Intellectual
Property
Office

UNITED STATES PATENT AND TRADEMARK OFFICE
DEPARTMENT OF COMMERCE

Patent backlogs, inventories and pendency: An international framework

Working draft

June 26, 2013

A joint UK Intellectual Property Office and US Patent and Trademark Office report

Benjamin Mitra-Kahn, Alan Marco, Michael Carley, Paul D'Agostino, Peter Evans, Carl Frey, Nadiya Sultan

2013/25

A joint UK Intellectual Property Office and US Patent and Trademark Office report

This report was a collaborative effort of the UK Intellectual Property Office and the US Patent and Trademark Office economics teams, following a joint agreement between the two offices in 2011. The country-specific analyses benefited from collaboration as to methodology and presentation, but remain the product of their respective individual country teams and do not necessarily reflect the official views of any other country or patent office. Please note that this work uses research datasets which may not exactly reproduce the official statistics of the UK IPO or the USPTO.

Working draft - June 2013

This report is subject to further revisions.

Authors:
Intellectual Property Office - UK
Benjamin Mitra-Kahn, Peter Evans, Carl Frey, Nadiya Sultan
United States Patent and Trademark Office
Alan Marco, Michael Carley, Paul D'Agostino

The authors would like to thank:

From Intellectual Property Office - UK
Rich Corken, Tony Clayton, Glyn Hughes, Andy Bartlett, James Porter, Michael Prior, Carol Wheeler.

From United States Patent and Trademark Office
Stuart Graham, Charles Eloshway, Mark Guetlich, Daniel Hunter, Greg Mills, Meredith Schoenfeld, Camila Tapernoux, and Saurabh Vishnubhakat

From the IP Office economist group:
Ellias Collette, Sergio M. Paulino de Carvalho, Carsten Fink, Laurence Joly, Geoff Sadlier, Hansueli Stam, and Nikolaus Thumm

Attendees at the 2011 Patent Statistics for Decision Makers Conference

Members of the UK Patents Research Expert Group, especially Roger Burt

Suggested reference
Mitra-Kahn, B., Marco, A., et al., 2013, "Patent backlogs, inventories, and pendency: An international framework," IPO- & USPTO joint report, http://www.ipo.gov.uk/pro-ipresearch.htm.

ISBN - 978-1-908908-70-4
Published by the Intellectual Property Office
Patent backlogs, inventories and pendency:
An international framework

Published as Working Draft 26 June 2013

1 2 3 4 5 6 7 8 9 10

Any enquiries regarding this publication should be sent to:

The Intellectual Property Office
Concept House
Cardiff Road
Newport
NP10 8QQ

Tel: 0300 300 2000
Minicom: 0300 0200 015
Fax: 01633 817 777

e-mail: information@ipo.gov.uk

This publication is available from our website at www.ipo.gov.uk

Contents

Executive Summary

This is a joint report by the United States Patent Office (USPTO) and the UK Intellectual Property Office (UK IPO) on the economic and operational impacts of patent application backlogs. The report offers both a new and comparative methodology for measuring backlogs and pendency, as well as empirical findings on the relationship between application stocks and examination pendency in the UK and the US. Understanding these relationships is critical for better evidence-based policy making.

1. Introduction

- The term "backlog" as such is not well defined. To some it refers to all unexamined applications, to some all pending applications, and to some "excess" applications beyond office capacity.
- To cut through this ambiguity, the authors define three stocks (or inventories) of pending patent applications that should be able to be well identified by any patent office. Together the three stocks make up the cumulative "backlog" experienced by a patent office.
- By 2011, the UK had accumulated a backlog of 40,000 pending applications, including applications later abandoned. This represents no change over a period of ten years, although the composition of the backlog has changed. In the US, the backlog of pending applications has grown from about 400,000 in 1996 to over 1,200,000 in 2007.
- Large stocks of pending applications are important to the extent that they slow down the time to decision of incoming applications. That is, the primary cost associated with backlogs is increased delays in processing pending patent applications.
- Increased pendency can be attributed to delays in office actions, and also to delays in applicant responses.
- The analysis demonstrates that traditional pendency measures need revising. This report proposes a different measure, and tests how inventories impact pendency times.

2. Methodology

- To facilitate international comparisons and meaningful analysis, the total patent application inventory is split into three different stocks based on identifiable stages in the application process common to all offices:

    - Stock 1. Received at the office, but not yet ready for examination

    - Stock 2. Ready for examiner action, but not having completed its first examination where it receives a decision on the merits

    - Stock 3. After first examination, but still pending before final disposal (e.g., grant or abandonment)

- The application process entails a series of exchanges between applicants filing for a patent and the office processing the application. This is why the report also distinguishes between the share of the application inventory that is "inside" the office being processed and those pending applications that are "outside" in the sense that they are in the hands of applicants.

- We recognize that certain offices may find it useful to further refine the simple taxonomy above and define sub-stocks. For instance, in the US it is useful to divide Stock 3 into regular applications, and those that have sought continued prosecution through particular institutional means.

3. Pendency

We find it useful to define some terms that distinguish different measurements of pendency.

- **Exit pendency** is the commonly applied pendency measure, and is based on the time to terminal disposal (e.g., final grant or abandonment) for all applications terminating in the same month. It does not account for the observed pendency of applications still being processed.

- **Entry pendency** is based on the time to terminal disposal for a filing cohort of all applications that were filed in the same month. It does not account for the pendency of applications in that cohort that are still being processed.

- **Expected pendency** is the time to grant or abandonment that an applicant expects to experience (on average) on the day of filing. This actual expectation is unobserved, but it can be estimated using predictions from statistical techniques.[1]

1 In the analytical sections of this report we use survival time regressions, which test the influence of a set of independent variables on the duration of a process or event.

4. UK Analysis

Background

- The UK patent application process is an 'active' system where applicants act to progress their application at each stage by giving instructions or paying fees.
- Most patents receive a final decision within 4.5 years after filing and a large proportion of granted patents are pending for between 2.5 to 4 years.

Combined Search and Exam (CSE)

- Patentees can request a combined search and exam and these types of applications take only half as long to grant as standard patent applications (between 1-2 years).
- Doubling the proportion of applications that are CSEs is associated with a 12% reduction in the average pendency (or time to grant) that applicants can expect to experience when they file new applications.

The impact of backlogs on pendency

Using statistical regressions that test how long applications 'survive' before being granted, we find that the effect of backlogs on average pendency varies between stocks:

- A build up of applications in Stock 1 (filed applications awaiting a search request) and Stock 2 (applications that have requested a search but not received a first examination report) is most strongly associated with increases in average pendency time to grant.
- Especially applications in Stock 2 outside the office awaiting additional search requests or an initial examination request could significantly increase average pendency time. A doubling of this stock is associated with a doubling in expected pendency time across the board.
- By comparison, changes in the other application stocks are associated with only a minor effect on average pendency and increases in Stock 3 outside the office (applications awaiting amendments) are associated with reduced pendency time to grant.

What else is causing delays?

The report identifies five other factors that could be causing delays or could be utilised to reduce pendency

- The Patent Office: Examination pendency is more responsive to increases in examiner capacity than initially thought and this could be an effective way to address the large and growing Stock 2 held by the office.

- The Patent itself: Survival analysis suggests that patent complexity as usually measured in the processing system is associated with a small reduction in average pendency. This finding could be due to unmeasured factors so should be interpreted with caution.

- The Applicants and requests for additional examiner work: Applicant requests for amendments have grown from an average of 1.1 requests per application in 2000 to 1.3 requests in 2006. The statistical analysis shows that further amendment requests are associated with approximately 20 weeks in additional pendency.

- Flexible deadlines on amendment requests: On average, roughly two-thirds of applicants chose to extend the office's 4-month deadline on amendment requests by the allowable 2 months, and one third took longer than the statutory 6 months to file the request. Stock 3 could be reduced by several thousand applications if deadlines on amendments were enforceable.

- The type of applicant: Traditional patent applications filed by private applicants take 100-200 days longer to grant than represented applicants. This effect is amplified when considering CSE applications.

Conclusions

The report differentiates between potential short term and long term solutions to application inventories and pendency:

- Short term - To keep the backlog and pendency stable in the short run, Stock 1 and Stock 2 outside the office need to be kept stable, while managing the other processes efficiently. To reduce pendency in the short term, one needs to address the parts of the backlog that are biggest and those factors that could impact pendency the most. In the UK, this requires targeting Stock 2 inside the office, for instance through increasing examination capacity, and Stock 3 outside the office, by for instance enforcing deadlines already in existence.

- Long term - Reducing the backlogs could be achieved through changing the rules and regulations governing Stock 2. This is likely to be effective as Stock 2 inside the office is large and growing and Stock 2 outside the office is expected to have the biggest effect on pendency if reduced. However, reducing Stock 2 will not be an easy feat and will require careful consideration of the existing regulatory framework – national and international which affects applicant behaviour.

5. US Analysis

Descriptive Analysis

The report evaluates commonly mentioned causes for the growth of the US backlog, and finds:

- New applications have not grown rapidly over the last 16 years
- Hiring of patent examiners has kept pace with the growth in new applications
- The amount of work an examiner must do to terminally dispose of an application has increased between 1996 and 2011
- It is unclear whether application complexity can account for the growth in application stocks
- Applications with at least one Request for Continued Examination (a method to continue the prosecution of a rejected application) appear to disproportionately "crowd out" (displace) other applications with a ratio of approximately 1:2

Pendency Analysis

- **Disposal rates versus disposal types: Understanding cumulative patenting and abandonment.** This section explains how patent applications face two disposal types: patenting and abandonment. This fact complicates any computation of total pendency based on the time to disposal because abandonments and patent grants may behave very differently from one another.
- We find that the ultimate abandonment proportion of an incoming filing cohort is negatively correlated with the speed of patenting. In instances where the speed of patenting is high, the abandonment rate is low. The correlation between pendency and abandonment may cause post-first-action pendency to be understated.
- **Queuing time.** We find that median pendency has grown primarily due to the increase in first-action pendency, and that the second main contributor to total pendency is the greater intensity of RCE use.
- The relative importance of the two main contributors to total patent pendency can be approximated by analysing the total pendency for the average patent grant. Over the 2000 to 2008 time period, average total pendency increased by 1.05 years (from 2.25 years to 3.3 years). About 86% of the increase can be attributed to the increase in first-action pendency. The remaining 14% can be attributed to an increase in RCE usage.

Survival time analysis

- **Data and methodology.** For applications filed between 1996-2011, survival time regressions are performed on: (a) pendency from filing to first action (first-action pendency), and (b) pendency from first action to patent grant or abandonment (post-first-action pendency).

- Variables for the regressions include monthly stocks for S1, S2, S3, and the workforce at the broad technology level (chemical, electronics, and mechanical). Stock 3 is further sub-divided into applications that receive at least one RCE, and those that do not. We also control for observable application characteristics.

First-action pendency. Based on 2012 averages, we show:

- One additional unexamined application increases the predicted first-action pendency of each new application by about 39 seconds. Over the course of a year, this amounts to an additional 5.6 months of cumulative first-action pendency experienced by incoming applications.

- Filing an RCE on an application that would otherwise not have filed an RCE adds approximately 66 seconds to the first-action pendency of each incoming application. Over the course of a year, this amounts to an additional 9.4 months of cumulative first-action pendency experienced by incoming applications.

- One additional junior examiner saves each new application about 1,191 seconds (19.8 minutes) in first-action pendency. Over the course of a year, this amounts to a total cumulative savings of about 170.7 months of first-action pendency experienced by incoming applications.

- In the text, we also discuss the impact of observed application characteristics on first-action pendency.

Post-first-action pendency.

- Longer first-action pendency increases the rate of abandonment: Raising first-action pendency from one year to two years (for the 2003 first-action cohort), increases the predicted abandonment rate from 22.5% to 28.5%. Thus, increased application inventories and increased pendency can influence the observed allowance rate.

- An increase in stocks has a direct effect both on first-action pendency and on post-first-action pendency. Both of these impacts serve to change total pendency. Further, there is an indirect effect that operates through the first-action pendency. In particular, increases in patent application stocks have a direct impact on first-action pendency. But, first-action pendency itself affects post-first-action pendency.

- In general, we find that the largest contribution to changes in total pendency derives from the direct effect on first-action pendency. However, we find that stocks have a significant impact on the overall abandonment rate.

Concluding remarks

- The US analysis found that some traditionally cited sources of the "backlog" may not explain the growth in patent inventories over the 2000s.

- Survival time regressions demonstrate the ability of our methods to more precisely quantify the relationship between stocks, examiner capacity, and application characteristics with first-action pendency and post-first-action pendency.

- It is important to recognize that not all applicants bear the same costs with respect to delay; some may even prefer it. Policies regarding continued prosecution must balance the direct costs and benefits to the applicant and examiner with the indirect costs that are imposed on other applicants who must incur additional wait time.

- While we find continued prosecution (in the form of RCEs) to be a major source of delay for other applications, it is important to note that RCEs can be a useful method to correct errors, or for applicants to sufficiently narrow and clarify their claims, or to buy additional time in the examination process. USPTO policy responses should address the question of pendency without unduly restricting access to quality examination. The office has sought public comment about RCE use through several recent RCE Roundtables.[2]

- The USPTO has responded to the backlog challenge with a variety of responses, including hiring additional examiners, changing the examiner count system, changing the examiner docket management system, initiatives related to the America Invents Act (e.g., expedited review and fee-setting), and new pilot programs designed to streamline patent prosecution (e.g., Quick Path Information Disclosure Statement and the After Final Consideration Pilot).

- The policies implemented to date by the USPTO appear to be having the desired effect of reducing pendency. Official numbers put the quantity of unexamined patent applications below 600,000 in February of 2013, with first-action pendency (exit pendency) at 19.2 months.[3]

---

2 http://www.uspto.gov/patents/init_events/rce_outreach.jsp
3 http://www.uspto.gov/dashboards/patents/main.dashxml accessed 11 April 2013.

6. Conclusion

- Patent offices face the challenge of balancing several important policy goals: innovation and technological growth, a well-functioning market for technology, quality examination, and patent application pendency. Thus, pendency goals alone cannot drive policy.

- Offices must look carefully to identify the sources of delay, and to identify policies that may be able to address pendency without sacrificing examination quality.

- The results from the UK and US analyses should serve as examples of how policy-makers can identify contributors to pendency, and can suggest ways forward.

- In both countries, the ability of applicants to extend the prosecution of rejected or denied applications is one of the drivers of total pendency. Extended prosecution by itself is neither good nor bad *per se*. It can serve as a useful method to correct errors, or for applicants to sufficiently narrow and clarify their claims. Policy responses should address the question of pendency without restricting access to the benefits of continued prosecution. Offices should seek institutional methods to incentivize early stage claim specificity and narrowing.

- In both offices, we find that pendency is sensitive to similar types of inventory. In the UK, pendency is particularly sensitive to the size of Stock 1 (new applications undergoing clerical tasks and awaiting applicant requests plus fees to be filed) and Stock 2 outside the office (applicants taking time to file search or examination requests). In the US, Stocks 1 and 2 affect the first-action pendency of incoming applications and the post-first-action pendency of applications entering Stock 3.

- The results suggest avenues for future research, because they enable policy-makers to assess the quantitative impact of targeting specific inventories through increase examination capacity, institutional changes, or by changing fees. For instance, in 2010, the UK reduced the applicant's period of reply to certain examination reports from four months to two months. The US recently introduced tiered pricing for RCEs: the second and subsequent RCE cost more than the first.[4]

- The methodology presented here is not meant to be comprehensive. However, our findings suggest that the framework is useful for bringing more precision to the discussion of patent application inventories and pendency, and to the estimation of expected pendency.

- Most importantly, we hope that this report will suggest avenues for continued research and greater collaboration among patent offices in coordinating research and sharing data.

4 See http://www.uspto.gov/web/offices/ac/qs/ope/fee031913.htm.

1. The problem with patent backlogs and pendency

1.1 Many ways of defining patent backlogs

There is much discussion about patent backlogs in the press,[5] in policy circles,[6] and amongst academics,[7] but there is almost no shared empirical base for analysing backlogs across offices. Patent offices around the world have different administrative systems for granting patents in part due to different legal constraints. This means that a straight comparison of the grant process is difficult and possibly misleading, particularly to the national offices who think about the backlog differently. In the US, the backlog is typically defined as the quantity of applications waiting for the first substantial review by an examiner, and in the UK the backlog is generally considered to be the number of applications that are awaiting an office action in response to an applicant request. Other offices may think of the backlog in terms of the number of applications that remain unexamined after a certain time period or the portion of pending applications that exceed the steady state capacity of the office. Attorneys often treat as backlog those applications for which search and examination has been requested, but which have not been granted or refused.

Each of these definitions has its own operational logic, but the conceptual differences and lack of empirics, must be overcome to arrive at commonly agreed upon solutions. In fact, due to the ambiguity of the term, in this report we recommend a move away from the term "backlog" and instead discuss patent application "inventories" or "stocks." The aim of this report is to provide a first step towards defining a common framework for measuring patent application inventories. The framework facilitates international comparisons on the causes and consequences of patent application inventories in our offices: the UK Intellectual Property Office (UK IPO) and the United States Patent and Trademark Office (USPTO). In our respective country analyses, we pay particular attention to the relationship between the size of the inventory and the patent application pendency of incoming applications. Our analysis should provide policy recommendations geared towards reducing inventories and reducing the costs of delay, both to the wider economy and within patent offices themselves. This research represents the first in-depth international comparison of patent application inventories. As a consequence, this report will not provide an analysis of each factor that may be important to national offices with respect to inventories and pendency. Rather, it sets out a framework for making international comparisons, and applies that framework to the particular cases of the UK and US patent offices. We are able to identify some of the major factors related to increased pendency of patent applications, both from the patent office and patent applicant perspectives. More detailed work will need to be done by individual offices on these issues, but this framework should provide offices with a tool to do that work.

5 See for example *The Economist* "The spluttering invention machine" on 17 March 2011, or "Patently Absurd" on 5 May 2011.

6 See for example the joint statement by the UK-IPO and USPTO in April 2011 (http://www.ipo.gov.uk/about/press/press-release/press-release-2011/press-release-20110405.htm); WIPO (2009), "WIPO Symposium Concludes Global Patent Application Backlogs Unsustainable", Geneva, Switzerland (http://www.wipo.int/pressroom/en/articles/2009/article_0035.html); London Economics (2010), "Patent Backlogs and Mutual Recognition" (http://www.ipo.gov.uk/p-backlog-report.pdf); and IP5 (2011), "IP5 Statistics Report" (http://www.fiveipoffices.org/stats/statisticalreports/ip5-statistics-2011.pdf)

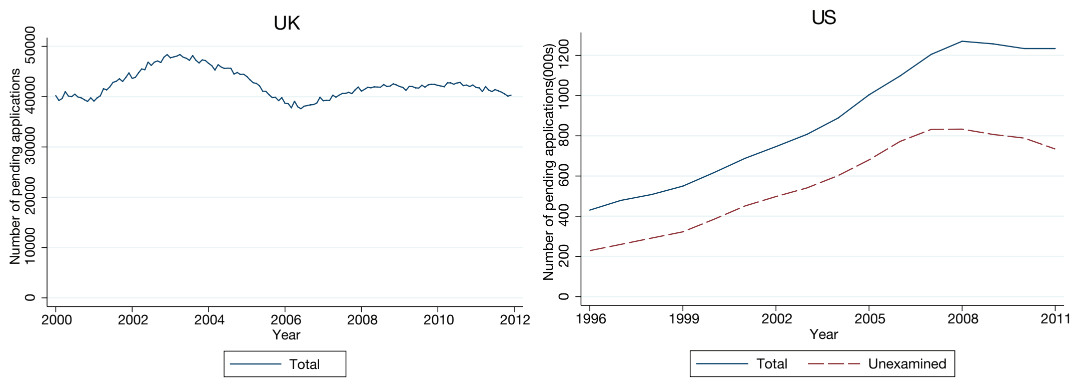
7 See for example Mejer, M.; van Pottelsberghe de la Potterie, B. 2011. "Patent backlogs at USPTO and EPO: Systemic failure vs. deliberate delays." *World Patent Information* **33**(2), pp. 122-7

The framework has been developed jointly by the UK Intellectual Property Office and the US Patent and Trademark Office, with the support of WIPO's Patent Economist Group[8]. Our ambition is to build a method that facilitates comparisons across very different patent systems experiencing very different market conditions.

1.2 Preliminary analysis of patent application inventories at the UK IPO and USPTO

By December 2011, the UK IPO had a total stock of 40,000 pending applications. As shown in Figure 1.1, this inventory has been relatively stable over the past decade, with a small hump of just under 50,000 pending applications in December 2002 and a more gradual increase from a low of roughly 38,000 pending patent applications in December 2006.

Figure 1.1 Application inventory in the UK and US



Source: UK IPO 2013, USPTO 2013

In the US, the total stock of pending utility applications[9] (including unexamined applications, and those in active examination) has risen steadily since the 1990s, as shown in Figure 1.1. It plateaued in 2008 at just over 1.2 million applications. Since that time the total inventory has remained steady through 2011, and indeed through the end of 2012 (not shown). The more traditional backlog (unexamined applications) followed roughly the same trend through 2007 when it peaked at just over 800 thousand applications. Since that time it has fallen to 750 thousand in 2011, and near 600 thousand by the end of 2012.[10]

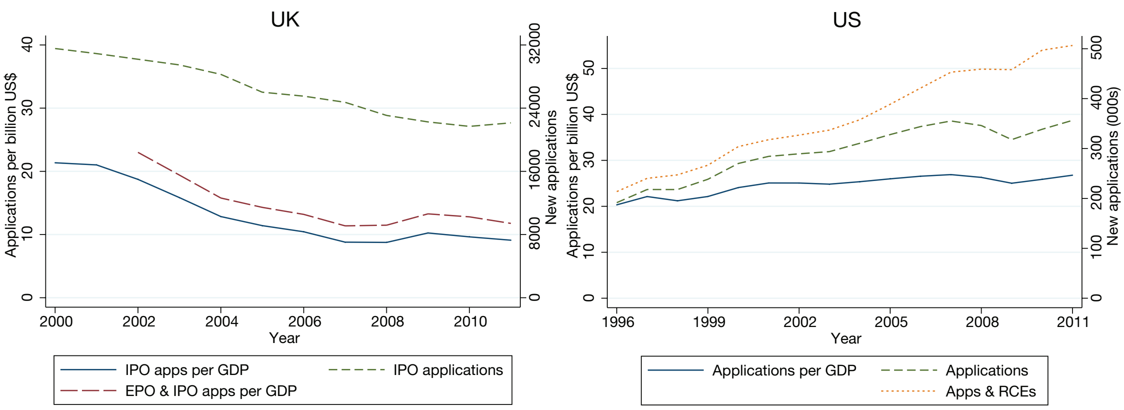
8 See http://www.wipo.int/econ_stat/en/news/2010/news_0001.html

9 Throughout the report, we analyse only regular utility applications. Thus, we exclude design, plant, and reissue applications, as well as re-examinations.

10 These figures may differ from official USPTO statistics due to slight differences in counting methods, and the definition of applications included.

In the chapters that follow we disaggregate the total inventory and analyse it in more detail, in the context of the individual office's procedures, institutions, and market conditions. In the UK, annual applications have plateaued at 20,000 applications over the last decade, following a long decline from the 1970s through the 1990s due to the creation of the European Patent Office. The US office, on the other hand, appears to have had a steady increase in demand over the last decade, getting just over half a million applications in 2010 (although a significant proportion of that growth comes from a particular type of patent application called a request for continued examination (RCE), which essentially extends the prosecution of the current application.[11] These are treated as new applications in many official statistics. Figure 1.2 shows the absolute counts of new applications as a proportion of GDP. The US data are shown with and without RCEs.

Figure 1.2 Applications per GDP in the UK and the US

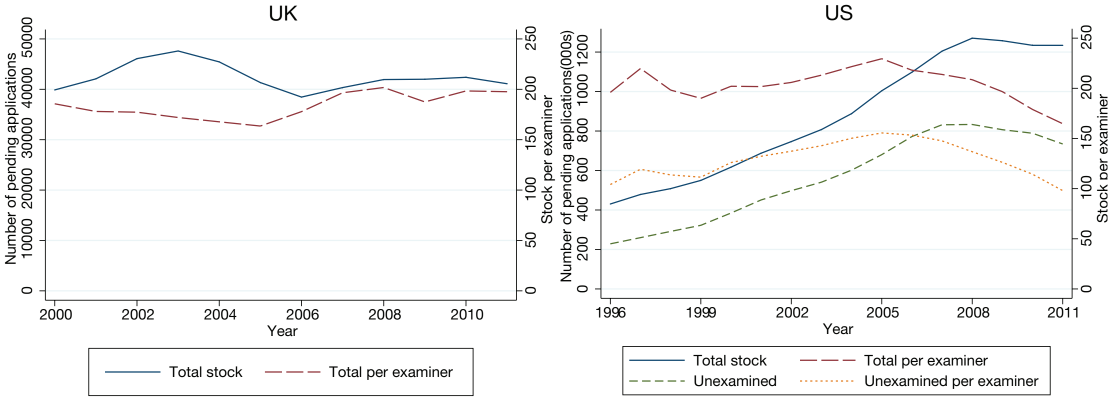


Source: UK IPO 2013, USPTO 2013, UK GDP figures from World Bank, US GDP from Bureau of Economic Analysis

Figure 1.3 illustrates how one can quickly compare both trends and sizes of stocks within the UK and US office. These graphs show the total number of applications which were pending at any given month and the ratio of that stock to active examiners. In the US we see a steady increase in the total application inventory to just over 1.2 million applications in 2008, and then a flattening out as the inventory-per-examiner begins to drop from almost 250 in 2005 to 160 in 2011. In the UK, the application inventory has been relatively stable but the ratio of inventory to examiners has increased from around 160 in 2005 to 200 in 2011. What appears striking is the degree of similarity in the application stock per examiner, ranging between 150 to 250 for the period under investigation in both offices.

11 RCEs are one form of "non-serialized continuations." Such continuations do not receive a new serial number and are not substantially different from the original application. RCEs (instituted in 1999) are by far the most common form of non-serialized continuations. Thus, for the purposes of this paper, we include all forms of non-serialized continuations in the RCE category. For more information, see Appendix D.

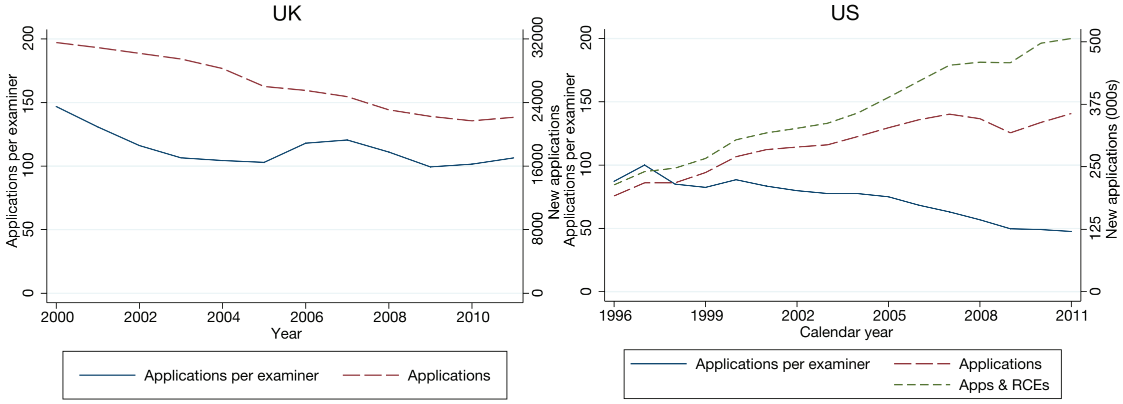
Figure 1.3 Application inventory per examiner in the UK and US



Source: UK IPO 2013, USPTO 2013

Any framework for international comparison needs to accommodate significant differences in examination procedures, economic conditions, and office size across countries. In examining patent application trends, it is also useful to consider the office's capacity for dealing with new applications. In this respect the UK and US offices have seen surprisingly similar trends as suggested by Figure 1.4, which shows how the ratio of applications per examiner has fallen over the last decade.

Figure 1.4 Patent applications per annum per examiner in the UK and US



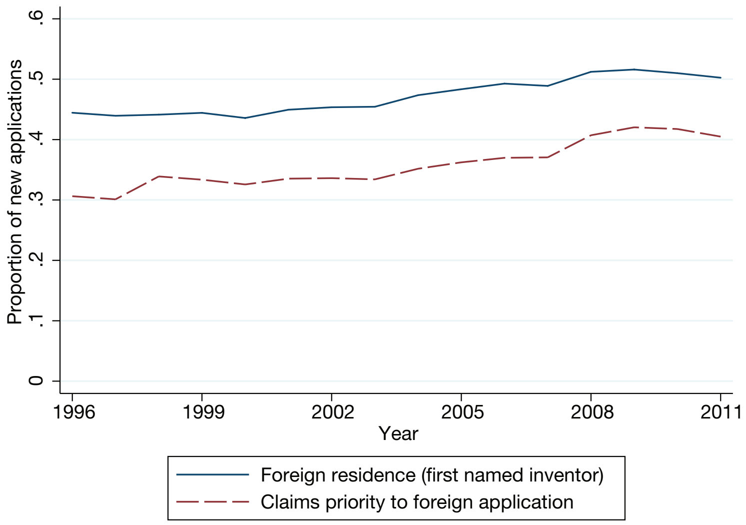
Source: UK IPO 2013; USPTO 2013

For the UK this means that in 2011 there were approximately 100 new applications for every examiner in the office, down from 150 ten years ago, while for the US the ratio has fallen from 100 in 1997 to around 50 in 2011. Without a better way to compare inventories, this would be a typical-but misleading-way to illustrate the capacity of an office. In the US the act of applying includes an implicit examination request, so all applications require examiner input. This is not the case in the UK. In the UK an application needs no examiner input until search/exam is requested. As shown in Figure 1.3 above, the offices also have a remarkably similar ratio of patent application inventory per examiner-suggesting that the number of examiners used to deal with active applications in each office is very similar.

The profile of applications has been changing in the US in terms of origin.[12] Applications from non-US inventors have outnumbered those from domestic inventors since 2008, based on first-named inventor (see Figure 1.5). There has been a similar increase in applications that claim foreign or international priority, although the percentage remains under 50%.

12 Traditionally, origin is determined by the country of residence of the first-named inventor.

Figure 1.5 Origin of new applications in the US



Source: USPTO 2013

There have been suggestions that foreign applications may be harder to process than domestic ones in the US. The UK office has seen a relatively stable 40% of its applications coming from foreign or multi-national firms not based in the UK since before 2000. (This includes both direct filed priority applications and international applications filed under the Patent Cooperation Treaty or PCT). There is no suggestion that the distinction between direct or international applications impact process speeds, at least not as much as those applications which come from unrepresented applicants – applicants who do not use a patent lawyer or specialist drafter. These issues are examined in more detail in the focused country analyses in Chapters 4 and 5.

One has to examine the details of pendency on a national basis precisely because the office processes are different, and meet somewhat different local needs because of the ways in which patent systems have developed. However, at a more aggregate level, our framework attempts to identify specific points of comparison common to all offices, and then uses the resulting data to allow each office to drill down in the areas they feel are relevant.

2. Defining and measuring patent application stocks

As mentioned above, defining the patent backlog poses significant analytical problems, because different patent offices have different systems. Moreover, conceptual issues surround the definition of the term “backlog.” For the purposes of analysing the backlog one must be able to quantify it, requiring a precise definition which is universal in application. Because of the ambiguity and connotation of the term backlog, we default to calling the entire stock of pending applications in a given office the application inventory. Our methodology breaks this inventory into several component stocks. Throughout the rest of the report, we generally refer to inventory and stocks rather than backlogs.

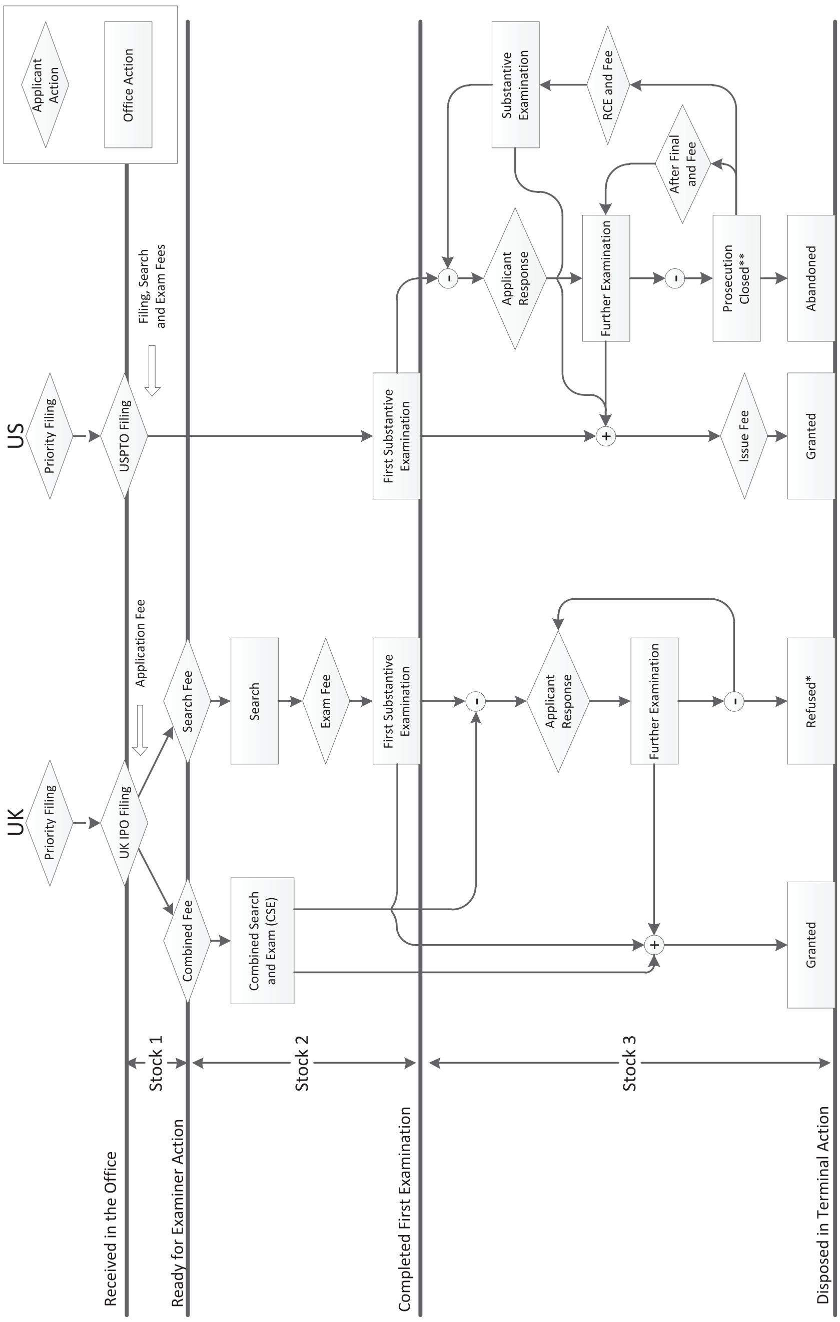
Our first step is therefore to define a concept which will work across the UK and US offices, despite their institutional differences. The US charges all basic application fees up front (including filing, search, and examination fees) and requires a payment upon issuance of the final patent.[13] The UK charges at each office action to search and examine, but nothing upon grant. In the US, applicants must expressly abandon to opt out of initial search and examination. In the UK, applicants must opt in to examination by affirmatively requesting search and examination.

