observe patent applicants after the AIPA, we need to wait five years to identify new entrants with reasonable certainty. Fee changes 1–5, occurring at the beginning of the study period, are mainly CPI adjustments, so we do not expect significant effects on new entrants. We can thus ignore these fee changes in the econometric analysis. However, fee change 6, occurring about four years after AIPA, is one of the largest changes in the study period. Therefore, we introduce a second definition of a new entrant, namely an entity that did not file a U.S. utility patent application in the previous three years. We will explore whether the results significantly differ with this slightly different definition of new entrants.

2.2 *Identifying new entrants*

In theory, identifying new entrants is straightforward. Given a disambiguated dictionary of applicants for all patent applications, a new entrant is simply an applicant who has not filed patent applications for a sufficiently long time to be considered 'new' for the purpose of the study. Unfortunately, such a dictionary does not exist, significantly complicating our task.

Before proceeding, two important distinctions must be introduced. The first is between 'harmonization' and 'disambiguation.' The former involves standardizing the spelling or format of names, making similar terms appear consistent across records (*e.g.*, 'Inc.' and 'Incorporated'). This process helps in reducing inconsistencies in data entries but does not necessarily identify relationships between entities. In contrast, disambiguation involves identifying and clarifying the true identity of entities that might appear under different names but refer to the same organization (*e.g.*, recognizing that 'IBM' and 'International Business Machines' refer to the same company). It also includes recognizing distinct entities that may belong to the same corporate group, such as identifying 'Alphabet Inc.' as the ultimate owner of 'Google LLC.'

The second distinction is between 'applicants' and 'assignees.' The former refers to the entity (individual or organization) formally submitting the patent application to the USPTO. Under U.S. law, inventors are typically the initial applicants unless they are obligated to assign, or have already assigned, their rights to a company or other entity, which can then apply as the applicant. The latter refers to a person or entity that has acquired ownership rights to the application or patent by way of an agreement. The applicant and assignee may be the same or different entities. An assignee may be designated at the time of filing or anytime during or after the patent application process through an assignment document filed with the USPTO. There is no obligation under U.S. law for an assignee to disclose its ownership rights in an invention to the USPTO or the public.

Most available patent datasets offer harmonized patent assignee names instead of disambiguated patent applicant names. However, accessing disambiguated applicant data is critical to accurately identify first-time applicants. For that reason, we have spent considerable effort in processing and cleaning applicant names. This task was complicated by the fact that, as a general rule, we are interested in identifying the inventors' employers instead of the inventors who applied for the patent applications.[8] Specifically, we have leveraged existing harmonization and disambiguation efforts and developed our ad-hoc machine-learning algorithms. Appendix A describes the process we have followed.

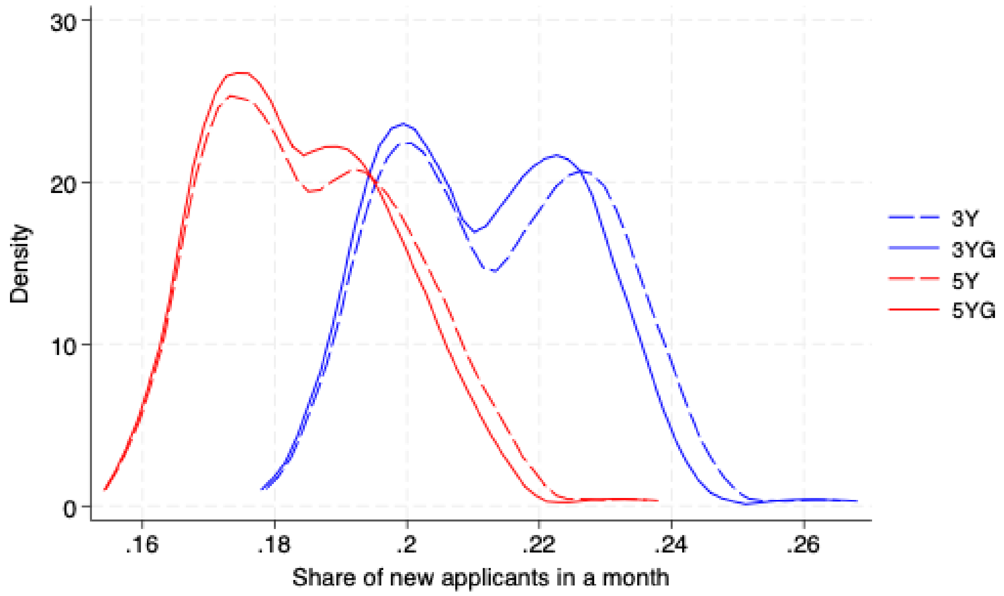
We introduce four ways of measuring new applicants, defined along two dimensions: the time elapsed since the last known patent application (at least three years or at least five years) and whether the data exploit group-level (that is, ownership) information. We label the four measures as 3Y, 3YG, 5Y, and 5YG, with labels '3Y' and '5Y' capturing applicants who did not file a patent in the last three years and five years, respectively, and the suffix 'G' indicating that applicants have been disambiguated using group-level information whenever possible.

Figure 1 depicts the distributions of the monthly proportions of new applicants for each different measure, jointly considering three entity categories. The distributions have similar shapes but are shifted to the left as we implement stricter criteria. The largest shift occurs when

[8] Note that inventors may or may not be obligated to assign their inventions to their employers. As Appendix A explains, we focus on employers when the employer is known because they are an assignee on one or more applications.

considering a three-year versus a five-year time window, with the proportion of new applicants being (expectedly) higher with the three-year criterion. Disambiguating applicants using group-level information (expectedly) shifts the respective distributions to the left, but only slightly so. Recognizing the limitations of our measure, our empirical strategy is to perform the relevant econometric analyses with the different definitions to test the sensitivity of the results to the definition used.

Figure 1. *Distribution of the monthly share of new applicants*

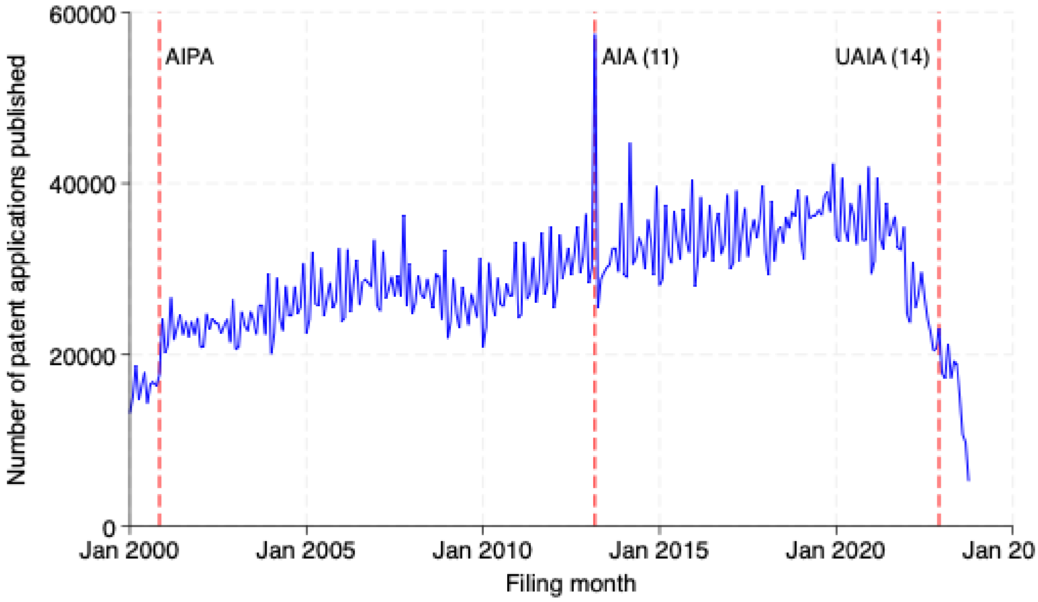


Notes: Kernel density estimations of the monthly proportion of new applicants between January 1st, 2006, and December 31st, 2019. All entity categories combined (undiscounted, small, and micro).

3. DESCRIPTIVE STATISTICS

This section provides a first look at the data by discussing selected descriptive statistics. Figure 2 presents the monthly number of utility patent applications filed since January 1st, 2000, highlighting three notable events. The first event was the coming into force of key provisions of AIPA. The figure depicts a discrete jump in patent application counts, as recorded in the data, after AIPA due to the publication of patent applications. (Before AIPA, patent numbers only included granted patents.) This jump illustrates why we started tracking applicants in November 2000. The second event was the March 2013 effective date of key AIA provisions, associated with fee change 11 from Table 1. It also corresponds to the U.S. patent system transitioning from a first-to-invent to a first-inventor-to-file system. We see a clear jump in patent applications just before the coming into force of the first-inventor-to-file provisions of the AIA, followed by what appears to be a lower-than-expected volume of patent applications in the subsequent months. Finally, the figure also clearly illustrates the truncation bias in recent months, typically observed in patent studies due, among other reasons, to the time lag between filing and publication. The UAIA, associated with fee change 14 from Table 1, entered into force in that period, making the analysis of its effect particularly challenging.

Figure 2. *Monthly number of utility patent applications published*



Notes: Since AIPA, most nonprovisional utility patent applications are published 18 months from the earliest filing date. Most applications are, therefore, published sometime after the filing date. The Figure includes all nonprovisional utility applications published through late January 2024 and consequently severely undercounts applications filed after July 2022. However, the truncation bias extends beyond the most recent 18 months due to delays in processing and publishing applications, as well as requests for non-publication, potentially affecting data for the last 2–3 years.

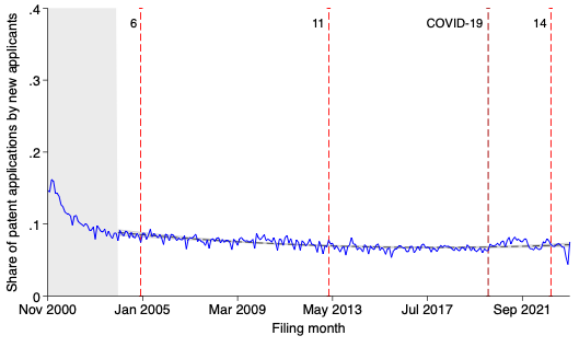
The next set of figures presents various metrics related to new applicants. Figure 3 depicts the proportion of utility patent applications filed by new applicants. Panel A considers that new applicants are applicants who did not file patent applications in the last three years, whereas Panel B considers a five-year time lag. Both panels rely on group-level disambiguation, as explained in Section 2.2, and start in November 2000.

Using proportions instead of raw counts allows us to illustrate the 'initialization phase' of the data, as captured by the shaded areas. The proportion of patent applications by new applicants is initially high and slowly decreases as new applicants file patent applications. The two series look very similar, indicating that the identification of new applicants is robust to the time lag considered. The series in Panel B has slightly lower values than in Panel A, which is expected given that applicants filing four or five years after their last patent application will not be considered 'new' in Panel B. They evolve between six and nine percent in Panel A and between five and eight percent in Panel B. The fact that the two series are similar suggests that the patenting activity is a regular activity for most entities.

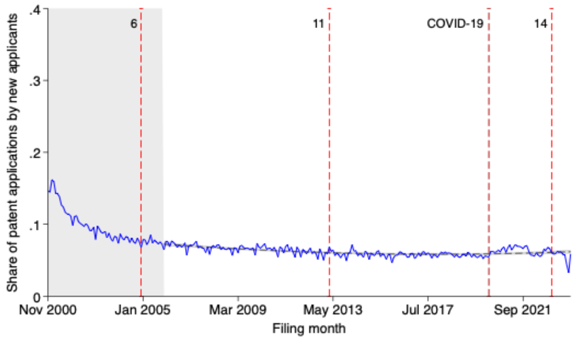
Figure 3 reveals that the trend of patent applications by new applicants slowly decreases over time. The most significant departure from the trend occurred in March 2020, which coincided with the onset of the COVID-19 pandemic. The proportion of patent applications filed by new applicants increases, which is consistent with the burst of patenting activity on COVID-19-related inventions (WIPO 2023), as well as a possible generalized increase in new business attempts during the first months of the COVID-19 pandemic.

Figure 3. *Proportion of nonprovisional utility patent applications filed by new applicants*

A. 'New' as per the 3-year-group rule



B. 'New' as per the 5-year-group rule



Notes: Shaded areas indicate the initialization phases of the variables. The dashed black lines represent the fitted trends. Numbers next to the vertical dashed red lines indicate the fee change numbers (corresponding to numbers in Table 1).

