Early disclosure law reduces duplication in the US and European patent systems

Sonja Lück [a], Benjamin Balsmeier [b, c], Florian Seliger [c], and Lee Fleming [d]

a) University of Paderborn: sonja.lueck@uni-paderborn.de
b) University of Luxembourg: benjamin.balsmeier@uni.lu
c) ETH Zurich: seliger@kof.ethz.ch
d) University of California, Berkeley: lfleming@berkeley.edu

May 2019

Abstract
Much research assumes or theorizes that competition in innovation elicits duplication of research and that disclosure decreases such duplication. We validate this empirically using the American Inventors Protection Act (AIPA), three complementary identification strategies, and a new measure of blocked future patent applications. We show that AIPA -- intended to reduce duplication, through default disclosure of patent applications 18 months after filing – reduced duplication in the US and European patent systems. Use of blocked future applications, rather than the oft-used measure of future prior art citations, affords greater nuance in empirical investigation of positive and negative externalities of innovation.

JEL-Classification: O31, O33
Keywords: AIPA, Duplicative Research, Knowledge Disclosure, Patents

* The authors thank Guan Cheng Li for invaluable research assistance. We are grateful for detailed and very helpful advice on an earlier related research project from Stu Graham, and we thank Bronwyn Hall, Gaetan de Rassefosse, Christian Helmers, Hong Lou, Deepak Hegde, Olav Sorenson and three anonymous Science referees for very helpful comments and suggestions. We gratefully acknowledge financial support from The Coleman Fung Institute for Engineering Leadership and the National Science Foundation (1360228). Errors and omissions remain the authors'.

The premise and grand bargain of patent law is to provide just enough incentive to an inventor to disclose their invention; in return for a temporary monopoly (typically 20 years), an inventor makes enough details public, such that other inventors can either avoid duplicating the effort and/or build upon the idea more easily. The inventor calculates that a patent will prove more valuable than a trade secret. In turn, society benefits from less duplication of research, and faster and more differentiated follow-on research, following publication of the patent. Following this logic, the American Inventors Protection Act (AIPA) of 2000 was intended to encourage faster disclosure of inventions; aligning the US Patent and Trademark Office (USPTO) with the rest of the world, AIPA stipulated that by default all applications would be published 18 months after first filing, rather than at issuance. Here we analyze newly available administrative data (both American and European) that enables a direct measure of duplicated research effort, by observing applications that are rejected or "blocked" because a patent examiner declares the application to be obvious or not novel, relative to the original and explicitly identified "blocking" patent.

This new measure of "blockings" fits into the rich theoretical and nascent empirical literature on patent racing. Most of the research on patent racing proceeds from the assumption that competition in innovation inevitably elicits duplicated research and development investment (Loury, 1979, Reinganum, 1985, Lee and Wilde, 1980, Dasgupta and Stiglitz, 1980, Gilbert and Newbery, 1982; Scotchmer, 1991; Roin, 2005; Baker and Mezzetti, 2005; Bar, 2006; Bessen and Maskin 2009, Thompson and Kuhn, 2017). Perhaps the most famous anecdote used to motivate this literature is Elisha Gray's loss, by two hours, of his patent race for the telephone to Alexander Graham Bell. While races may elicit greater effort on the part of competitors, the lack of disclosure can waste societal resources in duplication; secrets are kept longer, rivals waste resources re-inventing the metaphorical wheel, and the societal benefits of technological improvement are delayed.

Inventors in a race, like Bell and Gray, are often confronted with a Prisoner's Dilemma type of problem. They cannot coordinate and/or credibly commit themselves towards early disclosure of all their efforts, which could make both parties (and third parties) better off. If only one party discloses, the other will often see no benefit in disclosing anymore, so no disclosure happens in the first place. The market may thus fail to provide the welfare maximizing disclosure system, giving reason for the policy maker to step in.

Disclosure regulation intends to solve this dilemma and balance the interest of inventors and society (Hall and Harhoff, 2012; Williams, 2017). Empirical work has validated many of the assumptions of this intent. Early disclosure increases (more citations by future patents), accelerates (faster appearance of future citations) and improves (higher difference relative to prior art) follow-on invention (e.g. Hegde, Herkenhoff, and Zhu, 2019; Baruffaldi and Simeth, 2018; Graham and Hegde, 2015). Nuanced measures of technological proximity also illustrate how early disclosure makes "close" patents less similar and "distant" patents more similar, relative to the disclosed patent (Hegde, Herkenhoff, and Zhu 2019), assumedly illustrating how inventors move on more quickly from already claimed territory. Furman, Nagler and Watzinger (2018) show that the opening of patent libraries increased geographically proximal patenting activity. Empirical work has also found benefits for the initial patent holders that are associated with (voluntary) disclosure, including more licensing agreements, venture capital interest, and positive reputation and network effects (Hegde and Lou, 2018; Khashabi and Mohammadi, 2016; Muller and Pénin, 2006, Pénin, 2007).

It remains to be established, however, whether earlier disclosure actually reduces duplicated research, as predicted by theory, and hoped for by AIPA policy makers. Surveys have provided qualitative evidence, indicating that inventors indeed peruse patent documents as a valuable resource of knowledge that influences their inventive efforts and direction (Ouellette, 2012, 2017). To establish that early disclosure does reduce duplication, we use newly available data and the American Inventors Protection Act (AIPA) as a quasi-natural experiment. This enables measurement of how earlier patent disclosure through AIPA had the intended impact of reduced duplication, as measured by fewer blockings of post AIPA patent application claims.

The contribution of the paper is twofold. First, we add to the patent racing and disclosure literature by establishing a first order but empirically unaddressed assumption that earlier knowledge disclosure should lead to reduced duplication. We demonstrate this with three complementary estimations that return consistent results, including a regression discontinuity design, twins matching between U.S. and European patents, and a difference in differences estimation. Second, we apply the new measure of blockings and demonstrate its usefulness for empirical research. The count of blockings provides a more nuanced measure of the positive and negative externalities of invention, thus complementing more widely used measures such as future prior art citations

(Trajtenberg, 1991), knowledge spillovers to close or distant technologies (Jaffe, 1986, 1989), and a focal firm's stock price reaction to a patent publication (Kogan et. al., 2017).

Measuring blocked USPTO patent applications

The newly available USPTO Office Action Dataset provides a way to measure blocked inventions by a focal US patent.[1] It contains detailed information derived from office actions issued by examiners to applicants during the patent examination process. An "office action" is a written notification to the applicant of the examiner's negative decision on patentability and generally discloses the grounds for a rejection, the claims affected, and the pertinent prior art. This initial release consists of three files derived from 4.4 million office actions mailed during 19 March 2008 to 11 July 2017 period from USPTO examiners to the applicants of 2.2 million unique patent applications (Lu, Myers and Beliveau, 2017). Within these data we identified the action types '102' and '103' as stipulated by section 35 USC. Essentially, the patent examiner expresses doubts on novelty (102 action) or obviousness (103 action) regarding the patent application at hand and refers to explicitly identified patents as a reason for this decision.[2]

We defined the number of blockings by a focal patent three different ways: (1) the number of 102 and 103 office actions generated by subsequent patent applications that refer to the focal patent as the blocking patent, (2) the number of subsequent patent applications that generated at least one office action referring to the focal patent, and (3) the number of subsequent patent applications that generated at least one office action referring to the focal patent and which were not granted eventually. Presented results are based on (1), and are very similar if alternatives (2) or (3) are chosen (please see Appendix A, Tables A2 and A3, and Tables A6 and A7 for separate estimations of 102 and 103 actions).

[1] USPTO Office Action Research Dataset: https://www.uspto.gov/learning-and-resources/electronic-data-products/office-action-research-dataset-patents.

[2] The USPTO provides the following definitions:

- 102 „not novel": A claimed invention may be rejected under 35 U.S.C. 102 when the invention is anticipated (or is "not novel") over a disclosure that is available as prior art.
  Source: https://www.uspto.gov/web/offices/pac/mpep/s2131.html
- 103 „obviousness": A patent for a claimed invention may not be obtained, notwithstanding that the claimed invention is not identically disclosed as set forth in section 102, if the differences between the claimed invention and the prior art are such that the claimed invention as a whole would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to which the claimed invention pertains. Patentability shall not be negated by the manner in which the invention was made.
  Source: https://www.uspto.gov/web/offices/pac/mpep/s2141.html.

The US blocking data has an important drawback though. It covers only office actions from March 2008 onwards, meaning the data is left-censored, and hence provides very few blocked patent applications that were filed before and after AIPA was enacted (all patents that were either granted or abandoned before March 2008 and have been blocked remain unobservable). To avoid bias, we use and find consistent results with European Patent Office (EPO) data -- which are available since 1977.

Measuring blocked EPO patent applications

To model the number of blocked inventions at the EPO by a focal US patent we exploit two unique (and unavailable in the US data) features of the uncensored European data. EPO standard examination practice is to determine all relevant prior art and explicitly classify each citation according to its relevance. The resulting examiner's search report contains on average fewer citations than a typical USPTO patent, reflecting explicit rules of selectivity and justification in prior art citation. Two categories, labeled X and Y by the EPO, are relevant (please see Appendix G for more details and an example of an EPO search report).[3]

Analogous to the US blockings, we define the number of EPO blockings as the number of X and Y citations referring to the focal US patent as the blocking patent (results are robust to counting X and Y citations separately). As with 102 and 103 actions in the US this may not imply an abandoning of the whole patent application in question as there could still be claims that are novel and broad enough to warrant patent protection, just with a smaller scope than initially requested. Presented results are based on the sum of X and Y citations. Analogous to the variety of US patent measures, we also estimated the number of EPO patent applications of which at least one claim was blocked, and the number of eventually abandoned patent applications, and found very similar results.[4]

[3] The EPO defines an X citation as: "…where a document is such that when taken alone, a claimed invention cannot be considered novel or cannot be considered to involve an inventive step", and a Y citation as, "…applicable where a document is such that a claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other documents of the same category, such combination being obvious to a person skilled in the art." Source: https://www.epo.org/law-practice/legal-texts/html/guidelines/e/b_x_9_2_1.htm.

[4] Harhoff and Wagner (2009) show that the number of X and Y citations a patent receives strongly correlates with patent value.

The European blocking data also allow the calculation of the fraction of X and Y citations out of all citations made by EPO examiners to a given US patent as a dependent variable. This measure should be more robust to potential changes in examiner citation behavior -- assuming that behavioral changes were consistent across citation categories.[5] An analogous measure of the fraction of US blockings does not exist however the number of 102 and 103 office actions scaled by future prior art cites reveals similar results (see Appendix A8 to A10, and Figure A4).[6] As the European data are not censored, we are also able to increase comparability to US data by counting blockings within a time window of 8 years since application of the blocking US patent (results are robust to not truncating the data and a smaller time window of 5 years).[7] The research required the integration of many data sources; for a complete list and summary statistics please see Appendix A, Table A5, and Appendix F.

Three complementary identification strategies

Aligning the US Patent and Trademark Office (USPTO) with the rest of the world, AIPA stipulated that by default all patents would be published 18 months after first filing of the application instead of at issuance (inventors could still choose to keep their application unpublished until grant, if they filed only in the U.S. and did not seek foreign protection). This law went into effect on November 29, 2000. We pursue three complementary identification strategies that exploit this regulatory change; a 1) regression discontinuity design (RDD), 2) twin study (TW), and 3) difference in differences (DiD) estimation, each of which have caveats but also address different threats to a causal interpretation. With consistent findings across the three measures, we present results and descriptive data of our preferred RDD method first, and provide a somewhat shorter overview of the latter two approaches, plus a placebo test (analysis details in the Appendix B - D). Across all three identification approaches, we find that earlier patent disclosure induced by AIPA reduces the number of duplicated claims in patent applications, by 2.9 to 14.3 percent (differences across technology classes are presented in Appendix E).

---

[5] For full list see: https://www.epo.org/law-practice/legal-texts/html/guidelines/e/b_x_9_2_1.htm.

[6] The fraction could of course also be driven by an increase in non-blocking cites, which other research has found (e.g. Hegde et al., 2019). It seems unlikely, however, that examiners changed their citation behavior in a way that non-blocking cites increase while blocking cites decrease without a connection to a real change in patenting behavior.

[7] To maximize comparability over time, we count blocking cites to *granted* patents that could always, before and after AIPA, be observed. To the extent that the publication of pre-grant documents changed the citations to granted patents, this change should be fully absorbed by our fractional measure, or cancelled out in the DiD estimation.

Regression Discontinuity Design (RDD)

Building on Davis (2008) and Lee and Lemieux (2010) we model a RDD that exploits the discontinuous jump to disclosure for US patents filed on or after November 29, 2000, which did not seek parallel foreign protection (hereafter 'US only'). Two arguments motivate the RDD. First, the hypothesized change in duplicated future patent claims is observable for patents filed immediately before and after AIPA.[8] Being able to restrict the analysis to a small time window limits potentially confounding influences from law, policy changes or the bust of the dotcom bubble that happened concurrent to AIPA or within the year change from 2000 to 2001. Second, we find no evidence of avoidance strategies by inventors, for example, filing as many patents as possible before or after the regime shift (the number of patent applications in the weeks before and after AIPA remained stable, see Appendix A, Figure A1), or a shift towards more or less patenting with or without parallel foreign protection (the fraction of applications with parallel foreign protection in weeks before and after AIPA remained stable, also Appendix A, Figure A1). Other patent characteristics also remained stable across windows of varying length (see Appendix A, Figure A2). Table 1 presents descriptive statistics for the one month and one year windows, where one month refers to 31 days before and after AIPA, and one year is 365 days before and after AIPA, respectively.

Table 1: Descriptive statistics

Blocking actions					
One month window	Obs	Mean	SD	Min	Max
Number of 102 & 103 blockings	17382	1.58	3.76	0.00	187.00
Number of 102 & 103 blockings (in log)	17382	0.54	0.78	0.00	5.24
Number of X & Y blockings	17382	0.23	0.85	0.00	21.00
Number of X & Y blockings (in log)	17382	0.11	0.35	0.00	3.09
Fraction of X & Y blockings	17382	0.06	0.18	0.00	1.00
One year window	Obs	Mean	SD	Min	Max
Number of 102 & 103 blockings	191220	1.50	3.46	0.00	189.00
Number of 102 & 103 blockings (in log)	191220	0.52	0.76	0.00	5.25
Number of X & Y blockings	191275	0.23	0.92	0.00	48.00
Number of X & Y blockings (in log)	191275	0.11	0.35	0.00	3.89
Fraction of X & Y blockings	191275	0.06	0.18	0.00	1.00

Notes: This table reports descriptive statistics on all eventually granted patents filed with the USPTO one month, respectively, one year before and after AIPA became effective on November 29, 2000, which had no parallel foreign filing. The number of blocking actions is the number of times a given patent was referenced as a blocking patent (102 and 103) in the 4.4. Million office actions the USPTO issued between March 2008 and June 2017. The number of X and Y blockings is the number of times a given US patent was cited and classified as such by an EPO patent examiner in his official search report within 8 years since the referenced US patent was filed.