A simplified model of the UK and US patent system is presented in Figure 2.1. This is an idealised and simplified path through both systems, not reflecting the often iterative nature of amendments and different options available. As indicated above, the list of differences goes on, and includes issues of language which can be important. For example, where the *London Economics* team (who produced the original version of the below diagram) write “Examination” for the US office, the USPTO would recognise this stage as a “first action.”[14]

13 Some optional payments occur throughout prosecution, e.g., for extensions of time, requests for continued examination, appeals.

14 Technically, the first substantive review by the examiner is called the ‘First Action On the Merits’ or ‘FAOM’ in the vernacular (pronounced “foam”). However, for the purposes of this report we refer to these actions simply as ‘first actions.’

Figure 2.1 A simplified comparison of the UK and US patent process



*UK major deadlines: (1) 12 month deadline from the priority or filing date to search request; (2) 6 month deadline to file an exam request after an application is first published; (3) 2 month deadline from the first exam decision not to grant to requesting an amendment (examiners may extend); and (4) statutory deadline that refuses an application which has taken more than 4.5 years to process if the examination is performed earlier than 42 months from the priority or filing date. If the examination report is issued later than 42 months from the priority or filing date, the compliance period expires 12 months from the date at which the exam report is issued.
These deadlines apply subject to extensions as-of-right.
\+ Notice of Grant (UK).
– Refused: application does not meet the requirements of the Act and Rules (UK).

** In the US, once prosecution is closed an applicant may file an after final submission, a Request for Continued Examination (RCE), or an appeal (not pictured). The availability of after final submissions is limited by a 6 month clock which begins when prosecution is closed.
The number RCEs is not limited, but, generally, must be filed within 6 months of prosecution closing.
\+ Notice of Allowance (US).
– Application not in condition for allowance (US).

Note: These are simplified diagrams and do not present all possible actions (UK & US).
Most applications publish 18 months after filing (UK & US).

Source: London Economics 2010 with edits by UK IPO 2013 and USPTO 2013[15]

15 London Economics. 2010. *Patent Backlogs and Mutual Recognition*. A report for the UK Intellectual Property Office, http://www.ipo.gov.uk/p-backlog-report.pdf; and the international event to accompany the launch on 10 March 2010, http://www.ipo.gov.uk/p-policy-backlog-participants.pdf

Our research endeavours to create a taxonomy of pending application stocks based on identifying a set of points where the two offices' processes overlap. After studying the two offices in depth, we arrived at a core set of examination milestones that appear to be common to both our systems, despite any institutional differences. Given the differences in office procedure – and nomenclature – we suspect that these points are common to the processes in other offices. Further discussion and presentations to the WIPO Patent Economist Group[16] confirm this thesis, which provides us with four common milestones in a patent application's life at any patent office:

1. **Received in the Office.** The application is received by the office on a given date. After this point there may be multiple procedural issues, but ultimately all patent applications must be delivered to an office. We call this receipt.

2. **Ready for examiner action.** The application is ready for investigation by a patent examiner. There must be a point in a patent application's life when it is ready for an examiner to open the file. This point in time represents the earliest time that an examiner could undertake a search or examination. In the US this point is defined when the application is first docketed.[17] In the UK this point is defined when the Office receives the search request for the application.

3. **Completed first examination.** An examiner makes the first decision which could have been an allowance. In the US, this is called the first action on the merits; in the UK, this is the first substantive exam. Again, different offices have different examination schedules and systems, but at some point a decision is made which could be (and sometimes is) an allowed application.

4. **Disposed in terminal action.**[18] The patent application is disposed of, and leaves the patent office. This can happen upon grant, or when the application is ultimately refused or abandoned (in the UK or US). This event can be simultaneous with milestone 3 above if the patent is allowed on the first possible occasion and the formalities are completed concurrently.

---

16 The first meeting was held in May 2010 (http://www.wipo.int/econ_stat/en/news/2010/news_0001.html) and follow-up presentations of this framework were given in March and November 2011.

17 Technically, the application may be docketed to a supervisor, who later will docket the application to an examiner. We consider the first docket date as "ready," because the application itself is ready for examination. Significant delay between docketing to the supervisor and docketing to the examiner is generally due to long queues on examiners dockets. Any distinction between the two queues is artificial.

18 Note that in both the UK and the US systems there is no true terminal rejection. Further, although the US system has an examiner action entitled "Final Rejection," it is merely nomenclature and does not, in fact, mean that it is necessarily terminal. In both systems, the terminal states are either grants or abandonments. However, the description allows for other systems to contain a true final rejection.

These four milestones must necessarily occur in a successful patent's route through an examining office. Unsuccessful applications may drop out before the end-point; possible exit points differ based on differences between offices. For instance, pre-decision abandonment is rare in the US, because examination fees are paid at the time of filing and those fees represent a request for examination. It is worth noting that incentives in the EPO process are rather different, because of the requirement for internal maintenance fees on pending applications starting in year 3 from application date - abandonment by failure to pay maintenance fees at EPO is relatively common.

If these four milestones can be identified within the office's administrative data, it is possible to generate comparable application stocks within any office, as described in the next section.

2.1 Defining inventory stocks

The prosecution milestones above allow us to compare offices like-for-like, and construct three different stocks of applications which account for the amount of workload pending between our four application events: Received in the Office, Ready for examiner action, Completed first examination, Disposed in terminal action. This matters in part because the shape and size of each stock appear to have differential impacts on pendency as we show in Chapters 4 and 5. They also allow us to identify bottlenecks and give us a way to compare and contrast the actions of different offices.

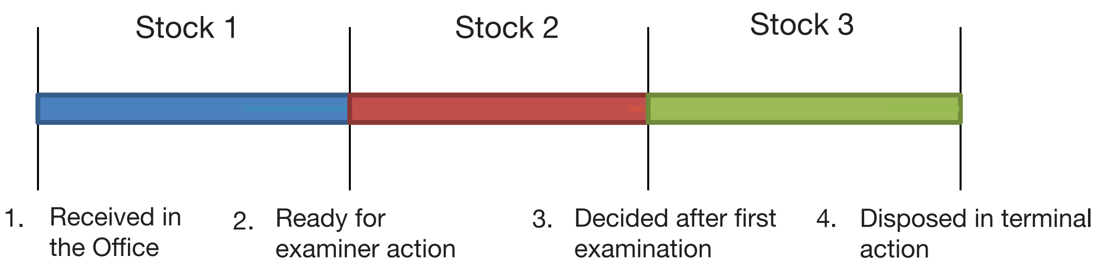
Our framework identifies three patent stocks, shown in Figure 2.2:

Stock 1 (S1). Stock 1 is the stock of applications received by an office but not yet available to any examiner, because it is not yet ready for examiner action. In many offices applications remain in S1 due to the processing time necessary to complete formalities and to allocate the patent according to the office's technological classification system. Applications exit S1 either by becoming ready for examiner action (and moving into Stock 2), or by exiting the office altogether, e.g., through express abandonment (in the US), or by withdrawal and re-filing at the EPO (in the UK).

Stock 2 (S2). Stock 2 is the stock of applications that are ready for examination but have not yet received a first decision. In other words, these applications are currently undergoing initial examination, or they are queued for examination on the examiners' desks or dockets. In the UK these are applications for which a search or combined search and examination has been requested. Within S2, it is difficult to distinguish between queuing time and active processing time by examiners, without knowing more institutional detail. For instance, in the US most of the time spent in S2 is queuing time, because docketing to an examiner occurs as soon as formalities are completed. The active processing necessary for an initial decision amounts to hours or days, relative to the months or even years of queuing time. If, however, offices docket applications to examiners only when examiners become available, then the time spent in S2 may be very short and will consist primarily of processing time. Some offices would consider the sum of S1 and S2 as the "backlog", because these are the patent applications for which no first decision has yet been made.

Stock 3 (S3). Stock 3 is the stock of applications having received a first decision, but not having been terminally disposed of. These applications may be bouncing between the patent office and the applicant, or being reviewed by the office. This can be a considerable part of an office's workload, and is often subject to different behaviour by applicants who aim to delay decisions, or offices which prioritize amendment decision times over other timelines. Even in the case of a positive first decision, there may be a delay between the date the application is allowed and the date it is granted. For instance, in the US, applicants must pay an issuance fee prior to grant.

Figure 2.2 Three stocks of patent application inventories



Source: UK IPO 2013

The exact transition points for the UK and US offices are summarised in Table 2.1 below. It is worth noting that despite differences in the patent system and office actions, our four common milestones can be identified for both the UK and US offices. The most complex of the points is by necessity 'Disposal,' as offices will have different policies for terminating applications that are unsuccessful. The UK does not reject applications as such. If grant is refused on grounds of non-compliance with the Patent Act, applicants are given the opportunity to amend their application. If applicants do not respond to this opportunity within the deadline, the Office does not terminate the application until after the overall compliance date. This can unintentionally inflate the overall application inventory. Similarly, while the US rejects applications, applicants have a variety of possible responses, including amendments and requests for continued examination. Thus the application is never disposed via rejection, but rather via abandonment. In the UK, a granted patent is immediately issued (and disposed of) at the time of allowance, while the US requires an issue fee before making a grant official. The differences in the systems are discussed in more detail in the separate country chapters. In practice, the US "Notice of Allowability" and the UK "Notification of Grant" letter may trigger the same applicant thought process in each country – in the UK it is whether to arrange for payment of renewal fees and in the US it is the actual payment of the issuance fee.

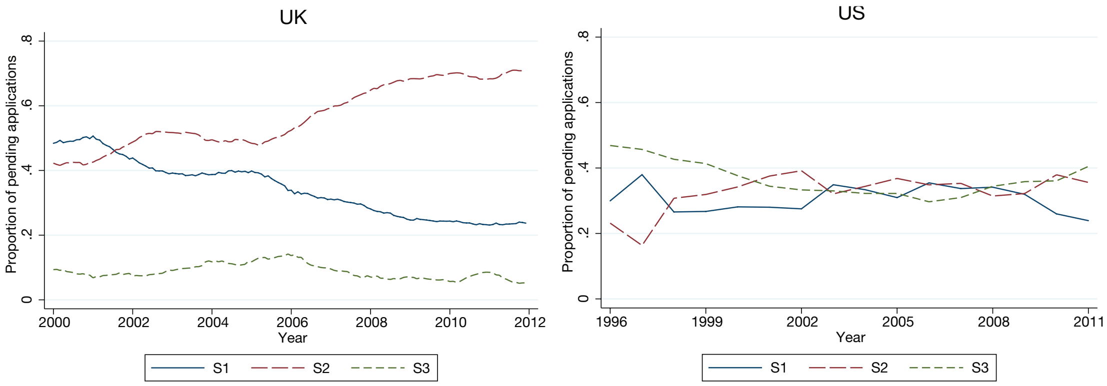
Table 2.1 Mapping the taxonomy onto each office

	UK	USA
1. Received in the Office	Date of priority if applied for in the UK first, or otherwise date of receipt in the office.	Date of receipt in the US.
2. Ready for examiner action	Date when a search request (form 9) is filed and paid.	Date the application is docketed to an examiner and entered into the examination system as actionable for examiners.
3. Completed first examination	At completion of first substantive exam which follows the request for examination.	When First Action On the Merits is completed (typically a non-final rejection or an allowance).
4. Disposed in terminal action	Either on granting a patent (B publication). Or if an application is abandoned through missing statutory deadlines (4.5 years from filing, or other intermediate deadlines).	Upon publication of the patent, after having received the issuance fee. Or, the date of abandonment (after a failure to respond to an office action, or at the time of an express abandonment).

Source: UK IPO 2013, USPTO 2013

This simple taxonomy can aid different offices in different ways. For instance, the UK office has many discrete actions that it performs while applications are within Stock 2, and those are of particular interest to the office. On the other hand, the US has few of these and is more interested in the actions taken by examiners and applicants on applications in Stock 3. This is reflected in the relative size of each stock shown in Figure 2.3. The UK office has an increasing amount of its inventory in Stock 2, while the US until 2000 had the majority of its inventory in Stock 3, although the stocks now appear more evenly distributed. For each office it may be useful to create sub-stocks based on features that are important to that office. For instance, the US office has a particular interest in those applications that have been taken to appeal, or those applications that have been subject to a request for continued examination (RCE). The main point is that the framework allows both offices to have a super-structure for making international comparison, and it also allows each office to drill down where they suspect a lot of activity is occurring.

Figure 2.3 Different structures in UK and US application inventory



Source: UK IPO 2013, USPTO 2013

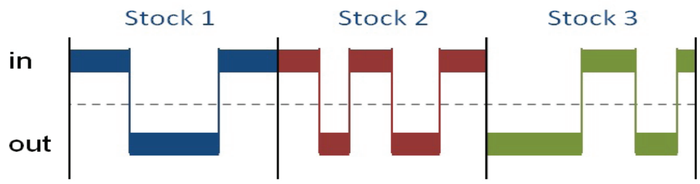
Figure 2.3 shows that there has been a relatively stable relationship between Stock 1 and 2 in both the UK and US. As applications leave Stock 1, Stock 2 should grow unless the flow from Stock 2 to Stock 3 matches. Indeed for all these stocks there should be a relatively straightforward relationship as applications move from one stock to the next (save those applications that exit each stock through abandonment). But, there are more complex dynamics at work. In the US we have seen evidence for bottleneck effects from Stock 3, which affects the magnitude of Stock 2. We explore these effects in more depth in Chapter 5.

2.2 Queuing time within and outside the office

While total pending stocks are relatively easy to count in our offices, they do not, on their own, provide much insight into where possible bottlenecks exist in the process, or how much queuing any given patent application does – and queuing is a major issue. One major consideration is whether applications are queuing within the office, or queuing with the applicant.

To answer this question, we need to identify the distinct points when a patent application is returned to the applicant, and then again identify when the applicant returns the required response. For example, the UK office requires that applicants submit a form to request examination after a patent search is completed (and usually published). Traditionally the key dates are recorded on published and granted patents; and, with electronic management systems, receipts are generated automatically and logged on a database. Therefore, we know that the UK has records for when an exam request was filed, and – along with a record of when the search report was sent to the applicant – we can calculate the time each patent spent with the applicant at this stage. In theory, it should be a simple exercise to identify the major 'nodes' for exchanges and thereby identify when an application is within or outside the office (See Figure 2.4).

Figure 2.4 Determining Applicant and Patent Office stocks



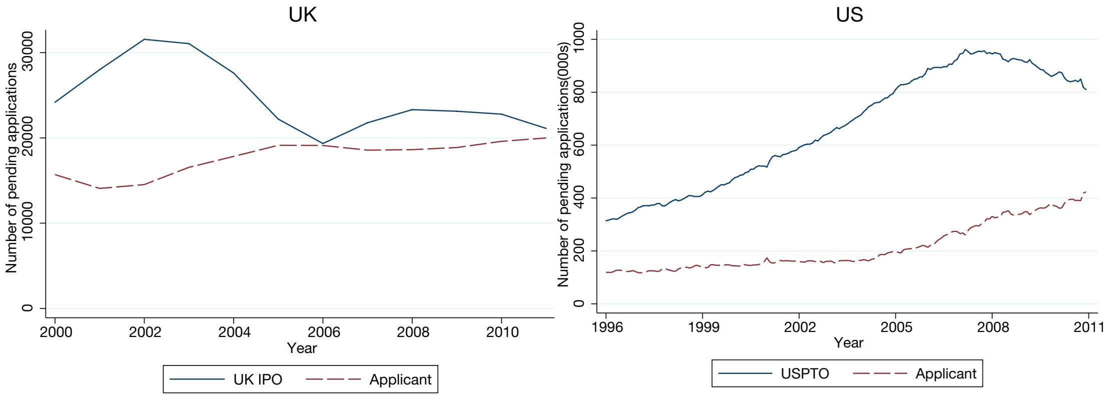
Source: UK IPO 2013, USPTO 2013

In practice, however, this effort can quickly become very complicated, depending on the structure of internal databases. The USPTO tracks patent prosecution using more than 2000 event codes and 300 application status codes. Identifying the most common transition points between applicants and examiners is straightforward; however, exceptions abound.

For the UK data we did not find any major points in Stock 1 where a large portion of applications go back and forth between the office and applicants. In Stock 2 on the other hand, there is an active exchange from the time a search has been requested to the time the first exam report is issued. There are a number of deadlines, and applicants can file requests for additional searches, can combine their search and exam, and may reply to the office search report. Stock 3 can contain applications which are awaiting amendment to claims as a result of adverse examiner decisions, or applications which applicants intend to withdraw or allow to lapse.

In the US data, there were some interactions in Stock 1 as applicants need to complete a legal filing before it can move to Stock 2. However, those incidences represent a very small fraction of applications. Within Stock 3, on the other hand, there is a significant amount of back and forth between examiners and applicants. Each transition comprises queuing time (either on the examiner's or applicant's desk) and processing time. As Figure 2.5 suggests, there has been a growing stock of patent applications sitting outside both patent offices. This has implications for the application inventory itself, because offices cannot actively do anything about these applications. Additionally, it has a bearing on resource management because a proportion of applications that go out will inevitably come back into the office and re-enter the queue.

Figure 2.5 Application inventory stocks within and outside the office in the UK and US



Source: UK IPO 2013, USPTO 2013

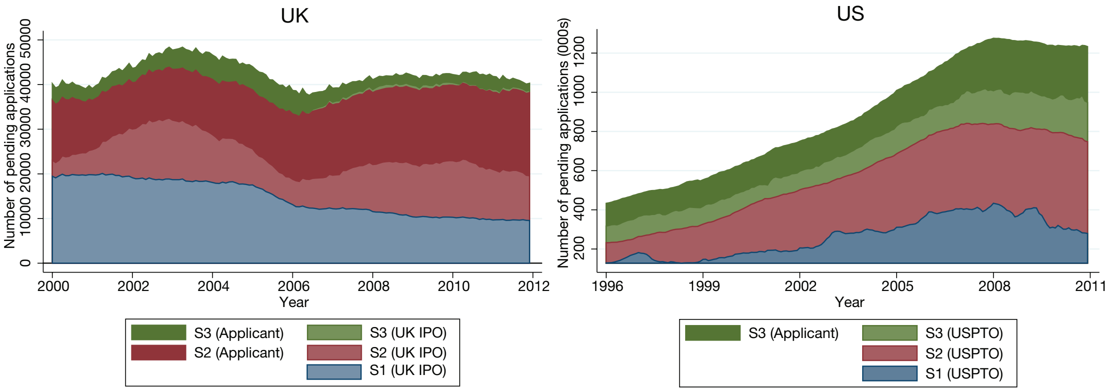
It is worth noting the size difference between offices to highlight that our framework delivers comparable results regardless of the type of office. There is also a story of how the US composition of the stocks has changed as the office appears to have increasingly shifted applications into the hands of applicants. As illustrated in Figure 2.5 the UK office came close to balancing the two in 2006, and has in the last year apparently shifted the majority of its inventory into the hands of applicants. In terms of 'dealing' with the inventory of applications there are different elements to consider – for example, the UK could change deadlines. But, stricter deadlines would mean that more work is returned to the office by applicants on a faster timescale. The US proposal to have three tracks for applications has some scope to allocate examiner resources to those applications where speed of prosecution is more important.[19]

Looking at the US office in Figure 2.5, a third of the application inventory sits with applicants. Figure 2.6 demonstrates that the vast majority of outside applications are within S3, and the magnitude has been rising steadily.[20] However, as a proportion of S3, applications that are outside the office have varied: 56% of S3 was with applicants in 1996 compared to 58% in 2011. It peaked in 2002 at over 70%.

19 To date, only two tracks have been implemented: the standard track, and the fast track, which allows applicants to have expedited prosecution. The third track was intended to be a "slow" track, which would prevent applications from entering stock 3 for a designated period of time. See USPTO Press Release, 10-24. http://www.uspto.gov/news/pr/2010/10_24.jsp

20 Applicant stocks within S1 and S2 are so small as to not be graphed. Time in S1 is almost exclusively office processing time, and time in S2 represents almost entirely queuing time.

Figure 2.6 UK and US application inventory inside and outside the office by stock



Source: UK IPO 2013, USPTO 2013

There are some significant trends in the US stocks overall as shown in Figure 2.3. Stocks 1 and 2 appear to be somewhat negatively correlated, which is especially noticeable between 1996-2003 and 2009-2011. During periods where initial processing speeds up, applications will move quickly from Stock 1 to Stock 2, causing the negative relationship. Further Stock 3 experienced a downward trend from 1996 to 2006 in terms of overall proportions. The trend has reversed itself between 2006 to 2011.

Looking at the UK office in Figure 2.3, there are vast differences in the quantities of Stock 2 and 3. Looking at the left hand diagram in that panel, it appears that Stock 3 has been relatively stable over the last ten years, while Stock 2 has expanded quite substantially, especially compared to Stock 1. But consider Figure 2.6; what matters is how many applications are outside the office – in applicants' hands – and in particular the large quantity of Stock 3, as well as the growing quantity of Stock 2 which is sitting with applicants (in the darker shades below) and which the office cannot actively action.

In Figure 2.6, Stock 2 within the UK office has been going up and down over time, but the quantity of Stock 2 outside the office has been growing steadily, and jumped in early 2010. There are perhaps two reasons for this: either Stock 2 applications are staying for longer or the UK is getting more applications which go through to that stage. The answer seems to be a bit of both. The number of search requests filed at the IPO as a proportion of applications has been rising over the last decade. But part of this increase can be explained by policy decisions to target search requests after 2005/06. Another part of the jump in 2010 is probably in response to new targets to reduce the number of older pending applications, but there is much left unexplained, and we suggest some answers to this in Chapter 4. Stock 3 is also quite telling, in that these are post examination actions, and here the office appears to hold very little stock, while applicants hold quite a lot. This could be a reflection of pendency issues which is what the next chapter will discuss.

For the UK it is likely that some applicants leave priority UK applications (which are relatively cheap to search and examine) in Stock 2 or 3 while they pursue a European Patent (EP) or PCT application in parallel using the UK priority date. They may then abandon the UK application later to pursue wider protection. Such applications will appear in the UK IPO application inventory, but will not necessarily be seen as “backlog” by an IP attorney handling the application.

In constructing application stocks we looked to extract the relevant dates from patent applications. When we identify the nodes marking office-applicant exchanges for each patent application, we are in a position to analyse the stocks as they change in the office. The common framework allows our two offices to discuss the application inventory in common terms, and to measure directly how the inventory in a particular office responds to policy changes.

For other offices, so long as computer records are retained, the process of making comparisons would consist of identifying these points in the local patent application’s life. We believe this framework can usefully be applied to other offices interested in analysing their own application inventories and pendency, and wanting to cooperate on multi-lateral research. Because backlogs and pendency are so inter-related, it is important to examine pendency in detail, which we do in the following chapter.

3. Pendency

Arguably one of the largest concerns with respect to patent inventories is that larger inventories result in longer pendency. For applicants, that is the time to grant of a patent which can be used to protect an invention in the market, to raise funding, or to collect licensing revenues. There is also an important impact on non-applicants from excessive delays: they may face uncertainty over the technology they can use or over uses to which it can be put if pending applications cover part of the technology space in which they are active. If large inventories could be examined expediently, then the damaging effects of uncertainty would be reduced.

Much of the political discourse with respect to backlogs is directed at pendency. In fact, it can be argued that large stocks of pending applications are significant primarily because they cause delay and uncertainty in securing patent rights for current and potential innovators. That is, the main cost associated with backlogs is increased pendency.

However, applicants do not uniformly benefit from decreased pendency. And, in fact, some applicants may prefer some delay. Prolonged examination may provide applicants with more time to obtain information about the commercial or competitive landscape, which may lead to more informed decisions with respect to continued prosecution.[21] Additionally, some applicants—for example, those in the pharmaceutical industry—may face regulatory delay at other agencies, so pendency at the patent office is not a bottleneck. These applicants may be less sensitive to patent pendency.

Nonetheless, for a large group of time sensitive applicants, faster prosecution is desirable. Applicants in quickly moving technologies—or those with short product life-cycles—may desire prosecution speeds that mirror the speed of technological progress. Entrepreneurs who rely in part on patents to secure capital may benefit from reduced pendency. Applicants with well-defined and easily commercialized inventions are harmed by delay, because each additional month in the queue means one less month of exclusivity unless potential competitors are deterred by a pending patent. And, applicants cannot be made worse off by having at least the option for faster prosecution. That is, increased pendency removes the possibility for faster prosecution and thus harms at least the time sensitive applicants.

21 However, the fact that a particular applicant may find delay to be privately beneficial is not sufficient to infer that delay is socially beneficial. As mentioned above, delay creates uncertainty for others operating in the same technology space. General economic principles suggest that as a rule, less uncertainty is better from a social perspective.

3.1 Measures of pendency

Unfortunately, just like backlogs, there is no consistent single definition of pendency. The USPTO Data Visualization Center (the "Dashboard")[22] lists no fewer than nine different measures of pendency. Our methodology for international comparisons is useful in examining pendency across offices, and also for helping offices to predict expected pendency in a consistent way. The general method of measuring pendency through entry or exit is discussed below.

We assert that an applicant's decision to enter the patent office is sensitive to the pendency it expects to experience. Long expected pendency has at least two direct impacts on applicants. First, it delays the patent grant, reducing the ability of a patent holder to appropriate value from exclusive rights to the innovation. Second, greater delay in obtaining a patent may lead to higher abandonment rates, whether due to financial constraints, applicant discouragement, missed commercial opportunities, or simply because more information about the commercial opportunity is available to the applicant by that point in time. The statistical analysis in Chapter 5 confirms this second hypothesis, although we cannot ascertain the specific reason for the higher abandonment rates. We should also expect long pendency to increase costs for others. For the patent process itself it is likely that repeated stopping and starting, and multiple interactions between offices and applicant spread over extended periods, will raise the overall cost of search and examination, which is ultimately borne by all applicants. In addition, the uncertainty caused to third parties in technology markets by the presence of large numbers of pending applications may create a barrier to entry even where the applications are ultimately refused.

In discussing pendency, we find it useful to define some measurements of pendency:

- **Entry pendency** is based on the filing cohort. It measures the average pendency for all applications filed in a given time period. Entry pendency is only accurate if sufficient time has elapsed to observe a large majority of disposals. It does not account for the omission of still pending applications from that cohort. However, if all applications have been disposed, entry pendency is a perfect measure for the actual pendency of all applications in that filing cohort.

- **Exit pendency**, the commonly applied pendency measure, is based on the exit cohort (all applications exiting in the same time period). That is, exit pendency measures the average pendency for all disposals in a given month or year. It does not account for the observed pendency of applications still being processed, which leads to a bias.

- **Expected pendency** is the pendency that an applicant expects to experience (on average) on the day of filing. This actual expectation is unobserved; however, we construct estimates of expected pendency based on the predicted pendency from survival time regressions, described in Chapters 4 and 5 below.

22 USPTO Data Visualization Center, http://www.uspto.gov/dashboards/patents/main.dashxml

Exit pendency and entry pendency are reasonable approximations for current expected pendency under some assumptions, and each is relatively easy to calculate with current data. However, each has recognized limitations.

Because entry pendency cannot be calculated prospectively, it is always several years out of date. In contrast, exit pendency may not reflect current conditions, especially during periods of patent office productivity change. For instance, throughout 2011 and 2012, the USPTO pursued the Clearing the Oldest Patent Applications (COPA) initiative.[23] A program like COPA may increase examination throughput or production, and, accordingly, should decrease the expected pendency of incoming applications. Yet, because COPA prioritizes clearing older applications, exit pendency may temporarily increase.

A further complication exists in terms of understanding pendency. From an accounting standpoint, abandonments count as "disposals" and may cause traditional measures of pendency to understate actual pendency. That is, suppose that the time to obtain a patent increases by 30% across the board. If all applicants continue to seek patent protection, then observed pendency would increase by approximately 30%. However, if a proportion of applicants (especially those suffering the greatest delay) give up, then the abandonment rate increases. Because abandonments count as disposals, an increase in abandonments may cause the observed pendency to decrease relative to the 30% increase. For instance, we may only observe a 20% increase in pendency. The US survival estimates in Chapter 5 provide a way to disentangle the two effects, by using the competing risk methodology.

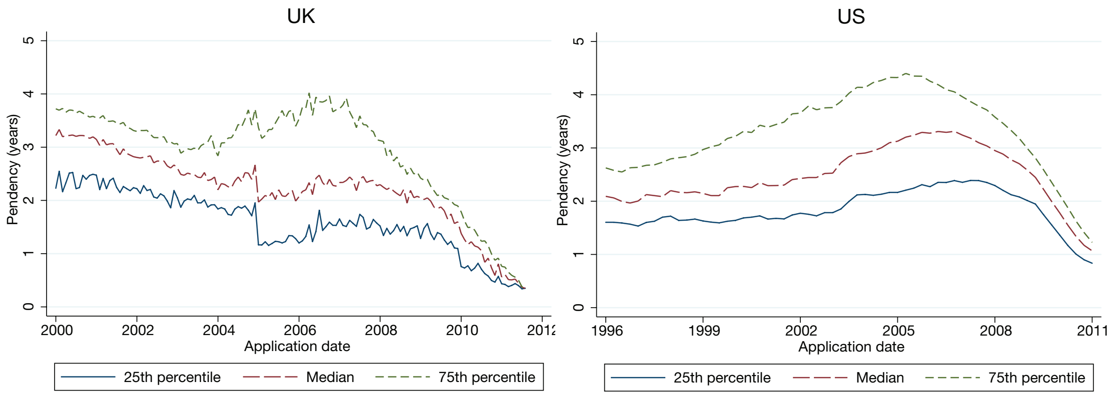
3.2 Entry pendency

As mentioned above, entry pendency groups together all the patents applied for at a particular point in time and measures how long they spent as pending applications. The major drawback with this method is that it is not an accurate measure of pendency until a large majority of the cohort is already disposed of. For instance, one might expect the US 2003 filing cohort to have been fully disposed by 2011, but in fact some applications are still pending. These issues mean that entry pendency measurements are several years old before they are accurate.

Looking at the median entry pendency over time in Figure 3.1, we can see that the UK entry pendency for granted applications was falling until 2004, and then dropped substantially – by some six months at the median after 2004. After then entry pendency appears to stabilize, and we begin to see truncation in the dataset from 2008 onwards. From that point onwards, if an application is granted it will have happened in less than three years (as the dataset ends in December 2011) so the indicator is quite biased. Similarly, in the US entry pendency based on applications filed after 2007 is biased downward due to censoring. Censoring in this case means that we cannot observe the pendency for those applications that have not experienced a terminal disposal (i.e., they are "censored"). If we ignore those observations, it leads to a downward bias because each of them will have a longer pendency than any of the previously disposed applications. Fortunately, survival regressions are able to account for censored observations using well established statistical techniques.

23 http://www.uspto.gov/news/pr/2011/11_08.jsp

Figure 3.1 Entry pendency in the UK and US

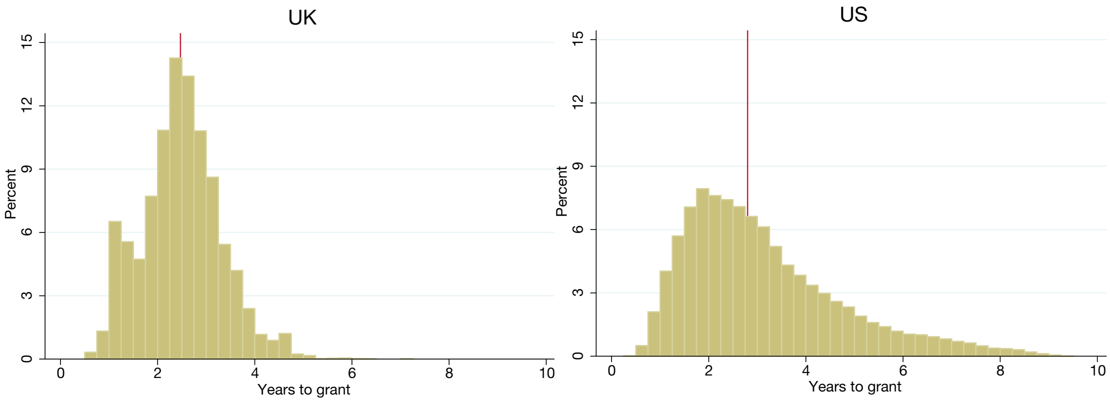


Source: UK IPO 2013, USPTO 2013

Note: The US and UK graphs present entry pendency for granted patents only. UK entry cohorts are monthly. US entry cohorts are quarterly.

Figure 3.1 gives us a trend for entry pendency, but it is important to understand that we are only looking at the median and interquartile range grant times, and it is important to consider the distribution of these applications. To get an idea of the distribution, Figure 3.2 considers the 2003 filing cohort, meaning all applications filed in 2003. In the US case, we see that the median pendency for 2003 filings is less than three years, with a single peak at two years and a relatively long tail. Some grants occur nine years after application (which is the end of the dataset, so there may even be some truncation here). In the UK data we see two peaks, the first just after year one and the second after two and a half years. Part of the explanation for that lies in the UK's dual track system where applicants can choose to split the application into search and exam requests or have them all at once – as discussed in Chapter 4. The median grant time for the 2003 cohort in the UK is a little lower than the US, at just under two and a half years, and the distribution is much more concentrated around the median.