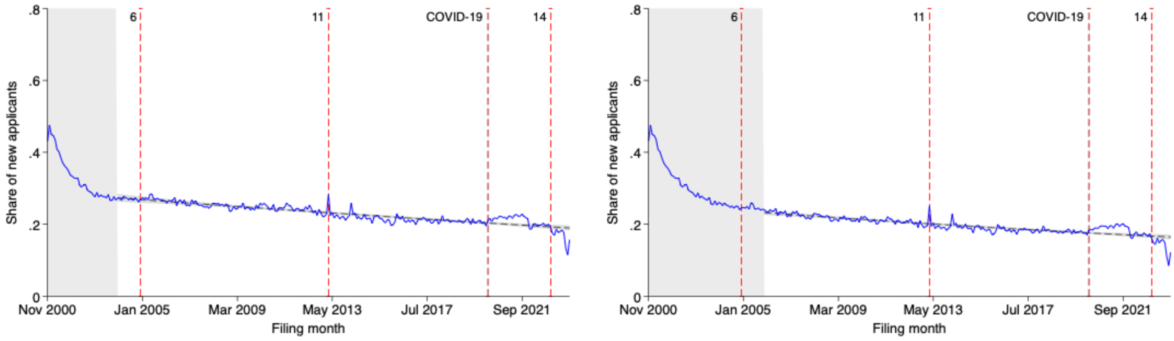
Another metric of interest is the proportion of new applicants among total patent applicants. The previous figure, capturing the proportion of *patent applications* filed by new applicants, may artificially smooth changes in the rate of new entrants if established applicants file significantly more patent applications at every period relative to new entrants. In addition, changing practices in the use of continuing type applications may also affect the interpretation of the figure.[9] At first glimpse, the series depicted in Figure 4 look similar to those in Figure 3. However, upon closer inspection, two notable deviations appear. First, the level of the series differs—it is below 10 percent in Figure 3 and in the 20–30 percent range in Figure 4. This difference is due to the fact that, at every period, established applicants file more patent applications (per applicant) than new entrants. Second, this visualization depicts a clear peak at the time of the AIA (fee change 11). The econometric analysis will dig deeper into the reaction of patent applicants around this time period.

[9] Specifically, recent years have seen an increasing number of continuing-type applications that claim priority to an earlier application. See <https://www.uspto.gov/sites/default/files/documents/20200813_PPAC_Pendency_Update-and-Continuation.pdf> (last accessed January 31st, 2025). By definition, applicants filing such applications cannot be new entrants.

Figure 4. *Proportion of new applicants*

A. 'New' as per the 3-year-group rule

B. 'New' as per the 5-year-group rule

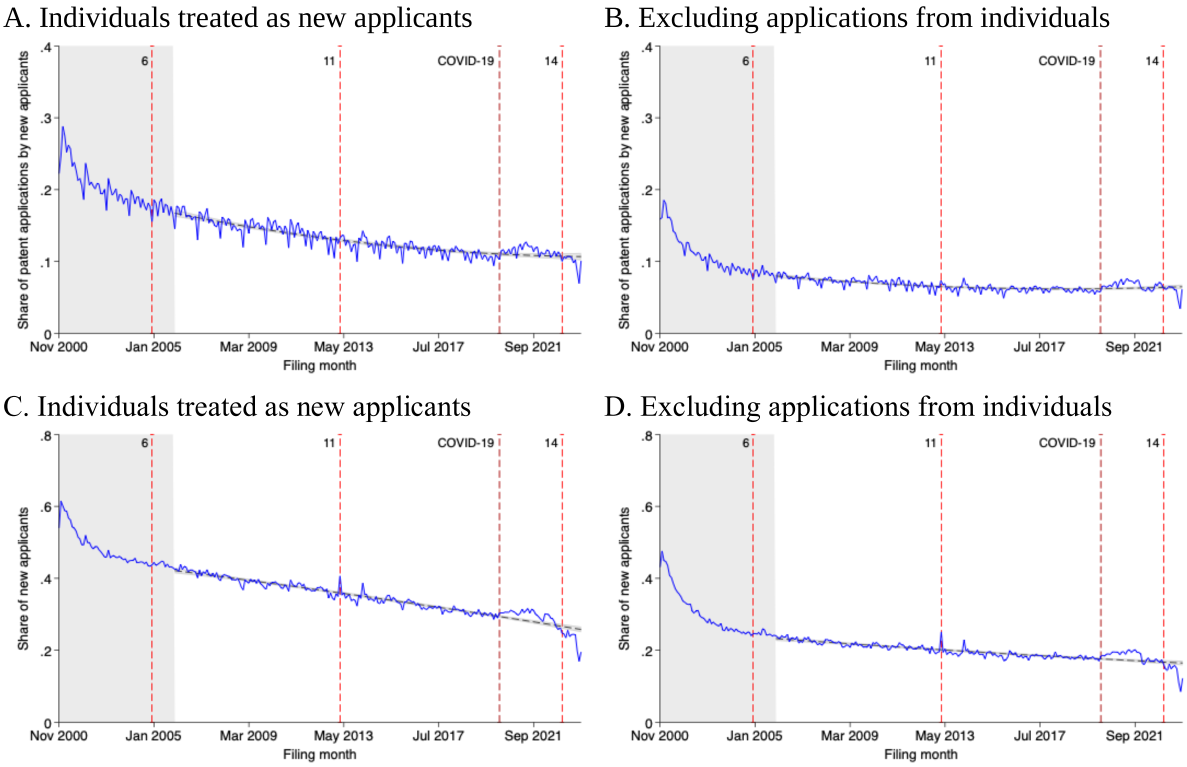


Notes: Shaded areas indicate the initialization phases of the variables. The dashed black lines represent the fitted trends. Numbers next to the vertical dashed red lines indicate the fee change numbers (corresponding to numbers in Table 1).

Another aspect of the data that deserves discussion relates to the treatment of applicants tagged as individuals. Since we do not observe who these applicants are, we cannot assess whether they are new to the patent system. Some patent applications are initially filed by individuals and are abandoned or rejected during the prosecution process without an assignment agreement being recorded. It is thus possible that a well-known company was behind the patent application. As explained in Appendix A, we made extensive efforts to guess the 'true' applicant whenever possible. Despite our best efforts, the data still contain a sizable proportion of individuals (7.24% of all patent applications)—they may well all be the owning entities, but we cannot know for sure.

There is no unique identifier that can be used to match applications filed by the same individual. Besides, harmonization efforts based on names for those applications identified in the dataset as having been filed by individuals would be error-ridden. Therefore, we cannot know whether a new entrant or an established applicant filed these applications. The previous figures treat patent applications from individuals as being from applicants with previous (recent) applications. To test the implications of this treatment, Figure 5 considers two alternative treatments. Panels A and C treat these applications as arising from new entrants, whereas panels B and D exclude these applications from the sample. All figures rely on the five-year-group rule. The levels of the series change compared to the original figures, but the patterns remain very similar. Therefore, the results of the analysis in this paper do not depend critically upon the treatment of the applications identified in the dataset as having been filed by individuals.

Figure 5. *Sensitivity to the treatment of patent applications by those identified in the dataset as 'individuals'*



Notes: Panels A and B depict the proportion of non-provisional utility patent applications filed by new applicants. Panels C and D depict the proportion of new applicants among total applicants. The dashed black lines represent the fitted trends. Numbers next to the vertical dashed red lines indicate the fee change numbers (corresponding to numbers in Table 1). In all panels, 'new' is defined as per the five-year-group rule.

Next, we break down the proportion of new applicants by entity status in Figure 6. Panel A treats individuals as 'established' applicants, whereas panel B treats them as new applicants. 'Individual' applicants who did not claim discounted fees may be large companies likely to file patent applications regularly.[10] Hence, panel A may better capture the proportion of new applicants paying undiscounted fees. By contrast, 'individual' applicants who claim the micro entity fee reduction are likely to be new applicants, such that panel B may better capture this category. Regarding small entities, the true proportion of new applicants is probably somewhere between panel A and panel B.

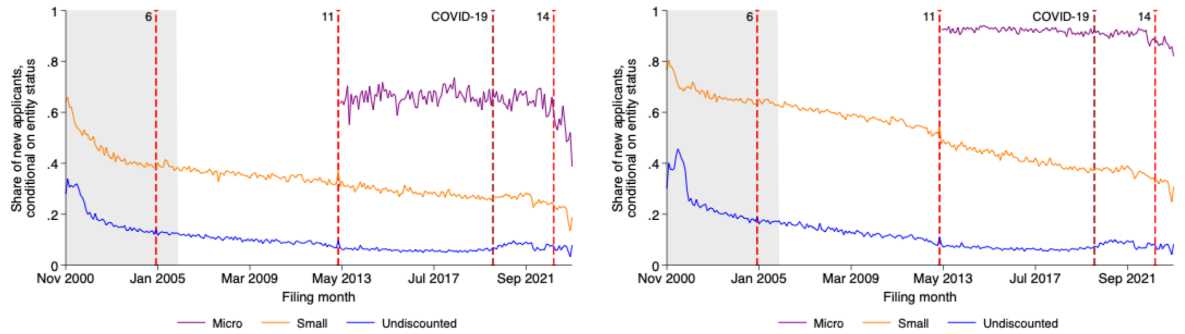
The first striking observation is the steady decline in the proportion of new applicants paying small entity fees, adding context to the trend observed in Figure 4. Because the 'new entity' variable is fully initialized after five years, how we measure new applicants cannot explain this decline. Only COVID-19 seems to have been able to interrupt this trend. A second striking observation is the high rate of new applicants among micro applicants (see panel B). This high rate is expected because there is usually a limit to the number of patent applications that applicants can file and still qualify for the micro entity fee reduction. When this limit is reached, applicants typically fall in the small entity category. Finally, we observe a drop in the proportion of small-entity applicants after the AIA (fee change 11 in Table 1), especially visible in Panel B. This drop is a logical consequence of the introduction of the micro entity status. Before the AIA, applicants who would qualify for the micro entity status for a particular

[10] These cases could also correspond to inventors who have licensed, or promised to license, the invention to another entity that does not qualify for discounts. Such cases are likely to be quite rare, though.

application had to pay the small entity fee and are, therefore, tagged as small entities in the data.

Figure 6. *Proportion of new applicants, by entity status*

A. Individuals treated as established applicants B. Individuals treated as new applicants



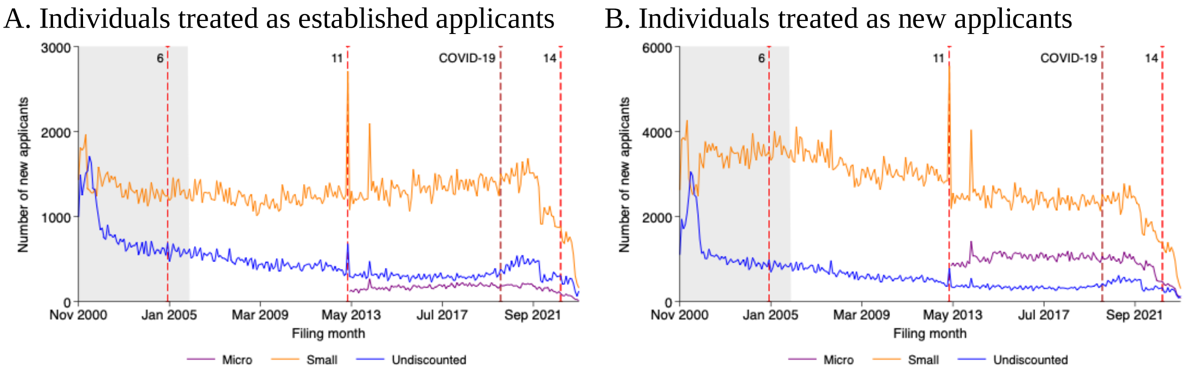
Notes: Shaded areas indicate the initialization phases of the variables. Numbers next to the vertical dashed red lines indicate the fee change numbers (corresponding to numbers in Table 1). In all panels, 'new' is defined as per the five-year-group rule.

Overall, the data indicate that the percentage of new entrants for applicants paying undiscounted fees has been below ten percent in the last decade, likely around 30–50 percent for small entities and above 90 percent for micro entities.

The next figure investigates the apparent decline in the rate of new entrants among small entities. It reports the absolute number of new applicants by entity status. The monthly number of new entrants is roughly steady for micro entities and, in the second half of the time period, for applicants claiming undiscounted fees. The results are less clear for small entities. The number of new entrants was flat in the first half of the time period and increased in the second half when considering individuals claiming small entity fees as established applicants (Panel A). However, when considering these individuals as new applicants, the series decreased in the first half and remained flat in the second half (Panel B).

Figure 7 also reveals two notable observations. First, the post-COVID-19 activity observed in the previous figures seems to be driven, at least partially, by new entrants—a trend particularly visible in Panel A. Second, the second-largest peak in the number of new entrants after the AIA occurred in March 2014. This peak is likely a ripple effect of the AIA, with the entry into the system of provisional and PCT applications filed prior to the effective date of the first-inventor-to-file provision of the AIA.

Figure 7. *Number of new applicants, by entity status*



Notes: Shaded areas indicate the initialization phases of the variables. Numbers next to the vertical dashed red lines indicate the fee change numbers (corresponding to numbers in Table 1). In all panels, 'new' is defined as per the five-year-group rule.