[8] Note the immediate change in blockings estimated and illustrated does not reflect an immediate change in applications, rather, the applications filed after Nov. 29, 2000 become less likely to block *future* patent applications.

To further minimize potential confounds and avoid adding endogenous variables to the regressions, we estimated models with all available data as well as a matched data set. For the balanced sample we matched each after AIPA patent to a corresponding patent before AIPA, using Coarsened Exact Matching (CEM), such that differences in observable patent characteristics before and after AIPA are minimized. We match on backward cites to capture potential differences in novelty, 6 NBER technology classes that capture differences in technological popularity, number of independent claims, dependent claims per independent claim, word count of the first claim, average number of words per claim, whether an attorney was involved, and patent pendency (time from application to grant date) to capture differences in patent scope and writing. Appendix A Table A5 shows the descriptive statistics of the matched and unmatched data, confirming few differences before matching and no significant differences in observables after matching. Notably, we balanced the sample of pre and post AIPA patents without a loss in common support. Though it is difficult to rule out concerns of strategic shifts in patenting behavior completely, these analyses should lessen these concerns (the twin study design below will address such concerns directly).[9]

Finally, patent examiners may have changed their citation behavior in a way that we might wrongly attribute a change in blockings to a change in innovative behavior instead of examiner citation behavior. To address this concern we modeled the fraction of X and Y citations out of all citations made by an EPO examiner (for similar results of US blockings see Appendix, Tables A8-A10 and Figure A4). This estimation should be robust to potential changes in examiner citation behavior as long as any change applies equally to all citation categories.

Our baseline specification is:

$$\log(blocked\ patents + 1)_i = \beta_1 \cdot Post\ AIPA_i + f(Days\ to\ AIPA)_i + \varepsilon_i\ , \qquad (1)$$

where $blocked\ patents_i$, refers either to the number of 102 and 103 office actions referencing patent *i*, or the number of times an EPO examiner references patent *i* with an X or Y in her search report, or the fraction of X and Y cites out of all references that patent *i* receives from EPO search

[9] One might argue that patent lawyers may have started to write patents in a different way than before AIPA. Given that observable characteristics of patent writings (number of words of the first claim, average number of words per claim, no. of independent and dependent claims) did not change (at least in the short run of one month or year, respectively), it seems unlikely that this could have had a significant impact on our results.

reports, respectively. $Post\ AIPA_i$ is an indicator indicating whether patent i was filed on or after November 29, 2000 and $Days\ to\ AIPA_i$ is the assignment variable, i.e. the difference in days between the filing date of patent i and November 29, 2000, and ε_i is the error term. Under the assumption that patents filed right before and after AIPA only differ in their filing date and disclosure, and f is correctly specified, β_1 will capture the causal influence of AIPA on the number of duplicated claims of future patent applications.

Alternatively, we estimate the same model as (1) but add technology fixed effects (6 NBER classes), δ_i , technology class specific trends, and firm fixed-effects (γ_i) that control for secular trends in technology popularity and unobserved time-invariant heterogeneity across firms. Our full model specification is thus:

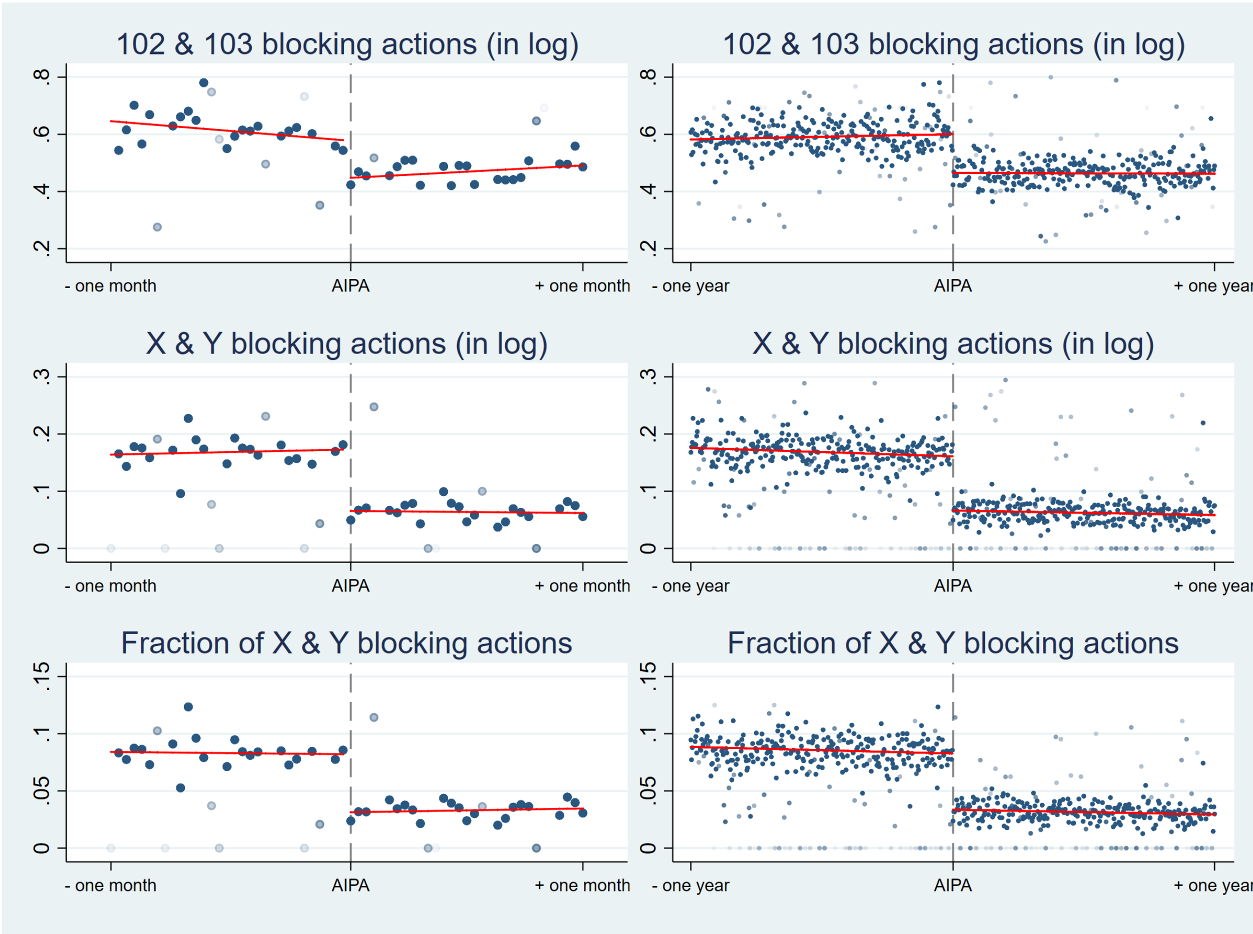
$$\begin{aligned} &\log(blocked\ patents + 1)_i \\ &= \ PostA\ IPA_i + \ f(Days\ to\ AIPA)_i + \delta_i + \delta_i \cdot \ Days\ to\ AIPA_i + \gamma_i + \ \varepsilon_i \end{aligned} \tag{2}$$

We present below results of six different specifications, starting with the simplest without further controls, and assuming the same linear slope f before and after AIPA (model 1). We then allow for differing linear slopes f before and after AIPA (2), add NBER technology class fixed effects (3), add technology class specific trends (4), add firm fixed effects (5), and finally allow for non-linear slopes of f before and after AIPA (6).[10]

All these models are first estimated with all US only patents applied for within one month before and one month after AIPA (Panel 1). We then re-estimate all models based on a matched sample as previously introduced (Panel 2). Table 2 presents the results across six specifications using three different dependent variables; first US blockings, second EPO blockings, third the fraction of X and Y cites from EPO search reports. Figure 1 illustrates the effects graphically. The year window sub-graphs (right-hand side) support the assumption of linear slopes. Corresponding results of the yearly models are presented in Appendix A Table A4.

[10] Modelling f more flexibly with higher order polynomials reveals similar results. The same is true if we disaggregate technology classes to NBER sub-class levels (36) and allow technology class specific trends to vary before and after AIPA.

Figure 1: RDD Graphs



Notes: Illustrates the discontinuous difference in blocking actions one month (year) before and after AIPA became effective on November 29, 2000. Dots represent the average amount of 102 and 103 office actions (log of X and Y cites, and fraction of X and Y cites, respectively) referring to patents filed at the USPTO on a given day. AIPA represents November 29, 2000. The red lines represent fitted values.

Table 2: RDD estimations

	Dependent variable: No. of 102 & 103 blocking actions (in logs)					
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL 1: Raw data, one month						
Post AIPA	-0.133***	-0.129***	-0.132***	-0.137***	-0.143***	-0.114**
	(0.023)	(0.023)	(0.023)	(0.023)	(0.033)	(0.046)
N	17382	17382	17382	17382	11083	11083
R^2	0.008	0.009	0.021	0.023	0.211	0.212
PANEL 2: matched data, one month						
Post AIPA	-0.121***	-0.121***	-0.115***	-0.118***	-0.114***	-0.104**
	(0.026)	(0.026)	(0.026)	(0.026)	(0.037)	(0.052)
N	13490	13490	13490	13490	8280	8280
R^2	0.008	0.009	0.021	0.023	0.218	0.218
	Dependent variable: No. of X & Y blocking actions (in logs)					
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL 1: Raw data, one month						
Post AIPA	-0.107***	-0.108***	-0.103***	-0.103***	-0.100***	-0.073***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.015)	(0.021)
N	17382	17382	17382	17382	11080	11080
R^2	0.023	0.023	0.030	0.031	0.236	0.237
PANEL 2: matched data, one month						
Post AIPA	-0.100***	-0.100***	-0.098***	-0.098***	-0.081***	-0.055**
	(0.012)	(0.012)	(0.012)	(0.012)	(0.017)	(0.022)
N	13481	13481	13481	13481	8277	8277
R^2	0.020	0.020	0.027	0.027	0.249	0.249
	Dependent variable: Fraction of X & Y blocking actions					
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL 1: Raw data, one month						
Post AIPA	-0.051***	-0.051***	-0.049***	-0.049***	-0.046***	-0.037***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)	(0.010)
N	17382	17382	17382	17382	11080	11080
R^2	0.020	0.020	0.025	0.025	0.210	0.210
PANEL 2: matched data, one month						
Post AIPA	-0.045***	-0.045***	-0.045***	-0.046***	-0.035***	-0.027**
	(0.006)	(0.006)	(0.006)	(0.006)	(0.008)	(0.012)
N	13481	13481	13481	13481	8277	8277
R^2	0.017	0.017	0.021	0.021	0.216	0.216
All PANEL	(1)	(2)	(3)	(4)	(5)	(6)
RDD time controls	linear (same slope)	linear (diff. slopes)	linear (diff. slopes)	linear (diff. slopes)	linear (diff. slopes)	quadratic (diff. slopes)
Technology class fixed effects	no	no	yes	yes	yes	yes
Technology trends	no	no	no	yes	yes	yes
Firm fixed effects	no	no	no	no	yes	yes

Notes: This table reports results of RDD models, equation (1) and (2). Technology class fixed effects are based on six NBER technology classes, and technology trends are linear time trends per technology class using filing days. Raw data are all US patent applications filed between one month before and one month after November 29, 2000, when AIPA became effective. Post AIPA is an indicator that equals one if the patent was filed on or after November 29, 2000. Matched data is a balanced sample, where patents before and after AIPA are (CEM) matched based on backward cites to capture potential differences in novelty, 6 NBER technology classes to capture differences in technological popularity, number of independent claims, dependent claims per independent claim, word count of the first claim, average number of words per claim, whether an attorney was involved, and patent pendency (time from application to grant date) to capture differences in patent scope and writing. Appendix A Table A5 shows the descriptive statistics of the matched data. Heteroscedasticity robust standard errors are presented in parentheses.

The RDD results suggest that AIPA reduced the number of blocked US patent claims between 9.9 percent (column 6, second model, matched data) and 13.3 percent (column 5, first model, raw data).[11] The effect appears consistent but smaller for European patents, where claims blocked by US patents were reduced by 5.4 (column 6, fourth model, matched data) to 10.2 (column 2, third model, raw data) percent. These numbers do not seem to be driven by a change in citation behavior as the effects translate into a reduction of the fraction of blocking cites by between 2.7 (column 6, sixth model, matched data) and 5.0 percentage *points* (column 1, fifth model, raw data).

Twins Design (TW)

Our second identification approach builds on Graham and Harhoff (2003) and follows Hegde, Herkenhoff and Zhu (2019) who propose a matched twins design by focusing only on US patent applications with a parallel EPO patent application of the same invention. These 'patent twins' are identified by a common patent family identifier that is assigned to each patent by an EPO patent examiner and available in the Patstat database. Identification comes from the difference between citations to the US and EPO applications, with the assumption that they cover the same invention and should therefore disclose the same knowledge.

The major advantage of this approach is that the inclusion of patent family fixed effects should control for unobserved differences in inventions. It should thus be robust to any unobserved changes in invention behavior, writing, and patent scope. It relies on the assumption that EPO patent examiners remain unaffected by the publication of a corresponding US patent, i.e. EPO examiners keep on citing the EPO patent application, and are not more likely to cite the parallel US patent once it is published. These assumptions should not be critical for two reasons. First, even if the assumption is not met, it should work against the hypothesized negative effect, because the bias would only increase citations to US patents relative to the EPO counterpart. Second, we also estimate a model where the fraction of X and Y cites serves as the dependent variable, which should be unaffected by a potential shift in citation behavior, assuming other citations are similarly affected as the blocking cites.

[11] Due to the log specification the economic magnitude of the effect is calculated as exp(β)-1.

It should also be noted that US patents with parallel foreign protection might be more valuable, at least from the view of the applicant, than US only patents, and there are likely even more filings in jurisdictions beyond the US and Europe. This means that knowledge about an invention is more likely to leak out before publication, regardless of AIPA, such that any potential AIPA effect would be attenuated within this group of patents. Taken together, this implies that the twin approach should provide a conservative estimate. The biggest downside is that we can only estimate these models for European blockings, since 102 and 103 references to non-US patents are not available from the USPTO.

We estimated 3 different specifications using OLS, starting with patent family fixed effects, which control for unobserved differences across inventions and absorb technology and firm fixed effects. We then add year fixed effects (2) and technology class specific trends. The number of X and Y cites, as well as the fraction of X and Y cites, serve as dependent variables. Detailed results are presented in Appendix B. The results confirm our previous estimations. The number of X and Y blockings decreases by 9.8 to 10.1 percent and the fraction of X and Y cites decreases by 5.9 to 6.3 percentage points, depending on specification.

Difference in Differences Design (DiD)

The third identification approach is a DiD estimation where US only patents serve as the treated group and all US patents with parallel foreign applications filed before or at the time of the US patent application serve as the control group (Graham and Hegde, 2015; Hegde and Lou, 2018). The latter group is supposedly unaffected or at least less affected by AIPA because the foreign invention is published by the foreign patenting entity 18 months after filing in its jurisdiction. The parallel trend assumptions appears valid (see figure A5 in Appendix C). The prior analysis as well as the graphical inspection of the data suggest, however, that US patents with parallel foreign protection are not unaffected by AIPA. Similar to their US only counterparts the number of blocked patents by US patents with foreign publication drops significantly after AIPA. This is most likely because patent publications in a foreign jurisdiction do not receive as much public attention as a USPTO publication, and patent examiners as well as applicants tend to search and cite mainly the US database. The DiD is thus likely to underestimate the true impact of AIPA. More precisely, it will estimate how much stronger the AIPA impact was for US only patent applications relative to US patent applications with parallel foreign applications.