Figure 3.2 Entry pendency of granted applications from the 2003 filing cohort in the UK and US



Source: UK IPO 2013, USPTO 2013

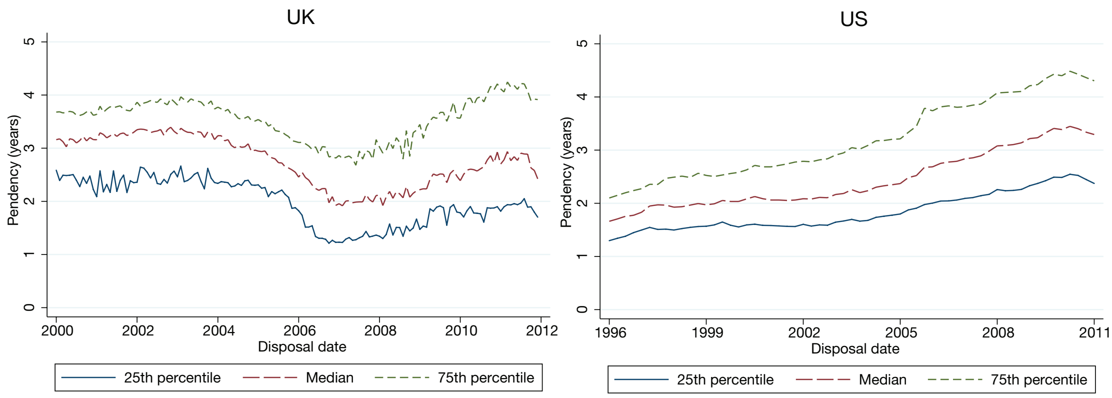
Note: The US and UK graphs present entry pendency for granted patents only. The vertical line represents median entry pendency.

3.3 Exit pendency

As discussed above, exit pendency measures pendency at the date of actual disposal. Instead of aggregating the cohort based on the time of filing, we instead calculate the age of all applications exiting at a given time. This measure of pendency is inherently retrospective, and is by far the most oft used measurement of pendency.

In Figure 3.3 we examine the median and interquartile range exit pendency over the last decade for the UK and the US. UK pendency for granted patents declined gradually: falling from a median pendency of 3 years in 2002, to 2 years in 2007. It has been slowly increasing since then, but some of that is shaped by the drive in the office to process the oldest applications first, and this is the same picture we see for the US office. At the USPTO, docket management changes in the fall of 2009 and the COPA program in 2011 and 2012 both incentivize working on older applications, thus exit pendency may increase even while throughput increases. Offices appear slower if they grant applications that have been sitting around the system for longer, which can cause incentive problems if offices are sensitive to the political fallout from "bad" pendency numbers.

Figure 3.3 Exit pendency in the UK and US



Source: UK IPO 2013, USPTO 2013

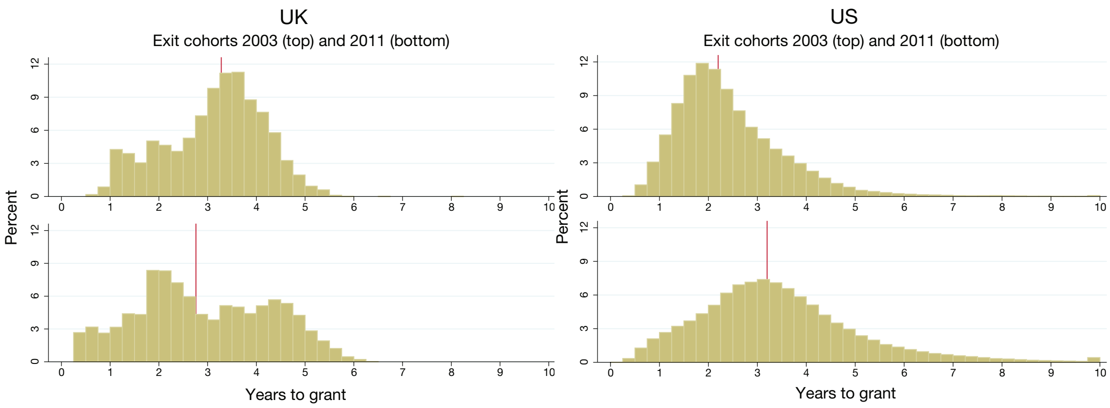
Note: The US and UK graphs present exit pendency for granted patents only. UK exit cohorts are monthly. US exit cohorts are quarterly.

As mentioned above, while exit pendency provides up to date measures, it is a potentially biased measure of expected pendency because it does not account for those applications that are still pending. Worsening the problem is the fact that this measure will not tell you if this year's disposed patents were the oldest in the queue. We see some of that effect in the 2009-11 data for both offices as both had programmes of processing the oldest application (by application date) first. The bias is particularly acute anytime there are sudden structural changes at the patent office with respect to arrival rates of new applications, the hiring of more examiners (or other institutional changes that would affect office throughput), or policy changes that focus on disposing of older applications.

The advantage of using exit pendency is that it can be observed right up to the end of the sample period, and it is easy to calculate. For this reason, it has become the default measure of pendency at most patent offices despite the known bias. However, it is well known at patent offices that exit pendency is an inaccurate predictor of future pendency, especially during periods of productivity change.

If we look at two different exit cohorts we can see these biases playing out. Figure 3.4 shows the exit pendency for applications granted in 2011 and 2003 in each office. What becomes apparent is that the UK office has two peaks, but the median time to grant has fallen from 3.3 years to 2.7 years over this period, with pendency being much more evenly distributed over a 1-6 year period. The US picture is somewhat different. First of all, it is single peaked. But, the median has shifted from two and a quarter years in 2003, to three and a quarter years in 2011, with pendency less concentrated around the median.

Figure 3.4 Exit pendency for filing cohorts 2003 and 2011 in the UK and US

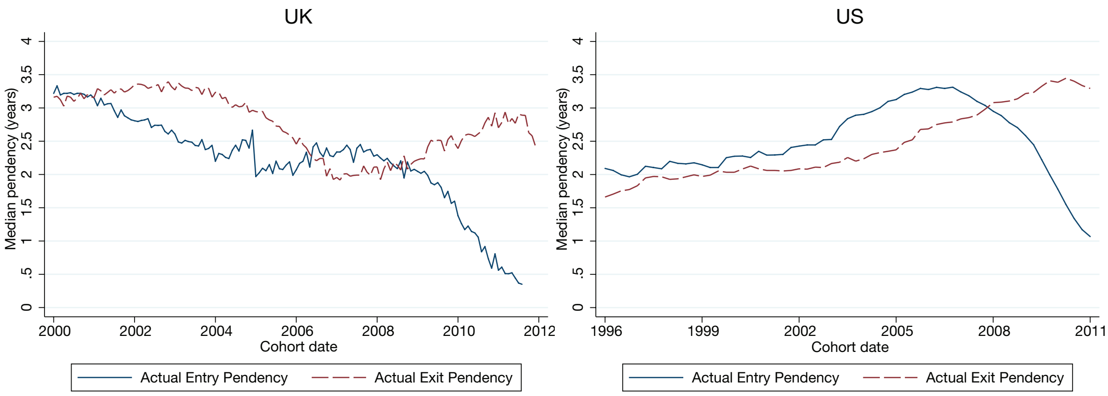


Source: UK IPO 2013, USPTO 2013

Note: The US and UK graphs present exit pendency for granted patents only. The vertical line represents median exit pendency

Figure 3.5 combines the median entry and exit total pendency measures from each office (Figures 3.1 and 3.3). In the UK, entry pendency and exit pendency begin at approximately the same level, before diverging from 2001 to 2007 with exit pendency frequently more than six months higher than entry pendency. UK entry pendency then rises above UK exit pendency until 2009, at which point there is significant censoring which leads to a downward bias in the observed entry pendency. The pattern for the US is different in that entry pendency is consistently several months higher than exit pendency from 1996 to 2003. The difference between entry and exit pendency grows to about nine months between 2003 and 2006, at which point entry pendency begins to fall, in part due to censoring.

Figure 3.5 Comparison of entry and exit pendency in the UK and US



Source: UK IPO 2013, USPTO 2013

Note: The US and UK graphs present total pendency measures for granted patents only. UK cohorts are monthly. US cohorts are quarterly.

Our point is that current methods for measuring pendency are inaccurate estimates of expected pendency, and do not reflect the information relevant to the applicant. It does not answer the basic question: “If I apply for a patent today, how long will it take before it is granted?”

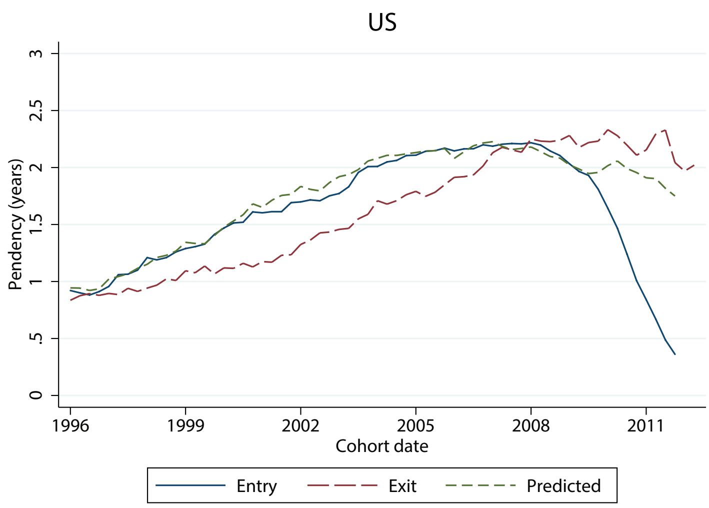
3.4 Expected pendency

The framework developed in Chapter 2 facilitates the estimation of expected pendency using survival analysis. This methodology is explained in more detail in the country-specific analyses in Chapters 4 and 5 and in Section 3.5, below. Pendency is measured as the time until patenting, using only observable characteristics at the time of filing. Survival regressions are used to predict the pendency of applications *at the time of filing.* Because survival analysis controls for the censored pending applications, the estimations can be used to calculate an unbiased measure of expected pendency. This allows us to illustrate the actual pendency of any cohort of patents that have already passed through the system, and use this information to predict expected pendency of new patents.

The detailed analyses in Chapters 4 and 5 delve more deeply into the relationship between stocks and pendency. Because of the different institutional structures, the specific analyses are quite different. For instance, in the US it is important to distinguish between the time between filing and first action (first-action pendency) and the time from first action to disposal (post-first-action pendency). However, before moving on to the separate analyses, it is useful here to demonstrate the benefit of estimating expected pendency using predictions from survival regressions.

In contrast to figures 3.1 through 3.5, Figure 3.6 graphs first-action pendency according to several different metrics for the US. We compare estimates of the time from filing to first action using exit pendency, entry pendency, and predicted (expected) pendency.[24] Predicted pendency tracks entry pendency quite well through about 2009. After this point, entry pendency based on applications filed after 2009 is biased downward due to censoring issues. After 2009, exit pendency is a better approximation for expected pendency; however, in part due to COPA, exit pendency is biased upwards in comparison to the expected pendency for newly arriving applications.

Figure 3.6 Entry, exit and predicted first-action pendency in the US



Source: USPTO 2013

Note: We compare predicted first-action pendency, estimated using survival regression methodology, with the actual observed mean entry and exit pendency for the same quarter.

To more precisely ascertain the relationship between stocks and pendency requires more careful statistical analysis in the form of survival time regressions. Those are described in detail in Chapters 4 and 5, and their related appendices.

24 First-action pendency lends itself more easily to predicted pendency. See Chapter 5 for a more detailed discussion.

3.5 Survival time methodology

In the focused country analyses (Chapters 4 and 5), we employ several different statistical methods for analysing the relationships between application characteristics, the various stocks of pending applications and other explanatory factors. Each method uses the data in a different way and can be used to answer different questions. We do not propose one method over another in any prescriptive way; each has its strengths and limitations. Rather, the approaches should be viewed as complementary.

Kaplan-Meier survival curves are a non-parametric method of summarizing duration data for one or more groups. A survival curve shows the proportion of subjects which have not yet experienced the event. In this context, “non-parametric” means that the survival curves are derived from observed disposition times, and they are analysed graphically. No particular mathematical relationship is assumed between observable characteristics and the shape of the survival curves. This facilitates the comparison of the actual duration experience across different years or groups. The comparison is typically done graphically, by plotting the survival curves for selected groups. See Chapter 4 for examples.

In contrast, parametric survival time regressions are used to estimate the impact of particular explanatory factors, while holding all other factors constant. Here “parametric” means that some mathematical relationship must be assumed between the explanatory factors and observed pendency. The parameters of that relationship can be measured, tested, and used to make predictions. For instance, one may want to know whether applications filed in 2005 and 2010 would have had the same duration, had all other observable characteristics (like application stocks and examiner counts) been the same. Regressions allow such a hypothetical comparison. However, regressions rely on assumptions about the data and the particular mathematical relationship. For instance, standard linear regressions assume that the relationship between the explanatory variable and the response variable is linear. Chapter 4 uses fully parametric survival-time regressions, but most results are reported graphically using the non-parametric survival curves. See Appendix C and Appendix E for detailed explanations of the parametric models employed.

Chapter 5 also uses fully parametric models, rather than the commonly used semi-parametric Cox proportional hazards model.[25] While the Cox proportional hazards model[26] makes fewer assumptions about the particular shape of the survival curve, which makes it quite flexible, the parametric models lend themselves more easily to making predictions, including those necessary for analyzing survival time when there is more than one outcome, as with patenting and abandonment. Such models are called “competing risk” models, which we use in Chapter 5 to analyse the competing risks of patenting and abandonment. The model examines whether there are significant differences between the two types of disposals. These models require more stringent assumptions to be met in order to ensure accuracy. The competing risk models are used to generate cumulative incidence functions (CIFs), which are similar to survival curves except that they show the proportion of all applications that have either been patented or abandoned.

25 For more details about the differences between semi-parametric and parametric methods for survival regressions, see Box-Steffensmeier, J. M.; Jones, B. S. (2004). *Event History Modelling: A Guide for Social Scientists.* Cambridge: Cambridge University Press. Cleves, M.; Gutierrez, R. G.; Gould, W.; Marchenko, Y. V. (2010). *An Introduction to Survival Analysis Using Stata* (3rd edition). College Station: Stata Press. Chapter 1, 17.

26 Both the UK IPO and USPTO, explored the Cox proportional hazards model, but found the reported models more readily interpretable and more suitable for the analysis.

4. UK analysis

This chapter sets out how the UK patent system operates and identifies some key characteristics which are important for understanding the backlog and pendency issue. It shows that the size of different patent application stocks, particularly Stock 2 outside the office (stocks that are with applicants waiting for an additional search or exam request) increases the pendency of patent applications in the UK. We investigate a number of discrete factors which can be addressed to reduce the inventory. Our analysis takes into account the characteristics of patent applications, the potential bottlenecks and capacity of the office, and finally the behaviour of applicants.

4.1 A brief introduction to the UK patent application process

The UK patent application process is an applicant driven system where applicants have to submit specific forms and payment at the application, search and examination stage. Applicants therefore have a discrete choice to continue with their patent application at each stage. See Appendix B for more details on the application process at the UK IPO.

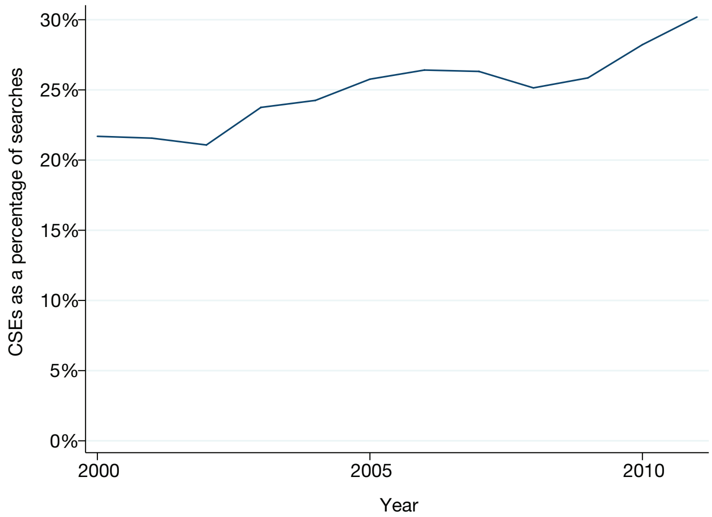
Throughout the system there are inbuilt deadlines, so for example if an applicant does not file an exam request within six months of the application being published, the office can consider the application abandoned.[27] Unlike the US process there is also a statutory cut-off where applications older than four and a half years are considered abandoned as the prosecution has taken too long (although there are extensions available and the office can extend the prosecution time).[28]

27 The bigger deadlines include a 12 month deadline from application to search request; applicants have six months to file an exam request after an application is published; there is a four month deadline from the first exam decision not to grant to requesting an amendment (although examiners have discretion to extend this); as well as a statutory deadline which kills an application which has taken more than four and a half years to process, unless extension requests have been filed.

28 For full and proper advice on the UK patent rules and procedures please consult the Manual of Patent Practice available on the IPO website: http://www.ipo.gov.uk/p-manual-practice

More importantly, UK applicants have the option of a ‘Combined Search & Exam’ or CSE application. CSEs have been an increasing proportion of UK applications, as illustrated in Figure 4.1 below, growing from just over 20% of applications in 2000 to 30% in 2011. With a CSE, the applicant files the search and examination request at the same time (often together with the application) and pays all the fees at once. This has the benefit of ensuring that the search and exam are done together. This is similar to the US standard process, but there is a marked difference in the pendency of CSE patents as compared to ‘standard’ patent applications in the UK. This is explored in section 4.3.

Figure 4.1 Percentage of UK search requests that are CSEs



Source: UK IPO 2013

4.2 A brief explanation of how we assess the impact of patent inventories on pendency

In order to assess the impact of patent backlogs and other features of the patent system on pendency time, we use a range of different statistical techniques. These are explained in Chapter 3.5 and the outputs are presented in Appendix C. The UK IPO's statistical analysis is done in two steps:

Linear regressions – In the first stage, we run linear regression analysis to identify any major influences on pendency. For this we only consider all the applications filed in 2001 because we are sure this entry cohort has made it through the whole examination process by the time of this report.[29] This approach allows us to observe not only time to grant but also time to abandonment or final rejection, which typically becomes known quite late in the examination process. The linear regressions also allow us to break-up pendency time and observe the impact that various explanatory factors have on time spent by an application in different stocks of the application inventory.

Survival time regressions – In the second stage we run survival time regressions for all patent applications filed between 2000 and 2011. Survival time regressions allow us to assess how a change in one of the various explanatory factors would affect expected pendency time. Running survival time regressions, for instance, helps us determine how a doubling of application inventories would affect the time we expect an application to spend in the system before being granted. Unlike the linear regressions, the survival time regressions compute expected pendency time to *grant* – not additional time spent in each sub-stock. We focus on pendency time to grant for two main reasons. We expect that at the outset of the process applicants intend to get a granted patent, and so what is of interest is how long it will take to successfully get through the patent system, not how long the average failed application takes.[30] Secondly, if we considered all patent applications in measuring pendency we would include all the applications that are abandoned. Because abandonment happens at the expiry of a deadline the data becomes heavily biased toward a few pendency numbers. Given the statutory cut-off for applications at four and a half years and the UK IPO's deadline for filing search requests within 12 months of applying, the vast majority of applications are pending for either 4.5 years or 12 months and are then automatically discarded. Including these applications would give us very uneven estimates for average pendency, and for what impacts pendency as the applications could have been abandoned for any number of reasons and at various points.

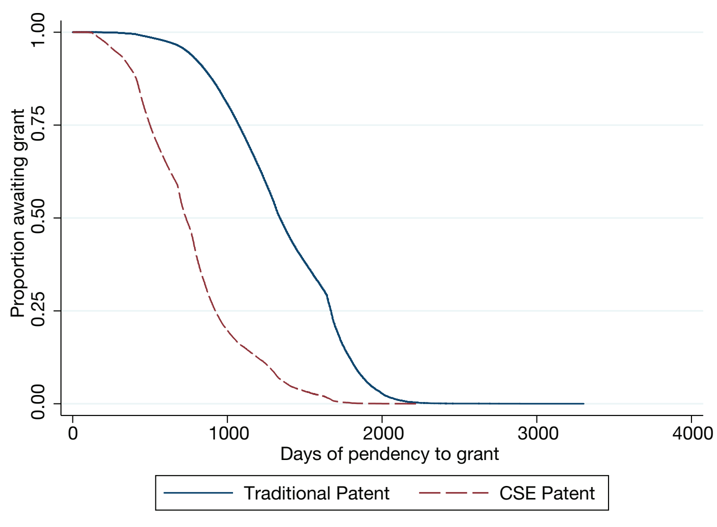
29 As applications filed in 2001 only represent one cohort of patents, we also performed linear regressions on applications filed after 2001 to test individual relationships. The results are similar to the results for 2001.

30 We know that, in the UK, applicants do lodge applications with the intention of abandoning them and pursuing elsewhere an application based on the same priority, this typically happens when an applicant decides to pursue their application at the EPO. But these applicants would not be concerned with the pendency time for the patent application as a whole as they are usually filed to establish priority or to receive a quick search report.

4.3 Combined Search and Examination "CSE" matters for pendency

When we separate applications into CSE and standard patents it quickly becomes clear that CSE applications take substantially less time to grant than standard patent applications. Figure 4.2 presents survival curves and these illustrate the length of time you would expect a CSE application (in red) or a standard patent application (in blue) to take before being granted. As shown, 100% of both CSEs and traditionally filed applications are still awaiting grant at 0 days from the filing date. However moving along the y-axis to the 0.75 mark, it becomes obvious that there are substantial differences between CSE and traditionally filed applications. More than 25% of CSE patents are granted within two years, while the first 25% of standard patents will only be granted within almost three years. Consider that 75% of all CSE patents in this cohort were granted within three years, and the difference is even starker. This provides a large part of the explanation for the dual-peaked UK entry and exit pendency graphs we showed in Chapter 3 (See Figures 3.2 and 3.4).

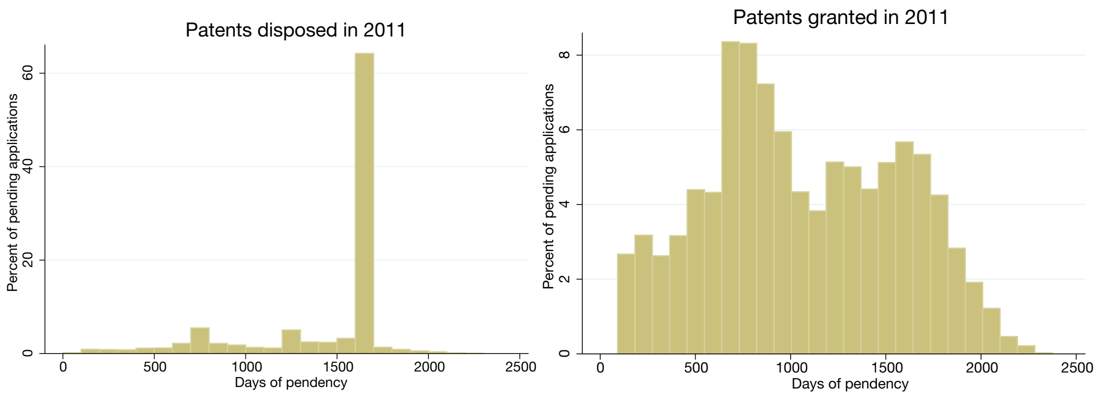
Figure 4.2 Survival analysis of CSE and regular UK patent applications, 2000-11



Source: UK IPO 2013

The figures below show how the average pendency of all applications (on the left) and just the applications which are granted (on the right) differ. The figures are histograms of total days pending, and count the number of days on the *x*-axis. In the right-hand graph, the spike between 1,600 and 1,700 days are the patents being terminated after four and a half years, which is the UK's statutory cut-off. For granted patents on the right we see the two spikes from Figure 3.4 around 8-900 days and again at 1,300-1,400 days. This suggests a much shorter turn-around and a very different distribution of pendency for granted applications compared to the full population of all applications (granted and abandoned applications together).

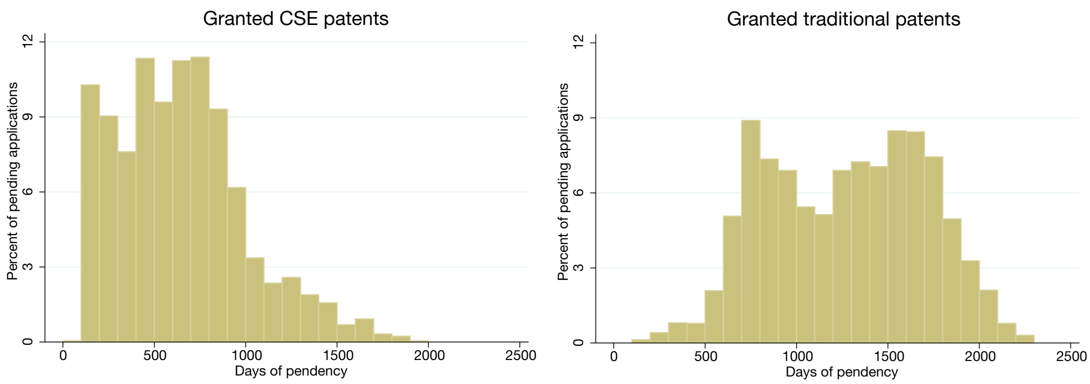
Figure 4.3 Pendency of patents disposed of (left) and patents granted (right) in 2011



Source: UK IPO 2013

The double-peak in granted patents above is only partly explained by the different pendency times of CSE applications. As illustrated in Figure 4.4 below, the most common exit pendency for CSE patents is between 1 and 2 years. Part of the reason for this peak lies in the fact that the office cannot grant a patent until after it is published, which is 18 months from priority, to allow for third party observations. In other words, the take away is that CSE and traditional patents are very different but CSEs do not necessarily help to explain the double peak in the distribution of pendency time for granted patents in the UK.

Figure 4.4 Pendency of patents disposed in 2011 – Granted CSE patents (left) versus granted traditional patents (right)



Source: UK IPO 2013

Note: that the distribution of CSE patents ends at 2,000 days, but the distribution of traditional patents ends at 2,500 days, suggesting a much longer pendency for the standard patent applications.

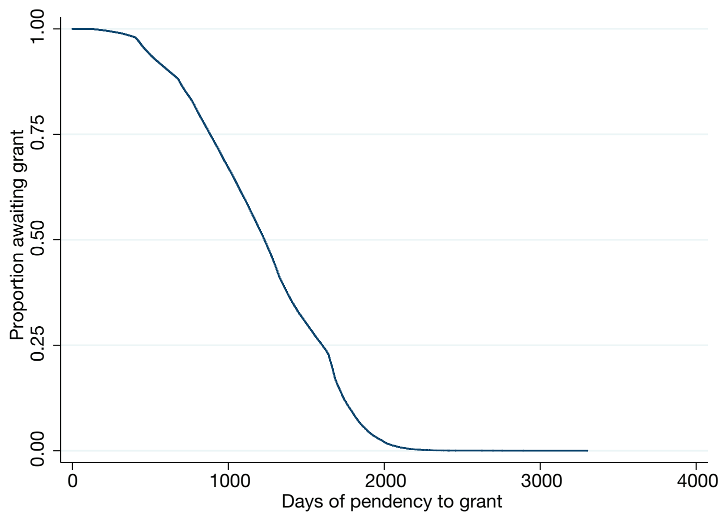
The linear regression analysis on all applications in 2001 confirms that there are particularly large differences in pendency between CSE and traditional patents. In Stock 1 we find that pendency of CSE applications is on average 72 days shorter than traditional patents. In Stock 2, where the search and the examination are conducted, we find that CSEs are substantially faster. On average, they were pending 343 days less inside the office than non-CSE applications. This result is in alignment with the office's target of 4 months to process CSEs. In addition, we find CSE patents to be 'processed' considerably faster by applicants outside the office in Stock 2 as well. Compared to traditional applications, CSEs were pending 113 days less outside the office. This is not surprising because once in Stock 2, there is no need for the applicant to do anything before first examination. As mentioned in Section 4.2, we have chosen to undertake linear regression analysis on the patents filed in 2001 as we are assured that the whole cohort of applications will have gone through the complete patent process. As part of this work we also ran the same linear regressions for individual years between 2000 and 2005, and found similar patterns and coefficients as those reported for 2001.

When we look at pendency in Stock 3, however, the overall picture becomes rather different. As expected, CSEs have no substantial or significant impact on the pendency inside the office in Stock 3, as they are processed in the same way as traditional applications at this stage. Outside the office however, CSE patents were pending an average 355 days longer than traditional applications. The likely reason for this is that applicants are given 2 years to reply to the first exam report on a CSE, and having gone through Stocks 1 and 2 quickly they have ample time before the 4.5 year statutory cut-off occurs. While applicants can reply sooner they very often chose to leave it until the deadline. This means that a lot of the time gained in Stock 2 is lost in Stock 3 due to applicants' choice. Such delays on the side of the applicant may suggest that the private cost of completing the patent process, perhaps added to the applicant's private value of uncertainty, is higher than the private value of granted patents. Similarly, delays in Stock 3 due to applicant behaviour may also suggest that applicants value the option of having an application that could either be left pending some time longer before statutory deadline kick in or could be progressed to a final decision (grant or rejection). But it also suggests that we need to understand how each stock impacts the pendency of incoming applications.

4.4 The impact of application stocks on pendency

We test whether the size of the different stocks and other factors impact pendency using survival analysis. The graph below shows the 2000-2011 application cohorts for the UK and how long patent applications spent in the system before being granted. As time goes on along the *x*-axis, more and more of the applications leave as granted patents. Where the *y*-axis reads 0.75, it means that 75% of all patents which will be granted are still being processed. Put differently, 25% of the patents in the represented cohort were granted before this time. The graph shows that 25% of patent grants in the UK occur before 900 days, or two and a half years from filing date. At around four and a quarter years only 25% of patents that will be granted have not yet been granted (and 75% have been). Of course applications filed in 2010 have not necessarily all passed through the system yet, so the information we have on these is limited. Nevertheless the survival curves for applications filed in 2000-2011 give us a clear way to illustrate pendency, and deliver the building blocks for predicting expected pendency through statistical regressions.

Figure 4.5 Survival analysis of UK patent applications, 2000-11



Source: UK IPO 2013

To determine whether the application inventory impacts on patent pendency, we run a series of survival regressions detailed in Appendix C and test for similar impacts as the linear regressions discussed in Section 4.3 above. The findings suggest that the size of Stock 1 inside the office can have a large impact on pendency. We estimate that if the size of Stock 1 were to double, pendency at the point of filing could increase by as much as 34% or an additional 209 days on top of the mean survival time of 616 days. By comparison, the number of arrivals – or new applications in a given month – are unlikely to have a big impact. A doubling of new applications is associated with a 7% or 42 day increase in expected pendency at the point of filing. While the potential effects of an increase in Stock 1 should be noted, the IPO's Stock 1 has actually decreased since the early 2000s. As a result we do not see extensions in Stock 1 as an imminent concern for the office.

An even bigger impact on expected pendency could come from increases in Stock 2, especially Stock 2 outside the office. The results presented in Appendix C suggest that if the size of Stock 2 outside the office were to double, then average pendency could double. Some caution should accompany this result as some additional tests suggested a lower impact. As Stock 2 is defined by the point in the process where the majority of interactions with applicants take place, we think it is worth looking at in some more detail at a later stage. However we suspect that the potential effect on average pendency of increases in Stock 2 outside the office may in part be due to applications that have not requested a combined search and exam. The linear regressions outputs in Appendix C show that CSEs are over a year quicker to pass through Stock 2 than traditional applications, as the combined search and exam eliminates some exchanges and interactions that could create delays for both the office and applicants. Indeed, when we add a CSE dummy to the survival time regression in Table C3 of Appendix C to see the impact of CSEs, we find that doubling the proportion of CSE applications would reduce expected pendency at the point of filing by 2 months or approximately 12% at the mean.

Compared to the share of Stock 2 held by applicants, an increase in Stock 2 inside the office would have relatively minor effect on expected pendency. Stock 2 inside the office is where the majority of the examination takes place, so we would expect an increase therein to prolong pendency time. However the fact that a doubling in Stock 2 inside the office would only lead to an expected 8% increase in pendency (or 51 pendency days) is perhaps surprising.

The findings from the survival time regressions coincide somewhat with the linear regressions, although they suggest a bigger impact on time to decision from Stock 2 inside. The linear regressions in Table C1 of Appendix C suggest that applications passing through Stock 2 spent an average 430 days more inside the office and an additional 287 days outside the office while in Stock 2.

The impact of Stock 3 outside is also surprising. These are the patents which have been examined but were not granted and so applicants are preparing amendment requests. The larger the size of this stock, the *faster* other patents travel through the system. If Stock 3 outside the office were to double we would expect pendency at the point of filing to fall by 6% or 35 days on average. While this might seem like a counter-intuitive result, it is worth bearing in mind that the regressions provide information on the likely effect of various explanatory factors on *expected* pendency time. As an application moves from Stock 2 into Stock 3, it moves off the examiner's desk and frees up space for examiners to attend to new applications.

While we can offer explanations for the effect of Stock 3 outside on application pendency coming out of the survival time-regressions, we have reason to treat these findings with caution. This is especially so as the linear regression on all patent applications filed in 2001 bear out slightly different findings. The linear regressions suggest that each additional amendment request results in an additional 120 days of pendency for the application in question, while the survival results show that getting the application out into Stock 3 speeds up the prosecution of other applications.

Stock 3 inside the office behaves as we would have expected in that it has a small impact on pendency, adding 3% or 19 days to average pendency at the point of filing if the stock were to double. The short turn-around periods for amendments mean the linear regressions show a small 18 day increase in pendency as a result of additional amendment requests, and this impacts both the application in question and applications elsewhere in the process as examiners time is spent re-working a case.

Having shown that the absolute size of different stocks has an impact on pendency, we now look to what else could be causing delays beyond the stocks themselves, or what has caused the build-up of these stocks.

4.5 What could be causing delays?

As we see it, there are three potential causes of delays beyond the stocks of patent applications examined above:

1) **The patent itself** might take longer to examine because it is more complex, longer or simply harder to understand;

2) **The patent office** may be taking longer due to systemic bottlenecks, inefficiencies or capacity issues; or

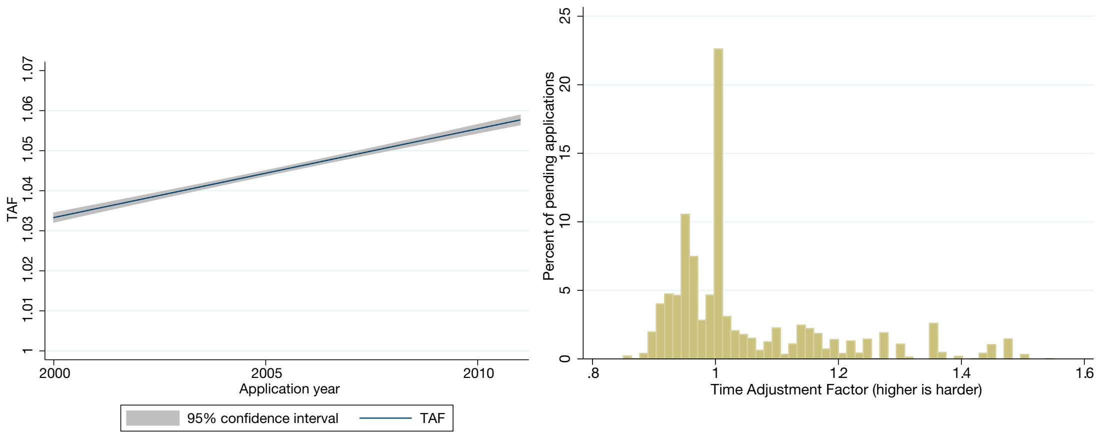
3) **The applicant** may be taking longer to file requests, or holding applications for longer for other reasons.