4. ECONOMETRIC APPROACH

Patent applicants are price-taker economic agents. The USPTO sets the fee levels based on policy and financial considerations, and no single applicant has the ability to influence these fees at the time of filing. This observation implies that fee changes are exogenous to any individual applicant.

Having noted this, patent applicants may exert some control over the fees they pay by advancing the filing in the case of a forthcoming fee increase or postponing it in the case of a forthcoming fee decrease. These intertemporal substitution effects are short-lived, given that they concern only inventions that are ready or almost ready to be submitted to the USPTO.

This discussion suggests that the use of observational data in an event-study setup will allow us to obtain causal estimates of the effect of fee levels on the rate of new entrants, provided that we control for possible short-term intertemporal substitution effects and other confounding factors. More specifically, we will estimate variants of the following regression model:

$$Y_w = c + \beta_1 \times Post\ change_w + \beta_2 \times Close\ pre\ change_w + \beta_3 \times Close\ post\ change_w + \varepsilon_w,$$

where Y_w is the number of new entrants Y in week w, the binary variable $Post\ change_w$ takes the value 1 in weeks w after the fee change and 0 otherwise, the binary variable $Close\ pre\ change_w$ takes the value 1 in the four weeks prior to the fee change and 0 otherwise, and the binary variable $Close\ post\ change_w$ takes the value 1 in the four weeks after the fee change and 0 otherwise. The error structure ε_w is assumed to be heteroskedastic and possibly autocorrelated up to four lags (Newey-West standard errors). We consider a window of 26 weeks before and after the fee change for the long-term effect. In practice, the sample will contain 53 observations per fee change, and the variable $Post\ change_w$ will take a value of 0 for the first 26 weeks and a value of 1 for the last 27 weeks.[11]

There are a couple of immediate variants to this specification. First, although fee changes occur at the same time for all entity types, the level of fees differs across entity types, so we need to compute the number of new entrants by entity type. Consequently, three variables

[11] All specifications will also include a dummy variable taking value 1 the week where the fee change took place, and 0 otherwise in order to absorb any contemporaneous effect.

capture new entrants: Y_w^U, Y_w^S, and Y_w^M, for applicants paying undiscounted, small, and micro entity fees, respectively. Second, remember that we have four ways of identifying 'new entrants,' such that there are four ways of constructing the dependent variable(s), labeled as 3Y, 3YG, 5Y, and 5YG (see Section 2.2). Finally, for the sake of robustness, we will also consider the daily number of new entrants, denoted by the subscript d on the dependent variable. In that case, we allow a lag up to 7 days in the error structure.

A close look at Table 1 suggests several restrictions to this baseline specification. First, the micro entity status was implemented on March 19th, 2013. Prior to that date, no micro entities were recorded in the USPTO systems. It follows that one can study the effect of fee changes on micro entities' entry into the patent system starting with the following fee change, occurring on January 16, 2018. In a related vein, we cannot study the effect of the March 19th, 2013, fee change on small entities because the composition of the sample of small entities differs before and after that date. Note also that, in March 2013, the U.S. patent system transitioned from a first-to-invent to a first-inventor-to-file system. This change in patent law, which is associated with the largest fee increase over the period under consideration, is a major one that may affect applicants' short-term patent filing behavior.[12] Put differently, any short-term intertemporal substitution effect that we observe may be attributed to the fee increase, the legislative change, or a combination of these two effects. Therefore, the study of fee change 11 will require particular caution.

5. RESULTS

We will perform detailed analyses of the three largest fee changes, adopting ad-hoc specifications to address the relevant concerns.

5.1 Fee increase '6' of December 8, 2004

We start by estimating the effect of the December 8, 2004, fee increase. As Table 1 shows, this increase is the second-largest increase in relative terms in the period under consideration.

Table 2 reports the OLS regression coefficients for applicants paying the full fee (variable Y_w^U in columns 1–4) and applicants paying the small entity fee (variable Y_w^S in columns 5–8). Because the micro entity fee reduction was introduced only in 2013, the latter category includes what we now call small and micro entities.

[12] This pattern is similar to that observed during another major law change, the TRIPS/GATT change to patent term that took effect in June 1995. This period saw a large increase in patent applications prior to the effective date (Lemus and Marshall 2018).

Table 2. *OLS estimates of fee change 6, weekly data*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DV:	Y_w^U	Y_w^U	Y_w^U	Y_w^U	Y_w^S	Y_w^S	Y_w^S	Y_w^S
Construction:	3Y	3YG	5Y	5YG	3Y	3YG	5Y	5YG
Post change	6.09	5.86	2.50	2.55	32.00	32.36	25.73	282.23
	(8.53)	(8.09)	(8.16)	(7.73)	(14.56)	(14.35)	(13.26)	(11.23)
Close pre change	-7.34	-8.61	-7.77	-7.80	8.34	9.73	8.20	282.23
	(9.60)	(9.00)	(8.58)	(8.02)	(17.68)	(17.79)	(16.42)	(11.23)
Close post change	1.32	0.02	0.48	0.66	-25.16	-23.89	-20.52	282.23
	(16.97)	(15.70)	(15.18)	(14.33)	(15.31)	(15.04)	(13.93)	(11.23)
c	**159.09**	**150.36**	**140.77**	**133.55**	**303.91**	**301.77**	**284.05**	**282.23**
	(7.50)	**(7.02)**	**(6.81)**	**(6.34)**	**(12.34)**	**(12.20)**	**(11.28)**	**(11.23)**
N	53	53	53	53	53	53	53	53
F-stat	76.35	81.43	81.64	79.34	3.75	4.09	4.28	3.61

Notes: The regression model is OLS with Newey-West standard errors. All regressions include a dummy variable capturing the week of the fee change (not reported). Results for fee change 6. Because the micro entity fee category did not exist at the time, small entities also include micro entities. Bold type indicates statistical significance at the 1 percent probability threshold.

At the first brush, the coefficients exhibit limited variations across columns in the respective panels, indicating that how we identify new applicants has only a limited impact on the final results. Interestingly, the coefficients for the constant term (c) are lower as we progress from column (1) to column (4) and from column (5) to column (8). This pattern is consistent with the fact that waiting five years to declare an applicant as 'new' places a higher bar than waiting only three years.[13] Similarly, the group-level disambiguation lumps applicants with different names, so we have a lower chance of observing new applicants when using group matching. Note that the change in coefficients from non-group to group matching is negligible with small entities—presumably because applicants who claim the small-entity reduction are less likely to belong to a group of companies than applicants paying the undiscounted fees.

Because the econometric regression model is OLS and the variables are dummies, the coefficients can be interpreted directly as the number of new applicants. For instance, a constant term of 133.55 in column (4) suggests that, before the change, there were 133.55 new applicants per week. The point estimates decreased to about 125.75 new applicants per week in the four weeks before the fee change (133.55-7.80), but this drop is not statistically significantly different from zero.

The variables of interest are $Close\ post\ change$ and $Post\ change$, which measure the short-term and long-term reactions, respectively. The coefficients are never statistically significantly different from zero in all specifications, reflecting no notable change in entry after the fee increase.

Two factors could affect the validity of this result. First, the number of new applicants is a count variable, and Poisson or negative binomial regression models are generally preferred over OLS (Hausman et al. 1984). Second, the use of weekly figures leads to small sample sizes (and, therefore, weak statistical power). The next set of regressions will address these issues.

[13] In practice, because this fee change occurred about four years after the start of the sample, new applicants captured by the five-year variables are applicants who did not file patent applications in the previous four years. See also the discussion in Section 3.

Table 3. *Poisson estimates of fee change 6, weekly data*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DV:	Y_w^U	Y_w^U	Y_w^U	Y_w^U	Y_w^S	Y_w^S	Y_w^S	Y_w^S
Construction:	3Y	3YG	5Y	5YG	3Y	3YG	5Y	5YG
Post change	6.20	5.96	2.66	2.71	**31.25**	**31.60**	**25.10**	**25.65**
	(3.93)	(3.82)	(3.74)	(3.66)	**(4.96)**	**(4.94)**	**(4.83)**	**(4.82)**
Close pre change	-7.72	-9.01	-8.17	-8.23	10.23	11.75	9.62	9.71
	(6.97)	(6.74)	(6.54)	(6.39)	(9.43)	(9.41)	(9.13)	(9.12)
Close post change	1.14	-0.07	0.29	0.53	**-22.91**	-21.58	-18.78	-18.19
	(7.01)	(6.78)	(6.65)	(6.51)	**(8.59)**	(8.56)	(8.40)	(8.38)
N	53	53	53	53	53	53	53	53
Deviance	312.8	292.6	297.4	278.3	543.9	538.6	491.1	487.8

Notes: Population-averaged Poisson regression model with AR(1) structure. All regressions include a dummy variable capturing the week of the fee change (not reported). Marginal effects at mean reported. Results for fee change 6. Because the micro entity fee category did not exist at the time, small entities also include micro entities. Bold type indicates statistical significance at the 1 percent probability threshold.

Table 3 presents marginal effects of Poisson regression models. Specifically, we estimate population-averaged Poisson regression models using a generalized estimating equations (GEE) approach with an AR(1) correlation structure to account for potential serial correlation in the data.[14] The coefficients are very similar to the OLS point estimates in Table 2, but they are more precisely estimated using Poisson. This new set of results suggests a statistically significant long-term *increase* in the number of small entities filing patent applications for the first time following the fee change. However, this result may not necessarily be causal: it may simply reflect an increasing trend in participation among small entities. (Put differently, the number of new entrants could have been even larger without the fee increase.)

To investigate this issue, Table 4 presents OLS and Poisson estimates that include a time trend. Poisson regression results suggest a positive trend in the number of new entrants for both groups of applicants (Undiscounted and Small). The increase in the number of applicants claiming the small-entity fee reduction does not depart from the trend after the fee change.

[14] For all estimated, a ‘classical’ Poisson regression model that does not assume autocorrelated error terms leads to very similar findings (not reported).

Table 4. *Weekly estimates with time trend*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DV:	Y_w^U	Y_w^U	Y_w^U	Y_w^U	Y_w^S	Y_w^S	Y_w^S	Y_w^S
Method	*OLS*	Poisson	OLS	Poisson	OLS	Poisson	OLS	Poisson
Construction:	3YG	3YG	5YG	5YG	3YG	3YG	5YG	5YG
Post change	-21.42	-21.66	-22.76	-22.91	-11.58	-11.68	-16.44	-16.84
	(21.96)	(9.81)	(20.69)	(9.38)	(38.72)	(12.75)	(35.19)	(12.38)
Close pre change	-20.06	**-19.26**	-18.41	**-17.64**	-8.70	-6.21	-9.64	-7.77
	(10.68)	**(7.00)**	(9.83)	**(6.61)**	(20.29)	(10.03)	(18.71)	(9.65)
Close post change	11.46	12.11	11.27	11.81	-5.46	-3.23	-2.09	-0.01
	(18.99)	(8.15)	(16.99)	(7.83)	(23.85)	(10.13)	(21.91)	(9.91)
Time trend	0.88	**0.88**	0.82	**0.82**	1.42	**1.39**	1.38	**1.36**
	(0.82)	**(0.29)**	(0.75)	**(0.28)**	(1.50)	**(0.38)**	(1.38)	**(0.37)**
N	53	53	53	53	53	53	53	53
F-stat	63.85	-	66.22	-	2.59	-	2.46	-
Deviance	-	283.6	-	269.5	-	527.5	-	476.6

Notes: Marginal effects at mean reported for Poisson models. All regressions include a dummy variable capturing the week of the fee change and a constant term (not reported). Results for fee change 6. Because the micro entity fee category did not exist at the time, small entities also include micro entities. Bold type indicates statistical significance at the 1 percent probability threshold.

Table 5 reports estimates of the daily effect of the fee change, with and without a time trend.[15] According to these estimates, the fee change did not affect the group of applicants paying undiscounted fees. However, the use of more granular data uncovers a significant short-term decrease in the number of new applicants paying small entity fees, in the range of 6–8 fewer new entrants per working day, depending on model specification. Note that the inclusion of a time trend (columns 6 and 8) annihilates the long-term response to the fee change for small entities visible in columns (5) and (7).

[15] Because the regression results include a dummy variable for weekend days, where filing numbers are usually much lower (though not null), they are not directly comparable to the weekly estimates.