The results are presented in the Appendix C. Similar to the RDD, we estimated 5 different specifications using OLS, starting with the 1) basic DiD without further controls, then adding year fixed effects (2), NBER technology class fixed effects (3), technology class specific trends (4), and firm fixed effects (5). We use all three previously used dependent variables, including the number of blocked US patent applications (102 and 103 references), number of X and Y cites, and fraction of X and Y cites out of all cites.

Based on these models the impact of AIPA appears to be less than the estimates based on the RDD and the twin study approach. Specifically, we find a reduction in blocked US patent claims (102 and 103) by US only patents between 4.5 and 3.7 percent, a reduction in EPO blockings by 2.9 to 4.0 percent, and a reduction in the fraction of X and Y cites by 1.3 to 1.8 percentage points. Given that these estimates reflect the additional effect of AIPA on US only patents over patents with parallel foreign protection, which was estimated previously to drop by about 10 percent, the small numbers remain consistent with our estimates based on the RDD.

Placebo test

Our placebo test draws on the opt-out option included in AIPA. Upon request and without additional costs, inventors could actively opt for pre-grant secrecy if they did not seek parallel foreign protection. These patents, about 15% of all US only applications (~7% in total), were thus treated by the USPTO as if they would have been submitted under the pre-AIPA regime; thus we would expect AIPA to have had no effect on this sub-group of patents. Not surprisingly, however, these patents are a highly selected group. Therefore, we (CEM) matched each patent that requested secrecy after AIPA to a pre-AIPA patent that does not significantly differ in terms of backward cites, technology class, number of independent claims, average number of words per independent claim, number of words in first claim, number of dependent claims to number of independent claims, an attorney dummy, and pendency. Re-running all previous regressions on this subset of patents reveals insignificant results throughout the specifications (see Appendix D).

If we run the regressions on the unmatched sample (Appendix D), we find positive effects of AIPA on future blockings, most likely reflecting the selection of particularly important patents into secrecy. Notably, this may cause a slight upward bias in our previous RDD and DiD regressions (the twin study is unaffected due to the restriction to patents with parallel foreign protection), where the secrecy patents were always included to get as close as possible to a treatment effect estimate. Since the number of secrecy patents is rather low in total, however, we only find slightly more negative coefficients that do not significantly differ from our previous regressions, when we either exclude the secrecy patents, or add an indicator for patent applications that requested secrecy (results available upon request).

Differences across technology fields

Appendix E describes heterogeneous effects across fields of technology (the 6 NBER categories: Computers and Communications, Drugs and Medical, Electrical and Electronics, Chemical, Mechanical and Others). This heterogeneity might result from differences in the strength and breadth of intellectual property (IP) protection, varying lead-times and difficulties of invention, and the degree of competition and likelihood of duplication (see Arora et al., 2008; Bessen and Maskin, 2009, Galasso and Schankerman, 2015; Sampat and Williams, 2019). The largest effects occur in Computer and Communications technologies, with weaker effects in Chemistry and Drugs and Medical.

Informal estimates of upper and lower bounds on impact

It remains difficult and speculative to empirically estimate any efficiency gains that might have resulted from AIPA. However, even if a lower number of blocked patents does not reduce duplication in performing research, the results at least imply a more efficient patent application process that avoids processing unsuccessful claims in the first place. In speculating about these bounds, we will use the RDD estimates of 9.9% and 5.4% for US and European patents, rounded to 10% and 5%, respectively. The USPTO reports an absolute amount of 12.3M 102 and 103 references in 4.4M office actions between March 2008 and July 2017 (USPTO, 2018), and the EPO reports 2.7M X and Y cites over the same period (EPO, 2017, often one cite corresponds to one office action in the EPO). This implies 1.37M avoided references, 0.49M avoided actions, and 0.15M avoided cites (10% and 5% less for the 8 year period, for US and EPO, respectively). Keep in mind the EPO grants fewer patents each year than the USPTO.

To provide a lower bound on increased efficiency, one could assume that R&D spending and allocation remained completely unaffected by AIPA (in other words, the prior amount of duplication continues unchanged). In that case, the gains would accrue only from decreased time and expenses for inventors, their lawyers, and patent examiners. Given the time to craft a claim and possibly respond to a rejection (or retain an expensive lawyer to do so) on the part of the inventor, and on the part of the examiner, to research, find a basis for rejection, and write up and send a rejection, each rejected claim probably takes multiple, probably tens, of professional hours. This implies millions of hours of saved effort over the time period for highly paid innovation professionals (e.g., 0.49M actions * 10 hrs/action). At the high end, if patenting entities were able to redirect the AIPA savings of 10% or 5% into new avenues of research, multiplied with an estimate of North American and European R&D of $668B in 2010, [12] this implies yearly savings in billions of dollars. The actual savings is surely lower and incorporates saved processing time as well as some redirected R&D.

Conclusion

The nature of competition for innovation confronts inventors with a dilemma. Keeping all their knowledge secret as long as possible often provides the dominant strategy to maximize the returns to their inventions, even though coordinating early disclosure could reduce wasteful duplication. Our results suggest that AIPA helped to alleviate this problem. The results confirm a long-standing theoretical literature that predicted the duplication of inventive efforts resulting from patent races. The empirical quantification of a first order effect of the patent system adds to the nascent literature on the costs and benefits of the patent system in general and its disclosure function in particular (see e.g. Roin, 2005; Williams, 2017; Sampat and Williams, 2019).

Although duplicated patent claims decreased after AIPA, we do not see a drop in R&D investments attributable to AIPA in nationwide statistics. This may not surprise, however, given prior results that future patent applicants seem to build more frequently and quickly on earlier disclosed knowledge and diverge more significantly from prior research (Hegde, Herkenhoff and Zhu, 2019). Ideally, resources are not just saved but spent differently and more efficiently. At the very least, innovation professionals spend less time processing failed patent applications, and ideally, inventors can redirect what would have been wasted and duplicated efforts into new fields of inquiry.

[12] 2010 spending in 2010 dollars, see http://uis.unesco.org/apps/visualisations/research-and-development-spending/, keeping in mind that not all of this investment aims for a patent or would be influenced by AIPA.

Recent studies have begun to use the blocking measure (Thompson and Kuhn, 2017) and this study provides an additional example and characterization. The measure provides a useful complement to the very well (and perhaps over) used measure of patent citations (Trajtenberg, 1991) as well as a more recent measure from the finance literature that assigns a financial value to a firm's patent, based on the stock market reaction in the days around its publication (Kogan et. al., 2017). Perhaps the most promising opportunity for the blocking measure is its ability to quantify negative externalities. Whereas citations are normally interpreted as positive (though possibly unwanted) knowledge diffusion, and an increase in a firm's stock price only considers the focal firm -- and does not take account of the potentially negative impact on competitors' stock prices -- blocking counts give an indication of innovative effort that was inhibited.

Much theoretical work, starting with Schumpeter's image of creative destruction, and the patent race literature, has described negative externalities of innovation. Few non-theoretical studies, however (e.g., Kogan et al., 2018; Bloom, Schankerman and van Reenen, 2013), have been able to model it empirically. Research suggests that positive spillovers dominate business stealing effects but where the latter actually happen and what drives them remains less clear. The blocking measure may thus help to shed more light on important but insufficiently tested hypotheses, e.g. on the relationships between competition, innovation, and creative destruction. One might be able to characterize the intensity of competition between inventors, firms, regions, and within and across industries. This characterization would enable a much better understanding and more accurate empirical modelling of innovation dynamics, and in particular, its antecedents and consequences.

Measures of blocking will also enable research into the costs and benefits of the patent system. By relying solely on patent counts or future cites as a dependent variable, empirical research might miss a large part of the negative side effects that many patents might generate. Taking blocked patent applications into account will enable researchers to draw a more complete picture and should thus provide better insight and a more realistic evaluation of policies and their impact on innovation and the economy. As we can in principal also observe who is negatively affected we might also better understand where negative externalities occur and how to minimize them. We hope that future work will thus benefit from being able to quantify and differentiate between the positive and negative externalities of patented inventions.

References

Arora, A., Ceccagnoli, M., & Cohen, W.M. (2008). R&D and the patent premium. International Journal of Industrial Organization, 26(5), 1153-1179.

Baker, S. & Mezzetti, C. (2005). Disclosure as a strategy in the patent race. Journal of Law and Economics, 48(1), 173-194.

Bar, T. (2006). Defensive publications in an R&D race. Journal of Economics & Management Strategy, 15(1), 229-254.

Baruffaldi., S. & Simeth, M. (2018). Patents and Knowledge Diffusion: The Effect of Early Disclosure. Available at SSRN: https://ssrn.com/abstract=3160533.

Bessen, J. & Maskin, E. (2009). Sequential innovation, patents, and imitation. The RAND Journal of Economics, 40(4), 611-635.

Bloom, N., Schankerman, M., & van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. Econometrica, 81(4), 1347-1393.

Cohen, W. M., Goto, A., Nagata, A., Nelson, R. R., & Walsh, J. P. (2002). R&D spillovers, patents and the incentives to innovate in Japan and the United States. Research policy. 31(8), 1349-1367.

Daiko T., Dernis H., Dosso M., Gkotsis P., Squicciarini M., Vezzani A. (2017). World Corporate Top R&D Investors: Industrial Property Strategies in the Digital Economy. A JRC and OECD common report. Luxembourg: Publications Office of the European Union.

Davis, L. W. 2008. The Effect of Driving Restrictions on Air Quality in Mexico City. Journal of Political Economy. 116(1), 38–81.

Dasgupta, P., & Stiglitz, J. (1980). Industrial Structure and the Speed of R&D. The Bell Journal of Economics. 11(1), 1-28.

EPO (2017). https://www.epo.org/law-practice/legal-texts/html/guidelines/e/b_x_9_2.htm

Furman, J., Nagler M, & Watzinger, M. (2018). Disclosure and Subsequent Innovation: Evidence from the Patent Depository Library Program. NBER Working Paper No. 24660.

Galasso, A. & Schankerman, M. (2015). Patents and Cumulative Innovation: Causal Evidence from the Courts. The Quarterly Journal of Economics, 130(1), 317-369.

Gilbert, R. J. & Newbery, D. M. (1982). Preemptive Patenting and the Persistence of Monopoly. The American Economic Review, 72(3), 514-526.

Graham, S., Hall, B., Harhoff, D., & Mowery, D. (2003). Patent quality control: A comparison of US patent re-examinations and European Patent oppositions. Patents in the knowledge-based economy, 74, 84-85.

Graham, S., & Hegde, D. (2015). Disclosing patents' secrets. Science, 347(6219), 236-237.

von Graevenitz, G., Wagner, S., & Harhoff, D. (2013). Incidence and growth of patent thickets: The impact of technological opportunities and complexity. The Journal of Industrial Economics, 61(3), 521-563.

Harhoff, D., & Wagner, S. (2009). The Duration of Patent Examination at the European Patent Office. Management Science, 55(12), 1969-1984.

Hegde, D., & Luo, H. (2018). Patent Publication and the Market for Ideas. Management Science, 64(2), 652-672.

Hegde, D., Herkenhoff, K., & Zhu, C. (2019). Patent Publication and Innovation. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3158031.

Hall, B. H., & Harhoff, D. (2012). Recent Research on the Economics of Patents. Annual Review of Economics, 4(1), 541–565.

Jaffe, A. B. (1986). Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits and Market Value. American Economic Review, 76(5), 984-1001.

Jaffe, A. B. (1989). Characterizing the 'technological position' of firms, with application to quantifying technological opportunity and research spillovers. Research Policy, 18(2), 87–97.

Khashabi, P., & Mohammadi, A. (2016). Embracing the sharks: The impact of information exposure on the likelihood and quality of CVC investments. Academy of Management Proceedings, 428.

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

Kogan L., Papanikolaou, D., Schmidt, L. D. W., & Song, J. (2018). Technological Innovation and the Distribution of Labor Income Growth. Working Paper.

Lee, D. S. & Lemieux, T. 2010. Regression Discontinuity Designs in Economics. Journal of Economic Literature, 48(2), 281–355.

Lee, T. & Wilde, L. L. (1980). Market Structure and Innovation: A Reformulation. The Quarterly Journal of Economics, 94(2), 429-436.

Loury, G. C. (1979). Market Structure and Innovation. The Quarterly Journal of Economics, 93(3), 395-410.

Lu, Q., Myers, A. & Beliveau, S. (2017). USPTO Patent Prosecution Research Data: Unlocking Office Action Traits. USPTO Economic Working Paper No. 2017-10.

Muller, P. & J. Pénin (2006). Why do firms disclose knowledge and how does it matter? Journal of Evolutionary Economics, 16(1-2), 85-108.

Ouellette, L. L. (2012). Do Patents Disclose Useful Information? Harvard Journal of Law and Technology, 25(2), 545.

Ouellette, L. L. (2017). Who reads patents? Nature Biotechnology 35(5), 421–424.

Pénin, J. (2007). Open knowledge disclosure: An overview of the evidence and economic motivations. Journal of Economic Surveys, 21(2), 326-348.

Reinganum, J. R. (1985). Innovation and Industry Evolution. The Quarterly Journal of Economics, 100(1), 81-99.

Roin, B. N. (2005). The disclosure function of the patent system (or lack thereof). Harvard Law Review, 118, 2007-2028.

Sampat, B. & Williams, H. L. (2019). How Do Patents Affect Follow-On Innovation? Evidence from the Human Genome. American Economic Review, 109(1), 203-236.

Scotchmer, S. (1991). Standing on the Shoulders of Giants: Cumulative Research and the Patent Law. The Journal of Economic Perspectives, 5(1), 29-41.

Trajtenberg, M. (1991). A Penny for Your Quotes: Patent Citations and the Value of Innovations. The RAND Journal of Economics, 21(1), 172-187.

Thompson, N. & Kuhn, J. (2017). Does Winning a Patent Race lead to more follow-on Innovation? SSRN: http://dx.doi.org/10.2139/ssrn.2899088.

USPTO Office Action Research Dataset: https://www.uspto.gov/learning-and-resources/electronic-data-products/office-action-research-dataset-patents.

Wagner, S. & Wakeman, S. (2016). What do patent-based measures tell us about product commercialization? Evidence from the pharmaceutical industry. Research Policy, 45(5), 1091-1102.

Williams, H. (2017). How Do Patents Affect Research Investments? Annual Review of Economics, 9, 441-469.