Following the methodology laid out above, and detailed in Appendix C, we examine the main drivers of patent pendency by looking at the impact of patent complexity, patent office capacity, and applicant behaviour on average pendency. Because patent pendency can result from actions of both applicants and examiners, we distinguish between time spent inside and outside the patent office. This enables us to assess whether delays occur inside or outside the patent office, providing different implications for policy considerations.

4.5.1 The patent itself

Patent complexity is driven by several factors including inherent technology complexity and claim structures. When more complex patent applications enter the system, it seems logical that average patent pendency will increase due to the additional time investment required by examiners. In the UK, patent examiners can assign a number to an application describing its complexity. This “Time Adjustment Factor” or “TAF” represents the expected time investment required for examination of the application. An application will be assigned a TAF rating on the basis of its International Patent Classification (IPC), with a rating of 1 being an application expected to take the standard amount of time to examine. If TAF is less than 1, the application should take less time, and if it is more than 1, it should take longer. The left-hand graph in Figure 4.6 below shows that TAF has risen slightly from a mean of 1.03 in 2000 to 1.06 in 2011. This represents the average TAF for all UK patent applications which were assigned an International Patent Classification (IPC).

Figure 4.6 Technology Adjustment Factor (TAF) over time (left) and TAF distribution for UK patent applications (right)



Source: UK IPO 2013

The right hand histogram in Figure 4.6 shows the distribution of TAF for UK patents applied for between 2000 and 2011, and as expected there is a peak around 1. Note that the range of TAF is from 0.8 to 1.6, meaning that examiners are allowed between 20% less time, and 60% more time to examine an application that is deemed more or less complex.

Using information on each patent, we analyse the impact of complexity on patent pendency by means of linear regressions and survival time regressions.

The survival time regressions suggest that when controlling for CSE applications, the size of the stocks and the number of examiners, a higher TAF is associated with shorter pendency time. This result is surprising, but is not particularly strong, meaning we expect at best a modest correlation between a higher TAF and lower pendency time. As illustrated in Appendix C, if TAF goes up by 100% (i.e. from 1 to 2), then expected pendency at the point of filing would drop by 10% or 64 days. More realistically, a 10% increase in TAF (i.e. from 1 to 1.1) is expected to reduce pendency time by a mere 1.5% on the mean or 6.4 days. Why this is the case needs to be explored, but it seems – on first inspection of the data – to be connected with applicants coming into the process better prepared as more complex applications tend to be handled by an attorney, while the office appears to be slightly slower on balance to process complex applications.

Our confidence in this association between complexity and pendency time however is limited, as the linear regressions bear out slightly different results and show that at certain stages in the backlog, complexity actually prolongs pendency time.

In line with the survival time regressions, the linear regressions suggest that complexity reduces pendency in Stock 1, meaning that more complex applications pass through Stock 1 faster. As above, the expected reduction in pendency time is modest. On average, a one unit increase in TAF complexity from a score of 1 to 2 reduces process time in Stock 1 by 29 days in the 2001 cohort. In Stock 2, however, where the examination is conducted, we find that complexity is associated with substantially longer pendency inside the office. A one unit increase in TAF complexity is associated with 207 days additional processing time. There are two contributory factors at play here. An application for complex technology takes longer to examine, but there are also likely to be fewer examiners qualified to actually do that examination. In other words, the queue may be longer in a complex technology sector as spare examiner capacity in other technology areas can't be deployed for it. Outside the office, pendency in Stock 2 falls by 20 days for the 2001 cohort when TAF increases from 1 to 2, implying that applicants place higher value on faster response to office actions.

Finally, in Stock 3 we find complexity to have a more modest impact on pendency. Unlike in Stock 2, complexity seems to cause additional pendency mainly outside the office in Stock 3. On average an increase in complexity from 1 to 2 is found to be associated with an additional 43 days of pendency here. Possibly amendments made by the applicant take longer for complex patents, but do not have any substantial impact on the time investment required of examiners.

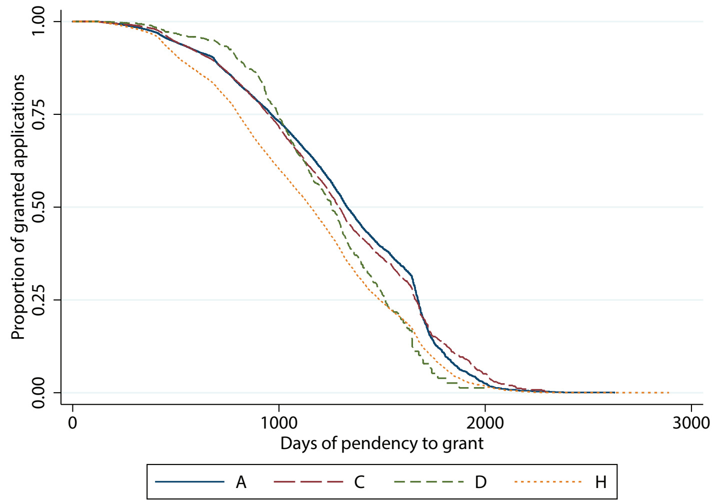
Looking across the data from 2000 to 2005 we see this pattern repeat itself. Inside Stock 2 and outside Stock 3, complexity slows the application down, while outside Stock 2 and inside Stock 3 there is a small positive effect from complexity.

Given that the survival time regressions and linear regressions show slightly different results for the effect of TAF on pendency time, we chose to examine an alternative indicator for patent complexity. In this step, we turned to the main International Patent Classifications (IPCs)[31], which is another proxy for patent complexity. Investigating the survival time of applications with different IPCs allows us to check our interpretation of TAF and its effects on pendency.

31 Note that up until 2008 the UK IPO used the domestic technology classification, and we have converted these into IPC classifications using the office's own equivalence tables.

Looking just at the letter of classification (A-H) there does not appear to be much difference between the slowest and fastest moving technology through the UK patent office (see figure 4.7). We have highlighted classes A, C, D and H which are the extremes, but D should be disregarded as they are less than 1% of the population. So overall, we conclude that complexity as indicated by the TAF and IPC main classifications has little effect on pendency, but further investigation is needed.

Figure 4.7 Survival analysis of UK patent applications, by IPC

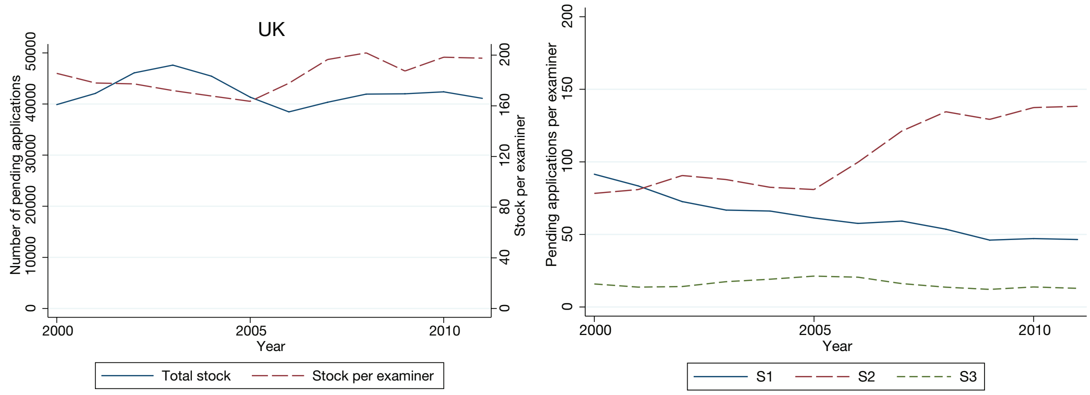


Source: UK IPO 2013

4.5.2 The patent office

Patent office capacity can be a cause of growing application inventories. Theoretically, the declining trend in applications to the UK may support the conclusion that capacity would be improving, but this effect could be counteracted by the rising proportion of search requests from those applications. Turning to the actual data, when we plot the total application inventory per examiner in the left hand graph of Figure 4.8 below, we see that examiner capacity has indeed improved over the last decade. If we split that out by the three main stocks on the right hand graph of Figure 4.8 the picture appears a little more erratic. But considering the small numbers (and how similar they are to US proportions) it seems that capacity is not an over-riding issue if the aim is to keep the application inventory constant.

Figure 4.8 UK application inventory and stocks by examiner

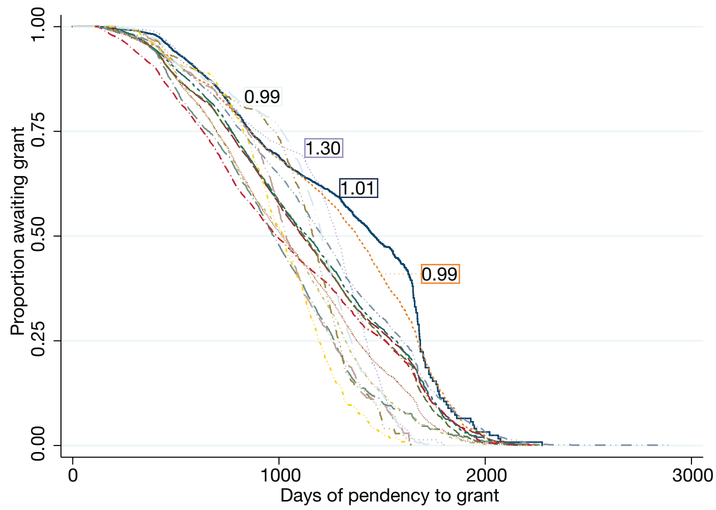


Source: UK IPO 2013

Reducing the application inventory around Stocks 1 and 2, on the other hand, does have a clear impact on pendency as shown by our survival regression results in Appendix C, and reducing some of the work in Stock 3 appears important according to the linear regressions. Additional patent examiners in turn speed up the patent process and reduce pendency. While this result is expected, the magnitude of the impact (as reflected by the coefficient on the number of examiners) is strong. The coefficient is similar to the US results, and suggests that examination pendency is more sensitive to increases in examiner pendency than previously thought. The survival regressions suggest that doubling the number of examiners would reduce average pendency by 9% or 57 days across the board. This is just direct effect of examiners on expected pendency at the point of filing. More examiners are also expected to have an indirect effect on expected pendency throught Stock 2. This could be quie significant as discussed in Section 4.4.

While the aggregate number of examiners appears to be of some importance, there is a question of whether particular examination groups have a greater effect on pendency than others. In the UK, the search and exam work is divided between different groups who specialise in a particular set of technologies, and data is available for which groups processed which patents. Admittedly, our investigation was limited to the final examiner on a given grant – so we attribute all time to the one group. Under this analysis, there were some divergences among groups as illustrated in Figure 4.9 below. Using the same survival type analysis where each line is a different examination group, we can see that there are four groups, two of which appear to take longer to grant the first 25% of patents, and two of which take longer to grant patents after the median application has been granted. Figure 4.9 below is colour coded, and does not identify the exam groups for confidentiality reasons. It shows the mean TAF for the patents granted by each group. The mean TAF may explain the relative to average slowness of the purple group because they have a high TAF of 1.30, but it does not address the relative slowness of the other three groups where the TAF is around the historical average of 1.03 to 1.06. It is possible that there are odd factors at play here. For instance, low complexity technology is often used as training material and examination by trainees can take longer; or it can be the case that certain groups had staffing shortages, certain groups may receive fewer CSEs than others or some groups may already have been changed due to performance issues.

Figure 4.9 Survival analysis of UK patent applications, by examiner group



Source: UK IPO 2013

To sum up, there are two points to take away here. First, in a sample of 30+ examination groups only three appear to be relatively slower than others (the fourth being explained by the higher average TAF). Secondly, more work is needed to identify whether increasing examination capacity in certain groups will have a bigger impact on pendency time.

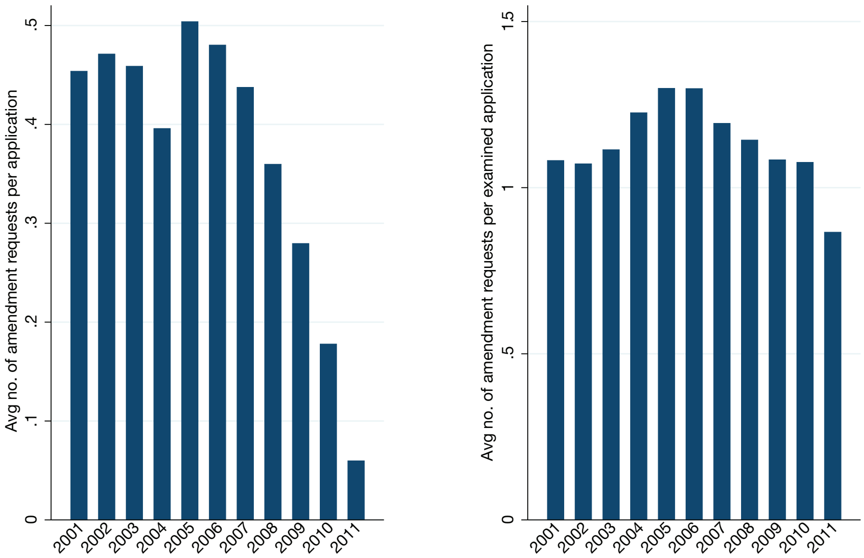
4.5.3 The applicant, and requests for additional examiner work

Applicants interact with the patent office throughout the application process where they make several active choices, especially in Stocks 2 and 3. Accordingly, their behaviour may have an impact on pendency. For example, applicants may file additional search requests and amendments.

The left panel of Figure 4.10 below shows average number of amendment requests per patent application. As with search requests, amendment requests are rarely filed for any given patent application as most filings do not make it to exam and are therefore not eligible for amendments. However while low, the average number of amendment requests per application does show an increase from 0.45 in 2001 to just above 0.5 in 2005. The fact that the mean number of requests surpassed 0.5 means that in 2005, an application was more likely to be amended than not. The sharp decrease after 2005 is likely to be due to timing issues, as many of the newer cohorts of patents would not have completed most of the stages leading up to amendment as yet.

Another way to look at amendments is to check how many requests were filed on examined applications. These are the applications eligible for amendment requests. As illustrated in the right panel of Figure 4.10, on average 1.3 amendment requests were filed on applications that completed their exam in 2006. This represents an increase from the average 1.1 requests filed on applications examined in 2000. Since 2006 this trend has reversed itself but this could in part be due to censoring in the data.

Figure 4.10 Average number of amendment requests per UK filing and examined application, 2001-11



Source: UK IPO 2013

Note: The right hand side graph of the above Figure 4.10 illustrates average amendment requests for applications filed in a particular year. By contrast, the left hand graph shows average amendment requests filed on applications examined in a particular year.

As expected, we find that an amendment request is associated with additional pendency. In the 2001 cohort, one additional amendment was associated with 18 days of additional pendency inside the office and 120 days outside the office. This may be due to relatively generous deadlines given to applicants and their attorneys and the Patent Law Treaty (PLT) requirement that an applicant can extend any period for reply as of right by 2 months.

In sum, we find that additional amendment requests are a major driver of patent pendency. Their biggest effect is outside the office, with an additional amendment request adding about four months in additional pendency. Therefore it would make sense to provide incentives which have the effect of avoiding repeat work, and limiting the number of amendments made. This would be through the imposition of deadlines as the time to requesting or amendment is adding pendency outside the office. Within the office, amendment requests add 18 days to pendency.

4.5.4 Enforcing deadlines on amendment requests could reduce pendency

There are deadlines for how long an applicant has to respond with an amendment request after receiving an exam report or an amendment completion report by the IPO. For all examination reports, the patent examiner specifies a suitable period for response; this period is set at the examiner's discretion, but examiners generally follow standard periods, as set out in the Manual of Patent Practice. Until recently, the standard period for response for most first examination reports has been 4 months from issue of the report. However, applicants are entitled to extend this period by 2 months if the applicant submits a written request.[32]

If we consider only the deadline for how long an applicant has when responding to an exam report we find that applicants often miss that deadline. If the rules were enforceable to the letter, that would make an application null and void, but the system has a level of discretion built in. Furthermore the rules as to what may be deemed a legally late submission have shifted over the last decade.

Between 2006 and 2010 there was effectively a six-month deadline, as applicants were limited to four months but could exercise their right for a two month extension on application. Table 4.1 below summarizes the number of first amendment requests filed with the IPO in a given year, and then the percentage and absolute number of requests filed more than four and six months after the exam report was sent to the applicant. The 2011 figure should be treated with care as all exam reports sent out in 2011 would not have received their amendment requests before the end of the data analysis, so we would expect these figures to go up.

Table 4.1 Amendment requests on UK patent applications

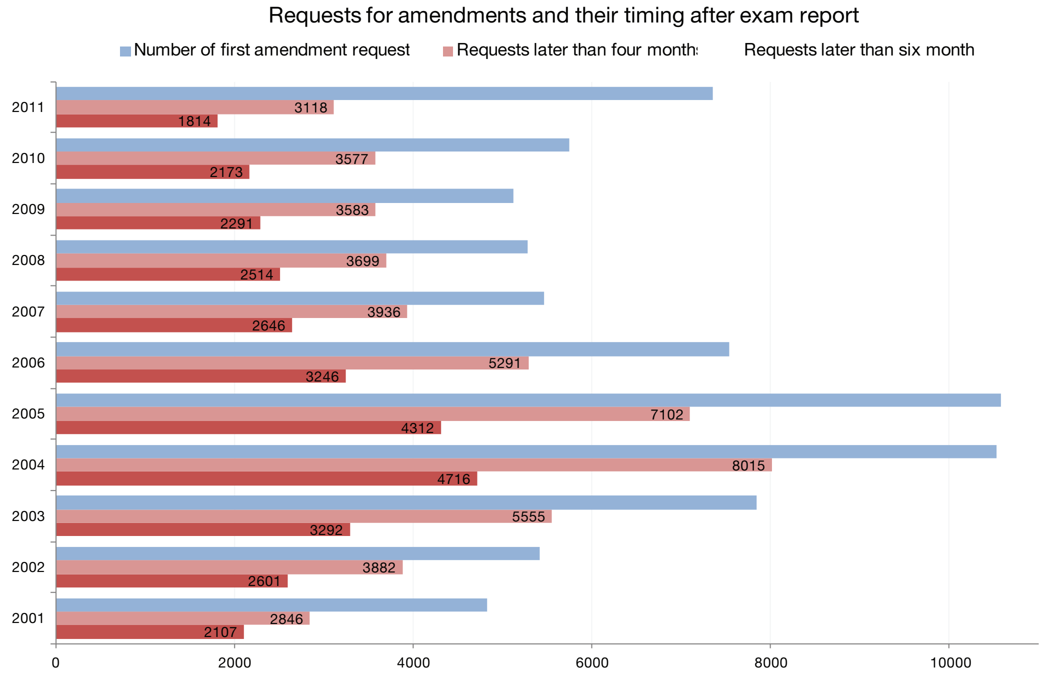
Year of request	Number of first amendment requests	Requests later than four months	Requests later than six months
2011	7,355	(3,118) 42%	(1,814) 25%
2010	5,754	(3,577) 62%	(2,173) 38%
2009	5,126	(3,583) 70%	(2,291) 45%
2008	5,279	(3,699) 70%	(2,514) 48%

Source: UK IPO, 2011

32 The latter via form 52 and paying the appropriate fee.

On average, roughly two-thirds of amendment requests were filed after four months, and a third were filed after six.[33] If the existing deadline on the first amendment request were enforceable, the UK IPO could have rejected a high number of applications on the grounds of late responses. This would have reduced the size of Stock 3, and therefore the application inventory by potentially thousands of applications if we look back over the last decade as in Figure 4.11 below.

Figure 4.11 Requests for amendments on UK patent applications and their timing after the exam report



Source: UK IPO, 2011

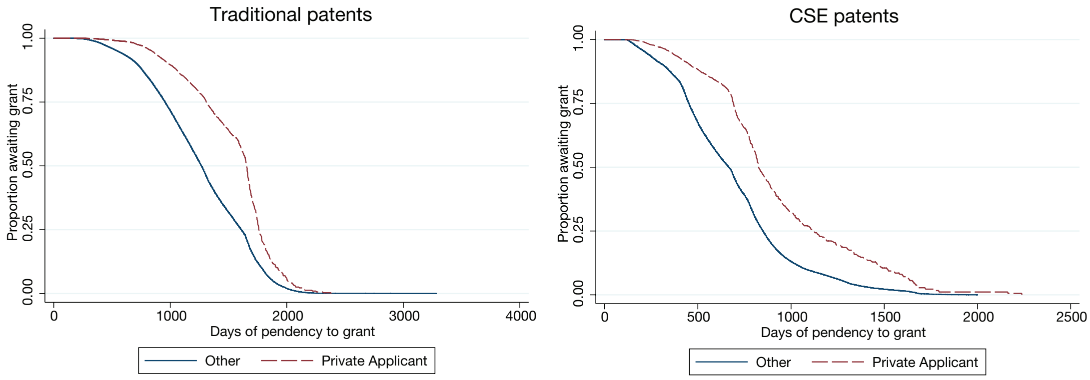
These are only the deadlines missed on the first amendment requests and many applications have several such requests, so there is more to investigate in how applicants respond to deadlines.

33 Note that the responses to first exams may include responses to first CSE report. The latter have a response date of 2 years so lots of responses to those will have been filed quite legitimately after 6 months.

4.6 The type of applicant matters

There has been some discussion among patent offices to suggest that non-domestic applications may take longer to process, and while we have not undertaken a deeper investigation of this here, the UK experience has generally been that multi-national firms who apply are not necessarily more time consuming. On the other hand, it has been suggested that costs to office of processing applications filed by private applicants – applicants who do not have a patent attorney – far exceed those of represented applicants. We have good evidence that only a small percentage of private applicants actually receive a granted patent, and even fewer renew beyond ten years. Figure 4.12 furthermore shows that private applicants do take longer to receive both traditional and CSE patents. Budget constraints and the lack of comprehensive attorney support and expertise may be part of the explanation for why private applicants take longer and are unable to maintain granted patents.

Figure 4.12 Survival analysis of granted patentsiled by private and represented applicants, traditional vs. CSE patents



Source: UK IPO 2013

Figure 4.12 shows that applications filed by private applicants are granted roughly 100-200 days later than traditional patent applications. Even more extreme, CSE applications filed by private applications are granted 200-300 days after CSE applications by represented applicants are granted. Given this, we can confirm anecdotal evidence that applications filed by private applicants take longer to process, and that is just for those who will eventually be successful. There are probably further effects on workload from the 80% or more unsuccessful private applicants. This is another issue for which further investigation is needed to understand how the eventually successful private applicants can be helped through the system.

4.7 Concluding remarks on the UK analysis

We find in this first round of analysis that it matters whether applications are filed by represented or private applicants, and whether they are traditionally filed or CSE applications. Represented applicants and CSEs go through the system faster, while traditional applications and private applicants are slower. But CSEs then slow down in Stock 3 and lose almost a year on average as compared with traditional patents. So there are possibilities of tweaking some of the deadlines and procedures in the interest of reducing pendency.

This is particularly the case when it comes to amendment requests, which are late up to 70% of the time and could, if deadlines were enforceable, cut Stock 3 by several thousand applications. Extensions are by way of right, but the standard period specified for response to the first substantive examination report issued in relation to a patent application is set by the Office. This standard period has in fact already been cut to 2 months[34]. Given that the size of Stock 3 outside the office has a negative impact on pendency, i.e. it speeds up the pendency of other patents, the aggregate effect is not certain. We expect that the majority of this effect comes from the fact that examiners are free to undertake examination when applications move out of Stock 2 and into Stock 3 outside the office.

Patent complexity does not appear to influence pendency in a substantial manner. Additional examiners, on the other hand, influence pendency significantly, even if there are a few exam groups where potential bottlenecks exist or existed (as we are looking at both current and past groups). The take-away here is that additional examiners impact on the aggregate size of the application inventory, and should therefore reduce pendency time.

The biggest single factor for expected pendency is the size of Stock 2 outside the office, and reducing this should reduce pendency. The issue with Stock 2 outside the office is the sheer volume of applications. This has grown from 13,429 in January 2000, to an estimated 18,598 in December of 2011. Managing even a five per cent reduction in this stock – around 1,000 applications - could only be done by changing the behaviour of many applicants to file exam requests sooner or not file additional search requests. Regulatory or behavioural changes in this space would be a positive move, but implementing such change could be a longer process than what is offered by policy levers available in Stock 3 around internal enforcement of deadlines, and in the ability to recruit examiners.

Therefore, the conclusion we draw from the regressions and data differentiate between short term and long term solutions to backlogs and pendency:

- Short term - Firstly, to keep the backlog and pendency stable in the short run Stock 1 and Stock 2 outside the office should be kept stable, while managing the other processes efficiently. Secondly, to reduce pendency, one needs to address the parts of the backlog that are biggest and those factors that would impact pendency the most. In the UK, this requires targeting Stock 2 inside the office, for instance through increasing examination capacity. Reducing Stock 3 outside the office would reduce the backlog, by for instance enforcing deadlines already in existence.

34 For more details, see: http://www.ipo.gov.uk/p-pn-period.htm

- Long term - Reducing the backlog could be achieved through changing the rules and regulations governing Stock 2. This is likely to be efficient as Stock 2 inside the office is large and growing and Stock 2 outside the office is expected to have the biggest effect on pendency if reduced. While efficient, reducing Stock 2 will not be an easy feat and will require careful consideration of the existing system of rules and regulations and the different options going forward.

5. US Analysis

In this chapter, we examine the stocks of pending patent applications at the USPTO using the methodology developed in Chapter 2. As we discussed above, several factors may contribute to the growth in patent application inventories. We seek to identify which of these factors are the primary drivers of the inventory and to what extent greater pendency results from each. The analysis highlights the usefulness of the taxonomy and demonstrates how it can be altered or adapted to particular institutional structures at different patent offices. A primary focus of the analysis is to better understand the relationship between stocks and the pendency of incoming applications by using survival analysis.

Before delving into that analysis, it is important to understand some details of the USPTO examination system, especially as they relate to pendency and the ability to make international comparisons in stocks of patent applications. Appendix D provides a detailed primer on US patent examination. As described in Chapter 2, the appendix explains that up until the time of first action, the US is an "opt-out" system. From that point forward, it becomes primarily an opt-in system. The appendix also describes important elements in patent prosecution, such as the first action on the merits, restriction practice, and the use of amendments and RCEs to extend prosecution.[35] The distinction between RCEs and traditional (serialized) continuations is discussed in Section D.4, along with continuations-in-part and divisional applications. Lastly, Appendix D describes the production system or "count system" used to evaluate, monitor, and compensate patent examiners.

The substantive analysis in this chapter consists of three parts. First, Section 5.1 employs a descriptive analysis to examine the growth of US patent application stocks and pendency over the past 16 years. Section 5.2 looks more closely at pendency in particular. To more precisely ascertain the relationship between stocks and pendency requires more careful statistical analysis in the form of survival time regressions. Those results are summarized in Section 5.3, with detailed regression results in Appendix E. Section 5.4 concludes with some implications of the analysis for future research.

The descriptive results suggest that some traditionally-cited factors, such as rapid increases in new applications, greater application complexity, and examiner attrition may not be major contributors to the US backlog. Further, the descriptive results suggest that growth in requests for continued examinations (RCEs) has disproportionately diverted resources away from other applications, resulting in a net increase in pendency.

35 Figure D1 is useful for understanding the primary paths through patent prosecution, as well as for understanding the frequency of different paths. For instance, the cycle of "non-final rejection, final rejection, RCE, non-final rejection" is common.

The results of the survival time regressions support those of the descriptive analysis. The survival analysis is similar to that in Chapter 4 for the UK, except that in our statistical regression analysis, we add a competing risk model in order for us to quantify the relationship between stocks and pendency and the relationship between stocks and disposal type (patenting versus abandonment). In summary, we find the following:

- Unsurprisingly, larger stocks of pending applications tend to increase the first-action pendency of *incoming* applications.
    - One additional unexamined application increases the predicted first-action pendency of each new application by about 39 seconds. Over the course of a year, this amounts to an additional 5.6 months of cumulative first-action pendency experienced by incoming applications.
    - Doubling the size of S1 and S2 (which is equivalent to moving from 2000 to 2009, in terms of all unexamined applications) is associated with a 45% increase in the first-action pendency of incoming applications.
- Pending RCEs tend to increase the first-action pendency of incoming applications by more than do regular pending applications.
    - Filing an RCE on an application that would otherwise not have filed an RCE adds approximately 66 seconds to the first-action pendency of each incoming application. Over the course of a year, this amounts to an additional 9.4 months of cumulative first-action pendency experienced by incoming applications.
    - Doubling the uptake of RCEs (as a proportion of actively examined applications) from 0.15 to 0.30 increases the first-action pendency of incoming applications by 16%, on average. This is equivalent to moving from 2000 to 2009, in terms of the RCE usage.
- Conversely, increasing the number of examiners tends to decrease first-action pendency for incoming applications.
    - One additional junior examiner saves each new application about 1,191 seconds (19.8 minutes) in first-action pendency. Over the course of a year, this amounts to a total cumulative savings of about 170.7 months of first-action pendency experienced by incoming applications.
    - Doubling the number of examiners (which is equivalent to moving from 2000 to 2008 levels) is associated with a 17% decrease in first-action pendency.
- Increased stocks have a second, indirect, effect: longer first-action pendency tends to be correlated with a decrease in the rate of patenting and an increase in the rate of abandonment.
    - Raising first-action pendency from one year to two years (which is equivalent to moving from 2000 to 2008 levels), increases the abandonment rate from 22.5% to

28.2% and conversely lowers the patenting rate from 77.5% to 71.8%. In other words, abandonment increases by 5.5 percentage points (or, a 24.4% growth in abandonment).

We should make clear at the outset that the results by themselves are not an indictment of RCE practice, or other applicant or office practices. They reflect only the costs—and not the possible benefits—of the practices.

5.1 Descriptive analysis

In this section, we discuss some of the commonly cited causes of the patent application backlog. Before turning to the more sophisticated statistical analysis of pendency, we explore the growth in application stocks using a purely descriptive analysis. The results from this analysis can be summarized as follows:

- New applications have not grown rapidly over the last 16 years
- Hiring of patent examiners has kept pace with growth in new applications
- Work per disposal has increased
- It is unclear whether application complexity accounts for the growth in application stocks
- RCEs appear to crowd out other applications with a ratio of approximately 1:2

Below, we provide greater detail on each of these results.

As discussed in Chapter 2, we identify three different stocks of pending applications, S1, S2, and S3, based on identifiable transitions common to all patent offices. For analysing the US, it is useful to subdivide Stock 3 into two parts:

S3a. Stock 3 applications with no previous RCEs or other non-serialized continuations filed.

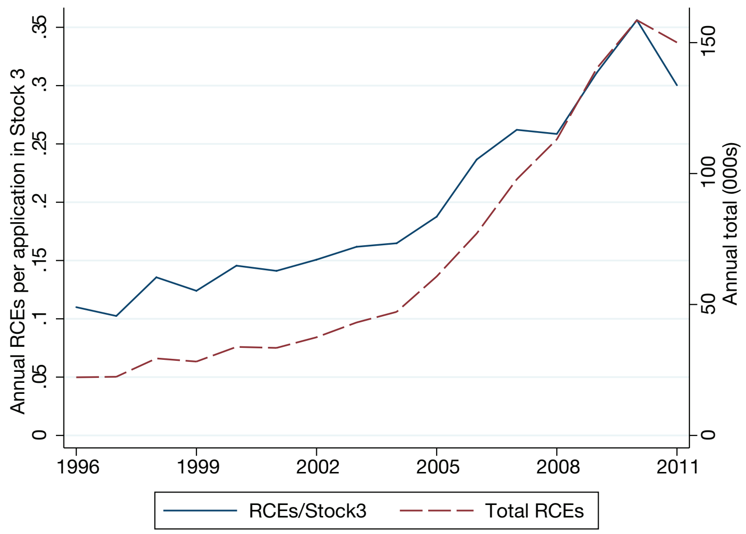
S3b. Stock 3 applications with at least one RCE or other non-serialized continuation. S3b applications do not re-enter S3a.[36]

36 Recall that we use the convention “RCE” to refer to any application with at least one RCE. Technically, each RCE is a new filing. Thus, official statistics will differ from our counts based on this distinction. Our definition is designed to reflect the transition point from a non-RCE to an RCE.

5.1.1 New applications have not grown rapidly over the last 16 years

As discussed in Chapter 1, annual filings have doubled since the late 1990s, exceeding half a million applications since 2010 (see Figure 1.2). However, a substantial portion of this growth is attributed to burgeoning RCEs. Figure 5.1 shows RCEs growing in total quantity and as a proportion of all applications in Stock 3. The growth in RCE filings has obscured the sensitivity of new application inflows to the 2007-09 recession. Without RCEs, growth in new applications has been flat relative to GDP for the last decade and total filings in the US actually decreased during the recession. New applications per GDP dollar have increased only 2.1% per year from 1996 to 2011.

Figure 5.1 Requests for continued examination (RCE) in the US



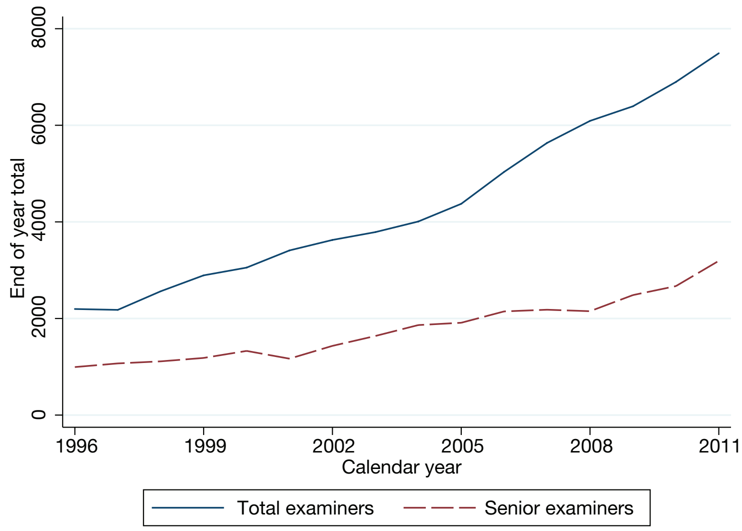
Source: USPTO 2013

Note: RCEs include all non-serialized continuations.