Table 5. Poisson estimates of fee change 6, daily data

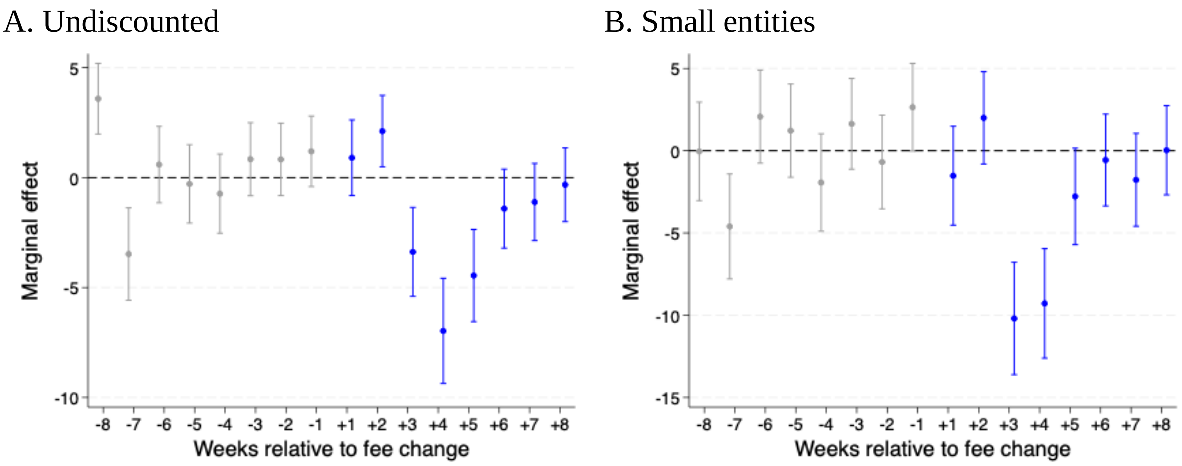
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DV:	Y_d^U	Y_d^U	Y_d^U	Y_d^U	Y_d^S	Y_d^S	Y_d^S	Y_d^S
Construction:	3YG	3YG	5YG	5YG	3YG	3YG	5YG	5YG
$Post\ change$	0.35	-1.88	-0.03	-2.28	**3.80**	0.91	**2.94**	0.13
	(0.60)	(1.54)	(0.57)	(1.45)	**(0.76)**	(1.96)	**(0.73)**	(1.86)
$Close\ pre\ change$	0.72	-0.24	0.50	-0.46	0.27	-0.96	0.16	-1.04
	(1.09)	(1.21)	(1.02)	(1.12)	(1.41)	(1.57)	(1.33)	(1.49)
$Close\ post\ change$	-2.00	-1.09	-1.65	-0.73	**-8.09**	**-7.01**	**-7.21**	**-6.16**
	(1.03)	(1.21)	(0.98)	(1.16)	**(1.24)**	**(1.44)**	**(1.19)**	**(1.38)**
$Time\ trend$		0.01		0.01		0.01		0.01
		(0.01)		(0.01)		(0.01)		(0.01)
N	365	365	365	365	365	365	365	365
Deviance	1097	1094	1012	1008	1626	1624	1516	1513

Notes: Population-averaged Poisson regression model with AR(1) structure. Marginal effects at mean reported. All regressions include a dummy variable capturing the day of the fee change, a dummy variable for Saturdays and Sundays, and a constant term (not reported). Results for fee change 6. Because the micro entity fee category did not exist at the time, small entities also include micro entities. Bold type indicates statistical significance at the 1 percent probability threshold.

To better understand the short-term response to the fee change, Figure 8 below reports the marginal effects of Poisson estimates of the week-by-week average daily number of new applicants. To obtain these estimates, we have created a set of sixteen dummy variables, capturing the eight weeks before and after the fee change (using regression models including a time trend as in columns 4 and 8 of Table 5). Thus, a point estimate of *p* in week *w* indicates that, in that week, there were *p* more new applicants per day compared to the long-term response.

To our surprise, the pattern looks very similar between undiscounted and small entities: a near-zero reaction in the first two weeks after the fee change, followed by a one-month dip before returning to zero. The response of applicants subject to small entity fees is more pronounced than those subject to undiscounted fees, which could explain the lack of statistical significance in columns (1)–(4) of Table 5.

__Figure 8.__ Short-term response to fee change 6 (event study)



Notes: Marginal effects of Poisson estimates. The dependent variable is Y_d^U (5YG) in Panel A and Y_d^S (5YG) in Panel B. 95-percent confidence intervals reported. See main text for details.

However, the third week corresponds to the last week of the calendar year (12/24 to 12/31), and weeks four and five correspond to the first two weeks of 2005. Hence, the dip in new applicants that we observe could be explained by a general slowdown in patent applications during this time period.

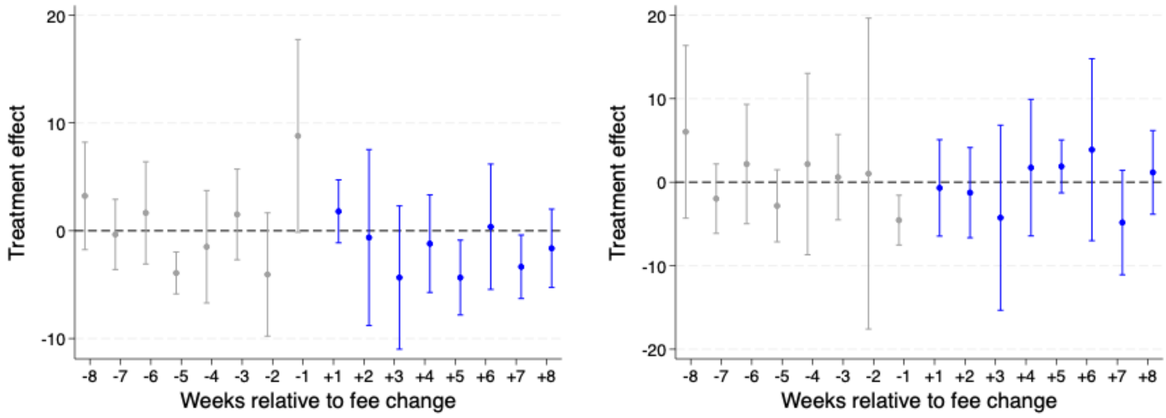
To account for this explanation, we implement a difference-in-differences (DiD) regression model using observations falling outside the time window as controls. Specifically, the treatment window is a 12-month window centered on December 8, 2004, and the control window is a 12-month time window centered on December 8, 2005. The DiD estimator works under the assumption that the trend in the number of new applicants observed in the control period can be used as a baseline for what the trend would have been in the treatment period absent of the treatment. For this assumption to be credible, one typically needs to observe a trend in the treatment period before the treatment that is similar to the trend in the control period—the so-called 'parallel trend' assumption.

Figure 9 depicts the coefficients associated with the treatment effect around the treatment date. Every point estimate reports the difference in the daily number of new entrants in week *w* in the treatment period versus the control period. The first eight point estimates (depicted in gray) allow us to check the parallel trend assumption visually, whereas the next eight point estimates depict the treatment effect. The point estimates for the treatment effect are significantly closer to zero compared to those in Figure 8. In other words, this analysis suggests that the changes observed in Figure 8 are not caused by the fee increase—they simply reflect a seasonal pattern.

Figure 9. *Treatment effect of the short-term response to fee change 6 (DiD)*

A. Undiscounted

B. Small entities



Notes: Results from an OLS DiD regression model with Newey-West standard errors. The dependent variable is Y_d^U (5YG) in Panel A and Y_d^S (5YG) in Panel B. 95-percent confidence intervals reported.

Considering all the results produced so far, we find no conclusive evidence that the fee increase affected the number of new entrants. We find no significant departure from the long-term time trend after the fee change, nor do we find significant short-term reactions (besides what appear to be seasonal effects).

5.2 Fee decrease '14' of December 29, 2022

Studying the effect of the December 29, 2022, fee decrease is particularly challenging because of the data truncation visible in Figure 2. We start by reporting regression results for the baseline model and then present a DiD regression model for small and micro entities using large entities as the control group. Unlike in Section 5.1, data from the year prior to the fee change would not provide a valid counterfactual because of the truncation issue—data truncation may affect the trend in both periods differently. However, fee change 14 concerned only small and micro entities; the undiscounted fee remained the same. We can exploit this fact and compare changes in the number of new entrants paying discounted rates—which decreased with the enactment of the UAIA—relative to the change in the number of new entrants that paid undiscounted rates—which were not impacted. We will also report changes in the proportion of new entrants that are small and micro entities among the population of small and micro entities in an attempt to control further for the truncation bias (inspired by Figure 6).

We obtain the results in Table 6 using a Poisson regression model. We impose a five-year threshold for identifying new applicants and rely on the disambiguation at the group level. Although the fee change concerned only small and micro applicants (columns 3–6), we also report estimates for applicants subject to undiscounted fees for comparison purposes (columns 1–2). First, we note a decrease in new applicants claiming undiscounted fees. This result holds with and without a time trend (columns 1 and 2, respectively). Since the fee did not change for these applicants, this result has to be an artifact of the truncation bias. Second, the long-term effect for small and micro entities is not significantly different from zero once we control for the trend (columns 4 and 6, respectively). Taking this result at face value would suggest that the fee reduction did not affect the number of new applicants. Finally, we observe a strong negative effect immediately following the fee change for small entities (column 4). However, the fee change occurred in late December, and we cannot exclude the possibility that we are capturing a seasonal trend or any other spurious effect.

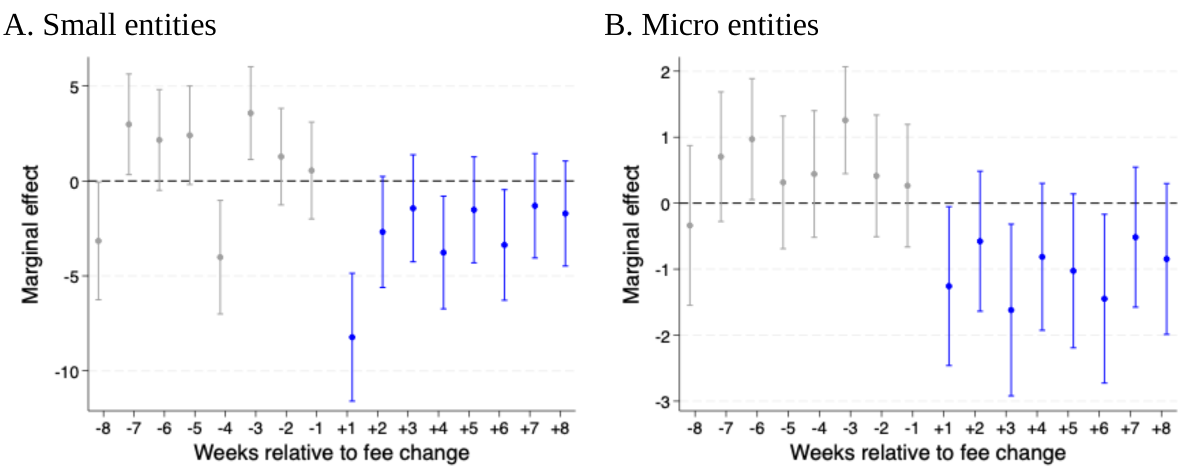
Figure 10 supports that explanation. It presents week-by-week estimates of the daily effect for the eight weeks before and after the fee change. The strongest response occurs during the first week after the fee change, which corresponds to the first week of the year. Also, note that the points estimates after the fee change are generally lower than the point estimates before the fee change. This is counter-intuitive since the fee was decreased by 20 percent. A logical explanation is that these results reflect the truncation bias.

Table 6. *Poisson estimates of fee change 14, daily data*

	(1)	(2)	(3)	(4)	(5)	(6)
DV:	Y_d^U	Y_d^U	Y_d^S	Y_d^S	Y_d^M	Y_d^M
Construction:	5YG	5YG	5YG	5YG	5YG	5YG
Post change	**-1.61**	**-2.47**	**-6.25**	0.07	**-1.51**	-0.29
	(0.36)	**(0.92)**	**(0.65)**	(1.64)	**(0.21)**	(0.53)
Close pre change	-0.10	-0.45	-1.83	0.94	-0.58	-0.08
	(0.61)	(0.68)	(1.07)	(1.33)	(0.32)	(0.42)
Close post change	**-1.78**	-1.46	-2.07	**-4.47**	-0.20	-0.65
	(0.61)	(0.71)	(1.19)	**(1.21)**	(0.42)	(0.40)
Time trend		0.00		**-0.03**		-0.01
		(0.00)		**(0.01)**		(0.00)
N	365	365	365	365	365	365
Deviance	511	510	810	790	410	403

Notes: Population-averaged Poisson regression model with AR(1) structure. Marginal effects at mean reported. All regressions include a dummy variable capturing the day of the fee change, a dummy variable for Saturdays and Sundays, and a constant term (not reported). Results for fee change 14. Bold type indicates statistical significance at the 1 percent probability threshold.

Figure 10. *Short-term response to fee change 14 (event study)*



Notes: Marginal effects of Poisson estimates. The dependent variable is Y_d^S (5YG) in Panel A and Y_d^M (5YG) in Panel B. 95-percent confidence intervals reported. See main text for details.