Appendix A: Descriptive statistics and robustness checks

Table A1: Descriptive statistics – Blocked applications and blocked and abandoned applications

Blocked applications					
One month	Obs	Mean	SD	Min	Max
Number of blocked US applications	17382	0.82	1.89	0.00	135.00
Number of blocked US applications (in log)	17382	0.40	0.56	0.00	4.91
Number of blocked EP applications	17382	0.82	1.88	0.00	135.00
Number of blocked EP applications (in log)	17382	0.40	0.56	0.00	4.91
One year	Obs	Mean	SD	Min	Max
Number of blocked US applications	191220	0.79	1.63	0.00	135.00
Number of blocked US applications (in log)	191220	0.39	0.55	0.00	4.91
Number of blocked EP applications	191275	0.79	1.63	0.00	135.00
Number of blocked EP applications (in log)	191275	0.39	0.55	0.00	4.91
Blocked and abandoned applications					
One month	Obs	Mean	SD	Min	Max
Number of abandoned US applications	17382	0.25	0.95	0.00	91.00
Number of abandoned US applications (in log)	17382	0.15	0.34	0.00	4.52
Number of abandoned EP applications	17382	0.25	0.95	0.00	91.00
Number of abandoned EP applications (in log)	17382	0.14	0.34	0.00	4.52
One year	Obs	Mean	SD	Min	Max
Number of abandoned US applications	191220	0.24	0.71	0.00	91.00
Number of abandoned US applications (in log)	191220	0.14	0.33	0.00	4.52
Number of abandoned EP applications	191275	0.24	0.71	0.00	91.00
Number of abandoned EP applications (in log)	191275	0.14	0.33	0.00	4.52

Notes: This table reports descriptive statistics on all eventually granted patents filed with the USPTO within one month, respectively, and one year before and after AIPA became effective on November 29, 2000, which had no parallel foreign filing. The number of blocking actions is the number of times a given patent was referenced as a blocking patent (102 and 103) in the 4.4. Million office actions the USPTO issued between March 2008 and June 2017. The number of X and Y blockings is the number of times a given US patent was cited and classified as such by an EPO patent examiner in his or her official search report within 8 years since the referenced US patent was filed. Blocked abandoned applications refer to the number of US and European patent applications, respectively, which elicited a 102 or 103 office action (X and Y citation), and were eventually not granted.

Table A2: Patents with at least one 102 or 103 action

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: No. of blocked US applications (in logs)					
PANEL 1: Raw data, one month						
Post AIPA	-0.100***	-0.098***	-0.101***	-0.104***	-0.112***	-0.088***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.024)	(0.033)
N	17382	17382	17382	17382	11083	11083
R^2	0.010	0.010	0.021	0.024	0.214	0.214
PANEL 2: matched data, one month						
Post AIPA	-0.090***	-0.090***	-0.086***	-0.088***	-0.091***	-0.077**
	(0.019)	(0.019)	(0.019)	(0.019)	(0.027)	(0.038)
N	13490	13490	13490	13490	8280	8280
R^2	0.009	0.010	0.021	0.023	0.216	0.217
PANEL 3: raw data, one year						
Post AIPA	-0.102***	-0.102***	-0.104***	-0.104***	-0.108***	-0.105***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.009)
N	191220	191220	191220	191220	139111	139111
R^2	0.008	0.008	0.016	0.018	0.182	0.182
PANEL 4: matched data, one year						
Post AIPA	-0.102***	-0.102***	-0.102***	-0.103***	-0.106***	-0.106***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.009)
N	174518	174518	174518	174518	125398	125398
R^2	0.008	0.008	0.016	0.017	0.180	0.180
	Dependent variable: No. of blocked EP applications (in logs)					
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL 1: Raw data, one month						
Post AIPA	-0.102***	-0.100***	-0.102***	-0.106***	-0.113***	-0.089***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.024)	(0.033)
N	17382	17382	17382	17382	11080	11080
R^2	0.010	0.010	0.022	0.024	0.214	0.215
PANEL 2: matched data, one month						
Post AIPA	-0.092***	-0.092***	-0.088***	-0.090***	-0.093***	-0.078**
	(0.019)	(0.019)	(0.019)	(0.019)	(0.027)	(0.038)
N	13481	13481	13481	13481	8277	8277
R^2	0.009	0.010	0.021	0.023	0.216	0.217
PANEL 3: raw data, one year						
Post AIPA	-0.103***	-0.103***	-0.104***	-0.105***	-0.109***	-0.107***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.009)
N	191275	191275	191275	191275	139133	139133
R^2	0.008	0.008	0.016	0.018	0.182	0.182
PANEL 4: matched data, one year						
Post AIPA	-0.102***	-0.102***	-0.102***	-0.103***	-0.106***	-0.106***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.009)
N	174461	174461	174461	174461	125356	125356
R^2	0.008	0.008	0.016	0.017	0.180	0.180
All PANEL	(1)	(2)	(3)	(4)	(5)	(6)
RDD time controls	linear (same slope)	linear (diff. slopes)	linear (diff. slopes)	linear (diff. slopes)	linear (diff. slopes)	quadratic (diff. slopes)
Technology class fixed effects	no	no	yes	yes	yes	yes
Technology trends	no	no	no	yes	yes	yes
Firm fixed effects	no	no	no	no	yes	yes

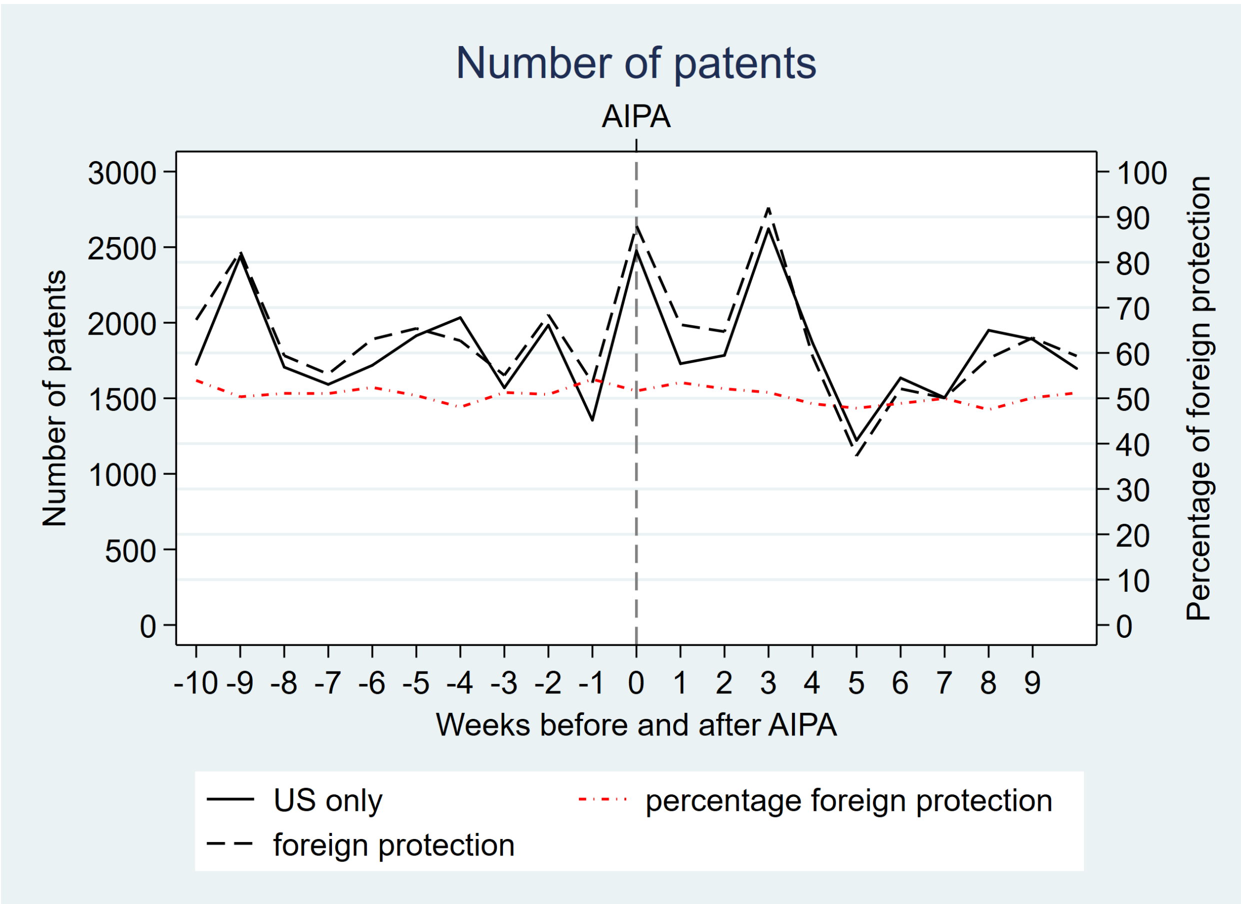
Notes: This table reports results of RDD models, equations (1) and (2). Technology class fixed effects are based on six NBER technology classes, and technology trends are linear time trends per technology class using filing days. Raw data are all US patent applications filed between one month before one month after November 29, 2000, when AIPA became effective. Post AIPA is an indicator that equals one if the patent was filed on or after November 29, 2000. Matched data is a balanced sample, where patents before and after AIPA are matched (CEM) based on backward cites to capture potential differences in novelty, 6 NBER technology classes to capture differences in technological popularity, number of independent claims, dependent claims per independent claim, word count of the first claim, average number of words per claim, whether an attorney was involved, and patent pendency (time from application to grant date) to capture differences in patent scope and writing. Appendix A Table A5 shows the descriptive statistics of the matched data. Heteroscedasticity robust standard errors are presented in parentheses.

Table A3: Blocked applications (at least one 102 or 103 and eventually abandoned)

	Dependent variable: No. of blocked abandoned US applications (in logs)					
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL 1: Raw data, one month						
Post AIPA	-0.050***	-0.049***	-0.048***	-0.049***	-0.039***	-0.020
	(0.010)	(0.010)	(0.010)	(0.010)	(0.014)	(0.020)
N	17382	17382	17382	17382	11083	11083
R^2	0.006	0.006	0.013	0.014	0.187	0.187
PANEL 2: matched data, one month						
Post AIPA	-0.041***	-0.041***	-0.040***	-0.040***	-0.030*	-0.011
	(0.011)	(0.011)	(0.011)	(0.011)	(0.016)	(0.023)
N	13490	13490	13490	13490	8280	8280
R^2	0.006	0.006	0.013	0.014	0.185	0.186
PANEL 3: raw data, one year						
Post AIPA	-0.044***	-0.044***	-0.044***	-0.044***	-0.043***	-0.044***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)
N	191220	191220	191220	191220	139111	139111
R^2	0.004	0.004	0.010	0.010	0.149	0.149
PANEL 4: matched data, one year						
Post AIPA	-0.042***	-0.042***	-0.042***	-0.042***	-0.042***	-0.044***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.006)
N	174518	174518	174518	174518	125398	125398
R^2	0.004	0.004	0.010	0.011	0.150	0.150
	Dependent variable: No. of blocked abandoned EP applications (in logs)					
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL 1: Raw data, one month						
Post AIPA	-0.051***	-0.049***	-0.048***	-0.049***	-0.039***	-0.019
	(0.010)	(0.010)	(0.010)	(0.010)	(0.014)	(0.020)
N	17382	17382	17382	17382	11080	11080
R^2	0.006	0.006	0.013	0.014	0.187	0.187
PANEL 2: matched data, one month						
Post AIPA	-0.042***	-0.042***	-0.041***	-0.041***	-0.031*	-0.011
	(0.011)	(0.011)	(0.011)	(0.011)	(0.016)	(0.023)
N	13481	13481	13481	13481	8277	8277
R^2	0.006	0.006	0.013	0.014	0.185	0.185
PANEL 3: raw data, one year						
Post AIPA	-0.044***	-0.044***	-0.044***	-0.044***	-0.043***	-0.045***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)
N	191275	191275	191275	191275	139133	139133
R^2	0.004	0.004	0.010	0.011	0.149	0.149
PANEL 4: matched data, one year						
Post AIPA	-0.042***	-0.042***	-0.042***	-0.043***	-0.042***	-0.044***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.006)
N	174461	174461	174461	174461	125356	125356
R^2	0.004	0.004	0.010	0.011	0.150	0.150
All PANEL	(1)	(2)	(3)	(4)	(5)	(6)
RDD time controls	linear (same slope)	linear (diff. slopes)	linear (diff. slopes)	linear (diff. slopes)	linear (diff. slopes)	quadratic (diff. slopes)
Technology class fixed effects	no	no	yes	yes	yes	yes
Technology trends	no	no	no	yes	yes	yes
Firm fixed effects	no	no	no	no	yes	yes

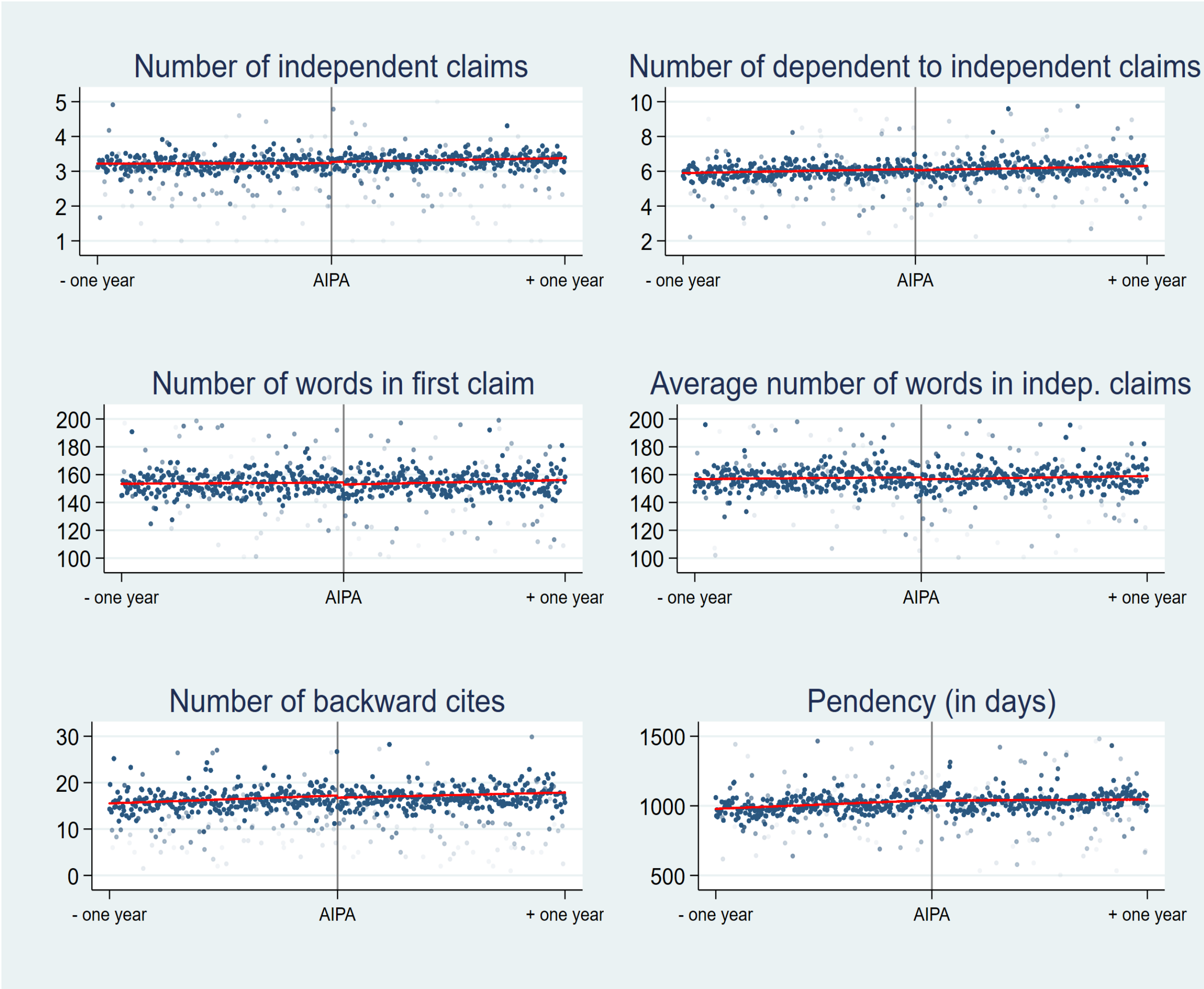
Notes: This table reports results of RDD models, equation (1) and (2). Technology class fixed effects are based on six NBER technology classes, and technology trends are linear time trends per technology class using filing days. Raw data are all US patent applications filed between one month before one month after November 29, 2000, when AIPA became effective. Post AIPA is an indicator that equals one if the patent was filed on or after November 29, 2000. Matched data is a balanced sample, where patents before and after AIPA are matched (CEM) based on backward cites to capture potential differences in novelty, 6 NBER technology classes to capture differences in technological popularity, number of independent claims, dependent claims per independent claim, word count of the first claim, average number of words per claim, whether an attorney was involved, and patent pendency (time from application to grant date) to capture differences in patent scope and writing. Appendix A Table A5 shows the descriptive statistics of the matched data. Heteroscedasticity robust standard errors are presented in parentheses.