5.1.2 Hiring of examiners has kept pace with growth in new applications

Figure 5.2 shows the number of active examiners by standardized GS-12 full time equivalent (FTE).[37] There has been significant growth in hiring over the last 16 years. Over the same period, average new workload, as measured by the number of new applications (excluding RCEs) per examiner, has been cut in half (see Figure 1.4). As annual filings recovered to near the pre-recessions peak in 2007, the number of new applications per examiner has held steady at about 50. Simply put, new applications have not outstripped capacity growth.

Figure 5.2 Examiner staffing in the US



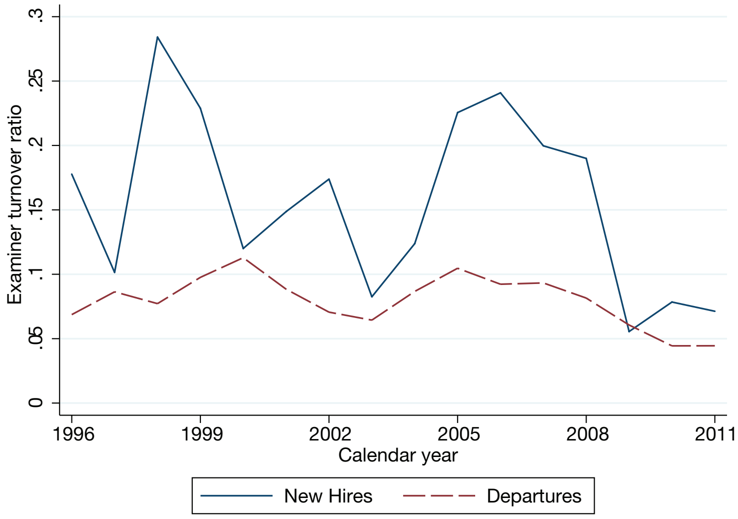
Source: USPTO 2013

Note: Using full time equivalents (FTE) based on examining time and expected production. Senior examiners are GS14 and above.

Similarly, average examiner workload in terms of pending applications has also eased since the peak in 2005, as shown in Figure 1.3 in Chapter 1. While total application inventory peaked in 2008, and has levelled off at about 1.2 million applications, it is now at a 15-year low on a per examiner basis (160 applications per examiner).

37 We count examiners on the basis of fulltime equivalents (FTEs) for a GS-12 examiner. The General Schedule is a promotion and pay scale common to most federal employees in the US. The grade of patent examiners ranges from GS-5 (Bachelor's degree with no work experience) to GS-15 (the top grade in the GS system). See http://usptocareers.gov/Pages/Misc/SalaryRates.aspx for information on examiner compensation. Primary examiners ("Primaries") are either GS-14 or GS-15. An examiner's production expectancy is based on a multiplier of the benchmark GS-12 expectancy. Thus, a GS-14 or 15 has a higher production expectancy, and is counted as more than one GS-12. Likewise, a GS-7 is counted as a fraction of a GS-12. Unless otherwise noted, we make all examiner calculations based on GS-12 equivalency. However, counts based purely on the number of examiners yield almost identical results.

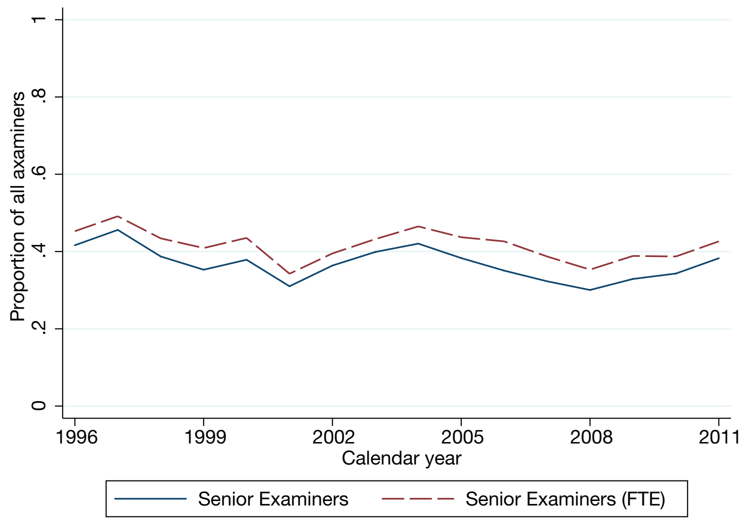
Figure 5.3 Examiner turnover ratio in the US



Source: USPTO 2013

On average, the hiring rate and the retention of experienced examiners have augmented capacity. Figure 5.3 shows that examiner hiring and attrition rates have varied annually, with hiring outstripping attrition. The attrition rate has remained more stable than the hiring rate. Attrition has ranged between 5% and 10% per year, and is currently at a historic low. Hiring, on the other hand, has varied between 6% and 28% per year. During periods of increased hiring, productivity tends to be lower due to the significant amount of time and resources devoted to training new employees. Despite the year to year changes in hiring and attrition, the proportion of the workforce at the senior level (GS-14 and GS-15) has remained stable, at approximately 35-45% of the examination capacity as shown in Figure 5.4. Turnover tends to be higher at more junior levels, so that the average examiner tenure has not changed as dramatically as one might expect.

Figure 5.4 Senior examiner staffing as a proportion of all examiners in the US



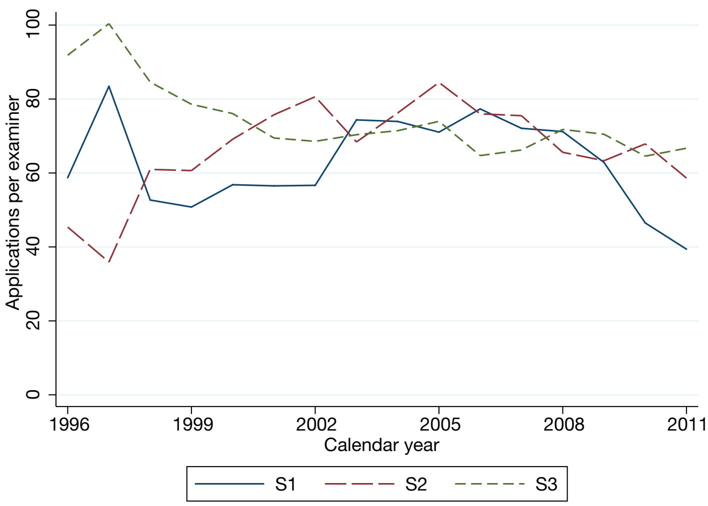
Source: USPTO 2013

5.1.3 Work per disposal has increased

Figure 5.5 shows each stock relative to the number of examiners.[38] Interestingly, the number of active examinations (S3) per examiner has steadily fallen over the last two decades, indicating a decrease in the "carrying capacity" of examiners. Technically, the examiner's docket includes some applications in S2 that are awaiting first action. We define the "active docket" as those applications in S3, having already received a first action. Thus, carrying capacity reflects the capacity of examiners to handle ongoing cases.

38 While examiners do not actively examine any applications in S1 or S2, the ratio serves as a useful benchmark.

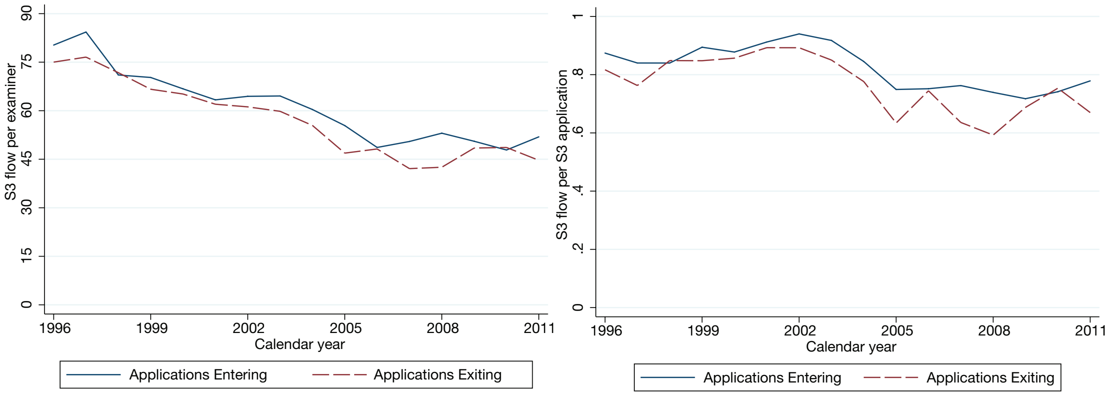
Figure 5.5 Stocks relative to examiner staffing in the US



Source: USPTO 2013

Figure 5.6 shows the flow of applications into and out of S3, on a per examiner basis and relative to the current stock of S3. Inflows into S3 reflect first actions on new applications while outflows reflect disposals.

Figure 5.6 Stock 3 inflows and outflows

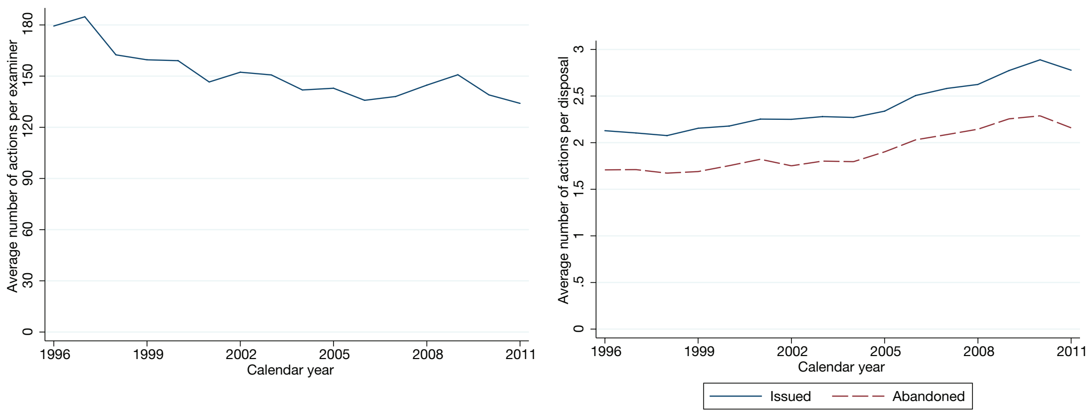


Source: USPTO 2013

Note: The left panel shows Stock 3 flows relative to examiner staffing and the right panel shows Stock 3 flows relative to the total inventory in Stock 3.

Inflows and outflows per examiner have both steadily fallen over the last 16 years. This indicates that examination turnover has slowed. Comparing inflows and outflows to the size of S3, as in the right panel, shows that throughput relative to the total quantity of active examinations has also decreased. The decrease is less distinct than on a per examiner basis, but it demonstrates that the overall throughput in S3 has decreased from the levels seen in the late 1990s. Both comparisons serve to indicate that examiners are doing more interim work per application, between the first action and disposal. We can investigate this directly by looking at the number of actions per examiner and per disposal.

Figure 5.7 Major actions in the US



Source: USPTO 2013

Note: The left panel shows major actions per examiner and the right panel shows major actions per terminal disposal. Major actions include final rejections, non-final rejections and allowances.

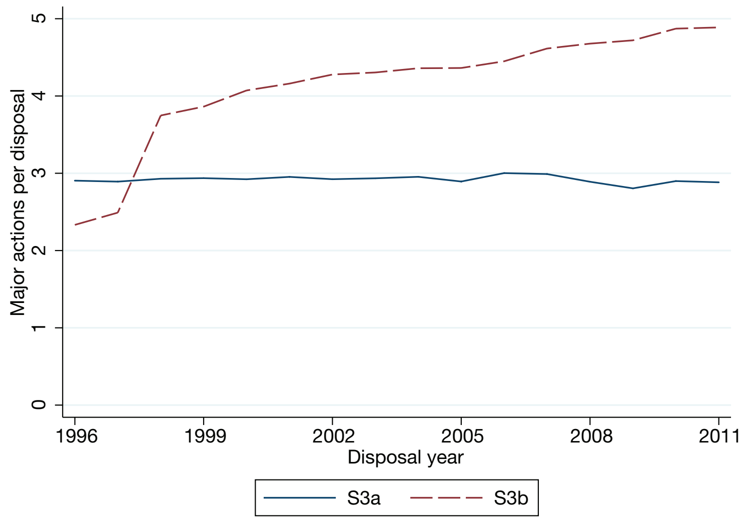
Figure 5.7 graphs the number of major actions[39] per examiner per year and per disposal. The left panel shows that the number of major actions per examiner per year has fallen somewhat over the past 16 years, while the number of major actions per disposal has risen by about 0.5 actions.[40] Note that patented applications have about 0.5 more major actions than abandonments, at the mean. The growth in actions per disposal is due primarily to increasing RCE usage and more actions required to dispose of an application that has received at least one RCE. In particular, the number of major actions for the average RCE disposal has been steadily increasing over the period (see Figure 5.8);[41] however, major actions for non-RCEs have remained constant. Thus, the driving force in actions per disposal is the uptake in RCEs.

39 Major actions are defined as any decision on the merits for an application. Those include non-final rejections, final rejections, and notices of allowance. Examiners do other work that involves substantial time, such as advisory actions, and responding to appeal briefs.

40 We measure the average number of actions for each application at the calendar date of disposal, whereas actions per FTE are measured based on contemporaneous counts of actions and FTEs on each calendar date. Another measurement of workload used by the PTO is the contemporaneous count of actions divided by the count of disposals on that calendar date.

41 RCEs are primarily available after a final rejection. Thus, in order to make the comparison on equal terms, we exclude applications that abandoned prior to a final rejection or that received a first-action allowance.

Figure 5.8 Major actions per post-final rejection disposal in the US



Source: USPTO 2013

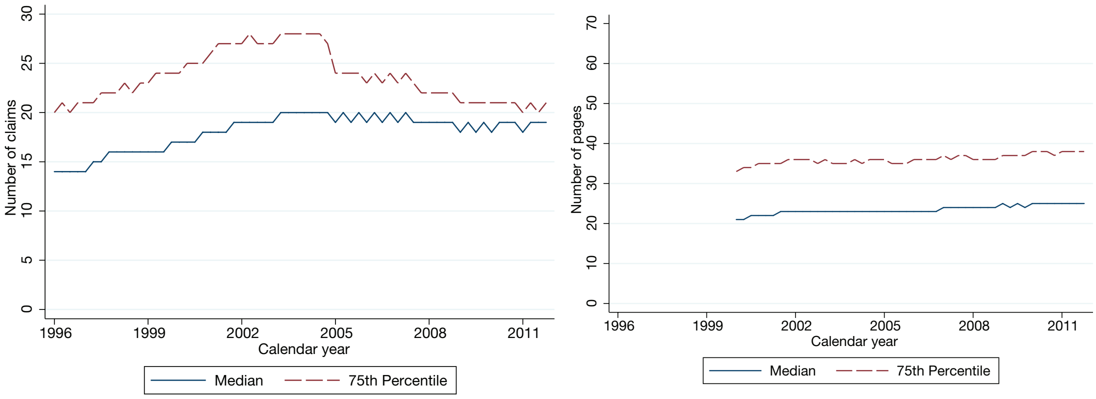
Note: Sample includes all disposals with at least one final rejection. Major actions include final rejections, non-final rejections, and allowances.

5.1.4 It is unclear whether application complexity can account for the growth in stocks

If the complexity of applications is increasing significantly, then the work required per disposal may be increasing simply because the job of examining is becoming more complex, on average. As such, applications would require more processing time, thus leading to increased backlogs. Unfortunately, "complexity" is subjective, and not easy to measure outside of a case-by-case basis. Individual applications may be getting more complex across the board, or the proportion of applications in more complex technologies may be increasing. To proxy for complexity, we examine the number of claims and pages at filing and the changing composition of technology and claimed priority. In the regression analysis, we control for the number of claims and we also use technology specific stocks of pending applications and examiners. Further, we account for differences in allowable examination time by using the office's GS-12 expectancy, which defines expected examination time for each technology.[42]

42 See Section 5.1.2 and Appendix D for more information on GS-12 expectancy.

Figure 5.9 US application complexity



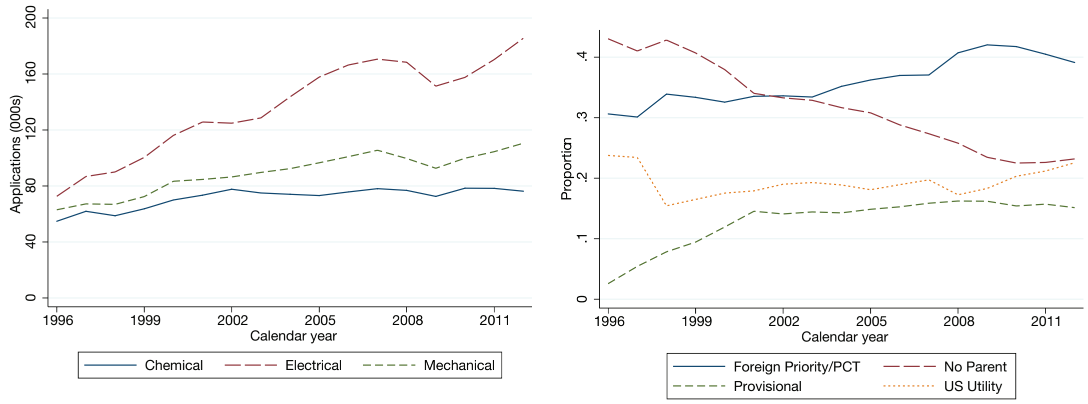
Source: USPTO 2013

Note: The left panel shows claims at filing and the right panel shows specification length at filing. Page data prior to 2000 are incomplete.

As Figure 5.9 indicates, the median number of (pre-examination) claims rose from 1996 through about 2004. At that time, the office instituted a higher fee for claims greater than 20. Subsequently, the median has held steady at 18-20 claims. The number of pages per specification has been largely stable for the median application.[43] Of course, simple claims and specification page counts cannot adequately measure application complexity or tell us whether complexity has increased sufficiently to account for all or some of the increase in patent inventories. The complexity of the claims themselves may be increasing in a way that is not captured by these simple counts.

43 Page counts are not as available during the early part of the period. Only 25% of applications contain page data for the 1996-1999 cohorts.

Figure 5.10 New US application characteristics



Source: USPTO 2013

Note: The left panel shows the number of new applications in the three disciplines and the right panel shows the proportion of new applications that claim priority to various parent types. Neither graph includes RCEs as new applications.

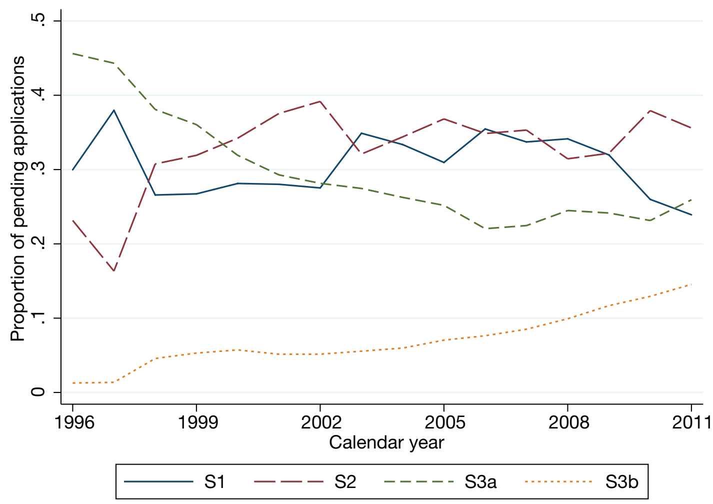
Other measures provide some indication that the composition of applications has changed over time. Electrical and Chemical applications have grown faster than Mechanical applications, which generally have the lowest complexity rating (see Figure 5.10, left panel). Further, applications with an international application as a direct parent (right panel) have accounted for growing proportions of new applications. At the same time, applications with US Utility parents have declined relative to Foreign, Provisional and Parentless applications. While not directly related to complexity, the change in composition may be correlated with more work per application to the extent that international applications may represent more work, and a child of a US Utility may be somewhat easier to examine than an application with little or no known prior history. Additionally, the survival results in Section 5.3 and Appendix E show that applications with a US Utility parent are fastest to first action.

To the extent that these factors are correlated with increasing complexity (and, more pointedly, workload), then it is possible that the changing composition of new applications may help explain the growth in patent application inventories.

5.1.5 RCEs appear to crowd out other applications

Examiners in the US have a production expectancy goal that serves as their major incentive in terms of performance review, bonuses, and promotion. Some actions serve to increase an examiner's production count against the expectancy goal and some do not. The primary contributors to counts are first actions and disposals (including abandonments). Of particular import here is that when an applicant files for an RCE, it is counted as a disposal in the examiner count system. That is, an RCE is considered an abandonment of the original application followed by the filing of a "new" continued application. As such, they contribute to examiners' production. Additionally, any major action after an RCE is filed (whether the first, second, third, etc.) is considered a first action and also contributes toward production. Changes in the count system have altered the relative benefit accruing from different actions, but the primary structure remains intact.

Figure 5.11 Stocks as proportions of all pending applications in the US



Source: USPTO 2013

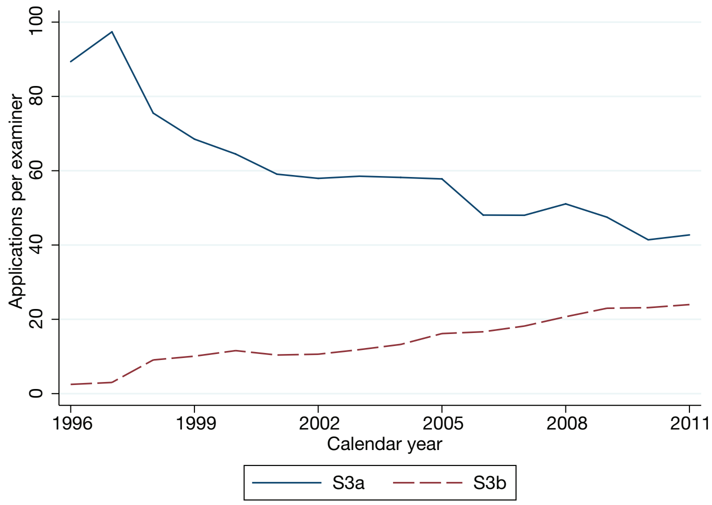
To further analyse Stock 3, we examine the two sub-stocks defined above.[44] Stock 3a is composed of applications that have received a first action, but have not filed an RCE. Stock 3b is composed of all applications that have received at least one RCE. Figure 5.11 shows that S3b has been a growing share of the entire inventory of applications. The adoption rate of RCEs from their inception partially accounts for the growth in S3b over the last 10 years.[45] S3b currently accounts for 15% of the entire inventory of applications and 30% of active examinations (Stock 3).

44 In making international comparisons, we recognize that the basic three stocks may not be sufficient for analysing all the idiosyncrasies across particular offices.

45 Although, recall that our definition of S3b includes all non-serialized continuations.

On a per examiner basis, RCEs have also grown. Figure 5.12 demonstrates this by graphing S3a and S3b stocks per examiner. While the average number of active non-RCE examinations has fallen on a per examiner basis, RCEs per examiner have grown. But, importantly, the reduction in non-RCE workload has been roughly twice as steep as the rise in RCEs. Implicitly, that means that each new RCE "crowds out" two regular applications. It also implies that examiner counts have been coming increasingly from RCEs relative to first actions or final disposals.

Figure 5.12 Stock 3a and Stock 3b relative to examiner staffing in the US



Source: USPTO 2013

These stylized facts should not be interpreted as an indictment of RCEs in general. It may be that RCEs serve a useful purpose in terms of error correction, or by providing a greater ability for applicants to narrow their claims in order to satisfy patentability standards. Nonetheless, they do come at a cost with respect to resource allocation.

5.2 Pendency

Chapter 3 discussed in detail the distinction between entry pendency, exit pendency, and expected pendency. This section will describe pendency in the US in greater detail, and will emphasize the importance of distinguishing between first-action pendency and post-first-action pendency in the US context. We also describe the competing risk between abandonment and patenting, and it's relevance for understanding pendency. Section 5.3 explores the relationship between patent inventories and pendency more precisely using survival time regressions.

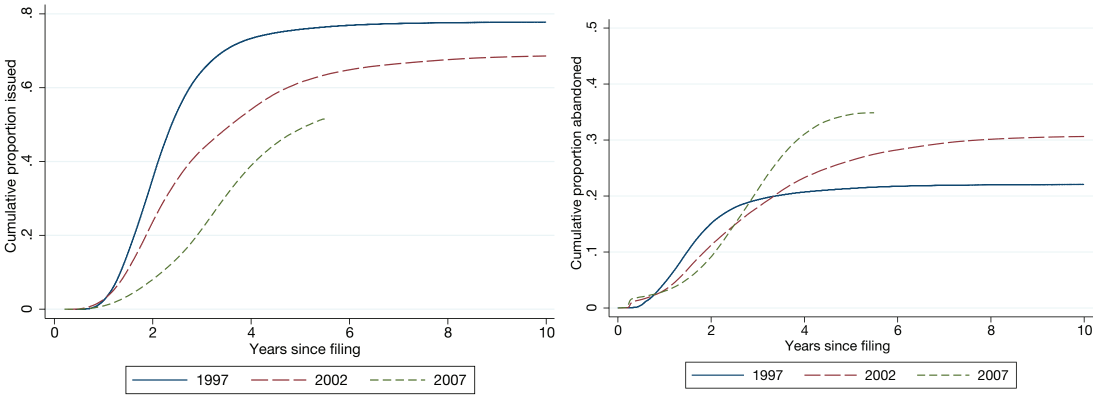
5.2.1 Disposal rates versus disposal types: Understanding cumulative patenting and abandonment

As discussed in Chapter 4, survival curves graph the proportion of applications that survive before a particular event occurs (such as disposal). The curves are useful for examining the entire distribution of pendency for a given filing cohort, or group of cohorts. However, they can become quite difficult to interpret when there are competing events that may lead to disposal. In the case of patent applications, the competing alternative to patenting—the competing risk—is abandonment.[46] Both patenting and abandonment lead to disposal, but they are very different outcomes, and may have different properties.

In the case of multiple outcomes—like patenting and abandonment—it is often useful to examine "cumulative incidence functions" (CIFs) rather than survival curves. CIFs are similar to survival curves, except that they graph the incidence of a particular outcome rather than those that have survived without experiencing that outcome (i.e., the "non-incidence"). The total proportion surviving will be exactly equal to the one minus the total proportion disposed. So, in some sense the two ways of graphing the data are identical. However, when there are multiple disposal types, CIFs can help to identify differences among them.

46 Recall that there are no terminal rejections, only abandonment of rejected applications.

Figure 5.13 Cumulative proportion of applications disposed over time by filing cohort



Source: USPTO 2013

Note: The left panel shows the patented applications and the right panel shows abandoned applications.

The left panel of Figure 5.13 graphs the proportion of all applications filed that are patented after a certain period of time, for three representative filing cohorts. Thus, the graph indicates both the speed of patenting and the ultimate patenting rate. Because the disposals can also occur via abandonment, abandonments should be examined in tandem (the right panel of Figure 5.13).

Numerically, after two years approximately 40% of 1997 filings were patented and 15% were abandoned. In comparison, the comparable numbers for 2002 and 2007 were about 20% and 10% patented, respectively, and 10% and 8% abandoned, respectively. Thus, the total disposal rate two years after filing was much higher for 1997 because the sum of the patent grants and abandonments was greater than the sum for 2002 and 2007. Graphically, this is indicated by the fact that the patenting and abandonment curves for 1997 are above their respective curves for the other cohorts at year two.

It is of particular import to note that the 1997 abandonment curve crosses the 2002 and 2007 abandonment curves at about year three. This does not mean that the disposal rate slowed, but rather, because the disposal rate was so high almost all the 1997 filings were disposed of by three or four years after filing. The ultimate patenting and abandonment proportions are indicated by the maximum that each curve attains. After ten years, there are almost no pending applications remaining, so the sum of the patented applications and the abandoned applications should be very close to one.

In this way, it can be seen that 1997 had an ultimate patenting proportion of about 0.79 (with the corresponding abandonment rate of 0.21). The ultimate patenting proportion was about 0.69 for 2003. Not enough data are available for 2007, but it appears that it will have a lower patenting proportion than either 1997 or 2002. We will return to this discussion in the regression analysis below.

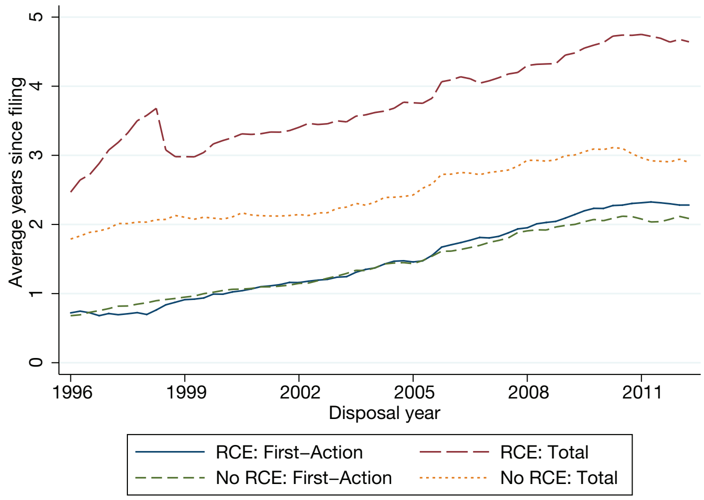
It is important to note that the ultimate abandonment proportion is negatively correlated with the speed of patenting, at least for the cohorts that are graphed. In instances where the speed to patenting is high, the abandonment rate is low. One possible explanation could be that some applicants are "discouraged" by greater pendency, and they become more likely to abandon, leading to a higher abandonment rate. An alternative explanation is that greater pendency may create an unintended opportunity for applicants to evaluate the commercialization opportunities of the innovation. We leave the motivation for abandonment for future research. In the present context it is sufficient to note that larger patent application inventories have at least two effects on incoming applications: longer pendency and more abandonment.

As noted in Section 3.1, the correlation between pendency and abandonment may cause post-first-action pendency to be understated. In fact, in Section 5.3 we show that higher first-action pendency leads to a higher post-first-action pendency for patenting. However, we also find that abandonments speed up. Thus, the net post-first-action pendency (including both patents and abandonments) may not reflect the true impact of higher inventories.

5.2.2 Queuing time

Figure 5.14 shows that the increase in average total pendency for granted patents is driven primarily by the queuing time for first-actions. Average first-action pendency more than doubled between 1996 and 2006, based on exit pendency at the time of grant,[47] and has remained at over two years beginning in 2009. Notably, there is not a significant difference in first-action pendency between those applications that will ultimately file an RCE and those that will not.

Figure 5.14 Exit pendency of issued applications by RCE status



Source: USPTO 2013

Note: reminder that this is not first action on the RCE filing, but first action on cases that will eventually have an RCE.

47 In other words, the first-action pendency in 2011 does not track applications with a first action in 2011. This was done in order to ensure that the mean difference between total pendency and first-action pendency reflects the actual difference for a consistent set of applications, in this case those that were granted in 2011.

In contrast, the mean time between first action and disposal (post-first-action pendency) differs significantly between applications that file an RCE and those that do not. Obviously those with RCEs will tend to have a longer post-first-action pendency in part because of the increase in transactions. In addition, the difference in post-first-action pendency between RCEs and non-RCEs has been diverging since 2006.[48] Part of that difference may be due to changes in docket management implemented in November 2009,[49] which gave RCEs a lower priority relative to other applications. Note that the difference in post-first-action pendency diverged more significantly after 2010.

We can make two observations about the growth in pendency. First, the major contributor to growth in total pendency is first-action pendency. That is, the processing time for active cases has not increased (and it may have even decreased for non-RCEs). Second, the greater intensity of RCE use is the second major contributor to total pendency.

The relative importance of the two main contributors to total patent pendency can be approximated by calculating the total pendency for the mean patent grant. For patents issued in 2000, the mean total pendency was 2.25 years, and for 2008 it was 3.29 years. Of the 1.05 year increase, about 86% can be attributed to the one year increase in queuing time. The remaining 14% can be directly attributed to an increase in RCE usage.

5.3 Survival time analysis

Because of the multiple factors involved in analysing the effects of application stocks on the pendency of incoming applications, regression analysis is appropriate. In particular, we utilize survival time regression to estimate the impact of pending inventories and application characteristics on the realized pendency of incoming applications. Survival analysis is well-suited for analysing patent pendency for several reasons.[50] First, it deals appropriately with "censored" data. In this case, censored observations include all pending applications for which patenting or abandonment cannot yet be observed. Computing pendency only for disposed applications will lead to a bias if many applications are still pending, as discussed in Chapter 3. This is the primary problem with using entry pendency as a measure of expected pendency. Second, survival regressions enable us to better understand the relationship between different observable variables and pendency, holding all other factors constant. For instance, it enables us to determine whether the size of Stock 1 or Stock 2 has a greater impact on pendency. Third, the estimated models allow for the prediction of pendency based only on variables that are observable as of the date the application is filed. Last, survival regressions enable us to account for the dual outcomes of patenting and abandonment, by using the competing risk methodology.

48 The peak in non-serialized continuations in the late 1990s is an artifact of the introduction of, and transition to, RCEs from other forms of non-serialized continuations.

49 See http://www.uspto.gov/patents/init_events/CountSystem.jsp#heading-3 and http://www.uspto.gov/patents/law/notices/rce_docket.pdf.

50 For a more thorough treatment of survival analysis see: Box-Steffensmeier, J. M.;Jones, B. S. (2004). *Event History Modeling: A Guide for Social Scientists*. Cambridge: Cambridge University Press or Kalbfleisch, J. D., and Ross L. Prentice. 1980 (1st ed.), 2002 (2nd ed.). *The Statistical Analysis of Failure Time Data.* New York: John Wiley & Sons or Cleves, M.; Gutierrez, R. G.; Gould, W.; Marchenko, Y. V. (2010 [3rd ed.]). *An Introduction to Survival Analysis Using Stata.* College Station: Stata Press.