The following table presents DiD estimates using the number of new applicants subject to undiscounted fees as a control group. As seen in Table 1, the undiscounted fees did not change on December 29, 2022. Furthermore, the overlap in time periods between the control group and the treatment groups should alleviate concerns about data truncation. That is, the DiD set-up in Table 7 will provide valid estimates of the effect of the fee change *if* data truncation affects the control and the treatment groups to the same extent. In essence, we are estimating the causal effect of the fee change as the difference in coefficients between columns (3) and (1) and (5) and (1) of Table 6 for small and micro entities, respectively. (In practice, however, because we are dealing with non-linear models, we cannot simply subtract the coefficients; see Wooldridge 2023.) Because the sets of regressions include a treatment and a control group, the number of observations doubles compared to Table 6 (N = 730 in columns 1–4 of Table 7).

Table 7. *Poisson DiD estimates, daily data*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Time window (months):	[-6,+6]				[-5,+5]		[-4,+4]	
Construction	3YG		5YG		5YG		5YG	
DV:	Y_d^S	Y_d^M	Y_d^S	Y_d^M	Y_d^S	Y_d^M	Y_d^S	Y_d^M
$Post\ change$	-1.66	**-1.02**	-1.41	**-0.93**	0.97	-0.45	1.37	-0.31
	(1.59)	**(0.32)**	(1.49)	**(0.31)**	(1.47)	(0.31)	(1.69)	(0.34)
$Close\ pre\ change$	-2.98	-0.75	-1.92	-0.63	-0.16	-0.32	-0.43	-0.32
	(3.12)	(0.58)	(2.84)	(0.56)	(2.58)	(0.51)	(2.72)	(0.53)
$Close\ post\ change$	4.13	0.51	2.50	0.38	1.71	0.20	1.11	0.08
	(2.45)	(0.51)	(2.45)	(0.51)	(2.51)	(0.51)	(2.67)	(0.53)
N	730	730	730	730	606	606	486	486
Deviance	9318	3731	8061	3281	6597	2694	5226	2137

Notes: Population-averaged Poisson regression model with AR(1) structure. The table only reports the marginal effects of the interacted terms. Results for fee change 14. Bold type indicates statistical significance at the 1 percent probability threshold.

The results in columns (1) and (3) of Table 7 suggest that the fee decrease did not significantly affect the participation of small entities. By contrast, there seems to be a significant long-term *decrease* in the number of new micro-entities entering the system following the fee decrease; see columns (2) and (4). However, the effect vanishes when considering a time window of ten months around the fee change (columns 5–6) or eight months (columns 7–8) instead of twelve. Since data truncation is particularly severe towards the end of the sample, we suggest that the long-term decrease in the number of new entries observed in columns (2) and (4) is likely an artifact of the data.

As seen in Figure 6, the number of new entrants is sensitive to the treatment of applicants identified as ‘individuals,’ especially for micro entities. There are good arguments supporting both their treatments as established applicants (the default) or new applicants. To ensure that the results are not affected by this design choice, Table 8 presents the Poisson DiD estimates, this time treating individuals as new applicants. The results are broadly aligned with those presented in Table 7: there are some significant effects in the twelve-month time window and a loss of significance as we narrow the time window.

Table 8. *Poisson DiD estimates with 'individuals' treated as new applicants, daily data*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Time window (months):	[-6,+6]				[-5,+5]		[-4,+4]	
Construction	3YG		5YG		5YG		5YG	
DV:	Y_d^S	Y_d^M	Y_d^S	Y_d^M	Y_d^S	Y_d^M	Y_d^S	Y_d^M
$Post\ change$	-1.71	**-4.11**	-1.35	**-3.95**	2.03	-1.04	2.62	0.06
	(2.28)	**(1.03)**	(2.24)	**(1.07)**	(2.24)	(0.95)	(2.58)	(1.00)
$Close\ pre\ change$	-6.76	-4.10	-5.03	-3.39	-2.61	-1.10	-2.96	-0.43
	(4.69)	(1.96)	(4.47)	(1.97)	(4.16)	(1.64)	(4.37)	(1.57)
$Close\ post\ change$	7.14	**3.95**	5.20	3.58	3.91	2.68	3.09	2.16
	(3.49)	**(1.36)**	(3.61)	(1.46)	(3.73)	(1.40)	(3.94)	(1.40)
N	730	730	730	730	606	606	486	486
Deviance	11958	5731	10691	5286	8824	4143	6944	3256

Notes: Population-averaged Poisson regression model with AR(1) structure. The table only reports the marginal effects of the interacted terms. Results for fee change 14. Bold type indicates statistical significance at the 1 percent probability threshold.

The next set of regression results, presented in Table 9, adopts an entirely different approach. The dependent variable is the *proportion* of new applicants among the applicants in the day (and within the specific fee category). For instance, if five micro entities are filing (one or more) patent applications on a given day, and one of these applicants is new, the variable takes a value of 0.20 for that day (similar to the data used in Figure 6). The rationale for this approach is that, if patent applications by new applicants become observable at the same rate as those by established applicants, the use of proportion data may alleviate truncation bias. The regression model is a fractional logit model, which extends the traditional logistic regression to handle dependent variables that represent proportions bounded between 0 and 1.

Table 9 reports the marginal effects at the mean, and the coefficients can be interpreted as the change in the proportion of new applicants. For instance, a value of -0.09 associated with the variable $Post\ change$ in column (5) means that the proportion of new micro applicants was 9 percentage points lower after the change than before. However, that result is not robust to the inclusion of a time trend, as column (6) indicates. Overall, this approach suggests no noticeable change in the proportion of new applicants, confirming previous findings.

Table 9. *Fractional logit estimates, daily proportion of new applicants*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DV:	Y_d^S	Y_d^S	Y_d^S	Y_d^S	Y_d^M	Y_d^M	Y_d^M	Y_d^M
Construction:	*3YG*	3YG	*5YG*	5YG	3YG	3YG	5YG	5YG
$Post\ change$	-0.02	0.00	-0.02	-0.01	**-0.09**	-0.03	**-0.08**	-0.00
	(0.01)	(0.02)	(0.01)	(0.02)	**(0.02)**	(0.06)	**(0.02)**	(0.06)
$Close\ pre\ change$	0.01	0.02	0.02	0.02	-0.02	0.01	-0.02	0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)	(0.04)	(0.04)
$Close\ post\ change$	0.00	-0.01	0.00	-0.00	-0.00	-0.03	-0.01	-0.04
	(0.01)	(0.02)	(0.02)	(0.02)	(0.05)	(0.05)	(0.05)	(0.05)
$Time\ trend$		-0.00		-0.00		-0.00		-0.00
		(0.00)		(0.00)		(0.00)		(0.00)
N	365	365	365	365	364	364	364	364
Log-pseudolikelihood	-138	-138	-128	-128	-171	-171	-169	-169

Notes: The regression model is fractional logit. All regressions include a dummy variable capturing the day of the fee change, a dummy variable for Saturdays and Sundays, a constant term, and a dummy variable capturing imputed missing values, if any (not reported). Marginal effects at mean reported. Results for fee change 14. Bold type indicates statistical significance at the 1 percent probability threshold.

5.3 Fee increase '11' of March 19, 2013

Fee increase 11 was the first fee that the USPTO set after it received fee-setting authority under the AIA. The revised fees became effective on Tuesday, March 19, 2013, only three days after the U.S. patent system transitioned from a first-to-invent to a first-inventor-to-file system.[16] The AIA also introduced the micro entity status. However, because no micro entity applicants were officially recorded before the AIA, we cannot study the effect of the AIA on micro entities. Similarly, because the composition of the group of small entities changed after the AIA, we cannot study the effect of the AIA on small entities.

Accordingly, Table 10 focuses on applicants paying the undiscounted fees. Columns (1)–(4) present weekly Poisson estimates, and columns (5)–(8) present daily estimates. We adopt different definitions of new applicants and consider models without and with a time trend. Overall, the results suggest a strong increase in the number of new applicants in the period immediately preceding the fee increase and a long-term drop in the number of new applicants after the fee increase (although the latter result is not robust to the inclusion of a time trend, see even-numbered columns).

[16] The first-inventor-to-file amendment took effect upon the expiration of the 18-month period beginning on the date of the enactment of the AIA, which was on September 16, 2011.

Table 10. Poisson estimates of fee change 11, weekly and daily data

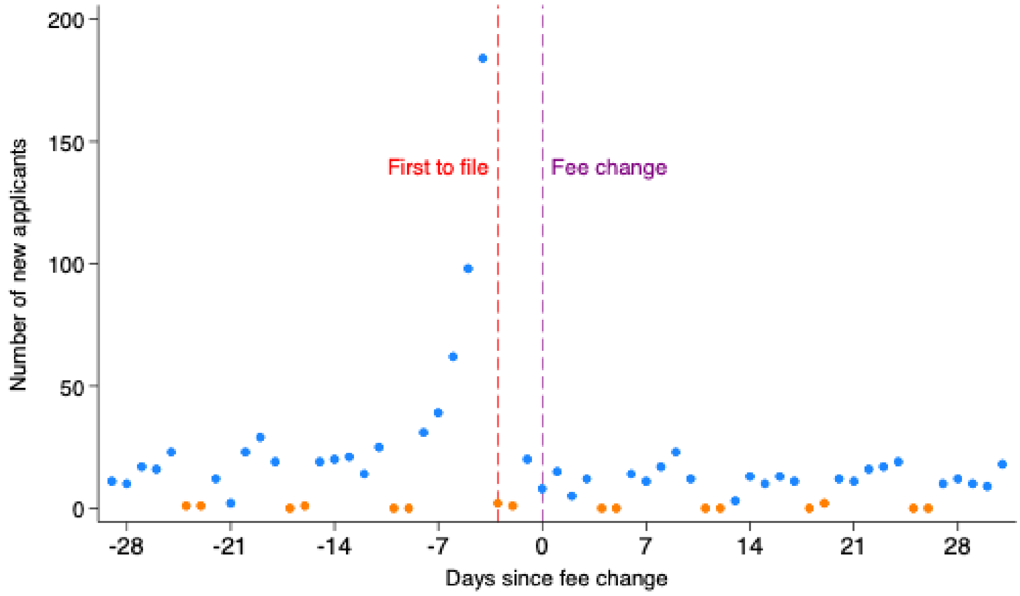
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DV:	Y_w^U	Y_w^U	Y_w^U	Y_w^U	Y_d^U	Y_d^U	Y_d^U	Y_d^U
Construction:	3YG	3YG	5YG	5YG	3YG	3YG	5YG	5YG
$Post\ change$	**-17.03**	-13.51	**-15.26**	-13.80	**-2.01**	-2.41	**-1.91**	-2.57
	(3.07)	(8.04)	**(2.82)**	(7.37)	**(0.59)**	(1.50)	**(0.53)**	(1.34)
$Close\ pre\ change$	**99.62**	**102.19**	**84.45**	**85.56**	**15.45**	**15.13**	**12.96**	**12.43**
	(7.24)	**(9.39)**	**(6.67)**	**(8.66)**	**(1.39)**	**(1.76)**	**(1.25)**	**(1.57)**
$Close\ post\ change$	-2.44	-3.90	-0.71	-1.32	-0.96	-0.80	-0.42	-0.15
	(5.80)	(6.46)	(5.37)	(6.03)	(1.06)	(1.20)	(0.97)	(1.12)
$Time\ trend$		-0.11		-0.05		0.00		0.00
		(0.24)		(0.22)		(0.01)		(0.01)
N	53	53	53	53	365	365	365	365
Deviance	664	664	576	576	1468	1467	1286	1286

Notes: Population-averaged Poisson regression model with AR(1) structure. Marginal effects at mean reported. All regressions include a dummy variable capturing the week or day of the fee change, a dummy variable for Saturdays and Sundays (columns 5–8), and a constant term (not reported). Results for fee change 11. Bold type indicates statistical significance at the 1 percent probability threshold.

Given that the fee increase occurred (almost) concomitantly with a landmark change in U.S. patent law, one might wonder whether the observed effects reflect this legal change rather than the fee increase. Specifically, applicants may have rushed to file patent applications to benefit from the outgoing first-to-invent rule, explaining the significant increase in applications immediately before the change. Subsequently, because many patent applications were accelerated, the 'stock' of patent applications almost ready to be filed may have been depleted, explaining the observed decrease in applications after the change. This alternative explanation could account for both the positive $Close\ pre\ change$ effect and the negative $Post\ change$ effect, potentially confounding our interpretation of the impact of the fee increase.

Figure 11 supports that explanation. It depicts the daily number of new applicants, marking the entry into force of the first-inventor-to-file rule and the fee change. We see a notable increase in new applicants the week preceding the first-inventor-to-file change. In light of this evidence, Table 11 presents regression results controlling for the activity in the week preceding the transition to the first-inventor-to-file system.

Figure 11. *Daily number of new applicants around fee change 11*



Notes: Number of new applicants paying the undiscounted fees. Orange dots represent weekend days. New applicants are defined as per the 5YG rule.