Figure A1: Illustration of little strategic change in applications pre and post AIPA



Notes: This figure illustrates the number of patents filed with the USPTO closely around AIPA with and without parallel foreign protection. AIPA became effective on November 29, 2000.

Figure A2: Observable patent characteristics before and after AIPA.



Notes: The characteristics of patents filed with the USPTO closely around AIPA do not appear to change much within one month, respectively, one year before and after AIPA became effective on November 29, 2000.

Table A4: RDD one year window

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: No. of 102 & 103 blocking actions (in logs)						
PANEL 3: raw data, one year						
Post AIPA	-0.135***	-0.135***	-0.137***	-0.138***	-0.146***	-0.141***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.013)
N	191220	191220	191220	191220	139111	139111
R^2	0.007	0.007	0.016	0.017	0.177	0.177
PANEL 4: matched data, one year						
Post AIPA	-0.136***	-0.136***	-0.136***	-0.137***	-0.144***	-0.143***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)	(0.013)
N	174518	174518	174518	174518	125398	125398
R^2	0.007	0.007	0.016	0.017	0.176	0.176
Dependent variable: No. of X & Y blocking actions (in logs)						
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL 3: raw data, one year						
Post AIPA	-0.095***	-0.095***	-0.094***	-0.094***	-0.096***	-0.097***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.006)
N	191275	191275	191275	191275	139133	139133
R^2	0.023	0.023	0.030	0.031	0.161	0.161
PANEL 4: matched data, one year						
Post AIPA	-0.094***	-0.094***	-0.094***	-0.094***	-0.094***	-0.096***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.006)
N	174461	174461	174461	174461	125356	125356
R^2	0.023	0.023	0.030	0.030	0.162	0.162
Dependent variable: Fraction of X & Y blocking actions						
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL 3: raw data, one year						
Post AIPA	-0.049***	-0.049***	-0.049***	-0.049***	-0.049***	-0.048***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
N	191275	191275	191275	191275	139133	139133
R^2	0.023	0.023	0.027	0.027	0.139	0.139
PANEL 4: matched data, one year						
Post AIPA	-0.049***	-0.049***	-0.049***	-0.049***	-0.049***	-0.047***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
N	174461	174461	174461	174461	125356	125356
R^2	0.023	0.023	0.026	0.026	0.139	0.140
All PANEL	(1)	(2)	(3)	(4)	(5)	(6)
RDD time controls	linear (same slope)	linear (diff. slopes)	linear (diff. slopes)	linear (diff. slopes)	linear (diff. slopes)	quadratic (diff. slopes)
Technology class fixed effects	no	no	yes	yes	yes	yes
Technology trends	no	no	no	yes	yes	yes
Firm fixed effects	no	no	no	no	yes	yes

Notes: This table reports results of RDD models, equation (1) and (2). Technology class fixed effects are based on six NBER technology classes, and technology trends are linear time trends per technology class using filing days. Raw data are all US patent applications filed between one year before one year after November 29, 2000, when AIPA became effective. Post AIPA is an indicator that equals one if the patent was filed on or after November 29, 2000. Matched data is a balanced sample, where patents before and after AIPA are matched (CEM) based on backward cites to capture potential differences in novelty, 6 NBER technology classes to capture differences in technological popularity, number of independent claims, dependent claims per independent claim, word count of the first claim, average number of words per claim, whether an attorney was involved, and patent pendency (time from application to grant date) to capture differences in patent scope and writing. Appendix A Table A5 shows the descriptive statistics of the matched data. Heteroscedasticity robust standard errors are presented in parentheses.

Table A5: Descriptive statistics – matched data

One month	Pre AIPA					Post AIPA					
	Obs	Mean	SD	Min	Max	Obs	Mean	SD	Min	Max	t-test
Number of 102 & 103 blockings	6745	1.82	4.48	0.00	187.00	6745	1.23	2.68	0.00	36.00	9.383
Number of 102 & 103 blockings (in log)	6745	0.60	0.81	0.00	5.24	6745	0.46	0.71	0.00	3.61	10.668
Number of X & Y blockings	6745	0.34	1.06	0.00	21.00	6745	0.13	0.60	0.00	11.00	14.159
Number of X & Y blockings (in log)	6745	0.17	0.42	0.00	3.09	6745	0.07	0.27	0.00	2.48	16.506
Fraction of X & Y blockings	6745	0.08	0.21	0.00	1.00	6745	0.04	0.15	0.00	1.00	15.562
NBER											
Chemical	6745	0.11	0.32	0.00	1.00	6745	0.11	0.32	0.00	1.00	0.000
Computers and Communications	6745	0.29	0.45	0.00	1.00	6745	0.29	0.45	0.00	1.00	0.000
Drugs and Medical	6745	0.13	0.34	0.00	1.00	6745	0.13	0.34	0.00	1.00	0.000
Electrical and Electronics	6745	0.18	0.39	0.00	1.00	6745	0.18	0.39	0.00	1.00	0.000
Mechanical	6745	0.13	0.34	0.00	1.00	6745	0.13	0.34	0.00	1.00	0.000
Others	6745	0.15	0.36	0.00	1.00	6745	0.15	0.36	0.00	1.00	0.000
Number of independent claims	6745	2.97	2.10	1.00	31.00	6745	2.99	2.09	1.00	29.00	-0.749
Number of dependent claims to independent claims	6745	5.98	4.86	0.00	60.00	6745	5.85	4.73	0.00	47.00	1.594
Number of words in first claim	6745	141.13	78.41	3.00	811.00	6745	140.53	78.83	5.00	875.00	0.440
Average number of words in independent claims	6745	145.23	77.08	4.67	811.00	6745	144.80	77.11	4.00	895.50	0.327
Backward cites	6745	13.26	14.48	0.00	197.00	6745	12.99	14.38	0.00	181.00	1.079
Pendency	6745	1008.14	584.61	193.00	4054.00	6745	1005.79	585.00	182.00	4008.00	0.233
Attorney (yes / no)	6745	0.96	0.19	0.00	1.00	6745	0.96	0.19	0.00	1.00	0.000
One year	**Pre AIPA**					**Post AIPA**					
	Obs	Mean	SD	Min	Max	Obs	Mean	SD	Min	Max	t-test
Number of 102 & 103 blockings	87259	1.75	3.97	0.00	189.00	87259	1.23	2.78	0.00	115.00	23.011
Number of 102 & 103 blockings (in log)	87259	0.59	0.80	0.00	5.25	87259	0.46	0.71	0.00	4.75	34.869
Number of X & Y blockings	87259	0.35	1.10	0.00	48.00	87259	0.13	0.68	0.00	36.00	49.717
Number of X & Y blockings (in log)	87259	0.17	0.42	0.00	3.89	87259	0.06	0.26	0.00	3.61	63.393
Fraction of X & Y blockings	87259	0.09	0.21	0.00	1.00	87259	0.03	0.14	0.00	1.00	63.517
NBER											
Chemical	87259	0.11	0.32	0.00	1.00	87259	0.11	0.32	0.00	1.00	0.000
Computers and Communications	87259	0.27	0.45	0.00	1.00	87259	0.27	0.45	0.00	1.00	0.000
Drugs and Medical	87259	0.13	0.33	0.00	1.00	87259	0.13	0.33	0.00	1.00	0.000
Electrical and Electronics	87259	0.19	0.39	0.00	1.00	87259	0.19	0.39	0.00	1.00	0.000
Mechanical	87259	0.14	0.35	0.00	1.00	87259	0.14	0.35	0.00	1.00	0.000
Others	87259	0.16	0.37	0.00	1.00	87259	0.16	0.37	0.00	1.00	0.000
Number of independent claims	87259	3.18	2.48	0.00	60.00	87259	3.22	2.48	0.00	60.00	-4.099
Number of dependent claims to independent claims	87253	5.89	5.12	0.00	102.00	87253	5.98	5.12	0.00	108.00	-3.534
Number of words in first claim	87259	150.49	95.02	1.00	2533.00	87259	150.48	94.87	1.00	2544.00	0.035
Average number of words in independent claims	87253	154.26	90.78	1.00	2533.00	87253	154.13	90.97	1.00	2692.50	0.292
Backward cites	87259	15.03	21.20	0.00	683.00	87259	15.19	21.22	0.00	683.00	-1.661
Pendency	87259	989.67	584.06	148.00	4089.00	87259	989.85	584.43	95.00	4045.00	-0.066
Attorney (yes / no)	87259	0.95	0.21	0.00	1.00	87259	0.95	0.21	0.00	1.00	0.000

Table A6: Only 102 actions

	Dependent variable: No. of 102 blocking actions (in logs)					
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL 1: Raw data, one month						
Post AIPA	-0.050***	-0.049***	-0.048***	-0.048***	-0.050***	-0.028
	(0.011)	(0.011)	(0.011)	(0.011)	(0.016)	(0.021)
N	17382	17382	17382	17382	11083	11083
R^2	0.003	0.003	0.006	0.006	0.166	0.167
PANEL 2: matched data, one month						
Post AIPA	-0.051***	-0.051***	-0.049***	-0.049***	-0.057***	-0.038*
	(0.012)	(0.012)	(0.012)	(0.012)	(0.017)	(0.023)
N	13490	13490	13490	13490	8280	8280
R^2	0.003	0.003	0.006	0.007	0.175	0.175
PANEL 3: Raw data, one year						
Post AIPA	-0.043***	-0.043***	-0.043***	-0.043***	-0.040***	-0.031***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.006)
N	191220	191220	191220	191220	139111	139111
R^2	0.003	0.003	0.006	0.006	0.128	0.128
PANEL 4: matched data, one year						
Post AIPA	-0.042***	-0.042***	-0.042***	-0.042***	-0.039***	-0.031***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.006)
N	174518	174518	174518	174518	125398	125398
R^2	0.002	0.002	0.006	0.006	0.131	0.131
All PANEL	(1)	(2)	(3)	(4)	(5)	(6)
RDD time controls	linear (same slope)	linear (diff. slopes)	linear (diff. slopes)	linear (diff. slopes)	linear (diff. slopes)	quadratic (diff. slopes)
Technology class fixed effects	no	no	yes	yes	yes	yes
Technology trends	no	no	no	yes	yes	yes
Firm fixed effects	no	no	no	no	yes	yes

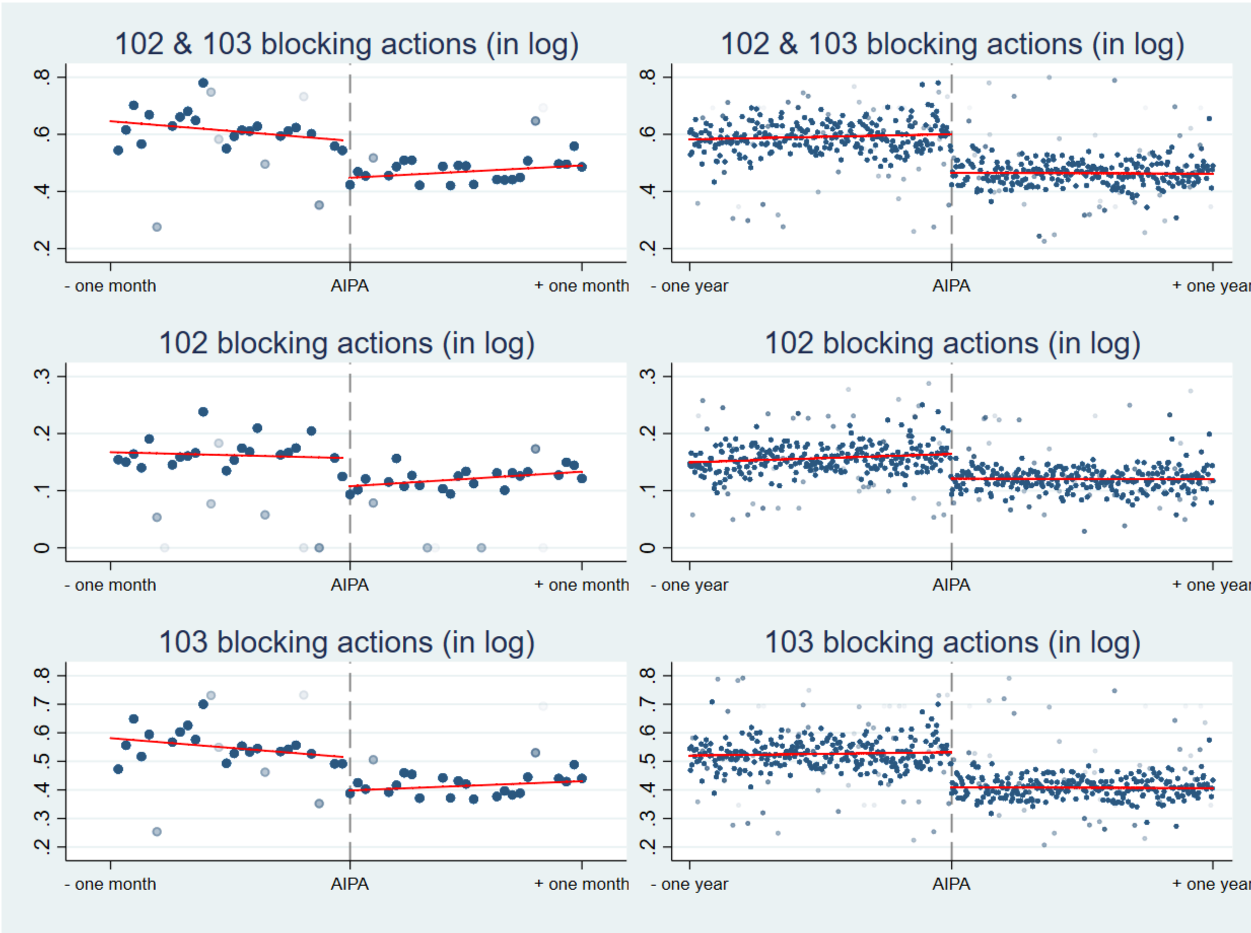
Notes: This table reports results of RDD models, equation (1) and (2). Technology class fixed effects are based on six NBER technology classes, and technology trends are linear time trends per technology class using filing days. Raw data are all US patent applications filed between one month before one month after November 29, 2000, when AIPA became effective. Post AIPA is an indicator that equals one if the patent was filed on or after November 29, 2000. Matched data is a balanced sample, where patents before and after AIPA are matched (CEM) based on backward cites to capture potential differences in novelty, 6 NBER technology classes to capture differences in technological popularity, number of independent claims, dependent claims per independent claim, word count of the first claim, average number of words per claim, whether an attorney was involved, and patent pendency (time from application to grant date) to capture differences in patent scope and writing. Heteroscedasticity robust standard errors are presented in parentheses.