The data and regression results are described in more detail below and in Appendix E. However, our main qualitative results can be summarized as follows:

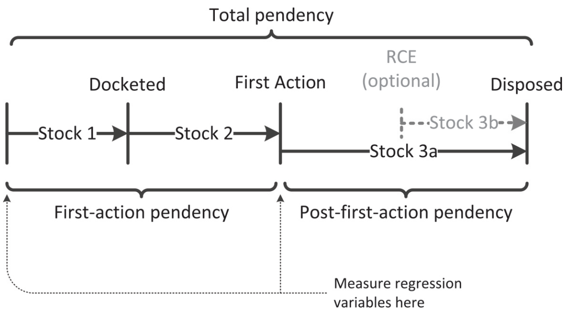
- Pending application stocks and the number of examiners have a significant impact on the first-action pendency for incoming applications
    - Stocks 1 and 2 increase the first-action pendency of incoming applications
    - Relative to other Stock 3 applications, the stock of applications that have filed at least one RCE significantly increases first-action pendency
    - The number of examiners decreases first-action pendency
- Stock counts have less of an impact on post-first-action pendency. However, there is a significant impact on the type of disposal
    - Stocks 1 and 2 tend to increase the rate of abandonment and decrease the rate of patenting (leaving the overall disposal rate about the same)
    - Stock 3 (both RCEs and non-RCEs) tends to increase post-first-action pendency (for both patents and abandonments), without significantly changing the rate of abandonment.
    - First-action pendency itself has an impact on the *type* of subsequent disposal: applications with high first-action pendency are much more likely to abandon.

In discussing the post-first-action pendency results, we rely on a graphical approach for the average patent application in a cohort. Appendix E provides more information about calculating pendency changes across the entire distribution of patent applications, rather than just at the mean.

5.3.1 Data and methodology

When examining pendency, we consider two distinct phases of total pendency, as shown in Figure 5.15. First-action pendency is the time from filing to first action (i.e., the application passes through S1 and S2); post-first-action pendency is the time the application passes through S3. We find that these two phases of total pendency are very different from one another. First, first-action pendency is not subject to the competing risk of abandonment,[51] and second, explanatory factors have a very different influence on first-action pendency in comparison to post-first-action pendency.

Figure 5.15 Phases of total pendency analysed in Chapter 5



Source: USPTO 2013

The data for analysing US patent pendency comprise a 10% sample of all regular utility patent applications filed at the USPTO from 1996-2011. The sample excludes reissues and re-examinations. It also excludes Patent Cooperation Treaty (PCT) applications, although it includes any regular US utility applications that matured as a national stage entry from a PCT application. Table 5.1 describes the final sample, which includes over 460,000 applications, of which approximately 100,000 were still pending as of June 2012. Electronics applications account for a large plurality of applications in the sample (46%), followed by mechanical (30%) and chemical (24%) applications.[52]

51 Technically, applicants may abandon prior to first action, but only through "express abandonment." There is no financial benefit for abandoning prior to first action, because the substantive examination is paid for at the time of filing. Applicants may wish to abandon prior to publication in order to keep the application secret. However, express abandonment is rare, comprising less than 2% of S2 applications.

52 Note that the mechanical discipline includes the transportation and business method art units.

Table 5.1 Summary of the regression sample

		Chemical	Electrical	Mechanical	Total
Applications		112,458	211,491	138,047	461,996
First Actions		97,615	186,368	121,279	405,262
	First-Action Pendency	1.63	1.91	1.60	1.75
Pending		22,370	47,944	29,538	99,852
	without First Action	9,523	22,583	13,934	46,040
	with First Action	12,847	25,361	15,604	53,812
Abandoned		36,471	41,210	39,170	116,851
	without First Action	5,320	2,540	2,834	10,694
	with First Action	31,151	38,670	36,336	106,157
	First-Action Pendency	1.78	2.13	1.75	1.90
	Total Pendency	2.85	3.05	2.59	2.84
Granted		53,617	122,337	69,339	245,293
	First-Action Pendency	1.43	1.76	1.38	1.58
	Total Pendency	2.77	2.95	2.54	2.79
Disposed after First Action		84,768	161,007	105,675	351,450
	First-Action Pendency	1.56	1.85	1.51	1.68
	Total Pendency	2.80	2.97	2.56	2.81

Note: Pendency figures reported above reflect the sample mean pendency (years) for applications with a first action.

Over half (53%) of the pending applications are currently within S3 (those with a first action). The remaining pending applications are in Stocks 1 or 2. We provide summary data on abandonments and granted patents separately, and we also provide the combined category of all disposals after first action. It can be seen that a minority of abandonments occur prior to a first action. Just from the averages in the table we can see that the first action pendency for abandonments tends to be greater than that of grants. Further, pendency tends to be highest for electronics, followed by chemical, and then mechanical.

The explanatory variables used to estimate pendency fall into two main categories: stock (count) variables, and application characteristics. The stock variables are environmental variables common to all applications within a technological discipline in a given month.[53] The application characteristics control for some of the observed heterogeneity among applications. Two of the characteristics are continuous variables (number of claims at filing and first-action pendency) and the rest are indicator variables (small entity status, etc.). For each incoming application, stock variables are measured as of time of filing for the purpose of estimating first-action pendency, and are measured as of the time of first action for the purpose of estimating post-first-action pendency, as shown in Figure 5.15. In other words, all variables are measured at the point the application begins its pendency period. This facilitates the estimation of predicted

53 Note that all stocks are measured at the discipline-month level. For incoming chemical applications, it makes little difference what the stocks of pending electronics applications are, or how many examiners there are for electronics applications.

pendency based only on variables that are observable at the time of entry. Table 5.2 provides some summary data for the variables used in the regressions.

Table 5.2 Summary statistics for survival regression variables

	First-action pendency regression			Post-first-action pendency regressions		
	Chemical	Electrical	Mechanical	Chemical	Electrical	Mechanical
Stocks (mean monthly counts)						
New applications	6,038	11,860	7,514	--	--	--
Stock 1	45,239	88,867	58,422	52,428	105,647	68,104
Stock 2	69,171	165,415	79,024	72,804	178,637	84,263
Stock 3a	65,023	109,905	68,829	65,845	117,222	71,408
Stock 3b (RCEs)	19,808	35,838	18,198	22,776	43,040	21,554
Junior examiners	477	1,468	580	506	1,615	642
Senior examiners	515	884	471	539	1,004	492
Application characteristics (means)						
First-action pendency (years)	--	--	--	1.63	1.91	1.60
Claims at filing	21.4	19.8	18.3	21.3	19.9	18.1
Application indicators (proportion)						
Small entity	0.29	0.18	0.47	0.28	0.18	0.47
International parent	0.39	0.39	0.30	0.38	0.38	0.29
No parent	0.20	0.33	0.38	0.21	0.34	0.39
Provisional parent[55]	0.15	0.11	0.15	0.15	0.10	0.14
US parent	0.26	0.17	0.17	0.27	0.18	0.17
Sample: 10% sample, 1996-2011	112,458	211,491	138,047	97,615	186,368	121,279

For our sample, chemical applications tend to have the most claims at filing. For both chemical and electronics applications, applications with an international priority are most common (representing a plurality of 39% of applications in the first-action pendency regressions). Mechanical applications appear to be different in at least two dimensions. They most commonly have no priority parent application (38%), and small entities are much more common in the mechanical arts.

54 For more information on US provisional applications see http://www.uspto.gov/patents/resources/types/provapp.jsp

We expect the application stocks to be positively correlated with the pendency of incoming applications, with the opposite to be true for the number of examiners. To the extent the number of claims is correlated with complexity, we expect the number of claims to be positively correlated with pendency because more complex applications may take longer to prosecute. However, this impact should largely be felt during examination, and not during queuing time. Similarly, we may expect that small entities have a slight disadvantage in prosecution time because they may not have well developed procedures for prosecuting applications.

The discipline, parent, and filing year indicators mainly serve as controls, but we discuss their impacts where appropriate. For instance, applications in electronics may, on average, be more complex than those in the mechanical arts. If we did not control for differences across these disciplines, some of the longer pendency due to more complexity would be erroneously attributed to other variables. In all regressions, the variable to be predicted (the dependent variable) is the observed duration until the given event (first action or final disposal).[55]

5.3.2 First-action pendency

Table 5.3 shows a summary of the regression results for the three pendency models estimated for the sample. The first-action pendency results are displayed in Column 1. To be able to make relevant comparisons across different variables, we evaluate the impact on pendency for meaningful changes in the explanatory variables. For the stock variables, we evaluate the impact at a large year to year change. The largest year-to-year change observed over the past 10 years tends to be around 20-50% for all the stock variables. Thus, for the purpose of comparison, we assume a 30% increase in the stock variables. In this way, we can compare the impact of the explanatory variables on the basis of a large-but possible-one year change. The formal regression coefficients can be found in Appendix E. The predicted response is shown in Column 1, in terms of a change in first-action pendency due to the change in the explanatory variable.

55 For all reported regressions, we employ the Weibull parametric form for the hazard function. We use the Weibull model because of the flexibility of the approach and the ease in predicting pendency times. However, we have estimated other models, including other parametric survival time models like the Lognormal, and the semi-parametric Cox proportional hazards model. All the models yield the same basic results with respect to the variables of interest. See Box-Steffensmeier and Jones (2004); Kalbfleisch and Prentice (2002); or Cleves et al (2010) for discussion of various parametric models.

Table 5.3 Summary of regression results: Predicted change in pendency

	Predicted change in pendency		
	First Action	Post-first action	
		Patent	Abandon
Stocks (30% change)			
New applications	1.5%	--	--
Stock 1	7.5%	n/s	-3.3%
Stock 2	7.0%	3.8%	-5.1%
Stock 3a	-11.5%	3.9%	n/s
Stock 3b (RCEs)	2.2%	0.8%	1.2%
Junior examiners	-3.6%	-4.5%	n/s
Senior examiners	-3.3%	-7.1%	n/s
Application characteristics (100% change)			
First-action pendency (years)	--	15.2%	-15.7%
Claims at filing	8.2%	5.4%	20.0%
Application indicators (discrete change)			
Small entity	-0.6%	2.7%	-48.2%
Discipline, electronics	9.8%	-12.6%	59.5%
Discipline, mechanical	-4.3%	-15.9%	17.7%
Parent, International	1.3%	n/s	1.8%
Parent, Provisional	5.0%	14.9%	2.8%
Parent, US CON/DIV/CIP	-17.8%	8.8%	7.3%

Notes: all changes are statistically significant at the 5% level, except where noted as n/s (not significant). Response in pendency for indicator variables represents a discrete change relative to the omitted category ("chemical" for Discipline, and "no parent" for Parent).

Of the inventory variables, the volume of new applications, Stock 1, and Stock 2 all serve to increase the first-action pendency of any newly arriving application. This is expected because examiners will generally need to work through current inventory before getting to newer applications. Quantitatively, S1 and S2 have roughly the same impact: a 30% increase is associated with a 7.5% or 7% increase in the first-action pendency of incoming applications. Because the impacts are so similar, it may not be necessary to distinguish between the two stocks of unexamined applications.

The effect of Stock 3 is more nuanced. Because S3 is really the main bottleneck to issuance, its effect is different from S1 and S2. In particular, S3a tends to be correlated with *lower* first-action pendency for incoming applications. In order for an application to enter into S3a, it must be that examiners have the capacity to make a first action. In this way, more applications in S3a reflect a higher carrying capacity for the examiner corps. Consequently, the size of S3a is correlated with lower first-action pendency for new applications. However, S3b has an effect similar to that of S1 and S2. The relative difference between S3a and S3b is important: shifting applications from S3a to S3b compounds the effects. Indeed, in order for an application to enter S3b, it must necessarily leave S3a. The relative difference between S3a and S3b is consistent with the descriptive analysis, which shows that RCE applications tend to crowd out regular applications.

The results for examiner counts are as one would expect: higher examiner counts tend to decrease first-action pendency. The effect of junior examiners is slightly greater for first-action pendency because newly hired examiners tend to work disproportionately on new cases.

One alternative way to understand the impacts of the stock variables is to consider the addition of a single new application or a single examiner, as described in the text box “Quantifying the impact of inventories and examiners.”

The calculations show that based on 2012 averages:

- One additional unexamined application adds approximately 39 seconds of first-action pendency to each incoming application, or around 5.6 months of total additional pendency across all incoming applications over the course of a year.

- Filing an RCE on an application that would otherwise not have filed an RCE adds approximately 66 seconds to the first-action pendency of each incoming application, or about 9.4 months of total additional pendency across all incoming applications over the course of a year.

- One additional junior examiner decreases the first-action pendency of incoming applications by 1191 seconds (19.8 minutes). Over the course of a year, that amounts to about 170.7 months of decreased first-action pendency across all incoming applications over the course of a year.

Quantifying the impact of inventories and examiners

To better understand the impact of stocks on the first-action pendency of new applications, we consider three different hypothetical situations. In the first hypothetical, we consider the contribution to first-action pendency of adding a single new application that remains unexamined for one year. Next, we consider the impact of converting one existing application from a regular application to an RCE application. Last, we consider the effect of adding one new junior examiner for one year. Recall that the impacts that we calculate reflect the effect of the additional hypothetical application or examiner on *other* applications.

Adding one unexamined application

A new unexamined application first enters the statistical model as part of the incoming cohort of new applications. After that initial month, it becomes part of Stock 1 and eventually Stock 2. For the purpose of this exercise, we assume that the new application is part of the incoming cohort for one month, then is in S1 for five months, after which it is in S2 for six months. In this way we account for the first year of the application's life. Quantitatively, each stock has a different impact on the 12 incoming cohorts of patent applications. Thus, we create a weighted average of the effect over 12 months reflecting one month spent as a new application, five months spent in S1, and six months spent in S2. The net result is that is that one additional unexamined application adds approximately 39 seconds of first-action pendency to all other incoming applications. Over the course of a year, that amounts to around 4126 hours (5.6 months) of additional first-action wait time added across all other incoming applications for that year, based on the 2012 monthly average of 31,418 new applications.

Adding one new RCE application

We make a similar computation for the hypothetical situation of converting one regular application in Stock 3 to an application with at least one RCE. That is, we remove one application from S3a and place it in S3b, indicating that it has received one or more RCE filings. For the purposes of calculating the annual impact, we assume that the application remains in Stock 3b for one year. Because S3b applications tend to increase the first-action pendency of incoming applications relative to S3a applications, the net effect will be an increase in first-action pendency for incoming applications. The calculation shows that one new RCE application adds approximately 65 seconds to the first-action pendency of each incoming application. It is important to recall that this does not mean that each RCE filing adds 65 seconds, because we define an RCE application as any application filing at least one RCE. Thus, we are comparing an RCE application to another application that otherwise would not have filed an RCE. Over the course of a year, the additional first-action wait time added across all incoming applications amounts to about 6879 hours (9.4 months) based on 2012 averages.

Adding one new junior examiner

In the last hypothetical situation, we compute the decrease in first-action pendency that would accrue from hiring one additional examiner. We assume that the additional examiner would be the average junior examiner; i.e., he or she would not be a brand new hire and would not require training (currently, the median junior examiner is a GS-11 or GS-12 on the General Schedule). We find that one additional junior examiner decreases the first-action pendency of incoming applications by 1191 seconds. Over the course of a year, that amounts to 124,722 hours (170.7 months) of decreased first-action wait time accumulated across all incoming applications in that year, based on 2012 averages.

The size of the effects may seem surprising. One examiner works for approximately 2000 hours per year, not all of which is spent on first actions. Thus, it may seem surprising (if not impossible) for one additional junior examiner to reduce total first action wait time by more than 124 thousand hours (a multiplier of 124 if examiners spend half their 2000 work hours on first actions). However, this sort of multiplier is a common feature of queuing models. To understand the effect of adding an additional examiner, consider a numerical example of opening one additional line at a grocery store.

Suppose that initially 40 customers are spread evenly across four lines with 10 customers each. For the sake of the example, each customer takes 1 minute to process (which is unrealistic, but simple). The wait time for the last person in each line will be 10 minutes (nine minutes waiting plus one minute checking out). The second to last person in each line will wait nine minutes, and so on. The total wait time for the customers in each line will be 10 + 9 + 8 + … + 1 = 55 minutes (even though all the waiting occurs over 10 minutes of real time). With four lines, the total wait time for all the customers will be 220 minutes, with the total labour time being 40 minutes (10 minutes for each cashier).

If the grocer adds a fifth cashier, what will the savings be in terms of total customer wait time? With five lines, there will be eight customers in each line. Each line will have a total wait time of 8 + 7 + … + 1 = 36 minutes, and with five lines the total is 180 minutes. So, by adding another cashier—working for eight minutes—we save 40 minutes of customer wait time.[56] The benefits arise from serving customers in parallel rather than in series.

In terms of additional applications in the queue, an analogy can be made to "rubber-necking" in traffic. When one driver takes 10 seconds to gawk at an accident by the side of the road, he adds 10 seconds to his own travel time. However, he also adds 10 seconds to each driver behind him meaning that the total increase in travel time far exceeds the 10 seconds of gawking. In urban economics similar calculations can be made for each additional driver on the road in terms of their contributions to overall congestion.

The quantitative impact of the stock variables and the application characteristics is described in more detail in Appendix E, including graphical representations about how changes in the explanatory variables affect the survival curves for first actions. Recall that the stock variables impact the pendency of incoming applications. In contrast, application-specific characteristics show how those characteristics affect the pendency of that application itself.

The impact of the indicator variables on pendency is measured for a change in the indicator from zero to one. Implicitly, this means that we are comparing the variable relative to the

56 Further, in this case total labour time is still 40 minutes (five cashiers working eight minutes each). To make the example more realistic, we would need to assume that eventually more customers will get into line and the cashiers will continue working. However, that unnecessarily complicates the arithmetic.

alternative. For instance, the effect of the small entity indicator of -0.6% means that a small entity has a first-action pendency that is 0.6% less than that of a large entity. For application characteristics that are continuous variables, we measure the impact of a 100% change in the explanatory variable. This quantity was chosen because it roughly corresponds to a change from the 25th percentile to the 75th percentile in terms of the value of the explanatory variable. For stocks, we measure the impact of a 30% increase in the stock. This corresponds to a large-but not unheard of-year-to-year change.[57]

It should be expected that application-specific characteristics should have little impact on first-action pendency. In fact, as a generality, it should be a policy goal that applications are treated equally. If applications are examined on a literal first-in/first-out basis, then application characteristics should be independent of first-action pendency. For instance, even though small entity status has a statistically significant effect on predicted first-action pendency, the magnitude is so small (-0.6% difference in first-action pendency) that there is no appreciable difference between the predicted first-action pendency of small entities versus large entities. The predicted survival curves in Appendix E (Figure E.5) confirm this.

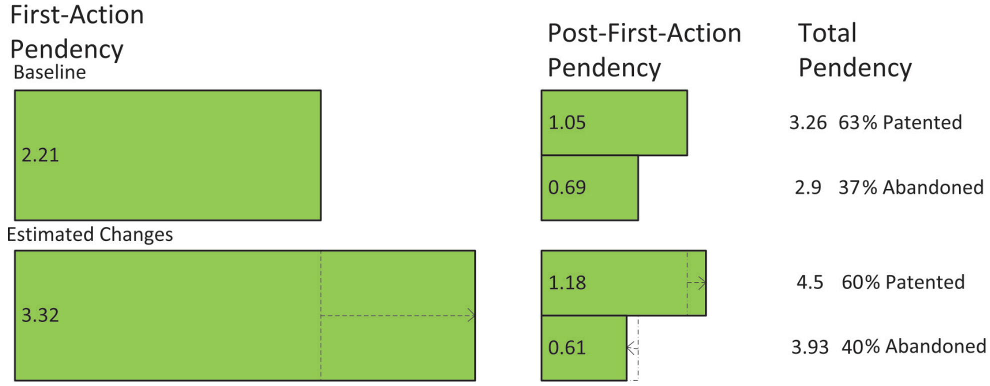
However, some application characteristics do seem to be correlated with differences in first-action pendency in an economically meaningful way, indicating a divergence from the first-in/first-out norm. Applications with more claims, electronics applications (broadly), and applications that claim priority to an international or US provisional parent have systematically longer first-action pendency. Conversely, applications with fewer claims, mechanical applications, and applications that claim priority to a previous US application have systematically lower first-action pendency. However, note that each discipline has its own queue, so that the discipline of the incoming application may affect first-action pendency for this reason alone. These results are detailed more fully in Appendix E.

57 Because stock variables are subject to a time trend, computing the 25th and 75th percentiles amounts to comparing earlier years to later years. Thus, we consider a large year-to-year change instead. With application characteristics, while there may be a time trend, most of the variation is due to differences between individual applications rather than between years. Thus, looking at the 25th to 75th percentiles is appropriate.

5.3.3 Post-first-action pendency

Columns 2 and 3 of Table 5.3 show a summary of the results of the two post-first-action pendency regressions.[58] Duration for these regressions is measured as the time between first action and disposal. However, because there are two possible disposal outcomes, we estimate the time to patenting and the time to abandonment separately. The results are combined to jointly determine the net post-first-action pendency and the abandonment rate. Note that all variables are measured at the time of first action. For that reason, we do not include newly arriving applications. Additionally, we include the observed first-action pendency for each application to determine whether it has an impact on post-first-action pendency. To understand the combined effects, one must look at the impact on both patenting and abandonment together. As an example, consider the effect of first-action pendency. Applications with longer first-action pendency tend to have longer wait times to patent (15.2% higher for a doubling in first-action pendency). In contrast, they have shorter wait times to abandonment (15.7% lower for a doubling in first-action pendency). That is, longer first-action pendency is correlated with faster (and more) abandonment and slower (and less) patenting. We describe the effects in more detail in Appendix E using cumulative incidence functions. However, the results are also quantified in Figure 5.16.

Figure 5.16: Estimated impact of increasing first-action pendency by 50%



Source: USPTO 2013

Note: Baseline is from means estimated for applications filed in 2007.

Figure 5.16 shows the impact of a 50% increase in first-action pendency on post-first-action pendency. Relative to a baseline application filed in 2007, first-action pendency increases from 2.21 years to 3.32 years. That influences post-first-action pendency by slowing down patenting: the time from first-action to patenting increases from 1.05 years to 1.18 years, a change of about 1.5 months. Additionally, the time to abandonment gets shorter, from 0.69 years to 0.61 years—a decrease of about one month.

58 More details are available in Appendix E. For more information about competing risks and cumulative incidence functions see Putter, H., M. Fiocco, R. B. Geskus. 2007. *Tutorial in Biostatistics: Competing Risks and Multi-State Models.* Statistics in Medicine, 26: 2389-2430.

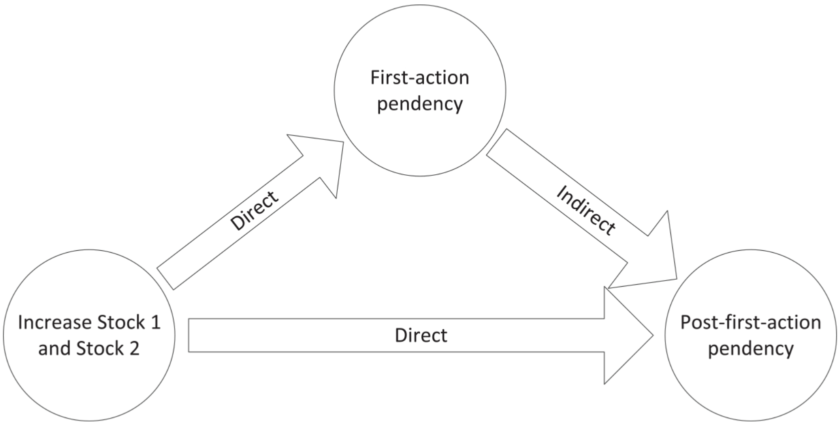
The figure shows the net effect on total pendency and disposal. Because the first-action pendency change is so large, overall pendency increases, even for abandonments. However, measured from the time of first action, abandonments occur at a faster rate, and patents occur at a slower rate. The net effect is to increase the frequency of abandonments (from 37% of applications to 40% of applications) and to decrease the frequency of grants (from 63% to 60%). In this way, the overall “allowance rate” at the USPTO can be influenced by pendency and the size of patent inventories, even when examination standards do not change. In fact, recent data shows a significant decrease in first-action pendency associated with an increase in the allowance rate, as measured by grants as a proportion of disposals.

Appendix E discusses in greater detail the way in which we combine the patenting and abandonment results. In some cases, like with first-action pendency, the effect on the abandonment rate is unambiguous: slowing patenting and speeding up abandonment will increase the abandonment rate. However, without more information, the effect on the average post-first-action pendency is ambiguous. The result depends on the relative proportion of patents to abandonments, and the quantitative impacts of the variable in question on patenting and abandonment.

In contrast, the effect on *pendency* may be unambiguous, but the abandonment rate could go either way. For instance, in the case of the number of claims, both patent pendency and the abandonment pendency are higher, so we know that average post-first-action pendency must increase. However, depending on which one slows more, we don’t know whether the net abandonment rate will increase or not. We consider each of the variables in depth in Appendix E. For the purposes of this section, we present the results of three hypothetical situations analogous to those from Section 5.3.1, in terms of the total result on first-action pendency and post-first-action pendency. First, we consider a significant increase in the number of unexamined applications. Second, we consider an increase in the usage of RCEs. Third, we consider an increase in the number of junior examiners.

In each hypothetical “experiment,” we analyse the impact of the variable(s) in question on first-action pendency (as described in Section 5.3.1) and also on post-first-action pendency. Figure 5.17 graphically shows the strategy in the example of an increase in Stocks 1 and 2. The increase in stocks has a direct effect both on first-action pendency and on post-first-action pendency. Both of these impacts serve to change total pendency. Further, there is an indirect effect that operates through the first-action pendency. In particular, S1 and S2 have a direct impact on first-action pendency. But, as described above in Figure 5.16, first-action pendency itself affects post-first-action pendency. In the analysis below, we describe both direct effects, as well as the indirect effect of first-action pendency on post-first-action pendency. In general, we find that the largest contribution to changes in total pendency derives from the direct effect on first-action pendency.

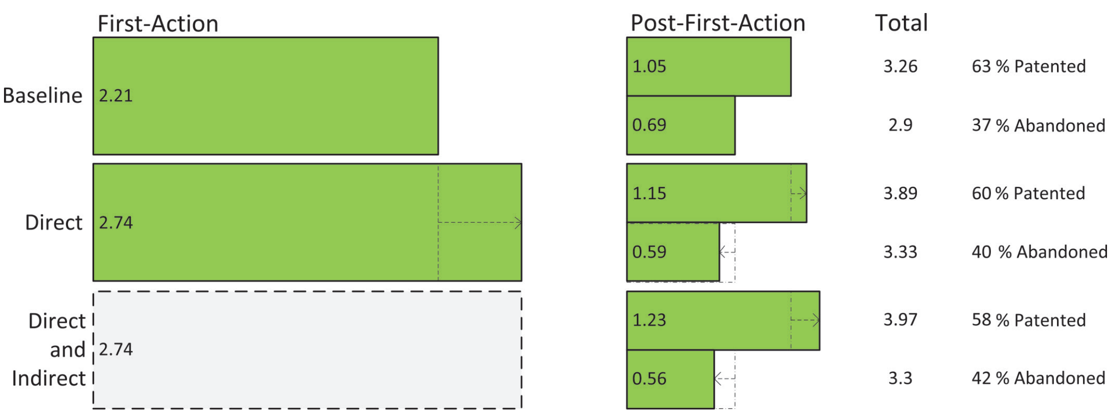
Figure 5.17: Influence of Stock 1 and Stock 2 on total pendency



Source: USPTO 2013

Figure 5.18 shows the impact of a 50% increase in Stocks 1 and 2 (representing an increase in unexamined applications). The second row shows the direct impact of this increase on first-action pendency and post-first-action pendency. Post-first-action pendency is shown separately for issued patents and abandonments. The third row incorporates the additional indirect effect of first-action pendency on post-first-action pendency, which tends to be relatively small in comparison to the direct effects. The net result is a 21.8% increase in total pendency for patented applications and a 13.8% increase in total pendency for abandonments. Further, the ultimate abandonment rate increases by five percentage points from 37% to 42%.

Figure 5.18: Estimated impact of increasing Stock 1 and Stock 2 by 50%

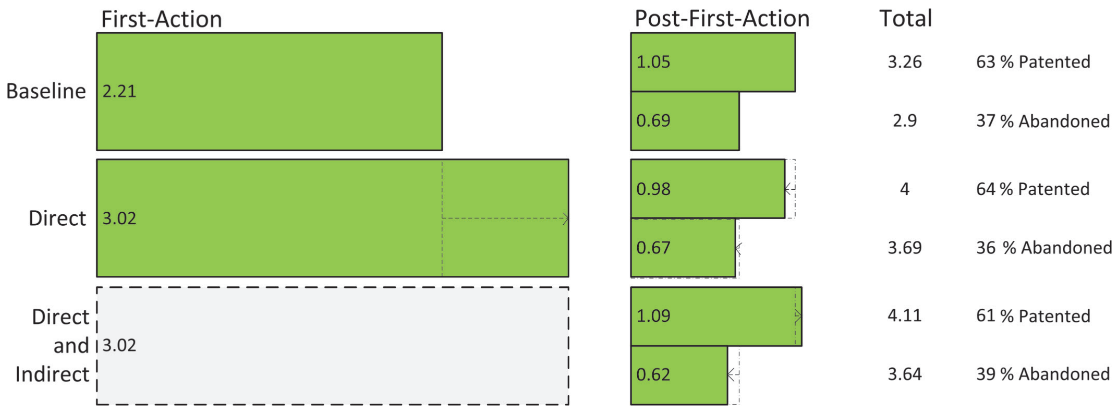


Source: USPTO 2013

Note: Baseline is from means estimated for applications filed in 2007.

Figure 5.19 shows the impact of increasing the uptake of RCEs. In particular, we consider doubling the proportion of Stock 3 applications that are in Stock 3b. For instance, the ratio of RCE applications to all of Stock 3 might go from 15% to 30% (as it did from 2000 to 2008). We assume that the overall size of S3 does not change, so that applications shift from S3a to S3b. The net result is primarily an increase in pendency, with total pendency increasing by 26.1% for patents and 25.5% for abandonments. The abandonment rate increases by two percentage points from 37% to 39%.

Figure 5.19: Estimated impact of a 100% increase in RCE uptake

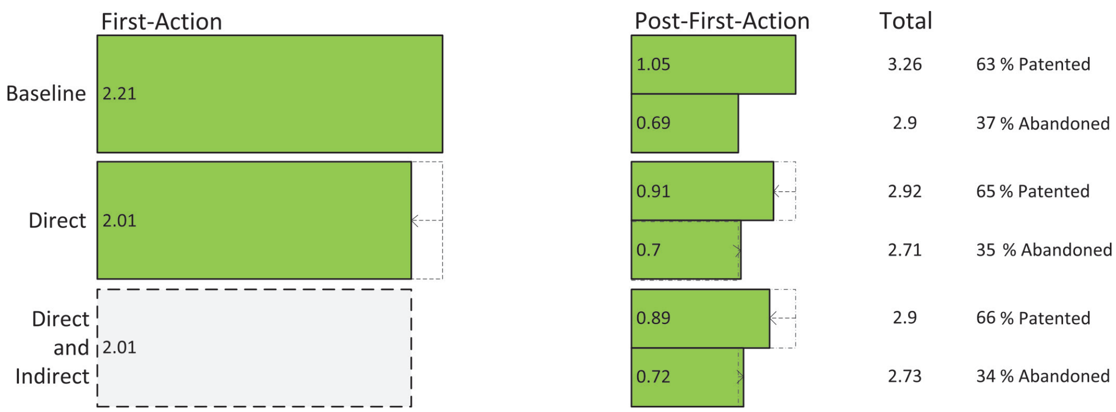


Source: USPTO 2013

Note: Baseline is from means estimated for applications filed in 2007.

Figure 5.20 shows the impact of doubling the number of junior examiners, which represents roughly a 50% increase in the number of examiners. This rather large increase in hiring is equivalent to moving from 2000 staffing levels to 2008 staffing levels. The net result is a decrease in total pendency of 11.0% for patents and 5.9% for abandonments. The abandonment rate decreases by three percentage points from 37% to 34%.

Figure 5.20: Estimated impact of a 100% increase in junior examiners



Source: USPTO 2013

Note: Baseline is from means estimated for applications filed in 2007.

Figures 5.16-5.20 show the predicted change in pendency for the mean patent application in the baseline year of 2007. Appendix E uses cumulative incidence functions (CIFs) to show the predicted change on the entire distribution of patent applications in the cohort. However, the qualitative results are not different from what is presented here. Our hope is that the methodology and the empirical results presented in this section will suggest avenues for future research.

5.4 Concluding remarks on the US analysis

The US analysis on pendency and patent applications inventory helps to shed light on several things. First, in Section 5.1, we found that some traditionally cited sources of the "backlog" may not explain the growth in patent inventories that we witnessed over the 2000s, and that USPTO hiring has kept pace with the numbers of incoming applications. In Section 5.2, we found that the growth in first-action pendency is the driving factor in the growth of total pendency.

The survival time regressions demonstrate the ability of our methods to more precisely quantify the relationship between, stocks, examiner capacity, and application specific characteristics and the first-action pendency of incoming applications. We also found that overall post-first-action pendency is not as sensitive to those same factors. However, the types of observed disposal (i.e., abandonment and allowance) are sensitive to stocks, examiner capacity, and application characteristics. Consequently, we find that abandonment rates (equivalently, allowance rates) vary significantly in ways that are due to pendency not necessarily with any changes in examination standards or examination quality.

However, it should be recognized that applicants do not bear the cost of pendency homogeneously. Nor are the costs of pendency without some potential benefit. For example, there can be value to applicants in the ability to manage the pace of patent prosecution and to buy extensions or continued prosecution. Applicants are able to weigh alternatives, assess their market, and assess competitors. Examiners are able to consider additional references, to interview applicants, and to require greater clarity or narrowing of claims. However, policies regarding continued prosecution must balance the direct costs and benefits to the applicant and examiner with the indirect costs that are imposed on other applicants who must incur additional wait time.

Much of our analysis found that the increasing use of requests for continued examinations (RCEs) seemed to be at the centre of the pendency issue, whether as a symptom or as a cause. For instance, we found that the conversion of applications from regular applications to RCEs disproportionately impacts the first-action pendency of incoming applications because RCEs compete with new applications for examiner time.[59] Furthermore, first-actions operate as a bottleneck, so that the greater use of RCEs serves to increase the stocks of unexamined applications (S1 and S2). At the same time, it is important to note that RCEs can be a useful method to correct errors, or for applicants to sufficiently narrow and clarify their claims. USPTO policy responses should address the question of pendency without unduly restricting access to quality examination[60]. Through RCE Roundtables the office has sought public comment about why RCEs are filed and how the need for filing an RCE could be reduced.[61]

To date, the USPTO has implemented several policies addressing the application inventory, either directly or indirectly, including: hiring additional examiners, changing the examiner count system, changing the examiner docket management system, initiatives related to the America Invents Act (e.g., expedited review and fee-setting), and new pilot programs designed to streamline patent prosecution (e.g., Quick Path Information Disclosure Statement and the After Final Consideration Pilot). Consequently, over the past two years there has been a significant decrease in the stock of unexamined patent applications (S1 and S2). Official numbers put the quantity of unexamined patent applications below 600,000 in February of 2013, with first-action pendency (exit pendency) at 19.2 months.[62] While we do not perform a detailed analysis of the efficacy of those policies in this report, the proposed methodology and survival time analysis can be used as a platform for future research in this area.