Columns (1)–(4) of Table 11 must be compared with columns (5)–(8) of Table 10. They report the marginal effects of Poisson estimates of the daily number of new applicants. The variable $Pre\ first\ to\ file$ takes the value of 1 in the seven days preceding the transition to a first-inventor-to-file system, and 0 otherwise. It absorbs the increase in applicant numbers visible in Figure 11. The coefficient associated with the long-term effect remains unaffected in specifications that do not include a time trend (columns 1 and 2). They reduce by about 40 percent in specifications with a time trend (columns 2 and 4) and are not statistically significantly different from zero. (The use of fractional logit models on proportion data also leads to insignificant coefficients, not reported.)

Table 11. *Poisson estimates of fee change 11, daily data*

	(1)	(2)	(3)	(4)
DV:	Y_d^U	Y_d^U	Y_d^U	Y_d^U
Construction:	3YG	3YG	5YG	5YG
$Post\ change$	**-1.97**	-1.21	**-1.82**	-1.45
	(0.55)	(1.41)	**(0.49)**	(1.25)
$Close\ pre\ change$	1.98	2.33	1.70	1.87
	(1.11)	(1.28)	(0.99)	(1.13)
$Close\ post\ change$	-1.20	-1.50	-0.62	-0.77
	(0.98)	(1.09)	(0.89)	(1.00)
$Pre\ first\ to\ file$	**22.02**	**22.07**	**18.07**	**18.10**
	(1.26)	**(1.27)**	**(1.11)**	**(1.11)**
$Time\ trend$		-0.00		-0.00
		(0.01)		(0.01)
N	365	365	365	365
Deviance	1019	1020	907	907

Notes: Population-averaged Poisson regression model with AR(1) structure. Marginal effects at mean reported. All regressions include a dummy variable capturing the day of the fee change, a dummy variable for Saturdays and Sundays, and a constant term (not reported). Results for fee change 11. Bold type indicates statistical significance at the 1 percent probability threshold.

The next regression table focuses on small and micro applicants. Although we cannot estimate the reform's effect on these two groups separately, as explained above, we can estimate the effect on them jointly. Fees *increased* for applicants qualifying as small entities but *decreased* for applicants qualifying as micro entities. Table 12 reports daily Poisson estimates considering different definitions of new applicants, models without and with a time trend, and models without and with a control for the activity in the week preceding the transition to the first-inventor-to-file system.

Overall, the fee change did not significantly affect the long-term number of new entries. Similarly to undiscounted entities, we observe a strong activity in the week preceding the first-inventor-to-file transition. Interestingly, we also observe consistently a drop in new entrants during the four weeks after the fee change ($Close\ post\ change$). However, the long-term effect of the fee change appears insignificant ($Post\ change$). In theory, the lack of significance could result from a combination of a decrease in new entries for small entities and an increase in new entries for micro entities. However, in light of the evidence accumulated thus far, the most likely explanation is that there was simply no effect on both groups.

***Table 12**. Poisson estimates of fee change 11, group small and micro entities, daily data*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DV:	Y_d^{SM}	Y_d^{SM}	Y_d^{SM}	Y_d^{SM}	Y_d^{SM}	Y_d^{SM}	Y_d^{SM}	Y_d^{SM}
Construction:	3YG	3YG	3YG	3YG	5YG	5YG	5YG	5YG
Post change	-0.66	-5.69	-0.42	2.11	0.17	-4.35	0.26	2.79
	(1.19)	(2.98)	(1.09)	(2.81)	(1.13)	(2.82)	(1.04)	(2.68)
Close pre change	**74.06**	**69.58**	**15.57**	**16.84**	**68.64**	**64.52**	**13.50**	**14.76**
	(3.06)	**(3.80)**	**(2.34)**	**(2.70)**	**(2.95)**	**(3.67)**	**(2.22)**	**(2.57)**
Close post change	**-11.19**	**-9.44**	**-10.54**	**-11.41**	**-11.57**	**-10.06**	**-10.71**	**-11.54**
	(1.89)	**(2.19)**	**(1.77)**	**(1.95)**	**(1.75)**	**(2.02)**	**(1.65)**	**(1.81)**
Time trend		0.02		-0.01		0.02		-0.01
		(0.01)		(0.01)		(0.01)		(0.01)
Pre first to file			**83.81**	**83.94**			**77.17**	**77.30**
			(2.38)	**(2.39)**			**(2.26)**	**(2.26)**
N	365	365	365	365	365	365	365	365
Deviance	4060	4062	2233	22311	3878	3880	2115	2112

Notes: Population-averaged Poisson regression model with AR(1) structure. Marginal effects at mean reported. All regressions include a dummy variable capturing the day of the fee change, a dummy variable for Saturdays and Sundays, and a constant term (not reported). Results for fee change 11. Bold type indicates statistical significance at the 1 percent probability threshold.

6. ADDITIONAL CONSIDERATIONS

The analysis has considered the entry of new applicants, regardless of their origin. This perspective matters to the USPTO for operational purposes. However, understanding the effect of fee changes on the entry of domestic applicants into the patent system is worth considering for U.S. patent policy presumably has its largest effect on the U.S. innovation ecosystem.

There are reasons to suspect that domestic applicants are more sensitive to fee changes than foreign applicants. Foreigners may typically first file a patent application in their own country and, if the invention is worthy of international protection, seek it in the United States (de Rassenfosse et al. 2013). In other words, on average, foreign applications have met a higher value threshold and faced a higher cost burden (*i.e.*, the cost of filing at home)—potentially leading to a lower sensitivity to U.S. fees.

The econometric analysis has considered the absolute number of new entrants, and the previous discussion suggests that most of these new entrants could be domestic applicants. In that sense, we do not expect that removing foreign applicants from the sample will fundamentally alter the study's conclusions. However, we can expect that removing foreign entrants from the count will lower the number of new entrants, reducing the point estimates of the effect of fee change. To find out, we have estimated our preferred specifications for each fee change, considering only domestic applicants.

The notion of a 'domestic' applicant is a tricky one when it comes to U.S. subsidiaries of foreign establishments. Should patent applications from these entities be counted as foreign or domestic? The complexity is compounded by the fact that the disambiguation of applicants may have grouped entities from the United States and abroad together. We have adopted a pragmatic approach, considering the country of residence of inventors. We have excluded from the sample all applications associated with at least one foreign inventor (in practice, limiting the sample to applications with only U.S. inventors). As an alternative approach, we have also limited the sample to applications with at least one U.S. inventor.

Relevant estimates for U.S.-invented patent applications are available in Appendix B. Figure B1 depicts the treatment effect of the short-term response to fee change 6 using the DiD specification, similar to Figure 9. Table B1 presents Poisson DiD estimates for fee change 14 with ‘individuals’ treated as new applicants, similar to Table 8. Finally, Table B2 presents Poisson estimates of fee change 11, similar to Table 11. The results are qualitatively similar to the reference results.

7. CONCLUDING REMARKS

This study analyzes the entry of new applicants into the patent system and the role of the patent fee policy therein. The descriptive statistics reveal that the rate of new entrants for ‘large’ entities (*i.e.*, paying undiscounted fees) has been consistently below ten percent. This figure means that less than one in ten applicants in any given month are first-time filers within the past few years. The rate of new entrants for small entities is substantially higher, around 30 to 50 percent. However, the data indicate that the rate of new entrants for small entities has been steadily declining, a trend that warrants attention.

The econometric analysis of three major fee reforms (2004, 2013, and 2022) consistently suggests that fee changes have had a limited effect on the number of new entrants. The most significant fluctuations in the number of new entrants observed during the study period are associated with non-fee-related events, such as the shift from a first-to-invent to a first-inventor-to-file system and the COVID-19 pandemic.

The limited impact of fee changes on new entrants likely stems from non-fee barriers to entry. The patent filing process can be daunting for new applicants, who may lack awareness of the patentability of their inventions or harbor misunderstandings about the patent system. The high costs associated with patent attorneys can also act as deterrents. Initiatives to lower these barriers, such as the PTO pro bono programs and the opening of regional outreach offices, have certainly an important role to play in broadening access.

Our results provide robust evidence that modest changes in existing fees are not likely to enhance or inhibit access to the patent system. However, we cannot infer from the present analysis what would happen if fees were changed significantly (Lucas 1976). Radical changes may have effects, the possibility of which is not tested by the empirical analysis, which looked only at marginal changes.

Future research should delve deeper into the long-term effects of broader access, focusing on the initial entry of new applicants and their subsequent innovation activities, such as increased R&D expenditure or entrepreneurial ventures. The impact of the 2022 fee reduction (UAIA) remains obscured by data truncation bias, suggesting the need for revisiting this analysis in a few years. Additionally, future studies could explore the effects of fee reforms on underrepresented groups, such as women and minority inventors, to ensure equitable access to the patent system. The broader issue of non-fee barriers to entry also merits further investigation, potentially examining events like the USPTO’s transition from paper to electronic filing or other USPTO or external changes in non-fee barriers. Such research endeavors would contribute to a more nuanced understanding of how to foster a truly inclusive and accessible patent system that supports innovation from all corners of society.

REFERENCES

Acemoglu, D., & Robinson, J. A. (2013). *Why nations fail: The origins of power, prosperity, and poverty*. Crown Currency.

Athreye, S. S., Fassio, C., & Roper, S. (2021). Small firms and patenting revisited. *Small Business Economics* 57(1): 513–530.

Baudry, M., & Dumont, B. (2006). Patent renewals as options: improving the mechanism for weeding out lousy patents. *Review of Industrial Organization* 28: 41–62.

de Rassenfosse, G., Dernis, H., Guellec, D., Picci, L., & van Pottelsberghe, B. (2013). The worldwide count of priority patents: A new indicator of inventive activity. *Research Policy* 42(3) : 720–737.

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

de Rassenfosse, G., & van Pottelsberghe, B. (2007). Per un pugno di dollari: A first look at the price elasticity of patents. *Oxford Review of Economic Policy* 23(4): 588–604.

de Rassenfosse, G., & van Pottelsberghe, B. (2012). On the price elasticity of demand for patents. *Oxford Bulletin of Economics and Statistics* 74(1): 58–77.

de Rassenfosse, G., & van Pottelsberghe, B. (2013). The role of fees in patent systems: Theory and evidence. *Journal of Economic Surveys* 27(4): 696–716.

Doody, P. (2012). Comprehensive legislative history of the Leahy-Smith America Invents Act compiled, edited, and abridged. Pillsbury Winthrop Shaw Pittman LLP, 793 pages.

Harhoff, D., Hoisl, K., Reichl, B., & van Pottelsberghe, B. (2009). Patent validation at the country level—the role of fees and translation costs. *Research Policy* 38(9): 1423–1437.

Hausman, J., Hall, B. H., & Griliches, Z. (1984). Econometric models for count data with an application to the patents-R&D relationship. *Econometrica* 52(4): 909–938.

House Report 112-98, (2011). Part 1, America Invents Act, 112th Congress.

Lemus, J., & Marshall, G. (2018). When the clock starts ticking: Measuring strategic responses to TRIPS's patent term change. *Research Policy* 47(4): 796–804.

Li, X. (2012). Behind the recent surge of Chinese patenting: An institutional view. *Research Policy* 41(1): 236–249.

Lucas, Robert (1976). Econometric Policy Evaluation: A Critique.. In Brunner, K.; Meltzer, A. (eds.). The Phillips Curve and Labor Markets. *Carnegie-Rochester Conference Series on Public Policy*. Vol. 1. New York: American Elsevier. pp. 19–46.

Masurel, E. (2002). Patenting behaviour by SMEs. *International Journal of Entrepreneurship and Innovation Management* 2(6): 574–583.

Mezzanotti, F., & Simcoe, T. (2023). Innovation and appropriability: Revisiting the role of intellectual property. *National Bureau of Economic Research Working Paper* No w31428.

Nicholas, T. (2011). Cheaper patents. *Research Policy* 40(2): 325–339.

Sichelman, T., & Graham, S. J. (2010). Patenting by entrepreneurs: An empirical study. *Michigan Telecommunications & Technology Law Review* 17: 111–180.

Thompson, M. J. (2017). The cost of patent protection: Renewal propensity. *World Patent Information* 49: 22–33.

United States Patent and Trademark Office (USPTO). (2013). Appendix to Setting and Adjusting Patent Fees. 78 FR 4212. Available at https://www.uspto.gov/sites/default/files/aia_implementation/AC54_Final_Elasticity_Supplement.pdf.

United States Patent and Trademark Office (USPTO). (2017). Appendix to Setting and Adjusting Patent Fees During Fiscal Year 2017. 82 FR 52780. Available at https://www.uspto.gov/sites/default/files/documents/Description_of_Elasticity_Estimates.pdf.

United States Patent and Trademark Office (USPTO). (2020). Appendix to Setting and Adjusting Patent Fees During Fiscal Year 2020. 85 FR 46932. Available at https://www.uspto.gov/sites/default/files/documents/Elasticity_Appendix.docx.

van Zeebroeck, N., Stevnsborg, N., De La Potterie, B. V. P., Guellec, D., & Archontopoulos, E. (2008). Patent inflation in Europe. *World Patent Information* 30(1): 43–52.