Table A7: Only 103 actions

	Dependent variable: No. of 103 blocking actions (in logs)					
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL 1: Raw data, one month						
Post AIPA	-0.118***	-0.115***	-0.119***	-0.124***	-0.128***	-0.110**
	(0.021)	(0.022)	(0.021)	(0.021)	(0.031)	(0.043)
N	17382	17382	17382	17382	11083	11083
R^2	0.008	0.009	0.023	0.026	0.215	0.215
PANEL 2: matched data, one month						
Post AIPA	-0.105***	-0.104***	-0.098***	-0.102***	-0.091***	-0.088*
	(0.024)	(0.024)	(0.024)	(0.024)	(0.035)	(0.048)
N	13490	13490	13490	13490	8280	8280
R^2	0.009	0.009	0.024	0.026	0.218	0.218
PANEL 3: Raw data, one year						
Post AIPA	-0.123***	-0.123***	-0.125***	-0.126***	-0.136***	-0.136***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.008)	(0.012)
N	191220	191220	191220	191220	139111	139111
R^2	0.007	0.007	0.018	0.019	0.179	0.179
PANEL 4: matched data, one year						
Post AIPA	-0.124***	-0.124***	-0.125***	-0.125***	-0.134***	-0.138***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.012)
N	174518	174518	174518	174518	125398	125398
R^2	0.007	0.007	0.017	0.019	0.178	0.178
All PANEL	(1)	(2)	(3)	(4)	(5)	(6)
RDD time controls	linear (same slope)	linear (diff. slopes)	linear (diff. slopes)	linear (diff. slopes)	linear (diff. slopes)	quadratic (diff. slopes)
Technology class fixed effects	no	no	yes	yes	yes	yes
Technology trends	no	no	no	yes	yes	yes
Firm fixed effects	no	no	no	no	yes	yes

Notes: This table reports results of RDD models, equation (1) and (2). Technology class fixed effects are based on six NBER technology classes, and technology trends are linear time trends per technology class using filing days. Raw data are all US patent applications filed between one month before one month after November 29, 2000, when AIPA became effective. Post AIPA is an indicator that equals one if the patent was filed on or after November 29, 2000. Matched data is a balanced sample, where patents before and after AIPA are matched (CEM) based on backward cites to capture potential differences in novelty, 6 NBER technology classes to capture differences in technological popularity, number of independent claims, dependent claims per independent claim, word count of the first claim, average number of words per claim, whether an attorney was involved, and patent pendency (time from application to grant date) to capture differences in patent scope and writing. Heteroscedasticity robust standard errors are presented in parentheses.

Figure A3: RDD Graphs for 102 and 103 actions combined and separately



Notes: This illustrates the discontinuous difference in blocking actions one month (year) before and after AIPA became effective on November 29, 2000. Dots represent the average amount of 102 and 103 office actions (first combined, then separately 102 and 103 only) referring to patents filed at the USPTO on a given day. AIPA represents November 29, 2000. The red lines represent fitted values.

Table A8: Fraction 102 and 103 in future cites

	Dependent variable: Fraction of 102 and 103 blocking actions to future cites					
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL 1: Raw data, one month						
Post AIPA	-0.023**	-0.022*	-0.021*	-0.021*	-0.011	-0.017
	(0.012)	(0.012)	(0.012)	(0.012)	(0.017)	(0.026)
N	16050	16050	16050	16050	10119	10119
R^2	0.002	0.002	0.006	0.007	0.138	0.138
PANEL 2: matched data, one month						
Post AIPA	-0.017	-0.017	-0.019	-0.018	-0.007	-0.020
	(0.014)	(0.014)	(0.014)	(0.014)	(0.022)	(0.032)
N	12406	12406	12406	12406	7520	7520
R^2	0.002	0.002	0.006	0.007	0.147	0.147
PANEL 3: Raw data, one year						
Post AIPA	-0.029***	-0.029***	-0.029***	-0.029***	-0.029***	-0.026***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.007)
N	176426	176426	176426	176426	127595	127595
R^2	0.000	0.000	0.004	0.004	0.104	0.104
PANEL 4: matched data, one year						
Post AIPA	-0.026***	-0.026***	-0.027***	-0.027***	-0.029***	-0.027***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.007)
N	161105	161105	161105	161105	115017	115017
R^2	0.000	0.000	0.005	0.005	0.107	0.107
All PANEL	(1)	(2)	(3)	(4)	(5)	(6)
RDD time controls	linear (same slope)	linear (diff. slopes)	linear (diff. slopes)	linear (diff. slopes)	linear (diff. slopes)	quadratic (diff. slopes)
Technology class fixed effects	no	no	yes	yes	yes	yes
Technology trends	no	no	no	yes	yes	yes
Firm fixed effects	no	no	no	no	yes	yes

Notes: This table reports results of RDD models, equation (1) and (2). The dependent variable is the number of 102 and 103 office actions referencing the focal patent as the blocking patent divided by the total number of future cites that the focal patent received from all USPTO patents granted until end of 2017. Technology class fixed effects are based on six NBER technology classes, and technology trends are linear time trends per technology class using filing days. Raw data are all US patent applications filed between one month before one month after November 29, 2000, when AIPA became effective. Post AIPA is an indicator that equals one if the patent was filed on or after November 29, 2000. Matched data is a balanced sample, where patents before and after AIPA are matched (CEM) based on backward cites to capture potential differences in novelty, 6 NBER technology classes to capture differences in technological popularity, number of independent claims, dependent claims per independent claim, word count of the first claim, average number of words per claim, whether an attorney was involved, and patent pendency (time from application to grant date) to capture differences in patent scope and writing. Heteroscedasticity robust standard errors are presented in parentheses.

Table A9: Only fraction 102 in future cites

	Dependent variable: Fraction of 102 blocking actions to future cites					
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL 1: Raw data, one month						
Post AIPA	-0.004	-0.004	-0.004	-0.003	-0.003	0.001
	(0.004)	(0.004)	(0.004)	(0.004)	(0.006)	(0.008)
N	16050	16050	16050	16050	10119	10119
R^2	0.000	0.000	0.004	0.004	0.143	0.143
PANEL 2: matched data, one month						
Post AIPA	-0.006	-0.006	-0.006	-0.006	-0.008	-0.008
	(0.004)	(0.004)	(0.004)	(0.004)	(0.006)	(0.009)
N	12406	12406	12406	12406	7520	7520
R^2	0.000	0.000	0.004	0.005	0.156	0.156
PANEL 3: Raw data, one year						
Post AIPA	-0.005***	-0.005***	-0.005***	-0.005***	-0.004***	-0.003
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
N	176426	176426	176426	176426	127595	127595
R^2	0.000	0.000	0.004	0.004	0.101	0.101
PANEL 4: matched data, one year						
Post AIPA	-0.005***	-0.005***	-0.005***	-0.005***	-0.004***	-0.003
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
N	161105	161105	161105	161105	115017	115017
R^2	0.000	0.000	0.004	0.004	0.105	0.105
All PANEL	(1)	(2)	(3)	(4)	(5)	(6)
RDD time controls	linear (same slope)	linear (diff. slopes)	linear (diff. slopes)	linear (diff. slopes)	linear (diff. slopes)	quadratic (diff. slopes)
Technology class fixed effects	no	no	yes	yes	yes	yes
Technology trends	no	no	no	yes	yes	yes
Firm fixed effects	no	no	no	no	yes	yes

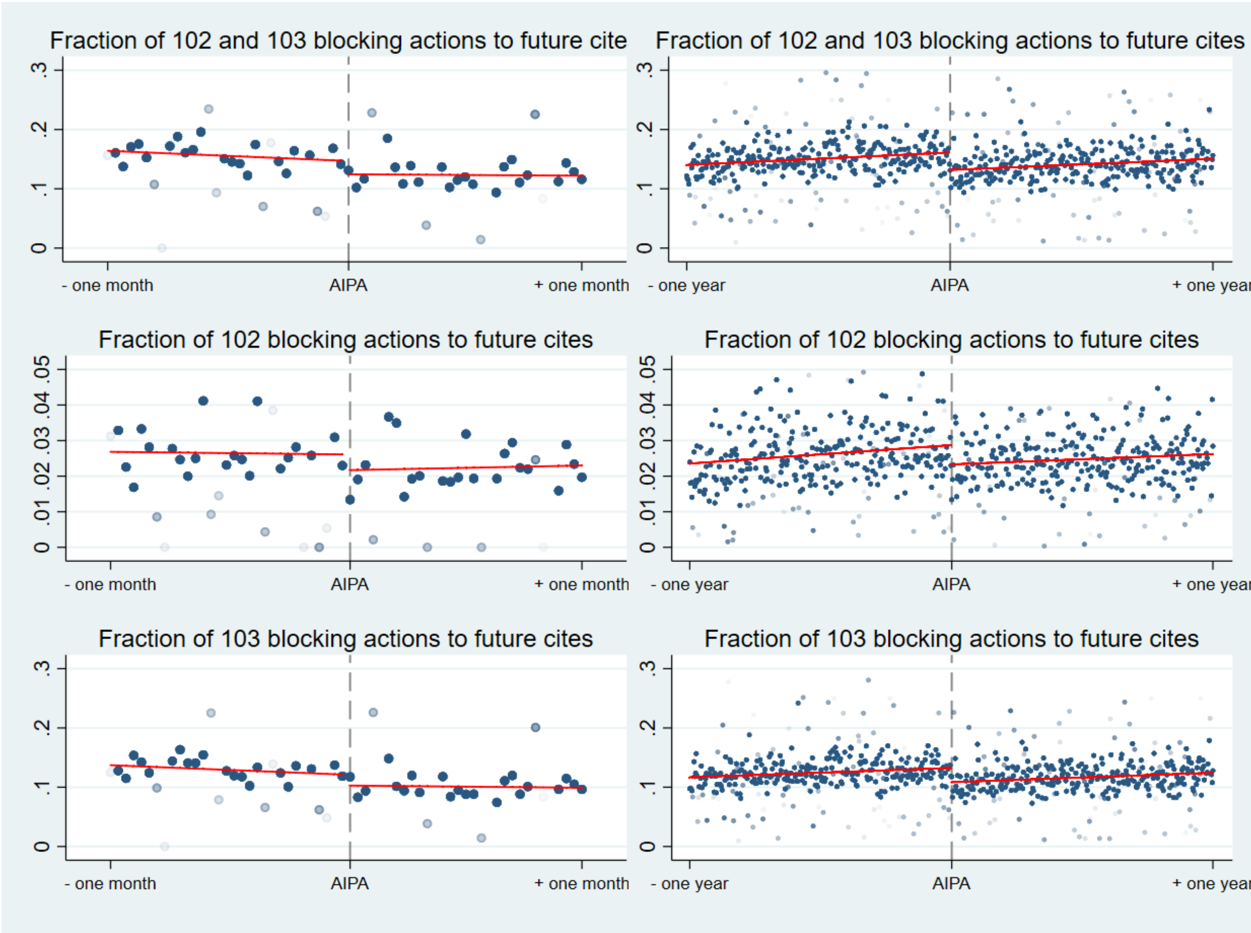
Notes: This table reports results of RDD models, equation (1) and (2). The dependent variable is the number of 102 office actions referencing the focal patent as the blocking patent divided by the total number of future cites that the focal patent received from all USPTO patents granted until end of 2017. Technology class fixed effects are based on six NBER technology classes, and technology trends are linear time trends per technology class using filing days. Raw data are all US patent applications filed between one month before one month after November 29, 2000, when AIPA became effective. Post AIPA is an indicator that equals one if the patent was filed on or after November 29, 2000. Matched data is a balanced sample, where patents before and after AIPA are matched (CEM) based on backward cites to capture potential differences in novelty, 6 NBER technology classes to capture differences in technological popularity, number of independent claims, dependent claims per independent claim, word count of the first claim, average number of words per claim, whether an attorney was involved, and patent pendency (time from application to grant date) to capture differences in patent scope and writing. Heteroscedasticity robust standard errors are presented in parentheses.

Table A10: Only fraction 103 in future cites

	Dependent variable: Fraction of 103 blocking actions to future cites					
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL 1: Raw data, one month						
Post AIPA	-0.018*	-0.018*	-0.017*	-0.017*	-0.008	-0.018
	(0.010)	(0.010)	(0.010)	(0.010)	(0.015)	(0.022)
N	16050	16050	16050	16050	10119	10119
R^2	0.002	0.002	0.005	0.006	0.140	0.140
PANEL 2: matched data, one month						
Post AIPA	-0.012	-0.012	-0.013	-0.012	0.001	-0.012
	(0.012)	(0.012)	(0.012)	(0.012)	(0.019)	(0.028)
N	12406	12406	12406	12406	7520	7520
R^2	0.002	0.002	0.005	0.006	0.143	0.143
PANEL 3: Raw data, one year						
Post AIPA	-0.024***	-0.024***	-0.024***	-0.024***	-0.025***	-0.023***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.006)
N	176426	176426	176426	176426	127595	127595
R^2	0.000	0.000	0.003	0.003	0.104	0.104
PANEL 4: matched data, one year						
Post AIPA	-0.021***	-0.022***	-0.022***	-0.022***	-0.025***	-0.023***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.006)
N	161105	161105	161105	161105	115017	115017
R^2	0.000	0.000	0.004	0.004	0.105	0.105
All PANEL	(1)	(2)	(3)	(4)	(5)	(6)
RDD time controls	linear (same slope)	linear (diff. slopes)	linear (diff. slopes)	linear (diff. slopes)	linear (diff. slopes)	quadratic (diff. slopes)
Technology class fixed effects	no	no	yes	yes	yes	yes
Technology trends	no	no	no	yes	yes	yes
Firm fixed effects	no	no	no	no	yes	yes

Notes: This table reports results of RDD models, equation (1) and (2). The dependent variable is the number of 103 office actions referencing the focal patent as the blocking patent divided by the total number of future cites that the focal patent received from all USPTO patents granted until end of 2017. Technology class fixed effects are based on six NBER technology classes, and technology trends are linear time trends per technology class using filing days. Raw data are all US patent applications filed between one month before one month after November 29, 2000, when AIPA became effective. Post AIPA is an indicator that equals one if the patent was filed on or after November 29, 2000. Matched data is a balanced sample, where patents before and after AIPA are matched (CEM) based on backward cites to capture potential differences in novelty, 6 NBER technology classes to capture differences in technological popularity, number of independent claims, dependent claims per independent claim, word count of the first claim, average number of words per claim, whether an attorney was involved, and patent pendency (time from application to grant date) to capture differences in patent scope and writing. Heteroscedasticity robust standard errors are presented in parentheses.