59 http://www.uspto.gov/patents/init_events/rce_outreach.jsp

60 For a more thorough discussion of patent quality, see: Kappos, D. J.; Graham, S. J. H. (2012). “The Case for Standard Measures of Patent Quality.” *MIT Sloan Management Review*, 53(3), pp 19-22.

61 http://www.uspto.gov/patents/init_events/rce_outreach.jsp

62 http://www.uspto.gov/dashboards/patents/main.dashxml accessed 11 April 2013.

6. Conclusion

All patent offices face the challenge of balancing several important policy goals. Any system of intellectual property rights must first serve the innovation marketplace, and must be responsive to the needs of innovators as well as to end users. A well-functioning system must ensure that a sustainable equilibrium can be maintained: balancing fees, costs, workload, pendency, all while subject to the budgetary constraints. At the core of patent policy is the issue of patent examination quality. Pendency goals alone cannot drive policy to the detriment of examination quality.

There are several avenues open to policy-makers to address pendency. Clearly, the most direct way to reduce pendency is through hiring, however various kinds of work-sharing between offices can partially substitute for this. However, as we found in Chapter 5, hiring in the US kept pace with the growth in new applications over the last 15 years, yet pendency grew significantly during part of that time. So, offices must look carefully to identify the sources of delay, and perhaps identify institutional changes that may be able to address pendency without sacrificing examination quality.

The results from the UK and US analyses should serve as examples of how policy-makers can identify contributors to pendency, and can suggest ways forward. For instance, the adoption of the combined search and examination (CSE) request in the UK provides a mechanism for applicants to significantly reduce pendency. In the US, changes instituted in 2009 in the docketing of Requests for Continued Examinations (RCEs), have given examiners a greater ability to prioritize first-actions rather than continued examination of RCEs. This small policy change may help to mitigate the deleterious effect that RCEs were shown to have on first-action pendency in Chapter 5.

In both offices, we find that the workload per application has increased. In the UK, the number of amendments required per application has increased since 2001 from 2000 - 2006, although this trend seems to be reversing itself in 2007-11. An additional amendment request is associated with about 20 weeks of additional pendency time. In the US, the number of major actions per disposal has been increasing over time, and the quantity of applications filing one or more RCE has been increasing as a share of all pending applications. Thus, in both countries, the ability of applicants to extend the prosecution of rejected or denied applications is one of the drivers of total pendency.

Extended prosecution by itself is neither good nor bad *per se*. It can serve as a useful method to correct errors, or for applicants to sufficiently narrow and clarify their claims. Indeed, some would suggest that in almost any patent application there is patentable subject matter, so long as it is narrowly claimed. At the same time, continued prosecution diverts resources away from other applications—in particular from first actions. In light of this, policy responses should address the question of pendency without restricting access to the benefits of continued prosecution. Offices should seek institutional methods to incentivize early stage claim specificity and narrowing.

Further, in both offices, we find that pendency is sensitive to similar types of inventory. In the UK, pendency is particularly sensitive to the size of Stock 1 (new applications undergoing clerical tasks and awaiting applicant requests plus fees to be filed) and Stock 2 outside the office (applicants taking time to file search or examination requests). In the US, we find that Stocks 1 and 2 affect the first-action pendency of incoming applications, and also the post-first-action pendency of applications that are already in Stock 3. The latter result shows that the pressure to do first-actions on incoming applications competes with the pressure to do continued examination on Stock 3 applications. The same result can be seen in the impact of Requests for Continued Examination on the first-action pendency of incoming applications. The results suggest avenues for future research, because they enable policy-makers to assess the quantitative impact of targeting specific inventories through increase examination capacity, institutional changes, or by changing fees. For instance, in 2010, the UK reduced the applicant's period of reply to certain examination reports from four months to two months. The US recently introduced tiered pricing for RCEs: the second and subsequent RCE cost more than the first.[63]

The methodology presented here is not meant to be comprehensive, and the results do not close the book on the subject of patent backlogs. However, our findings suggest that the framework is useful for bringing more precision to the discussion of patent application inventories and pendency. Further, the framework suggests a way to predict reasonable estimates of expected pendency in a way that is not subject to the biases of more traditional methods. Most importantly, we hope that this report will suggest avenues for continued research and greater collaboration among patent offices in coordinating research and sharing data.

63 See http://www.uspto.gov/web/offices/ac/qs/ope/fee031913.htm.

7.References

Bond, S.; Cummins, J. G. (2004). "Uncertainty and Investment: An Empirical Investigation using Data on Analysts Profit Forecasts". US Federal Reserve Bank, Finance and Economics Discussion Series 2004-20.

Box-Steffensmeier, J. M.; Jones, B. (2004). Event History Modelling: A Guide for Social Scientists. Cambridge: Cambridge University Press.

Cleves, M.; Gutierrez, R. G.; Gould, W.; Marchenko, Y. V. (2010 [3rd ed.]). *An Introduction to Survival Analysis Using Stata*. College Station, Texas: Stata Press.

The Economist (17 March, 2011). "The spluttering invention machine." Accessed online at: http://www.economist.com/node/18389167

The Economist (5 May, 2011). "Patently Absurd." Accessed online at: http://www.economist.com/node/18651194.

The Five IP Offices (IP5) (2011). IP5 Statistics Report 2011. Accessed online at: http://www.fiveipoffices.org/stats/statisticalreports/ip5-statistics-2011.pdf

Graham, S. J. H.; Mowery, D. C. (2005), "The Use of USPTO 'Continuation' Applications in the Patent of Software: Implications for Free and Open Source". Law & Policy, 27(1), p128-151.

Henkel, J.; Jell, F. (2011). "Patent Pending – why faster isn't always better." Patent Law e-Journal. Issue 01/201. Accessed online at: http://www.researchgate.net/publication/228173643_Patent_Pending__Why_Faster_Isnt_Always_Better

Kalbfleisch, J. D.; Prentice, R. L. (2002 [2nd ed.]). The Statistical Analysis of Failure Time Data. New York: John Wiley & Sons.

Kappos, D. J.; Graham, S. J. H. (2012). "The Case for Standard Measures of Patent Quality." MIT Sloan Management Review, 53(3), pp 19-22.

London Economics (January 2010). Patent Backlogs and Mutual Recognition. Accessed online at: http://www.ipo.gov.uk/p-backlog-report.pdf

Mejer, M.; van Pottelsberghe de la Potterie, B. (2011). "Patent backlogs at USPTO and EPO: Systemic failure vs. deliberate delays." World Patent Information 33(2), pp. 122-7.

Putter, H., M. Fiocco, R. B. Geskus (2007). Tutorial in Biostatistics: Competing Risks and Multi-State Models. Statistics in Medicine, 26: 2389-2430.

UK Intellectual Property Office (2013). The Manual of Patent Practice. Accessed online at: http://www.ipo.gov.uk/pro-types/pro-patent/p-law/p-manual/p-manual-practice.htm

World Intellectual Property Office (2009). "WIPO Symposium Concludes Global Patent Application Backlogs Unsustainable." Geneva, Switzerland: WIPO. Accessed online at: http://www.wipo.int/pressroom/en/articles/2009/article_0035.html.

Appendix A. Glossary

AIA (US) – On September 16, 2011 U.S. President Barack Obama signed into law (P.L. 112-29) the Leahy-Smith America Invents Act (AIA). For more information see http://www.uspto.gov/aia_implementation/global_impacts.jsp.

Censoring – When analysing the time between events, (right) censoring occurs when the analyst does not observe the event before the end of the observation period. Some statistical measures may be biased due to censoring, e.g. the mean total pendency of an entry cohort continues to increase until all applications in that cohort have been disposed. In this report, when we say something was censored, we mean an application had not yet experienced some event (first action or disposal) at the patent office when the data were analysed.

COPA (US) – Clearing the Oldest Patent Applications. An initiative begun in April 2011 to perform an initial examination for all applications filed on or before June 7, 2009. Since COPA prioritized conducting a first examination on older, unexamined applications, it likely increased the observed exit pendency beginning in April 2011 because more first actions were performed and a higher percentage of those were performed on older applications.

CSE (UK) – Combined Search and Exam. The full first examination is performed at the same time as the initial search and the traditionally separate search and exam fees are combined.

Competing Risk – An additional event which may occur before the event of interest, e.g. an application may abandon before it is determined to be allowable and issued, which precludes observation of the event of interest.

Cumulative Incidence Function – A function expressing the probability that a subject has experienced an event of interest at some point between an origin event and a certain time. When there are no competing risks, the cumulative incidence function = 1 – survival function. See also Section 3.5.

EP – European patent. The European Patent Convention (EPC) makes it possible to obtain patent protection in about 40 European countries on the basis of a single application. An application for a European Patent can be filed at the European Patent Office in Munich, the Hague or Berlin or at the central industrial property office of any contracting state. It can also be derived from an international application. The applicant selects the countries in which he wants protection. European patents are granted centrally by the European Patent Office and are renewed in individual member countries. For more information, please consult http://www.epo.org/applying/basics.html.

EPO – The European Patent Office is a body of the European Patent Organisation, set up on 7 October 1977 on the basis of the European Patent Convention (EPC) signed in Munich in 1973. More information about the EPO can be found online at http://www.epo.org/.

Entry Cohort – A group of applications which enter a stock or process in the same time period (e.g. within the same month, quarter, or year). In the UK analysis, there are entry cohorts based on the time of filing for a patent. In the US analysis, there are entry cohorts based on the time of filing and based on the time of first action.

Entry Pendency – Pendency measured for the same entry cohort. It measures the average (or median) pendency for all applications entering a stock in a given time period. Entry pendency is only accurate if sufficient time has elapsed to observe a large majority of disposals (see Censoring). It does not account for the omission of still pending applications from that cohort. However, if all applications have exited the stock, then entry pendency is a perfect measure for the actual pendency of all applications in that entry cohort.

Exit Cohort – A group of applications which exit a stock or process in the same time period (e.g. within the same month, quarter, or year). In the UK analysis, there are exit cohorts based on the time of disposal (patented or abandoned). In the US analysis, there are exit cohorts based on the time of first action and based on the time of disposal (patented or abandoned).

Exit Pendency – The most commonly applied pendency measure. It is based on the exit cohort (all applications exiting a process in the same time period). That is, exit total pendency measures the average (or median) time from filing to disposal for all disposals in a given time period. It does not account for the observed pendency of any pending applications, which leads to a bias.

Expected pendency – An estimate of the pendency that an applicant expects to experience (on average) as of the day of filing. Chapters 4 and 5 provide details on methodology for predicting pendency.

Final Rejection (US) – A rejection of an application on the same grounds as a preceding rejection which closes prosecution. There are limited avenues for an applicant to continue prosecution after a final rejection.

First Action (US) – The first time an examiner searches the prior art, examines the application, and issues a statement to the applicant regarding the status of the application. It is also the event which marks the transition from Stock 2 to Stock 3. We do not include restriction requirements as a first action, thus an application which receives a restriction requirement must wait for an applicant decision before the examiner will issue a first action. The main first action events in our data are the first time an application receives a non-final rejection or a notice of allowance.

First-action Pendency – The time between receipt and first action on the merits. Also, the length of time an application remains in Stocks 1 and 2.

FTE (US) – Full Time Equivalent. A measure of examiner staffing which accounts for expected productivity based on the technology being examined and the experience level of the examiner. It converts raw examiner counts to a normalized count based on the equivalent number of full-time GS-12 examiners, where GS-12 is based on the General Schedule for US federal government employment: https://www.opm.gov/policy-data-oversight/pay-leave/pay-systems/general-schedule/ .

IPC – The International Patent Classification. Established by the Strasbourg Agreement (1971) and provides for a hierarchical system of language independent symbols for the classification of patents and utility models according to the different areas of technology to which they pertain. For more information, see: http://www.wipo.int/classifications/ipc/en/.

Major Actions (US) – Non-final rejections, final rejections, and notices of allowance.

Manual of Patent Practice (UK) - The Manual of Patent Practice explains the UK IPO's practice under the Patents Act 1977. The Manual recites sections of the Patents Act 1977 and sections of the Copyright, Designs and Patents Act 1988 relating to patents. For more information, please see: http://www.ipo.gov.uk/pro-types/pro-patent/p-law/p-manual/p-manual-practice.htm.

MPEP (US) – Manual of Patent Examining Procedure. Documentation and guidance on the practices for the examination and prosecution of patent applications at the USPTO. For more information, please see: http://www.uspto.gov/web/offices/pac/mpep/.

Non-final Rejection (US) – A rejection of an application where the grounds are presented to the applicant for the first time.

Non-serialized continuation (US) – Non-serialized applications can be used to extend prosecution of the current application. They do not receive a new serial number and are not substantially different from the original applications. They do not go through classification and docketing as new applications do. Rather, they are immediately docketed to the same examiner, and involve different fees than do serialized continuations. While there have been other forms of non-serialized continuations, RCEs (instituted in 1999) are by far the most common. (See RCE).

PCT or "Convention" - The Patent Cooperation Treaty or convention makes it possible to seek patent protection for an invention simultaneously in each of a large number of countries by filing an "international" patent application. Such an application may be filed by anyone who is a national or resident of a PCT contracting State. It may generally be filed with the national patent office of the contracting State of which the applicant is a national or resident or, at the applicant's option, with the International Bureau of WIPO in Geneva. For more information, please see: http://www.wipo.int/treaties/en/registration/pct/.

Pendency – In general pendency refers to wait time or queuing time. With respect to patent applications, total pendency refers to the total time between receipt and disposal. In the US, we also examine first-action pendency and post-first-action pendency (see those entries). Patent pendency refers to total pendency calculated only for issued applications.

Post-first-action pendency – The time between first action and final disposal. This may be different for abandonments and patent grants.

Private Applicant (UK) – Applicants who file directly without representation by a patent attorney or IP professional. Similar to a pro se in the US.

Queuing Time – Portion of pendency time spent waiting for some action to take place; i.e. time when neither the examiner nor the applicant is actively working on the application.

RCE (US) – Request for Continued Examination. The American Inventors Protection Act was enacted November 29, 1999 and established this form of non-serialized continuation. It is by far the most common form of non-serialized continuation. An applicant may file for an RCE multiple times during prosecution. For the purposes of this paper, we include all forms of non-serialized continuations in the RCE category. The USPTO considers RCEs as new applications for some reporting purposes; however, for our purposes we treat them as S3 inventory because they have been in active examination and remain with the same examiner.

Stock 1 (S1) – The stock of applications received by an office but not yet available to any examiner, because it is not yet ready for examiner action. In many offices applications remain in S1 due to the processing time necessary to complete formalities and to allocate the patent according to the office's technological classification system. Applications exit S1 either by becoming ready for examiner action (and moving into Stock 2), or by exiting the office altogether, e.g., through express abandonment (in the US), or by withdrawal and re-filing at the EPO (in the UK).

Stock 2 (S2) – The stock of applications that are ready for examination, but not yet having received a first decision. In other words, these applications are currently undergoing initial examination, or they are queued for examination on the examiners' desks or dockets. In the UK these are applications for which a search or combined search and examination has been requested.

Stock 3 (S3) – The stock of applications having received a first decision, but not having been terminally disposed of. These applications may be bouncing between the patent office and the applicant, or being reviewed by the office.

Stock 3a (US) – Pending applications which have received a first action, but have not filed an RCE.

Stock 3b (US) – Pending applications which have filed at least one RCE. An application which files an RCE does not leave Stock 3b until it is patented or terminally abandoned.

Survival Curve/Function – A function expressing the probability that a subject will not have experienced the event of interest from an origin event up to a certain time. These curves may also be used to summarize the distribution of pendency for a cohort of applications. See also Section 3.5.

Survival Time Regression – Method of statistical analysis for time from an origin event to an event of interest. Chapter 4 uses survival time regressions to analyse total pendency and Chapter 5 uses survival time regressions to analyse first-action pendency and post-first-action pendency. See also Section 3.5.

TAF - Technology Adjustment Factor (UK). TAF represents the expected time investment required for examination of the application. An application will be assigned a TAF rating on the basis of its International Patent Classification (IPC), with a rating of 1 being an application expected to take the standard amount of time to examine. If TAF is less than 1, the application should take less time, and if it is more than 1, it should take longer.

Terminal Disposal (or Terminally Disposed) – The point at which an application leaves the Office permanently. In the some US statistics, an application which files a request for continued examination (RCE) is considered abandoned, but we are not counting it as a terminal disposal.

WIPO - The World Intellectual Property Organization. The United Nations agency dedicated to the use of intellectual property (patents, copyright, trademarks, designs, etc.) as a means of stimulating innovation and creativity. For more information, please see: http://www.wipo.int/about-wipo/en/

Appendix B

Primer on UK Patent Examination

1. The UK Intellectual Property Office

The Intellectual Property Office (IPO) is the official government body responsible for Intellectual Property (IP) rights in the United Kingdom and is an Executive Agency of the Department for Business, Innovation, and Skills. Not only is the IPO responsible for granting UK patents, it is also responsible for the registration of trademarks and designs in the UK, as well as having a copyright function. The IPO employs 1009 staff of which 241 are Patent Examiners[64]. Although the IPO has over recent years continued to action around 90% of searches within 4 months of request and also to reduce the pendency for first examination actions[65], an increase in input over the last two years[66] has meant that the amount of unprocessed work or "backlog" is increasingly of interest. The key aspects of the UK Patent system and existing means for applicants to accelerate their applications are discussed below and shown diagrammatically in Figure B1.

2. Filing an application and preliminary examination

The IPO requires applicants to pay an application fee within 12 months of the priority date if there is one or, if there is no priority declaration, within 12 months of the filing date[67]. If the application fee is not filed in time, the application is treated as withdrawn. By contrast, in the US system, the act of filing includes payment of fees. At the USPTO all three basic prosecution fees (application, search and substantive examination) are paid at the time of filing, whereas applicants in the UK pay the fees sequentially having made the decision at each stage to continue with the application.

Once an application fee is paid a preliminary examination is performed. This is an administrative action, performed by a Formalities Examiner, which involves checking for missing pages, the clarity of drawings etc. The Formalities Examiner reports the results of their examination to the applicant who is then given an opportunity to file new or replacement documentation if required.

---

64 Figures for March 2013

65 2011-2012 Annual report, www.ipo.gov.uk/about-anrep1112.pdf

66 2012-2013 has seen a 7% increase in search requests, and a 14% increase in exam requests.

67 This is the filing date of the earliest application from which the application claims priority under section 5(2) of the Act. This date is usually 12 months earlier than the filing date of the application being considered, though in some circumstances this may be up to 14 months earlier.

3. **Search**

In order for a search to be performed the applicant must file a “Request for a Search” (including fee). This request should usually be filed within 12 months of the priority date or, if there is no priority declaration, within 12 months of the filing date. If the request is not filed in time, the application is treated as withdrawn.

In most cases the Patent Examiner will complete a prior art search within 4 months of the request for search being received at the IPO. The results of the search are then reported to the applicant in a search report which indicates to the applicant any relevant prior art. There is no requirement for the applicant to respond to the search report. The applicant may use the results of the search to decide whether or not to continue with the application. The results of the search may also inform the applicant’s decision as to whether to apply internationally.

4. **Publication (A-pub)**

Provided a search report has been issued and all formalities requirements have been met, an application will be published. Publication takes place around 18 months from the priority date of the application or, if there is no priority declaration, around 18 months from the filing date[68]. Once a search report has been issued the only way to avoid publication is to withdraw the application. Publication can be accelerated in order to reduce this 18 month period (see **Acceleration** below).

5. **Substantive Examination**

If not already received, applicants wishing to continue with their application must usually file a ‘Request for Substantive Examination’ (including fee) within 6 months of the application being published. If the request is not filed in time, the application will be treated as withdrawn. Examination actions are then carried out in date order according to the earliest date associated with the application, e.g. the filing date or the priority date. At present examination of an application will usually take place around 42 months after the priority date or, if there is no priority declaration, around 42 months after the filing date.

The Patent Examiner carries out an examination of the application to determine whether it satisfies various criteria, such as novelty, inventive step, excluded matter, clarity of disclosure. In carrying out this assessment, the Examiner will use the results of the earlier search, the results of any updating of the original search, and any other relevant reports published by other Patent Offices, e.g. file wrappers for equivalent applications. In most instances the examiner will issue a first examination report raising objections to the application and setting a period for response.

68 This 18 month period is Prescribed by rule 26(1) of the Patents Rules 2007

6. **Combined Search and Examination**

By simply requesting combined search and examination (CSE), which is available as-of-right at no extra cost, the period to grant can be significantly reduced – since the examination process begins much earlier. To request combined search and examination the applicant simply needs to file requests for Search and Examination at the same time. At present Patent Examiners aim to complete CSEs within 4 months of request.

CSE essentially allows applications to "jump the queue" for examination. While this can have obvious advantages for the individual applicant it does of course take away Patent Examiner resources from elsewhere, notably away from performing older non-CSE examinations. Interestingly over the last year the IPO has seen an increase in the proportion of applications where CSE has been requested. For 2012-2013 the number of CSE actions performed by the IPO increased by 18% relative to the previous year[69].

7. **Amendment**

As discussed above a response from the applicant to an examination report is due by the date set[70]. This date for response is extendable by 2 months-as-of right. The applicant may respond with arguments and/or amendments which will be considered by the examiner. The IPO aims to re-examine applications where possible within 2 months of the applicant's response being received.

There is no limit on the number of rounds of amendment which can take place at the IPO. Nevertheless prosecution of an application at the IPO cannot continue indefinitely since the Patents Act 1977 ("the Act") and the Patents Rules 2007 ("the Rules") set out a date[71] by which an application must comply with the requirements of the Act and Rules (see **Compliance date below**). Additionally if the Patent Examiner is of the opinion that no saving amendments are possible, or if in spite of multiple amendment rounds the applicant has not succeeded in progressing the application towards grant, he may offer the applicant the opportunity to be heard by a Hearing Officer on the outstanding issues. The Hearing Officer may refuse the application if he finds it does not comply with the requirements of the Act and Rules.

If the applicant does not respond to an examination report the application is treated as refused at the end of the compliance period.

8. **Compliance date**

The compliance date (i.e. the date at which the application must comply with the requirements of the act and rules) is generally four and half years from the priority or filing date (whichever is earliest), however this is subject to there being a minimum of 12 months after the first exam report is issued. If by the compliance date the application is not in order for grant it is treated as having been refused.

A two month as-of-right extension to the compliance date is available upon request and payment of a fee. Further discretionary extensions are also available.

69 Internal IPO Statistic, February 2013.
70 At the beginning of the examination process this is usually 2 months after the date of the examination report.
71 Sometimes known as the compliance date, section 20 date or rule 30 period.

9. **Grant**

Once the examiner's objections have been overcome and the examiner is satisfied that the application complies with the requirements of the Act and Rules, the application will be marked in order and sent for grant. The applicant is notified by letter of the grant of their application[72]. The grant of the patent is then announced in the Patents Journal[73] and the granted patent is published (B-publication). There is no grant fee or further action required from the applicant. The applicant receives the Certificate of Grant shortly after B-publication.

The minimum period from filing to grant is constrained by the publication date of the application. The IPO waits until at least three months after the publication date before granting a patent, to allow third parties to make observations on the application under section 21 of the Act. While patents legislation does not set out the length of this period, three months has, over a number of years, become established by IPO practice.

As publication does not usually take place until 18 months from the filing or priority date the minimum period to grant is generally 21 months from the "earliest date" (the filing date or declared priority date).

Waiting until 21 months from the earliest date also allows the IPO to complete the search for "section 2(3) prior art" – that is, prior art in the form of published UK patent applications[74] with earlier priority dates or filing dates than the application being considered, but which were not published at the time of the initial search (also sometimes called "secret prior art"). Once published, these earlier applications can be used to show that the invention contained in the application being considered is not in fact new, as set out in section 2(3) of the Act.[75] The earlier applications should have been published by 18 months from their own earliest date (priority or filing), and waiting a further three months allows the Patent Examiner to be reasonably confident that these applications are present in the patent databases used for searching prior art.

If publication is accelerated (alongside CSE or the acceleration of other aspects of the application process) the minimum period to grant can be reduced, although the IPO will still wait until at three months after the publication date, to allow third party observations. If an application is sent for grant earlier than 21 months from the "earliest date" an additional search for "section 2(3) prior art" will be performed after grant. Revocation action under section 73(1) may be initiated if section 2(3) prior art is found.

72 For the purposes of sections 1-23 of the Act the patent is granted as of the date of this letter.
73 For the purposes of sections 24 onwards, the patent is granted as of the date of this announcement.
74 Or applications for European patents which designate UK.
75 The basis for this provision is that if more than one application is filed for the same invention, the earliest of them should take precedence.

10. Acceleration

Currently, both search and examination can be accelerated on request if the applicant provides an adequate reason. What constitutes an adequate reason will depend on the specific circumstances of the case in question, but typical reasons include the need for search results or a granted patent in order to secure an investor, or the need to expedite grant due to a potential infringer.

Search and examination can also be accelerated where the application qualifies for one of our special acceleration schemes:

- The **Green Channel** allows search and/or examination to be accelerated where the invention relates to environmentally-friendly technology.[76]

- The **Patent Prosecution Highway** (PPH) allows examination to be accelerated where the claims of an equivalent application have been found to be acceptable by another intellectual property office with which the IPO has a PPH agreement.[77]

- **PCT(UK) Fast Track** allows examination in the UK national phase to be accelerated if an international application has received a positive International Preliminary Report on Patentability (IPRP) in the international phase[78].

There is no charge for any of the above acceleration services.

As mentioned previously, combined search and examination (CSE) is available as-of-right on request at no extra cost, resulting in examination being brought forward to the time of the search (usually within four months of the date it is requested).

Early publication is also available as-of-right on request. If CSE and early publication are requested and search and examination are accelerated using the services currently available, it is possible for a patent to be granted in less than a year.

11. Examiner performance

Examiner performance is evaluated on a points system[79] which takes into account the amount of work the examiner produces (measured by a W/A or "output" value), the quality of that work and other considerations[80].

Each technology-related heading was associated with a time adjustment factor (TAF) to moderate the different demands on examiner time spent on cases in different subject areas. TAFs are used as a relative measure of the time-taken to complete substantive actions on patent applications[81]. The calculation of each examiner's W/A takes account of the TAF of each

76 www.ipo.gov.uk/p-green
77 www.ipo.gov.uk/p-pph
78 www.ipo.gov.uk/p-pn-fasttrack1
79 Patents Directorate Notice 03/2012
80 For example timeliness, attitude, achievement of personal objectives.
81 Higher TAFs are applied to more "difficult" subject matter.

case the examiner completes an action on. This means that TAFs directly affect the performance assessment of examiners.

An examiner's W/A is calculated on a monthly basis and is then averaged over the year. An examiner nominally receives 2 points for each substantive action (i.e. search or examination) which are corrected to take into account the TAF value for each action. At the end of the month the total number of points is divided by the number of days spent patent processing to give a W/A or "output" value. Senior examiners are expected to have a higher W/A than more junior examiners.

Examiners can claim time as "non-patent processing" to reflect annual leave and other duties (such as training others or project work), however actions such as re-examinations of amendments must be incorporated into the time allocated for patent processing.

12. **International filing options**

As with applications at the USPTO, UK applications can claim priority from applications filed at other Offices[82], or may be used to claim priority elsewhere. Where an application claims priority from an earlier application, the applicant may continue with both applications or may at any point decide not to proceed with one or other application.

An applicant may choose to apply for patent protection in the UK via the IPO directly, via an international application under the PCT, or via a European application. These three routes to a UK patent are summarised in Figure B2.

During the international phase of PCT processing the application is searched by an International Searching Authority[83] and published by WIPO. The applicant then nominates in which states the application will enter the national/regional phase[84]. When an application enters the UK national phase, the IPO republishes the application with a UK publication number and the application is then examined in the same way as any other UK patent application. Once the application complies with the Act and Rules it is granted as a UK patent.

An applicant may have two UK applications (e.g. where one application was filed as a national UK patent application and the other originated as a PCT application) with the same priority date[85] proceeding simultaneously before the IPO. In this circumstance, section 18(5) of the Act does not allow grant of both applications if the claims are for the same invention. In other words, if the applicant wants both applications to be granted one or other will need to be amended so that the two applications claim different inventions.

Applicants may decide to gain patent protection in the UK by applying for a European patent via the EPO (claiming priority if appropriate). With such an application the entire application process

82 In line with the Paris Convention.

83 In contrast to the USPTO, the IPO is not an International Searching Authority.

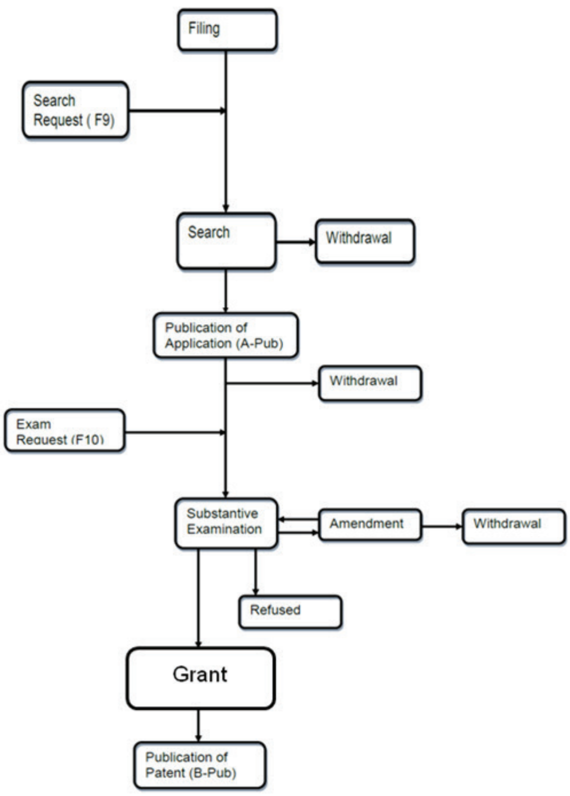
84 It is possible for a PCT application to enter the regional EP phase, thus becoming an EP application which designates UK.

85 It may be that one application claims priority from the other or that both applications claim priority from the same earlier application.

right through to grant is administered by the EPO. Grant by the EPO results in a bundle of patents valid in the various countries designated by the applicant. As soon as an EP(UK) patent (i.e. an European patent where the applicant has designated the UK) is granted by the EPO it takes effect in the UK[86].

It is possible for an applicant to have an application proceeding at the IPO and an application with the same priority date proceeding simultaneously before the EPO[22]. However if both patents are granted and the claims relate to the same invention, the IPO may commence revocation of the UK patent under section 73(2) of the Act. Revocation of the UK patent can be avoided by amending one or other of the patents so that they no longer claim the same invention, or by removing the UK designation from the European application prior to grant.

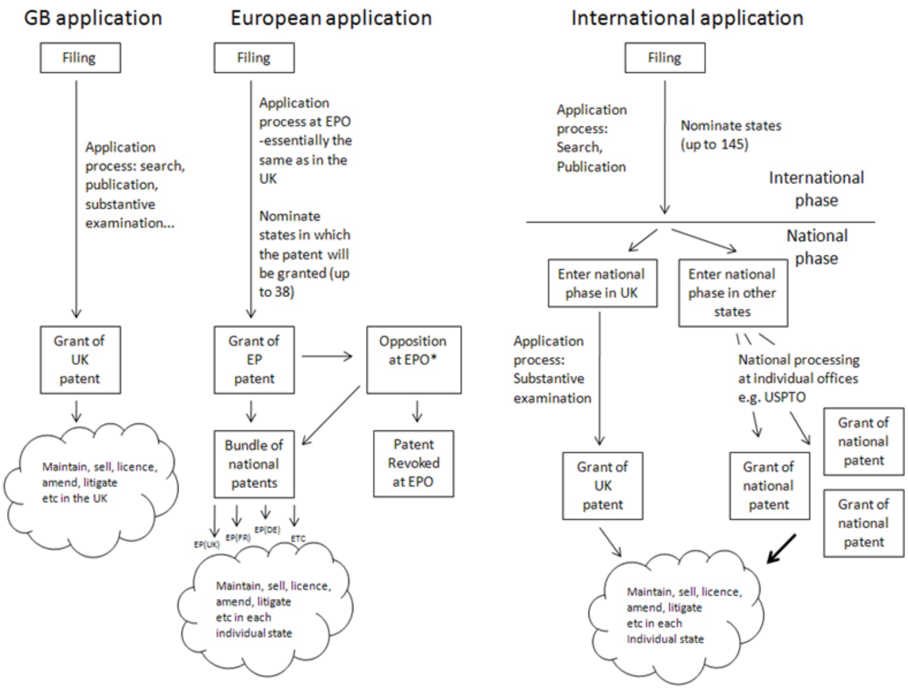
Figure B1: Overview of the standard application process at the IPO



Source: UK IPO 2013

86 Following the coming into force of the London agreement on 1 May 2008 there is no longer any requirement for the applicant to provide an English translation if the granted patent is in French or German. A translation of the claims into the other two languages must have been provided to the EPO prior to grant.

Figure B2: Overview of international application processes which can lead to grant of a Patent valid in the UK



Source: UK IPO 2013

Appendix C.

UK regression results

1. UK linear regression results

The linear regressions test the impact of a series of explanatory variables on the length of time (number of days) an application will spend in each stock of the application inventory. What we are trying to understand is how long applications spend in each stock, and what is associated with longer pendency time. We break the application inventory into five stocks, representing the days the application spends inside the office in Stocks 1, 2 and 3, as well as the time spent outside the office in Stocks 2 and 3.

The linear regressions are performed on applications filed in 2001 (2001 entry cohort). In other words, our dataset tracks all applications in the 2001 entry cohort through to their final outcomes (grant, abandonment or rejection). The 2001 data relates to all 30,820 patent applications filed in 2001 for which we were able to extract the underlying administrative data.

We test the pendency of the 2001 entry cohort against a series of explanatory variables. These include the number of searches requested by applicants, the number of amendments requested before a grant, the Time Adjustment Factor (or TAF) of a patent application, the number of examiners on staff in the month of filing, a dummy variable for whether the application successfully passed from one stock to the next, with a marker for grants and for CSE applications.