WIPO (2023). COVID-19 vaccines and therapeutics. Patent landscape report. 83 pages. ISBN: 978-92-805-3530-3.

Wooldridge, J. M. (2023). Simple approaches to nonlinear difference-in-differences with panel data. *The Econometrics Journal* 26(3): C31-C66.

Appendix A. Applicant disambiguation

This Appendix describes our process for disambiguating applicant names. Our starting point is the applicants listed in the 8,192,923 non-provisional published patent applications between January 2000 and late January 2024. About 80 percent of these patent applications are granted.

As explained in the main text, several harmonization efforts have been proposed, and we have sought to leverage them. Specifically, we consider twelve sources of information, listed in Table A1:

Table A1. *Source of information on applicants and assignees*

Source	Description
From Google Patents Public Datasets:	
1.	Harmonized assignees listed in the pre-grant publication(s)
2.	Original, non-harmonized assignees listed in the pre-grant publication(s)
3.	Harmonized assignees listed in the granted publication(s)
4.	Non-harmonized assignees listed in the granted publication(s)
From USPTO PatentsView:	
5.	Harmonized assignees listed in the pre-grant publication(s)
6.	Original, non-harmonized assignees listed in the pre-grant publication(s)
7.	Harmonized assignees listed in the granted publication(s)
8.	Non-harmonized assignees listed in the granted publication(s)
9.	Raw applicants listed in the granted publication(s)
10.	Raw applicants listed in the pre-grant publication(s)
From EPO PATSTAT:	
11.	Harmonized applicants
12.	Original, non-harmonized applicants

Our process to identify applicants follows these steps:

1. For each application number, collect the following data points from all sources:
    a. Assignees for published applications and granted patents and also applicants for published patent applications (12 sources, see Table A1);
    b. Inventors for published applications and granted patents (10 sources in total, similar to all but sources 9 and 10 in Table A1).
2. Remove all inventor names from a. using the information contained in b. for each focal application. This trick removes a large number of inventors listed as applicants. (If the inventor is the genuine applicant, the information will be recovered at a later stage in the process.)
3. Flag our preferred entity name for each focal application: harmonized assignees listed on the granted patents as provided by Google (= source number 3). These entries will correspond to our preferred name variation.
4. Build a 'dictionary' of all name variations associated with the preferred entity name. That is, we identify all entries from all but source 3 entities that correspond to source 3 entities. This step leads to a large dictionary of name variations from sources {1, 2, 4–12} that map to source 3 entries. For that purpose, we need to:
    a. Implement a machine learning (ML) classifier to build the dictionary.
    b. Adopt a conflict resolution mechanism in case the same name variation is matched to two distinct entities from source 3.

5. For each applicant number, we select the applicant names (from PATSTAT and PatentsView) and look in the dictionary for the preferred entity name.
6. Additional 'quick tricks' in case no applicant was found, as explained further below.
7. Final data-intensive tricks in case no applicant was found.

The rest of this Appendix provides more information on the main steps.

Overview of Step 1 to Step 4

Figure A1 illustrates the process for a hypothetical patent and seven sources of information, with source #7 being the preferred source. Step 1a lists all name variations from all sources associated with the patent. Step 2 removes all entries identified as inventors (using Step 1b, not shown). Step 3 flags the preferred name variation. Step 4 associates the relevant name variations with the preferred name. In this example, Step 2 does not eliminate all inventors based on data from the inventor table (see "Jane Liliane Doe"). Besides, there are two entities, and we need a way to link each name variation to the correct one. These are the reasons why the ML classifier comes into play in Step 4a.

Figure A1. *A patent invented by John Smith of Google and Jane Doe of Microsoft*

source	name
1	Google Inc
1	Microsoft Incorporated
2	Google llc
2	Microsoft Inc
3	John Smith
3	Jane Liliane Doe
4	GOOGLE
4	MIROSOFT
5	Google, Inc.
5	Microsoft, Inc
6	Smith J
7	Doe J
7	Google
7	Microsoft

Step 1.a

source	name
1	Google Inc
1	Microsoft Incorporated
2	Google llc
2	Microsoft Inc
3	~~John Smith~~
3	Jane Liliane Doe
4	GOOGLE
4	MIROSOFT
5	Google, Inc.
5	Microsoft, Inc
6	~~Smith J~~
7	~~Doe J~~
7	Google
7	Microsoft

Step 2

source	name
1	Google Inc
1	Microsoft Incorporated
2	Google llc
2	Microsoft Inc
3	~~John Smith~~
3	Jane Liliane Doe
4	GOOGLE
4	MIROSOFT
5	Google, Inc.
5	Microsoft, Inc
6	~~Smith J~~
7	~~Doe J~~
7	Google
7	Microsoft

Step 3

Dictionary

Google Inc	Google
Google llc	Google
GOOGLE	Google
Google, Inc.	Google
Google	Google
Microsoft Incorporated	Microsoft
Microsoft Inc	Microsoft
MIROSOFT	Microsoft
Microsoft, Inc	Microsoft
Microsoft	Microsoft

Step 4

We have nearly 96 million assignee or applicant names across the twelve sources, as Table A2 illustrates. Among these, 37 million are identified as inventors—that is, we can find a similar entry in the ten sources containing inventor names for the focal patent. Excluding these entries, we obtain 2,831,964 unique 'cleaned' names.

Table A2. *Total number of entries per source*

Assignees (N = 95,702,199)		**Inventors (N = 182,644,108)**	
Soure	*Non-null entries*	*Source*	*Non-null entries*
1	10,515,354	1	21,289,436
2	10,515,895	2	21,265,360
3	8,938,839	3	15,117,812
4	5,756,272	4	15,084,380
5	3,486,407	5	20,801,971
6	3,526,754	6	20,803,330
7	4,505,006	7	12,732,691
8	4,506,363	8	12,733,611
9	4,418,422	9	-

10	11,831,093	10	-
11	13,795,556	11	21,407,763
12	13,906,238	12	21,407,754

The next objective is to map each of these entries to one 'standard' name variation from source 3. To help us with this task, we manually labeled about 3000 pairs of names and trained an ML classifier. The classifier's output is the probability score that the two entities in a pair relate to the same applicant.

The classifier exploits a set of textual features, such as the proportion of common characters between the two text strings, the number of common words, the number of common n-grams, the average word length in each string, the longest word length in each string, etc. We have tested different ML models (including neural networks) and decided to deploy a random forest model, which exhibits the following performance: accuracy (proportion of correct predictions): 0.972; precision (proportion of positive identifications that were actually correct): 0.977; recall (actual positives identified correctly): 0.990; and F1 score: 0.983.

Overview of Step 5

In Step 5, we go sequentially over specific sources containing applicant information and gradually build the 'final' list of applicants. We start by imputing the preferred name associated with entries in source 11 (PATSTAT harmonized applicants). When entries are missing, we browse source 12 (PATSTAT original applicants), then source 10 (PatentsView raw applicants from the pre-grant publication), and finally, source 9 (PatentsView raw applicants from the granted publication). This step led to the identification of applicant information for 6,403,853 patent applications. Thus, there are 1,789,070 applications without an applicant.[17]

Overview of Step 6

Next, we resort to several 'tricks' to guess the applicants for the 1.7 million remaining patents:

1. Take the assignee information from the pre-grant publication (source 1). Changes in assignees rarely occur before the grant, such that the assignee listed in the pre-grant publication is most likely also the (non-inventor) applicant;
2. Then, search into the U.S. family members for an applicant, and when many are found, select the most common applicant;
3. Then, check if at least two inventors are always assigned to the same applicant and use that applicant's information;
4. Then, check if at least two assigned 'applicants' (who are also inventors) are always assigned to the same applicant and use that applicant's information.

These tricks collectively allow for the recovery of applicant information from 317,367 patent applications, meaning we still lack information for 1.5 million patent applications.

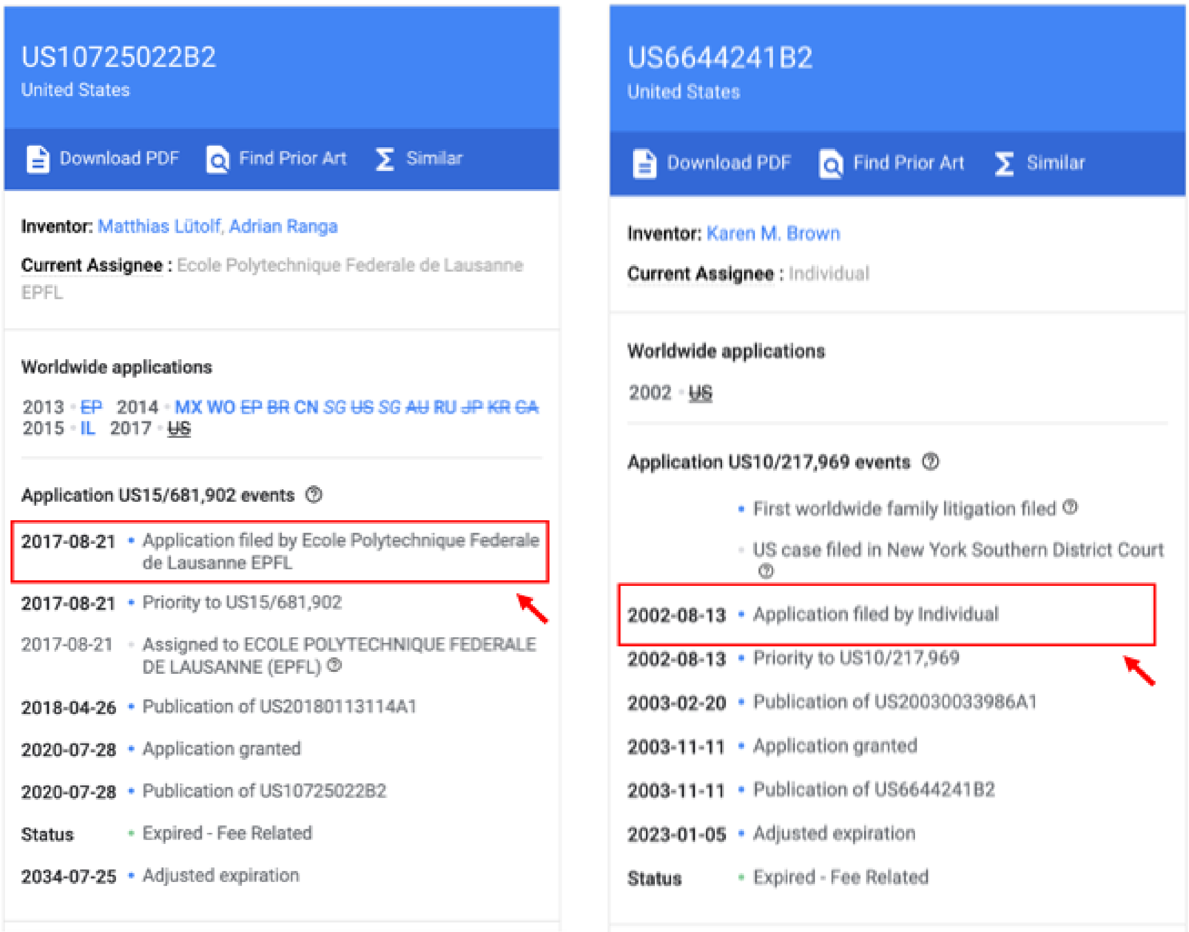
Overview of Step 7

In the final step, we use the information contained in source 11 (PATSTAT harmonized applicants) even if there is no corresponding entry in the dictionary we have built. This situation occurs because there is no "source 3" associated with this applicant's name (and, therefore, no

[17] In most instances, missing cases correspond to applicants who are also listed as inventors. The paper includes alternative treatments of these applications to ensure that how they are treated does not impact the results.

entry in the dictionary despite source 11 being available). The information is still missing for about 1 million patent applications. We visit https://patents.google.com and extract the applicant's name (most likely corresponding to individual inventors, but not necessarily). Figure A2 illustrates the field that we extract.

Figure A2. *Illustration of extracted field*



Source: https://patents.google.com/patent/US10725022B2/ and https://patents.google.com/patent/US6644241B2/

The main disadvantage of this approach is that the 1.5 million names recovered from (unmatched) source 11 and patents.google.com are not harmonized; they may not correspond to the format used for the other 6.7 million entries. Leaving them in a format different from the other group of patents is problematic because we want to track all patents by the same applicants. What would appear as new applicants in the data might simply be name variations of existing applicants.