Figure A4: RDD Graphs for fraction of 102 and 103 actions in future cites combined and separately



Notes: This illustrates the discontinuous difference in blocking actions one month (year) before and after AIPA became effective on November 29, 2000. Dots represent the average amount of 102 and 103 office actions (first combined, then separately 102 and 103 only) referring to patents filed at the USPTO on a given day scaled by the total amount of future cites. AIPA represents November 29, 2000. The red lines represent fitted values.

Appendix B: Twin study (TW) details

For the twin study approach, we select all US patents with a parallel foreign filing at the EPO. Parallel foreign filings are identified via the 'DOCDB' family identifier available in Patstat. "A patent family is a collection of patent documents that are considered to cover a single invention. The technical content covered by the applications is considered to be identical. Members of a simple patent family will all have exactly the same priorities." (For further details see https://www.epo.org/searching-for-patents/helpful-resources/first-time-here/patent-families/docdb.html).

We then collect all X and Y citations to each USPTO and EPO patent within a family (typically just 2, all results robust to restriction to just the two patent member case), where X and Y classifications come from EPO patent examiners as reported in their official publicly available EPO search report. We took all US patents into account that were filed 5 years before and 5 years after AIPA (EPO vs. USPTO application, before and after AIPA; restricting the sample to shorter time windows produces similar coefficients, coefficients are statistically significant but smaller in magnitude if the data is restricted to a one year window before and after AIPA). European parallel filings could happen earlier or later than the US application. We then estimate the following equation using OLS:

$$\log(blockings + 1)_i = \beta_1 \cdot PostAIPA_i + \beta_2 \cdot US_i + \beta_3 \cdot Post_i + \delta_i + \varepsilon_i \qquad \text{(A1)}$$

where $blockings_i$, refers either to the number of times an EPO examiner references patent i with an X and Y in his search report, or the fraction of X and Y cites out of all references that patent i receives from EPO search reports, respectively. $PostAIPA_i$ is an indicator indicating whether patent i was filed in the US on or after November 29, 2000, US_i is an indicator indicating a US patent, $Post_i$ indicates all patents filed after AIPA, δ_i represents family fixed effects, and ε_i is the error term. Assuming models assumptions are met, β_1 should capture the causal influence of AIPA on the number of blocked claims of future patent applications. Alternatively, we estimate the same model as (1) but add year fixed effects (2) and technology class specific trends (3). Table A11 presents the results.

Table A11: Twin match estimates

	X and Y (in log)			Fraction X and Y		
	(1)	(2)	(3)	(4)	(5)	(6)
Post AIPA	-0.106***	-0.103***	-0.102***	-0.064***	-0.061***	-0.061***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
N	781766	781767	781768	781769	781770	781771
R^2	0.431	0.432	0.437	0.434	0.435	0.439
Family fixed effects	yes	yes	yes	yes	yes	yes
Year fixed effects	no	Yes	yes	no	yes	yes
Technology trends	no	no	yes	no	no	yes

Appendix C: Difference in differences (DiD) details

For the DiD approach, we build on Graham and Hegde (2015) and define the treated group to be patents that have no parallel foreign filing and with the control group as all patents with at least one parallel foreign filing with the same or earlier filing date than the US counterpart. We took all US patents into account that were filed 5 years before and 5 years after AIPA. We then estimate the following equation using OLS:

$$\log(blockings + 1)_i = \beta_1 \cdot PostAIPA_i + \beta_2 \cdot US_i + \beta_3 \cdot Post_i + \varepsilon_i \qquad \text{(A2)}$$

where $blockings_i$, refers either to the number of 102 and 103 blocking actions patent i receives, the times an EPO examiner references patent i with an X and Y in his or her search report, or the fraction of X and Y cites out of all references that patent i receives from EPO search reports, respectively. $PostAIPA_i$ is an indicator indicating whether patent i was filed in the US on or after November 29, 2000 and has no parallel foreign application, US_i is an indicator indicating a US patent, $Post_i$ indicates all patents filed on or after AIPA, and ε_i is the error term. Alternatively, we estimate the same model as (1) then add year fixed effects, (2) NBER technology class fixed effects (3), technology class specific trends (4), and firm fixed effects (5). We use all three previously used dependent variables, including the number of blocked US patent applications (102 and 103 references), number of X and Y cites, and fraction of X and Y cites out of all cites. Table A12 presents the results.

Figure A5: Visual test of parallel trends assumption

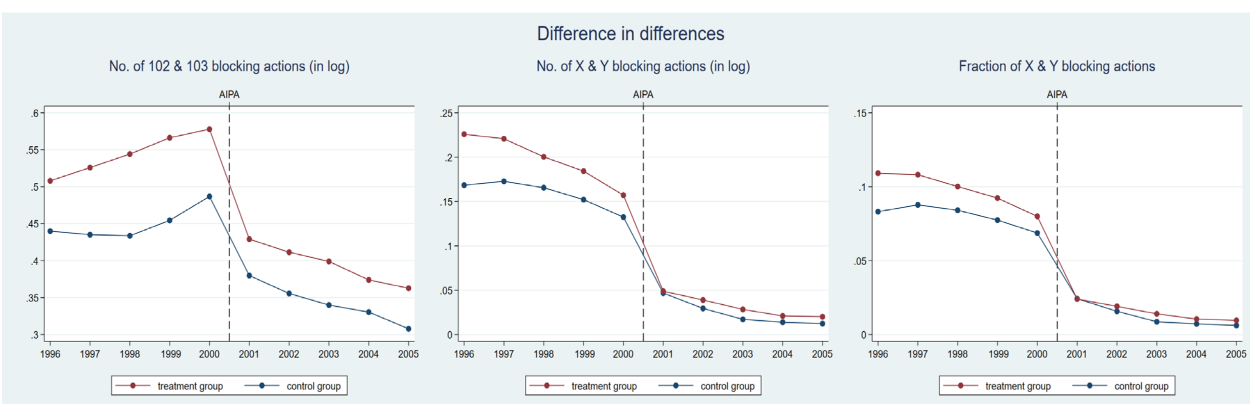


Table A12: Difference-in-differences estimates

Dependent variable: No. of 102 & 103 blocking actions (in logs)					
	(1)	(2)	(3)	(4)	(5)
Post AIPA	-0.039***	-0.046***	-0.044***	-0.038***	-0.045***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
US	0.093	0.099	0.091	0.089	-0.009***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Post	-0.113	-0.118***	-0.119***	-0.123***	-0.124***
	(0.002)	(0.006)	(0.006)	(0.006)	(0.007)
N	1351834	1351834	1351834	1351834	1099165
R^2	0.013	0.014	0.017	0.019	0.123
Dependent variable: No of X & Y blocking actions (in logs)					
	(1)	(2)	(3)	(4)	(5)
Post AIPA	-0.033***	-0.029***	-0.030***	-0.029***	-0.041***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
US	0.041***	0.036***	0.035***	0.035***	0.039***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Post	-0.134	-0.090	-0.089	-0.090	-0.087***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
N	1351822	1351822	1351822	1351822	1099155
R^2	0.057	0.059	0.062	0.063	0.147
Dependent variable : Fraction of X & Y blocking actions					
	(1)	(2)	(3)	(4)	(5)
Post AIPA	-0.015***	-0.013***	-0.014***	-0.014***	-0.018***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
US	0.018***	0.016***	0.016***	0.016***	0.019***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Post	-0.068	-0.046***	-0.045***	-0.045***	-0.045***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
N	1351822	1351822	1351822	1351822	1099155
R^2	0.056	0.058	0.059	0.060	0.134
	(1)	(2)	(3)	(4)	
Year fixed effects	no	yes	yes	yes	yes
Technology class fixed effects	no	no	yes	yes	yes
Technology trends	no	no	no	yes	yes
Firm fixed effects	no	no	no	no	yes

Notes: Technology class fixed effects are based on six NBER technology classes, and technology trends are linear time trends per technology class using filing days. Post AIPA is an indicator that equals one if the patent was filed on or after November 29, 2000. Heteroscedasticity robust standard errors are presented in parentheses.

Appendix D: Placebo test details

Table A13: Placebo Test Secrecy

	Dependent variable: No. of 102 & 103 blocking actions (in logs)					
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL 1: Raw data, one month						
Post AIPA	0.156***	0.127**	0.083	0.088*	0.045	0.075
	(0.039)	(0.053)	(0.052)	(0.053)	(0.075)	(0.101)
N	9535	9535	9535	9535	5671	5671
R^2	0.002	0.002	0.023	0.025	0.243	0.243
PANEL 2: matched data, one month						
Post AIPA	0.033	0.031	0.069	0.055	0.119	0.061
	(0.070)	(0.070)	(0.069)	(0.069)	(0.114)	(0.157)
N	2534	2534	2534	2534	1531	1531
R^2	0.000	0.001	0.038	0.045	0.230	0.231
	Dependent variable: No. of X & Y blocking actions (in logs)					
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL 1: Raw data, one month						
Post AIPA	-0.033*	-0.015	0.003	0.002	0.040	0.071*
	(0.019)	(0.025)	(0.025)	(0.025)	(0.032)	(0.040)
N	9541	9541	9541	9541	5673	5673
R^2	0.001	0.001	0.009	0.010	0.250	0.250
PANEL 2: matched data, one month						
Post AIPA	-0.015	-0.013	-0.013	-0.015	0.007	0.017
	(0.034)	(0.034)	(0.033)	(0.033)	(0.045)	(0.059)
N	2533	2533	2533	2533	1530	1530
R^2	0.001	0.002	0.019	0.020	0.191	0.191
	Dependent variable: Fraction of X & Y blocking actions					
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL 1: Raw data, one month						
Post AIPA	-0.006	-0.005	0.002	0.002	0.023	0.029
	(0.009)	(0.012)	(0.012)	(0.012)	(0.017)	(0.021)
N	9541	9541	9541	9541	5673	5673
R^2	0.000	0.000	0.004	0.006	0.225	0.225
PANEL 2: matched data, one month						
Post AIPA	-0.007	-0.007	-0.008	-0.010	0.014	0.008
	(0.017)	(0.017)	(0.017)	(0.017)	(0.024)	(0.033)
N	2533	2533	2533	2533	1530	1530
R^2	0.000	0.001	0.014	0.016	0.184	0.184
All PANEL	(1)	(2)	(3)	(4)	(5)	(6)
RDD time controls	linear (same slope)	linear (diff. slopes)	linear (diff. slopes)	linear (diff. slopes)	linear (diff. slopes)	quadratic (diff. slopes)
Technology class fixed effects	no	no	yes	yes	yes	yes
Technology trends	no	no	no	yes	yes	yes
Firm fixed effects	no	no	no	no	yes	yes

Notes: Technology class fixed effects are based on six NBER technology classes, and technology trends are linear time trends per technology class using filing days. Post AIPA is an indicator that equals one if the patent was filed on or after November 29, 2000. Matched data is a balanced sample, where patents before and after AIPA are matched (CEM) based on backward cites to capture potential differences in novelty, 6 NBER technology classes to capture differences in technological popularity, number of independent claims, dependent claims per independent claim, word count of the first claim, average number of words per claim, whether an attorney was involved, and patent pendency (time from application to grant date) to capture differences in patent scope and writing. Heteroscedasticity robust standard errors are presented in parentheses.

Appendix E: Heterogeneity in technologies

Different technologies are likely to be differently affected by early patent disclosure, due to differences in the strength and breadth of the intellectual property (IP) protection, difficulty of and lead-times in invention, and popularity and likelihood of duplication (see Arora et al., 2008; Bessen and Maskin, 2009, Galasso and Schankerman, 2015; Sampat and Williams, 2019).

In order to study potential differences with respect to disclosure in different technologies, we split the sample and re-estimate previous regressions separately for each of the 6 NBER main technology classes. Table A14 below presents the results of our monthly RDD regression models (TW and DiD designs reveal similar results, Table A15 shows yearly models).

The findings reveal that the main reduction of US blocked patents due to more timely publication of new knowledge actually occurs in Computer and Communication technologies. Estimated effect sizes amount to decreases ranging between 24.3 percent and 26.2 percent in the US. They are much higher compared to the decrease in blocked patents at the EPO, where they decreased by 5.3 to 9.0 percent. In these technologies, follow-on inventors might find it much easier to invent around a patent once new knowledge has been published. Usually, a product in these technologies contains multiple patented inventions which provides incentive to replace individual components. In addition, there is rapid technological obsolescence in these industries and abandoning an innovation project is much less costly than in other technologies. The stronger effect for the US might be driven by the intense competition in the US IT industry and/or the increased possibility of gaining software patents (which is more difficult at the EPO). AIPA's effect may well have been strongest in the American IT industry.[13]

In contrast, research in discrete technologies like Chemistry and Drugs and Medical Devices seems to be less affected by earlier patent disclosure of USPTO and EPO patents. This is consistent with the idea that research in these technologies is harder to adjust within short time-spans and that lead times of such inventions tend to be rather long, such that 18 months faster knowledge disclosure does not make a big difference.

[13] For an overall assessment, one would have to consider that abandoning innovation projects in ICT due to patent protection might inhibit cumulative innovation efforts where they might be desirable. However, a detailed analysis of the benefits and costs of cumulative innovation is beyond the scope if this paper.

The magnitude and significance of the effects appear significantly lower in the fields of Electronics and Mechanical Engineering within the US jurisdiction (effects are negative but insignificant in the one-month window model, but significant when estimated with the one-year window). European blocked patents are also significantly reduced, however, unlike the differences in the US, the magnitude of the effect is comparable to the effect in the Computer and Communication industries.