This discussion suggests that we need another layer of applicant harmonization. The objective is to match as many entries as possible from the 1.5 million list to the 6.7 million list, leading us to evaluate 10.05 trillion pairs of names. Fortunately, we can dramatically shorten this set in the following ways:

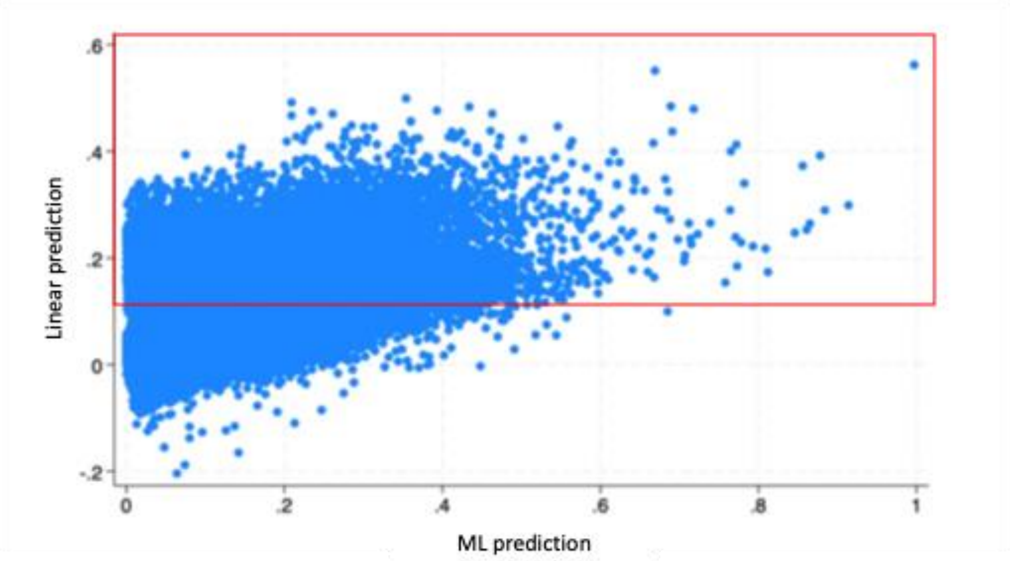
- Exclude 764,399 applications marked as filed by individuals (see right panel of Figure A2);
- Focus on unique names by dropping repeated entries in both lists;
- Perform some simple string manipulation on both lists to identify direct matches.

In total, there are 389,138 distinct applicants for 981,189 applications by non-individuals in the 1.5 million list and 203,613 distinct applicants in the 6.7 million list. A total of 25,265 applicants directly match across both lists, reducing the set of pairs to consider to 74 billion. This is still a very large number, and we seek to reduce it further by considering only applicants who share patents with at least one IPC code level 1 in common and focusing on applicants who appear at least twice in the 1.5 million list. This approach reduces the set of pairs to 12 billion, which is still too large.

To further reduce the set of pairs to analyze, we run our ML model on a subset of the data and compute the probability score that the two entities in a pair relate to the same applicant. We then develop a linear regression model (which is much faster to run) to predict which pairs from the 12 billion we can safely discard. In other words, we run a first pass on all pairs with a fast linear regression model to discard pairs that obviously relate to different entities. Then, we run the (slower) ML model on the subset of the data that we could not discard.

Figure A3 reports the correlation between the linear prediction and the ML prediction. Discarding all observations with a linear prediction score lower than 0.15 allows us to keep the vast majority of ML predictions with a prediction score greater than 0.50. This approach reduces the sample of pairs to 300 million, which becomes manageable. This step led to harmonizing 9750 applicants matched to 73,840 patent applications. Given that 1,024,671 applications in the 1.5 million list are not (identified by Google as) individuals, this final step resulted in the harmonization of 7.2 percent of the patents.

Figure A3. *Linear prediction vs. ML prediction*



Including group-level information

The work conducted thus far was solely concerned with harmonization, *i.e.*, identifying variations of the same name. We have also integrated group-level information in an effort to disambiguate applicant data. The two main disambiguation efforts that we are aware of are the UVA Darden Global Corporate Dataset, covering 3 million granted patents from 1980 to 2017, and Duke's DISCERN dataset, covering 1.3 million granted patents from 1980 to 2015. We use the UVA dataset due to its larger coverage.

This method is not perfect because it defines ownership at one point in time (*i.e.*, when the crawl was performed) and because the method seems to force ownership to companies (which is problematic for universities and public research organizations). Besides, it only includes data on granted patents from public companies. Nevertheless, it is the largest effort to date and we have therefore decided to exploit it.

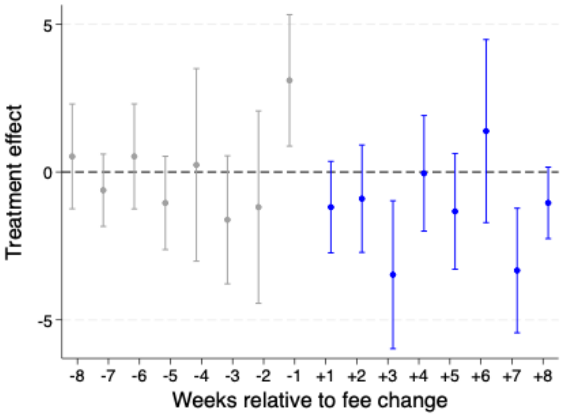
Ideally, we would have had many-to-one matches between our harmonized names and the UVA names. However, some entries have many-to-many matches, necessitating a rule to assign a harmonized applicant to the group owner. We have followed this process:

1. We exclude universities and public research organizations from the group-level harmonization as per the previous discussion;
2. We start by identifying harmonized names matched to only one UVA name (3 million records);
3. Next, when a harmonized name is matched to a pair more than 80 percent of the time in the UVA file and when the harmonized name appears more than 50 times in total, we consider that the match is 'almost unambiguous' (1.4 million records);
4. Next, when ambiguity remains, we use the direct record provided by the UVA file. That is, we match this time on patent numbers, and not on applicant names (50 thousand records);
5. Finally, we use the remaining 4 million harmonized names unmatched as the UVA 'group' name.

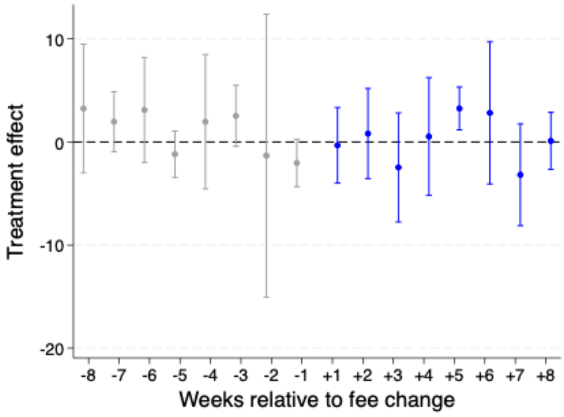
Appendix B. Additional results on U.S.-invented patent applications

Figure B1. *Treatment effect of the short-term response to fee change 6 (DiD) U.S. inventors*

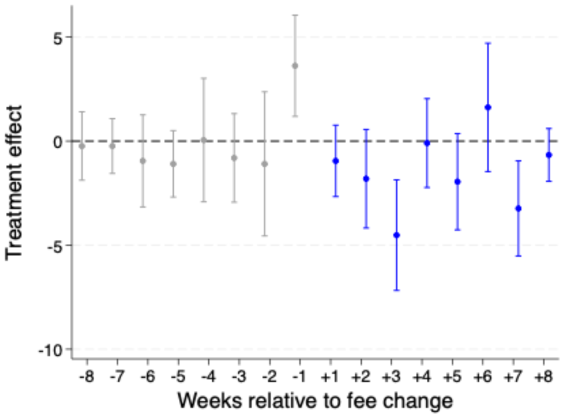
A. Undiscounted, all U.S. inventors



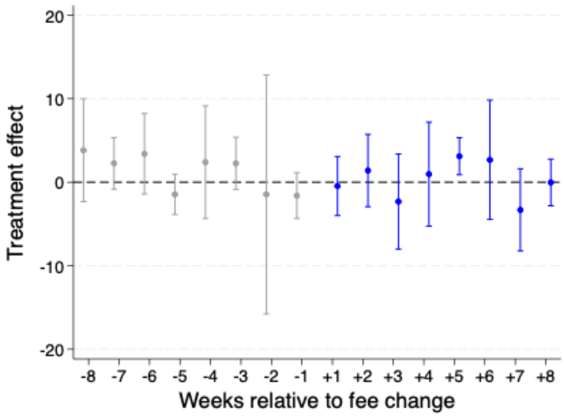
B. Small entities, all U.S. inventors



A. Undiscounted, at least one U.S. inventor



B. Small entities, at least one U.S. inventor



Notes: Similar to Figure 9. Results from an OLS DiD regression model with Newey-West standard errors. The dependent variable is Y_d^U (5YG) in Panel A and Y_d^S (5YG) in Panel B. 95-percent confidence intervals reported.

Table B1. *Poisson DiD estimates with 'individuals' treated as new applicants, daily data (U.S. inventors)*

Panel A: all U.S. inventors								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Time window (months):	[-6,+6]				[-5,+5]		[-4,+4]	
Construction	3YG		5YG		5YG		5YG	
DV:	Y_d^S	Y_d^M	Y_d^S	Y_d^M	Y_d^S	Y_d^M	Y_d^S	Y_d^M
Post change	-1.34	-0.67	-0.97	-0.52	-0.26	-0.37	0.46	-0.31
	(1.33)	(0.26)	(1.23)	(0.24)	(1.27)	(0.25)	(1.44)	(0.28)
Close pre change	-0.02	-0.01	0.88	0.10	1.40	0.17	1.15	0.13
	(2.19)	(0.37)	(1.84)	(0.33)	(1.72)	(0.32)	(1.89)	(0.34)
Close post change	2.82	0.12	1.98	0.05	1.85	0.01	1.01	-0.08
	(1.78)	(0.40)	(1.79)	(0.39)	(1.75)	(0.38)	(2.02)	(0.41)
N	730	730	730	730	606	606	486	486
Deviance	5074	1938	4369	1694	3558	1396	2806	1092

Panel B: at least one U.S. inventor								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Time window (months):	[-6,+6]				[-5,+5]		[-4,+4]	
Construction	3YG		5YG		5YG		5YG	
DV:	Y_d^S	Y_d^M	Y_d^S	Y_d^M	Y_d^S	Y_d^M	Y_d^S	Y_d^M
Post change	-0.83	**-0.66**	-0.40	-0.52	0.55	-0.34	1.16	-0.27
	(1.30)	**(0.25)**	(1.19)	(0.24)	(1.22)	(0.24)	(1.39)	(0.27)
Close pre change	-0.25	-0.11	1.03	0.05	1.59	0.14	1.36	0.13
	(2.29)	(0.39)	(1.87)	(0.34)	(1.75)	(0.32)	(1.90)	(0.34)
Close post change	2.58	0.10	1.50	0.00	1.20	-0.05	0.40	-0.12
	(1.91)	(0.41)	(1.96)	(0.41)	(1.97)	(0.41)	(2.21)	(0.43)
N	730	730	728	728	606	606	486	486
Deviance	5548	2117	4795	1845	3918	1525	3088	1189

Notes: Similar to Table 8. Population-averaged Poisson regression model with AR(1) structure. The table only reports the marginal effects of the interacted terms. Results for fee change 14. Bold type indicates statistical significance at the 1 percent probability threshold.

Table B2. *Poisson estimates of fee change 11, daily data (U.S. inventors)*

	All U.S. inventors				At least one U.S. inventor			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
DV:	Y_d^U	Y_d^U	Y_d^U	Y_d^U	Y_d^U	Y_d^U	Y_d^U	Y_d^U
Construction:	3YG	3YG	5YG	5YG	3YG	3YG	5YG	5YG
Post change	**-1.29**	-2.04	**-1.21**	-1.58	**-1.46**	-1.78	**-1.36**	-1.81
	(0.31)	(0.80)	**(0.27)**	(0.69)	**(0.34)**	(0.86)	**(0.30)**	(0.76)
Close pre change	0.05	-0.27	-0.18	-0.34	0.12	-0.02	0.00	-0.19
	(0.62)	(0.67)	(0.53)	(0.58)	(0.66)	(0.74)	(0.58)	(0.64)
Close post change	-1.29	-1.02	-1.14	-1.01	-1.31	-1.19	-1.03	-0.86
	(0.55)	(0.64)	(0.48)	(0.54)	(0.60)	(0.67)	(0.54)	(0.61)
Pre first to file	**11.78**	**11.73**	**9.75**	**9.72**	**13.20**	**13.17**	**10.70**	**10.67**
	(0.73)	**(0.73)**	**(0.64)**	**(0.64)**	**(0.78)**	**(0.78)**	**(0.69)**	**(0.69)**
Time trend		0.00		0.00		0.00		0.00
		(0.00)		(0.00)		(0.00)		(0.00)
N	365	365	365	365	365	365	365	364
Deviance	743	741	673	673	824	823	748	747

Notes: Similar to Table 11. Population-averaged Poisson regression model with AR(1) structure. Marginal effects at mean reported. All regressions include a dummy variable for Saturdays and Sundays and a constant term (not reported). Results for fee change 11. Bold type indicates statistical significance at the 1 percent probability threshold.