Table A14: Impact across technologies (one month window)

Dependent variable:	No. of 102 & 103 blocking actions (in logs)				No. of X & Y blocking actions (in logs)				Fraction of X & Y blocking actions			
	Panel 1: raw, one month		Panel 2: matched, one month		Panel 1: raw, one month		Panel 2: matched, one month		Panel 1: raw, one month		Panel 2: matched, one month	
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Chemical												
Post AIPA	-0.080	-0.044	-0.073	-0.104	-0.085**	-0.051	-0.067	-0.012	-0.030*	-0.013	-0.019	0.007
	(0.061)	(0.113)	(0.072)	(0.121)	(0.036)	(0.071)	(0.044)	(0.086)	(0.017)	(0.032)	(0.021)	(0.041)
N	1905	1028	1406	709	1905	1028	1404	709	1905	1028	1404	709
R^2	0.013	0.346	0.011	0.327	0.019	0.307	0.015	0.310	0.013	0.280	0.012	0.290
Computers and Communications												
Post AIPA	-0.279***	-0.290***	-0.304***	-0.280***	-0.094***	-0.073***	-0.089***	-0.054**	-0.048***	-0.036***	-0.044***	-0.028**
	(0.047)	(0.059)	(0.055)	(0.069)	(0.017)	(0.020)	(0.021)	(0.025)	(0.009)	(0.011)	(0.011)	(0.013)
N	5526	4292	3542	2700	5526	4291	3541	2699	5526	4291	3541	2699
R^2	0.026	0.195	0.020	0.191	0.030	0.190	0.029	0.177	0.028	0.184	0.030	0.179
Drugs and Medical												
Post AIPA	-0.077	-0.080	-0.042	0.047	-0.082**	-0.057	-0.057	0.010	-0.035**	-0.012	-0.026	0.027
	(0.065)	(0.097)	(0.076)	(0.120)	(0.035)	(0.064)	(0.042)	(0.077)	(0.016)	(0.026)	(0.019)	(0.034)
N	2259	1287	1604	825	2263	1288	1604	825	2263	1288	1604	825
R^2	0.004	0.400	0.003	0.444	0.013	0.283	0.011	0.337	0.009	0.266	0.009	0.312
Electrical and Electronics												
Post AIPA	-0.072	-0.061	-0.071	-0.028	-0.123***	-0.077**	-0.125***	-0.089**	-0.066***	-0.045***	-0.065***	-0.049***
	(0.056)	(0.070)	(0.065)	(0.081)	(0.027)	(0.031)	(0.030)	(0.036)	(0.013)	(0.016)	(0.015)	(0.019)
N	3070	2070	2126	1418	3070	2069	2125	1418	3070	2069	2125	1418
R^2	0.007	0.175	0.004	0.177	0.022	0.234	0.020	0.252	0.020	0.207	0.019	0.231
Mechanical												
Post AIPA	-0.042	-0.056	-0.000	-0.078	-0.174***	-0.218***	-0.181***	-0.265***	-0.076***	-0.096***	-0.073***	-0.098**
	(0.054)	(0.104)	(0.071)	(0.134)	(0.031)	(0.064)	(0.040)	(0.082)	(0.015)	(0.030)	(0.020)	(0.038)
N	2210	885	1196	442	2209	885	1194	442	2209	885	1194	442
R^2	0.002	0.232	0.001	0.274	0.028	0.297	0.029	0.330	0.024	0.263	0.024	0.261
Others												
Post AIPA	-0.105*	-0.121	-0.048	0.108	-0.071***	-0.136*	-0.065**	-0.050	-0.041***	-0.052	-0.037**	-0.010
	(0.059)	(0.161)	(0.068)	(0.207)	(0.023)	(0.073)	(0.025)	(0.081)	(0.014)	(0.043)	(0.015)	(0.049)
N	2412	578	1808	383	2409	576	1807	383	2409	576	1807	383
R^2	0.003	0.368	0.004	0.369	0.026	0.430	0.028	0.393	0.023	0.380	0.024	0.385
RDD time controls	linear	linear	linear	linear	linear	linear	linear	linear	linear	linear	linear	linear
Firm fixed effects	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes

Notes: This table reports results of RDD models, equation (1) and (2). Specifications (1) and (3) refer to equation (1), and specifications (2) and (4) refer to equations 2, including firm fixed effects. Raw data are all US patent applications filed one month before and on or one month after November 29, 2000, when AIPA became effective. Post AIPA is an indicator that equals one if the patent was filed on or after November 29, 2000. Matched data is a balanced sample, where patents before and after AIPA are matched (CEM) based on backward cites to capture potential differences in novelty, 6 NBER technology classes to capture differences in technological popularity, number of independent claims, dependent claims per independent claim, word count of the first claim, average number of words per claim, whether an attorney was involved, and patent pendency (time from application to grant date) to capture differences in patent scope and writing. Appendix A Table A5 shows the descriptive statistics of the matched data. Heteroscedasticity robust standard errors are presented in parentheses.

Table A15: Impact across technologies (one year window)

Dependent variable:	No. of 102 & 103 blocking actions (in logs)				No. of X & Y blocking actions (in logs)				Fraction of X & Y blocking actions			
	Panel 3: raw, one year		Panel 4: matched, one year		Panel 3: raw, one year		Panel 4: matched, one year		Panel 3: raw, one year		Panel 4: matched, one year	
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Chemical												
Post AIPA	-0.142***	-0.130***	-0.096	-0.029	-0.105***	-0.095***	-0.082**	-0.033	-0.050***	-0.045***	-0.028	-0.003
	(0.019)	(0.024)	(0.066)	(0.119)	(0.011)	(0.014)	(0.040)	(0.078)	(0.005)	(0.007)	(0.018)	(0.036)
N	21217	15233	1690	896	21224	15237	1690	896	21224	15237	1690	896
R^2	0.005	0.203	0.013	0.350	0.021	0.191	0.018	0.301	0.020	0.160	0.011	0.278
Computers and Communications												
Post AIPA	-0.224***	-0.215***	-0.287***	-0.320***	-0.080***	-0.075***	-0.094***	-0.072***	-0.044***	-0.042***	-0.048***	-0.035***
	(0.014)	(0.015)	(0.049)	(0.062)	(0.005)	(0.006)	(0.019)	(0.022)	(0.003)	(0.003)	(0.009)	(0.012)
N	55029	46512	4767	3653	55051	46524	4766	3652	55051	46524	4766	3652
R^2	0.017	0.162	0.026	0.205	0.025	0.137	0.028	0.195	0.025	0.117	0.027	0.195
Drugs and Medical												
Post AIPA	-0.110***	-0.103***	-0.107	-0.086	-0.107***	-0.092***	-0.084**	-0.055	-0.049***	-0.042***	-0.033**	-0.008
	(0.021)	(0.024)	(0.067)	(0.102)	(0.011)	(0.015)	(0.037)	(0.070)	(0.005)	(0.006)	(0.017)	(0.028)
N	24518	17832	2073	1162	24535	17835	2071	1161	24535	17835	2071	1161
R^2	0.007	0.296	0.003	0.400	0.021	0.204	0.012	0.288	0.021	0.189	0.008	0.269
Electrical and Electronics												
Post AIPA	-0.123***	-0.112***	-0.054	-0.063	-0.109***	-0.102***	-0.110***	-0.060*	-0.058***	-0.054***	-0.056***	-0.034**
	(0.015)	(0.018)	(0.058)	(0.076)	(0.007)	(0.008)	(0.028)	(0.034)	(0.004)	(0.004)	(0.013)	(0.017)
N	35049	28013	2560	1715	35048	28009	2559	1714	35048	28009	2559	1714
R^2	0.006	0.145	0.006	0.164	0.027	0.153	0.024	0.243	0.026	0.144	0.023	0.210
Mechanical												
Post AIPA	-0.074***	-0.076***	-0.053	-0.062	-0.106***	-0.133***	-0.175***	-0.204***	-0.055***	-0.068***	-0.078***	-0.094***
	(0.016)	(0.025)	(0.058)	(0.112)	(0.008)	(0.014)	(0.034)	(0.068)	(0.004)	(0.007)	(0.016)	(0.032)
N	25892	14162	1964	777	25896	14163	1961	777	25896	14163	1961	777
R^2	0.003	0.199	0.002	0.226	0.028	0.215	0.027	0.282	0.028	0.198	0.023	0.249
Others												
Post AIPA	-0.068***	-0.074**	-0.153**	-0.143	-0.075***	-0.103***	-0.079***	-0.140*	-0.041***	-0.048***	-0.044***	-0.055
	(0.018)	(0.034)	(0.061)	(0.166)	(0.007)	(0.016)	(0.024)	(0.075)	(0.004)	(0.008)	(0.014)	(0.044)
N	29515	10908	2188	505	29521	10912	2186	503	29521	10912	2186	503
R^2	0.002	0.258	0.004	0.336	0.021	0.251	0.026	0.399	0.020	0.243	0.023	0.392
RDD time controls	linear	linear	linear	linear	linear	linear	linear	linear	linear	linear	linear	linear
Firm fixed effects	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes

Notes: This table reports results of RDD models, equation (1) and (2). Specifications (1) and (3) refer to equation (1), and specifications (2) and (4) refer to equations 2, including firm fixed effects. Raw data are all US patent applications filed one year before and on or one year after November 29, 2000, when AIPA became effective. Post AIPA is an indicator that equals one if the patent was filed on or after November 29, 2000. Matched data is a balanced sample, where patents before and after AIPA are matched (CEM) based on backward cites to capture potential differences in novelty, 6 NBER technology classes to capture differences in technological popularity, number of independent claims, dependent claims per independent claim, word count of the first claim, average number of words per claim, whether an attorney was involved, and patent pendency (time from application to grant date) to capture differences in patent scope and writing. Appendix A Table A5 shows the descriptive statistics of the matched data. Heteroscedasticity robust standard errors are presented in parentheses.

Appendix F: Data sources

Disclosure dataset from Graham and Hegde *Science* 2015
Description: Graham, S. and Hegde, D. (2015): Disclosing patents' secrets. Inventors prefer to disclose know-how before patent grant. *Science*. Vol. 347, issue 6219, pp. 236-237.
Url: www.sciencemag.org/content/347/6219/236/suppl/DC1
Datasets: patentdisclosure_datafile_gh.dta
Purpose: secrecy – disclosure – foreign equivalent information, identify technology class, application date, granting date

USPTO dataset: Office Action Research Dataset for Patents
Description: Lu, Qiang and Myers, Amanda F. and Beliveau, Scott, USPTO Patent Prosecution Research Data: Unlocking Office Action Traits (November 20, 2017). USPTO Economic Working Paper No. 2017-10
Url: https://www.uspto.gov/learning-and-resources/electronic-data-products/office-action-research-dataset-patents
Datasets: citations.dta & office_actions.dta
Purpose: calculate the number of blocked patent applications

From citations.dta kept all office actions by application id

- Kept only actions with an action type (102 or 103) (Type of action raised, indicated by section of 35 USC or category (double patent).
- Retained only actions related to patent id (identified by the variable parsed).

From office_actions.dta kept

- date of the office action. Each office action can be exactly identified by the ifw_number (image file wrapper of the office action) in both datasets.

USPTO Patent Examination Research Dataset
Description: Graham, S., Marco, A. and Miller, R., The USPTO Patent Examination Research Dataset: A Window on the Process of Patent Examination (November 30, 2015).
Url: https://www.uspto.gov/learning-and-resources/electronic-data-products/patent-examination-research-dataset-public-pair
Dataset: application_data.dta
Purpose: identify attorneys and the disposal type (abandoned, issued or pending)

USPTO Patent Claims Research Dataset
Description: Marco, Alan C. and Sarnoff, Joshua D. and deGrazia, Charles, Patent Claims and Patent Scope (October 2016). USPTO Economic Working Paper 2016-04.
Url: https://www.uspto.gov/learning-and-resources/electronic-data-products/patent-claims-research-dataset
Dataset: patent_document_stats.dta
Purpose: identify claim statistics (no. of independent claims, dependent claims per independent claim, average number of words per claim)
Dataset: patent_claims_stats.dta
Purpose: identify claim statistics (word count of the first claim)

Patentsview: Assignee information
Description: Graham, S., Marco, A. and Miller, R., The USPTO Patent Examination Research Dataset: A Window on the Process of Patent Examination (November 30, 2015).
Url: http://www.patentsview.org/download/
Dataset: rawassignee.tsv
Purpose: identify assignee

European patent search database: 'PATSTAT Biblio - 2016 Autumn Edition'
Description + Url: https://www.epo.org/searching-for-patents/business/patstat
Dataset: us_ep_cites.dta
Purpose: identify blocked European patent applications

Appendix G: Example of an EPO search report

As an example of the EPO examination practice, we consider an EPO patent application filed by Samsung Electronics Co., Ltd. Samsung is a Korean company, with the largest patent portfolio among the biggest patent offices worldwide (Daiko et al., 2017). Of course, the size of the patent portfolio and origin country of the applicant does not play any role and the examination process is the same for all firms applying for EPO patent protection, irrespective of the country of the applicant.

The patent application with application number 'EP10150472' for an 'Image restoring apparatus and method thereof' was filed on January 21, 2010. The whole application and examination process is documented on https://register.epo.org/regviewer?lng=en where all relevant documents and legal details can be found for any EPO patent. The search report was already published on July 07, 2010, because the filing refers to a so-called priority filing at the Korean patent office from January 2009 (covering the same invention), and the EPO filing is a so-called subsequent filing in order to gain protection in European countries as well.

The search report is a list of 'documents considered to be relevant' (see next page for the first page of the search report). It thus contains citations of other patent documents and their relevance to claims of the patent under examination. Each citation has a category, the most important being X, Y and A citations. 'X' and 'Y' citations indicate whether references to other patent documents call the novelty or inventive step of the claimed invention into question. 'A' citations indicate that the cited documents only represent state of the art and are thus not critical. In our context, in order to define patents that are potentially blocked by cited patent documents, X and Y citations are the relevant category.

The EPO defines an X citation as: "…where a document is such that when taken alone, a claimed invention cannot be considered novel or cannot be considered to involve an inventive step", and a Y citation as, "…applicable where a document is such that a claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other documents of the same category, such combination being obvious to a person skilled in the art." Source: https://www.epo.org/law-practice/legal-texts/html/guidelines/e/b_x_9_2_1.htm.

The search report lists several X and Y citations, for example to US patent US2008238942 that is a Microsoft patent on ‘Object-Based Image Inpainting’. As a result, there were several amendments to the claims until the patent was granted on August 08, 2013.

Figure A6: Example of an EPO search report

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 10 15 0472

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X Y	LIN Y ET AL: "Fast image completion" VISUAL COMMUNICATIONS AND IMAGE PROCESSING; 12-7-2005 - 15-7-2005; BEIJING, 12 July 2005 (2005-07-12), XP030080885 * abstract * * Section 3 * * figures 1,2 * * tables 1,2 * -----	1,5-8, 12-15 2,3,9,10	INV. G06T5/00 G06T11/00
X Y	US 2008/238942 A1 (SUN XIAOYAN [CN] ET AL) 2 October 2008 (2008-10-02) * paragraph [0042] - paragraph [0049] * * figure 4 (d) * -----	1,5,6,8, 12,13,15 2,3,9,10	
X	HONGYING Z ET AL: "Image completion algorithm based on texture synthesis" CHINESE JOURNAL OF SYSTEMS ENGINEERING AND ELECTRONICS, SECOND ACADEMY MINISTRY OF AERO-SPACE INDUSTRY, BEIJING, CN, vol. 18, no. 2, 1 January 2007 (2007-01-01), pages 385-391, XP022936689 ISSN: 1004-4132 [retrieved on 2007-01-01] * abstract * * Section 2 * ----- -/--	1,4-8, 11-15	TECHNICAL FIELDS SEARCHED (IPC) G06T

The present search report has been drawn up for all claims

1

Place of search	Date of completion of the search	Examiner
Munich	13 April 2010	Salvador, Elena

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

page 1 of 2