TAF is only assigned to a patent once it enters the process of search and examination as it is an indicator that allows examiners to add more time to the search and exam. We report the results below with the TAF, but as only around half of applications go to search in the UK, the population is 16,876, not the 30,820 of the whole 2001 population. When running the regressions for the full population the results are comparable, but we only present one set of regression results in Table C1 below for ease of reading:

Table C1: Linear Regression to estimate factors affecting time spent in each stock

	Stock 1		Stock 2 inside		Stock 2 outside		Stock 3 inside		Stock 3 outside	
Searches	-69	***	79	***	99	***	0		28	***
Amendments	2		-8	***	-1		18	***	120	***
TAF	-29	*	207	***	-20	*	0		43	***
Examiners	0		0	***	0	*	0	*	0	*
S1 pass	116	***	13		-49	***	0		-7	
S2 Pass	5		430	***	287	***	-2		-190	***
S3 Pass	72		-30		-64		1		100	*
Grant Dummy	59	***	2		-19	***	-2	*	183	***
CSE Dummy	-72	***	-343	***	-113	***	3		355	***
Constant	25		-38		9		-8		-106	***
N	16876		16876		16876		16876		16876	

Stars indicate statistical significance at the 0.1% (***), 1% (**) and 5% (*) levels

Impact on Stock 1:

- The CSE dummy exhibits a strong negative impact on the time the patent spends in Stock 1, reducing the pendency by an expected 72 days.
- Patents granted, however, spend more time (59 days) in Stock 1, as suggested by the grant dummy.
- Patents passing Stock 1 were on average processed for 116 days longer (until being abandoned or granted than other applications), which is to be expected as a large proportion of applications do not have a search request and so do not proceed through to Stock 2.

Impact on Stock 2

- TAF has a large impact on the pendency in Stock 2: Inside the office, a unit increase in TAF would be expected to add 207 days to pendency, while it reduces pendency by 20 days outside the office. However, a one unit increase in TAF is rare, so the findings need to be interpreted with some degree of caution.[87]
- Notably, CSE applications spent 343 days less inside the office and an average 113 days less outside the office. CSE applications should not be able to leave the office during examination, except if the applicant requests it, so it is probable that the time saved outside the office is a simple function of the application never being outside the office.

87 As shown in Chapter 4, TAF varies from 0.8 to 1.4 over time, so a one unit increase and the implications thereof are hypothetical at best. The survival regressions provide a better impression of the impact TAF has. They show that a more realistic 10% increase in TAF is associated with a change of 6.4 days (or 1.5%) in processing time.

- Patents passing Stock 2 spent 430 days longer inside the office and 287 days longer outside the office, and given the big negative impact of CSE filings, this appears to be driven by traditional patent applications.

- Number of searches has a strong impact on the time an application spends inside the office (on average 79 days per search). The equivalent impact is 99 days outside the office.

Impact on Stock 3

- As expected, amendments drive the time the patent spends in Stock 3. A single amendment is associated with 120 days longer outside the office and 18 days inside the office.

- Granted patents spend about 2 days less inside the office than non-granted patents, but 176 more outside the office.

- Perhaps most striking is that CSE patents or applications spend an additional 3 days in the office during Stock 3, but almost a whole year, or 355 days, outside the office. The reason might be that having gotten through Stock 2 quickly – faster on average by 322 plus 126 days – the applicant now has a long time before the statutory 4.5 year deadline for completing the patent prosecution kicks in.

- Each application search is associated with 28 more days spent outside the office in Stock 3. They have no impact on the time the patent will spend inside the office in Stock 3. This could imply that patent searches are associated with delays by applicants or simply that applicants go back and look at the search results at this stage.

When we run the same set of regressions on applications filed in each year from 2000 to 2005, we get similar results. In particular the pattern on the TAF coefficients repeat consistently across the set. The pattern is that a higher TAF increases pendency inside Stock 2 and outside Stock 3 (so it takes longer to get through) but it speeds it up – marginally – outside Stock 2, with about zero effect inside Stock 3.

The coefficient on the number of examiners is effectively zero across this dataset, meaning that a one unit increase in examiners has little effect on the time that an application is pending. But this finding should be treated with caution, as it is probably a function of the data: We are only able to track number of examiners at entry and numbers of examiners do not vary much in the one year under investigation. The survival regressions below allow us to investigate in a more meaningful way the impact that examiners have on pendency time.

Correlation analysis reveals some further relationships not picked up by the regressions. Table C2 illustrates correlations between variables listed in rows 1-12 and variables listed in columns 1-7. Correlations are represented using correlation coefficients. These indicate the strength of the relationship between two variables. A correlation coefficient of 1 indicates a perfectly positive linear relationship and -1 indicates a perfectly negative linear relationship. A coefficient of 0 suggests that there is no linear relationship between the two variables.

Table C2 below suggests that there is a positive relationship between number of searches and subsequent amendments. This relationship can also be found between amendments and time spent in Stock 2. This could suggest delaying tactics where firms request several searches and several amendments but still drop the patent relatively early in Stock 3.

Table C2: Correlation between each variable

	Stock 1	Stock 2 inside	Stock 2 outside	Stock 3 inside	Stock 3 outside	Search requests	Amendment requests
S1time	1						
S2timeIN	0.133	1					
S2timeOUT	-0.0288	0.5879	1				
S3timeIN	0.135	0.2214	0.1423	1			
S3timeOUT	0.158	0.2287	0.1109	0.4085	1		
searches	0.2014	0.7289	0.4916	0.1271	0.0404	1	
amendments	0.2954	0.4567	0.314	0.5173	0.7266	0.2782	1
S1pass	0.3314	0.6761	0.3684	0.2416	0.4068	0.7616	0.4697
S2pass	0.3852	0.7319	0.5162	0.3426	0.5768	0.4163	0.6658
S3pass	0.3855	0.7315	0.5156	0.3428	0.5772	0.4163	0.6663
granted	0.3572	0.5737	0.3821	0.3966	0.6926	0.3365	0.7835
CSEdummy	0.1775	-0.0724	-0.1493	0.2031	0.6197	-0.2867	0.3448

A further relationship to bear in mind is the probability that an examined application will be granted. Once the exam is completed, the application can either be granted, or if it is not granted in the first instance, the applicant can either abandon the application or file an amendment. In 2001, of all the patent applicants that requested a first amendment, only 5% were eventually abandoned. This suggests that once applicants have started filing amendments they will eventually get something in a granted patent, and would go part of the way to explain why the number of search requests is associated with grants, as search requests are correlated with amendment requests which are a strong predictor of grants.

2. UK survival regression

The survival time regressions test the impact of select explanatory variables on the number of days that an application will be pending beyond the average time to grant of 616 days. The question we are trying to answer is, if we double one of the explanatory factors how much longer will an application stay in the system awaiting grant. Expected pendency is measured at the point of filing.

The data used for analysing UK pendency is the complete register for patent applications to the UK IPO between 2000 and 2011 where we could extract the relevant data. This is not a sample of the population, but the whole population of applications to the IPO through both direct and PCT routes from all origins. The data was extracted with a cut-off date for observations on the 1 January 2012, so all information after that point is not included in the dataset. The survival regressions considers applications that make it through the process to grant, while sorting pending applications and applications not granted separately, so the dataset under investigation includes 177,887 observed patents that had been granted by the 31 December 2011.

The explanatory variables for the regressions fall into two categories. The first category is stock variables, which count the number of applications arriving in a given month, or the stock of applications pending in either Stock 1 inside the office and Stock 2 or 3 inside or outside the office.

The second set of variables relate to the application itself, by considering the complexity of an application represented by the technology adjustment factor (TAF), and a dummy for applications that are Combined Search & Exam (CSE) filings. We also consider office capacity to handle the applications by counting the number of examiners on staff in a given month.

We would expect that the stock variables and complexity have a positive impact on pendency, meaning larger stocks or more complex applications lead to longer pendency. On the other hand we expect that the number of examiners and filing as a CSE would reduce pendency.

We run a pair of survival time regressions using a maximum likelihood estimation for a parametric regression model, assuming a log normal distribution (results 1 and 2 below). We run the regressions on pendency from filing to grant, so we estimate the impact of each factor on total pendency. The results are presented in Table C3 below. It shows the time-ratio results from the regression and the predicted percentage change in expected pendency from the baseline if there was a doubling in the particular explanatory variable. So, for example, if there were double the number of applications arriving at the office, specification 1 predicts that pendency from application to grant would increase by 8% on the average 616 pendency days.

Table C3: Survival Regression Results

	Survival Regression A (excl. CSE dummy)			Survival Regression B (incl. CSE dummy)		
	Time Ratio		**100 % increase from baseline**	**Time Ratio**		**100 % increase from baseline**
Arrivals (log)	1.12	***	8%	1.10	***	7%
S1 inside (log)	1.66	***	42%	1.53	***	34%
S2 inside (log)	1.13	***	9%	1.12	***	8%
S2 outside (log)	2.88	***	108%	2.74	***	100%
S3 inside (log)	1.04	***	3%	1.04	***	3%
S3 outside (log)	0.85	***	-10%	0.92	***	-6%
Examiners (log)	0.90	***	-7%	0.87	***	-9%
TAF (complexity)	0.75	***	-25%	0.90	***	-10%
CSE dummy	-		-	0.48	***	-12%[88]
Constant	0.00	***	-	0.00	***	-
N	177,887			177,887		

Stars indicate statistical significance at the 0.1% (***), 1% (**) and 5% (*) levels

[88] We do not increase CSEs by 100% as it is difficult to conceptualize a 100% increase in a fixed effect. Rather the 12% represents the increase in average pendency if the proportion of CSEs were to double from the mean of 20% to 40%. The corresponding test was performed on a limited sample size and so needs to be interpreted with some caution.

The second specification is more reliable as it includes the CSE dummy, which is a significant factor if applicants would choose to file more CSEs. CSEs pass through the system twice as fast as regular patent applications. Furthermore a doubling in the proportion of CSE applications could reduce expected pendency by 12% below the mean pendency of 616 days.

From the office side, a doubling of the number of examiners would reduce the expected pendency of all applications at the point of filing by 9% in specification 2, while a doubling of applications in Stock 3 outside the office would – somewhat surprisingly – also reduce the backlog considerably. The reason for the effect of Stock 3 outside on expected pendency could be twofold: Examiners have very short deadlines within which to turn around amendment requests (Stock 3 inside). When they complete an amendment and push something back into Stock 3 outside (or grant it) their time is freed up to undertake business as usual. This should allow for a faster overall pendency of other applications in the queue. Furthermore, because Stock 2 has such a large effect on pendency time, shifting applications from Stock 2 into Stock 3 (where they would arrive outside the office awaiting an amendment) should work to reduce expected time to grant of other pending applications. The precise size of the effect will depend on how much Stock 2 could be reduced by doubling numbers of examiners.

Appendix D.

Primer on US Patent Examination

The USPTO, headquartered in Alexandria, Virginia, is the Federal agency responsible for granting U.S. patents as mandated by Article I, Section 8 of the U.S. Constitution.[89] The USPTO has over 10,000 employees, including attorneys, analysts, computer specialists, and over 7,600 patent examiners.[90] Mainly to address pendency, staffing continues to be a key priority.[91] A persistent application inventory has resulted in high pendency, both for first actions and for final disposals.[92] The USPTO has also seen a dramatic increase in the use of continuations (largely in the form of RCEs) which provides an applicant greater latitude in further prosecuting a particular application. These and other pertinent trends are addressed as part of this report, primarily as they relate to pendency. However, before delving into that analysis, it is important to understand some details of the USPTO examination system, especially as they relate to pendency and the ability to make international comparisons in stocks of patent applications.

1. The opt-out system

As discussed above, examination in the US is an "opt-out" system[93] in that an application automatically receives a substantive examination unless the applicant expressly abandons prior to that point. The fee payment timeline reinforces this policy. The three basic prosecution fees—filing, search, and examination—are due at the outset of the patent application process. For all non-provisional utility applications for patents filed on or after December 8, 2004, an applicant must pay all three fees at the time of filing, or pay within a grace period with additional surcharges.[94] The applicant has two months from the filing date in which to pay the filing, search, and examination fees and any applicable surcharges. Thus, all three fees are collected before a filing is considered complete.

Given that US examination is automatic, some examiner time may be misallocated to examination of applications that are informally or implicitly abandoned by the applicant. Accordingly, time spent examining such applications may, in retrospect, have been better applied to other pending applications.

89 U.S. Constitution Article I, Section 8 Executive Branch *"to promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries."*

90 USPTO 2011-2015 Strategic Human Capital Plan page 24. See also, USPTO Data Visualization Center is located at http://www.uspto.gov/dashboards/patents/main.dashxml.

91 USPTO Strategic Plan 2010-2015 p. 4.

92 GAO Report 2005, 2007.

93 USPTO Utility Patent Application Process located at http://www.uspto.gov/patents/process/index.jsp.

94 *See generally* 37 C.F.R. § 1.16 (setting forth national application filing, search, and examination fees).

Following the initial examination and an office action by the examiner, applicants must affirmatively "opt-in" to continue prosecution or to receive a patent grant. Failure to do so results in abandonment. In the US, prosecution ends either with allowance and grant or with abandonment. That is, there no formally decisive rejection—only a failure to respond to a rejection. Applicant responses to office correspondence are due by statutory deadlines indicated in a form cover letter appended to the examiner's report. Applicants may petition for limited extensions of time (for a fee), which increase pendency.

In both the US and UK, applications are published 18 months after filing, with limited provisions to avoid publication.

2. Examination

U.S. examiners complete a prior art search and substantive examination in the same instance. In general, the first substantive correspondence that an applicant receives from the USPTO is a first action on the merits (FAOM, or simply "first action"). The first action includes a search report that supports a non-final rejection or an allowance, and a listing of relevant prior art identified by the examiner, if any. Another type of first action includes notification of missing parts (e.g., forms, fees, documents) typically issued by Legal Instrument Evaluators and restrictions issued by examiners.

USPTO restriction practice directly and indirectly affects pendency and application inventories. A restriction requirement applies to an application that is deemed to contain more than one invention, thereby imposing a search burden on the examiner. Correspondence is sent to the applicant requiring an election of a singular invention. The claims directed to the non-elected invention are thereby withdrawn and the examination of the elected invention continues. An applicant may then file a divisional application (DIV) for the withdrawn claims directed to the non-elected invention. Thus, restriction practice serves to decrease pendency while at the same time add to the stock of new DIV applications.

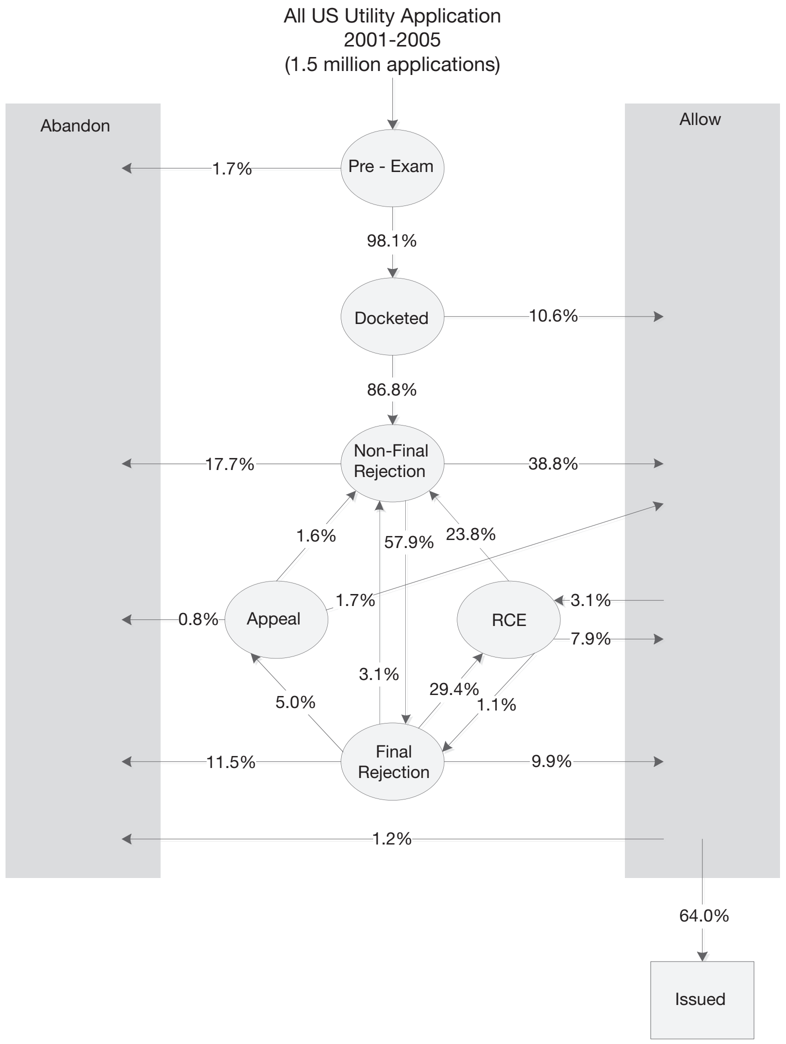
3. Prosecution, Amendments, and RCEs

Patent prosecution in the US provides applicants with many opportunities to address concerns raised by examiners. For the most part, the avenues available to applicants consist of amendments, requests for continuing examinations (RCEs), and appeals.

Figure D1 describes the basic patent prosecution system in the US, using data from the 2001-2005 cohorts of patent applications.[95] The data come from internal USPTO systems, and reflect the history for more than 1.5 million regular utility applications[96] filed between 2001 and 2005. The internal USPTO system—Patent Application and Location Monitoring (PALM)—contains no fewer than 2000 event codes corresponding to more than 300 status codes for ongoing applications. Thus, Figure D1 is greatly simplified, but it represents the most common events.

95 2005 was used as a cutoff in order to assure that most of the applications in the filing cohort had been disposed.
96 Regular utility applications exclude re-issues, re-examinations, plant, and design patents.

Figure D1: Simplified US patent prosecution process for applications filed between 2001 and 2005



Source: USPTO 2013

The percentage indicated for each transition reflects the percentage of the total filings that reach that stage.[97] For instance, approximately 1.7% of applications are abandoned (explicitly or implicitly) between filing and docketing. A cumulative total of 32.9% of applications were abandoned and 64.0% were issued, meaning that approximately 3.1% of those applications are still pending. The most common path for incoming applications is to be docketed and to receive a non-final rejection, followed by a final rejection or allowance.[98] The first action allowance rate was only 10.9% (10.6%/97.4%).

97 Percentages less than or equal to 0.5% were omitted.

98 Because applications can end up in the non-final rejection state multiple times, the transitions coming out of non-final rejection total to more than 100%.

From allowance, most applications issue, but in a small minority of those cases, applicants abandon or file an RCE.[99] In contrast, so-called “final rejection” still leaves many avenues open to the applicant, who may abandon, amend, file an RCE, or appeal. Amendments (unless accompanied by an RCE) may ultimately lead to an allowance or another non-final rejection. Appeals are expensive to applicants (in terms of USPTO fees, but also attorney fees, time, and effort), and are infrequently used relative to RCEs. Thus, the cycle of non-final rejection → final rejection → RCE → non-final rejection is common. From final rejection, abandonments and allowances are equally common, but both are less frequent than RCEs.[100] Of those applications that move from non-final to final (57.9%), fully half go to RCE after final (29.4%). This trend has increased since 2005.

Communication during the patent application process is predominantly in writing. However, interviews are also a part of prosecution: prior to or immediately after a first office action but before a final rejection, interviews are a matter of right for the applicant. After the final rejection, interviews may be granted at the examiner’s discretion if the examiner is convinced that the interview will dispose of the case with only minor consideration. If the application has been sent to issue or if applicant has filed an appeal brief, the application has left the jurisdiction of the examiner, and so interviews are prohibited.[101]

4. Continuations

In the US, continuations are one avenue open to inventors and their agents in prosecuting patent applications. There are two broad types of continuations: non-serialized and serialized continuations.[102] Non-serialized applications can be used to extend prosecution of the current application. They do not receive a new serial number and are not substantially different from the original applications. They do not go through classification and docketing as new applications do. Rather, they are immediately docketed to the same examiner, and involve different fees than do serialized continuations. While there have been other forms of non-serialized continuations, RCEs (instituted in 1999[103]) are by far the most common. Thus, for the purposes of this paper, we include all forms of non-serialized continuations in the RCE category. The USPTO considers RCEs as new applications for some reporting purposes; however, for our purposes we treat them as S3 inventory because they have been in active examination and remain with the same examiner.[104] An applicant may file for an RCE multiple times during prosecution.

---

99 RCE filings following allowance typically involve newly discovered prior art that is disclosed by applicants. The examiner must examine the prior art and “re-allow” the application, if appropriate.

100 The figure is intended to illustrate the main paths through prosecution. Obviously a full analysis of application prosecution would involve distinguishing between first, second, and third non-final rejections.

101 See Manual of Patent Examining Procedure (MPEP) sections 713.02. 713.03, 713.05, 713.09, 713.10

102 There have been several different incarnations of non-serialized continuations, including Continued Prosecution Applications (CPAs), Rule 129 continuations (R129s), and File Wrapper Continuations (FWCs). The most recent incarnation (and by far the most prevalent) is the Request for Continued Examination (RCEs). Throughout this section, we refer collectively to all these continuations as RCEs. Until November 2009, RCEs were put on the “amended docket,” which meant that the examiner had to respond within two months. Since that time, RCEs have gone on the “special new docket,” meaning that the examiner has more discretion as to when to respond (similar to newly docketed applications).

103 American Inventors Protection Act (AIPA), was enacted November 29, 1999, Public Law 106-113. See also MPEP §706.07(h).

104 In the statistical analysis we do not track RCE abandonments as true abandonments. However, it is worth considering how they influence workload.

In contrast, serialized continuations (or "traditional" continuations) are treated as new applications, in that they receive a new serial number, go through the classification process, and are usually docketed to an examiner without regard to the identity of the examiner on the parent application. The parent application is often abandoned after a continuation, but this is not required. Thus, the parent and child applications may proceed in parallel.[105] There are three types of traditional continuations: CONs, CIPs, and DIVs.

Applicants may file for a continuation[106] (CON) of a parent application to receive the benefit of the priority date of the parent. To claim priority, however, the CON must limit itself to the subject matter of the parent application. When new subject matter is introduced in a continuing application, the application is a "Continuation-In-Part" (or CIP). Claims relying on existing subject in the parent application receive priority benefit, but the new matter does not. Alternatively, following a restriction by the examiner, an applicant may file a divisional (DIV) application. A DIV is a child case that receives the priority date of the parent so long as the parent is still a pending application.[107]

5. Production System

Patent examiners are evaluated based on a process known as the "count" system. Each patent examiner must meet or exceed an annual production quota, which is based on several factors: (1) the "count" value of certain examiner actions, (2) the "complexity factor" of his art unit, (3) rank (based on the US General Scale for federal employees[108]), and (4) approved non-production time. Historically, the actions that generated counts were first actions on the merits,[109] and disposals (whether through abandonment or issuance). Each action contributed one count, so that a completed examination consisted of two counts (one for the first action and one for the final disposal). Recent changes to the count system have reallocated the counts, but the general rule of two total counts per disposal holds in most cases.[110]

To measure performance, the examiner's accumulated counts are compared to his annual production quota. Those examiners in more complex technologies have a lower quota based on the complexity factor of their art unit. Further, senior examiners have a higher quota than junior examiners. Examiners also can claim non-production time based on annual leave, administrative activities, or certain examination functions such as writing responses to appeal briefs.

105 For more detail on continuations see Graham, Stuart J. H. and David C. Mowery (2005), "The Use of USPTO 'Continuation' Applications in the Patenting of Software: Implications for Free and Open Source". Law & Policy, 27(1), p128-151. In cases where two (or more) jointly owned applications cover the same basic invention, a terminal disclaimer is required: for more information see http://www.uspto.gov/web/offices/pac/mpep/s1490.html

106 *See generally* MPEP § 201.07.

107 *See generally* MPEP § 201.06.

108 Special Rate Table Number 0576 at http://apps.opm.gov/SpecialRates/2012/Table057601012012.aspx.

109 While there are certain first actions that are not "on the merits," e.g., restriction requirements or examiner requests for more information, for parsimony we refer equivalently to first actions and FAOMs.

110 See Chapter 5 for more information.

In addition to the volume of work, an examiner is required to address identified applications and the various other applicant communications in a timely manner, prioritizing some work over others. This is known as "docket management". Combined, the count and docket management systems serve to balance the quantity of work and the mix of work subject to the office's objectives of quality and timely examination. The systems can be altered in order to incentivize specific examiner behaviours; and at the same time some unintended consequences may arise.

For instance, examiners may find one source of counts to be easier to accumulate than others, skewing the work towards that rewarding activity. Further, there is an incentive to skew work towards activities with compensation versus those without direct compensation, e.g., examining an RCE versus further considering an applicant's arguments in response to a response after a final rejection.

Patent examiners are covered by a collective bargaining agreement between management and members of the Patent Office Professional Association (POPA).[111] Because production expectancy is the result of bargaining between POPA and management, altering the count system and docket management systems are not trivial undertakings.

111 A full description of POPA and its mission can be found at http://www.popa.org.

Appendix E

US survival time regression results

1. **First-action pendency**

Recall that we estimate two sets of regressions, according to Figure 5.15: First-action pendency and post-first-action pendency. The post-first-action pendency regressions contain the competing risks of patenting and abandonment. In all regressions, the variable to be predicted (the dependent variable) is the observed duration until the given event (first action, patent grant, or abandonment).[112] In this section we focus on first-action pendency. The initial regression results are shown in Table E1, with the first-action pendency results in Column 1. Post-first-action pendency is subject to competing risks, so Columns 2 and 3 report the results of the post-first-action pendency regressions for patent grants (Column 2) and abandonments (Column 3).

In Column 1, the dependent variable is technically the duration until first action. The Weibull model is estimated using what is called the "accelerated failure time" metric. Consequently the regression coefficients in Table E2 are reported as *time ratios* for the variables of interest.

112 For more information about survival regression, see Box-Steffensmeier, J. M., and B. S. Jones. 2004. ***Event History Modeling: A Guide for Social Scientists.*** Cambridge: Cambridge University Press or Kalbfleisch, J. D., and Ross L. Prentice. 1980 (1st ed.), 2002 (2nd ed.). ***The Statistical Analysis of Failure Time Data.*** New York: John Wiley & Sons. For all reported regressions, we employ the Weibull parametric form for the hazard function. We use the Weibull model because of the flexibility of the approach and the ease in predicting pendency times. However, we have estimated other models, including other parametric survival time models like the Lognormal, and the semi-parametric Cox proportional hazards model. All the models yield the same basic results with respect to the variables of interest.

Table E1 Survival time regression results

	First action	Post-first action	
		Patent	Abandon
First-action pendency (log years)		1.226***	0.781***
Arrivals (log)	1.059***		
S1 (log)	1.318***	1.004	0.880***
S2 (log	1.293***	1.154***	0.819***
S3a (log)	0.629***	1.155***	1.107
S3b (log)	1.085***	1.032***	1.048**
Jr examiners (log)	0.869***	0.838***	0.995
Sr examiners (log)	0.880***	0.755***	0.985
Claims at filing (log)	1.120***	1.079***	1.301***
Small at filing	0.994**	1.027***	0.518***
Discipline, electronics	1.098***	0.874***	1.595***
Discipline, mechanical	0.957***	0.841***	1.177***
Parent, international	1.013***	1.000	1.018*
Parent, provisional	1.050***	1.149***	1.028**
Parent, con-div-cip	0.822***	1.088***	1.073***
Year Fixed Effects	Y	Y	Y
Constant	24.840***	14.335***	13.644***
Shape	2.139***	3.038***	2.207***
Frailty		1.932***	6.583***
N	461,996	405,262	405,262
LL	-362,030	-336,613	-295,552
AIC	724,122	673,289	591,169
BIC	724,465	673,638	591,518

Notes:

All coefficients are shown as time ratios, i.e., they are exponentiated.

S1 includes new arrivals for the post-first-action pendency regressions.

The omitted category for discipline is chemical.

The omitted category for parent is no parent.

Stars indicate statistical significance at the 0.1% (***), 1% (**), and 5% (*) levels.

Time ratios should be interpreted as multipliers with respect to pendency. For instance, having a US parent corresponds to a time ratio of 0.822 for first-action pendency. As a multiplier, that means that applications with a US parent (a continuation, continuation-in-part, or divisional) have a first-action pendency that is only 82.2% that of applications without parents. Because time ratios are multiplicative, a time ratio greater than one indicates an increase in pendency and a lower probability of first action per unit time. A time ratio that is less than one indicates a decrease in pendency and a higher probability of first action per unit time.

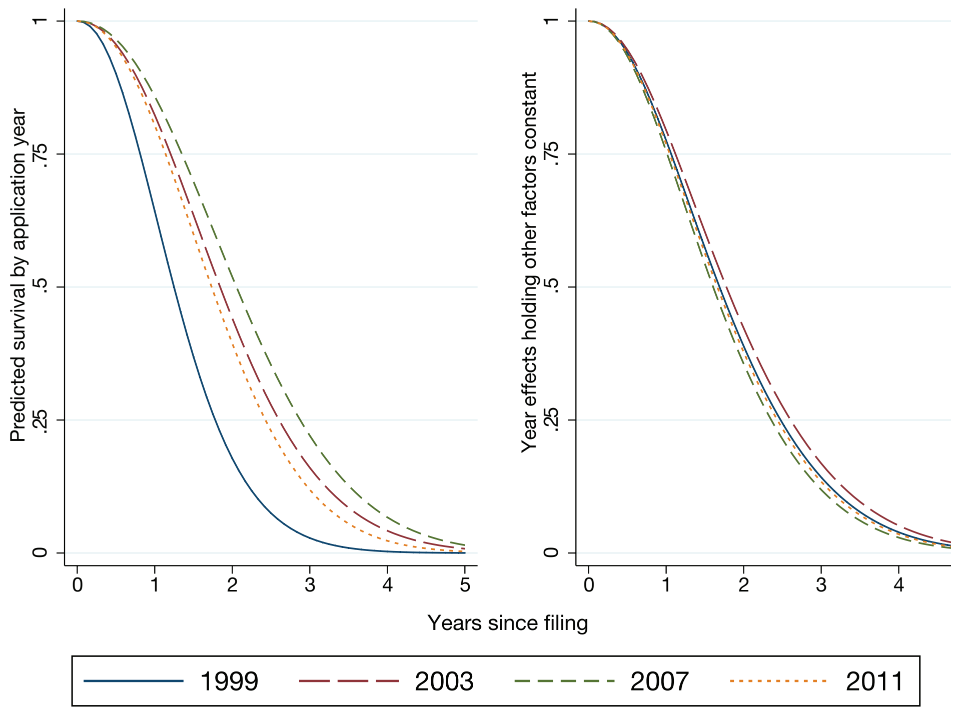
From the regression results, we can now assess the quantitative impact of pending application stocks on first-action pendency. In particular, new application arrivals, S1, and S2 all tend to increase pendency. These effects are expected. All of S1 and S2 are "in the queue" in front of any incoming applications. And, new arrivals join the queue at the same time as incoming applications. On the date of filing, a longer queue means greater pendency.

The effect of S3 is more nuanced. Because S3 is really the main bottleneck to issuance, its effect is different from S1 and S2. In order for an application to enter into S3, it must be that examiners have the capacity to make a first action. In this way, more applications in S3 reflect a higher carrying capacity for the examiner corps. However, there is a significant difference between S3a and S3b. While S3a appears to be correlated with higher carrying capacity (in that it reduces first-action pendency), S3b has a different effect. Relative to S3a, the RCEs in S3b significantly increase the first-action pendency of incoming applications. This is consistent with the descriptive analysis (Figure 5.12), which shows that RCEs crowd out regular applications.

Examiner counts tend to decrease first-action pendency, as we would expect. The effect of junior examiners is actually higher in this case, because newly hired examiners tend to work disproportionately on new cases.

Other factors that decrease first-action pendency include small entity status, mechanical applications, and those applications with US parents. Factors that increase first-action pendency include the number of claims, electronics applications, and applications with international or provisional parents.

Figure E1: First-action pendency by filing year

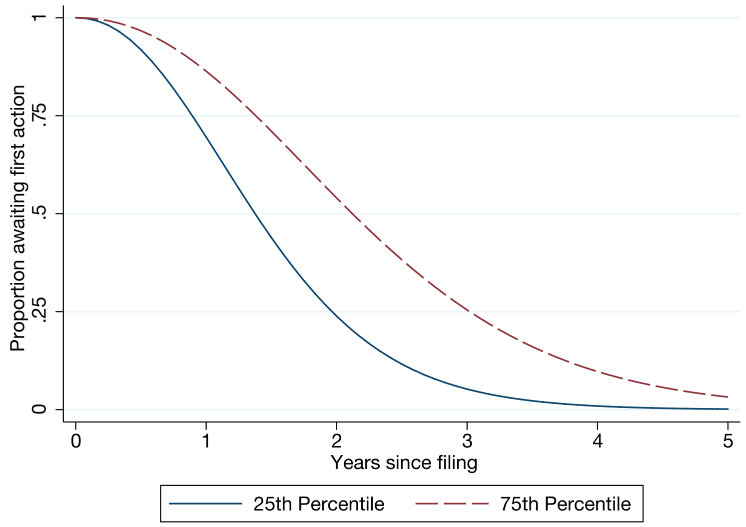


It is perhaps easier to see the results of the regressions using predicted survival curves. Figure E1 shows the benefit of the regression approach.[113] The first panel graphs the predicted survival curves, graphed for all applications filed in each year represented. Applications filed in 1999 had the shortest estimated first-action pendency and 2007 had the longest. However, much of the difference between 1999 and 2007 is due to the differences in other observable factors between the two years: current application inventory, the number of examiners, etc. The second panel graphs the estimated effects that are due *only to the years themselves* holding all other variables constant. In other words, the regression allows us to consider an application filed in two different years, assuming that the initial stocks were identical and all observable characteristics about the application itself (number of claims, etc.) are identical. The remaining differences are those that can be attributed to the filing years themselves. These differences may exist for institutional reasons; e.g., changes in examination procedures or overtime pay. It can be seen that those effects (in the right panel) are quite small, indicating that most of the year to year differences can be attributed to differences in stocks or other observable variables.

1.1 Stocks

Figure E2 graphs a hypothetical change in S1 and S2 from a relatively low level (the 25th percentile) to a relatively high level (the 75th percentile)[114]. Any shift in the survival curve to the right indicates a longer pendency. At the median (at the 0.5 level on the vertical axis), the increase in pendency is about nine months. Thus, increasing S1 and S2 causes a significant slowdown in first-action pendency for newly arriving applications.

Figure E2: Predicted first-action pendency from a hypothetical increase in Stock 1 and Stock 2



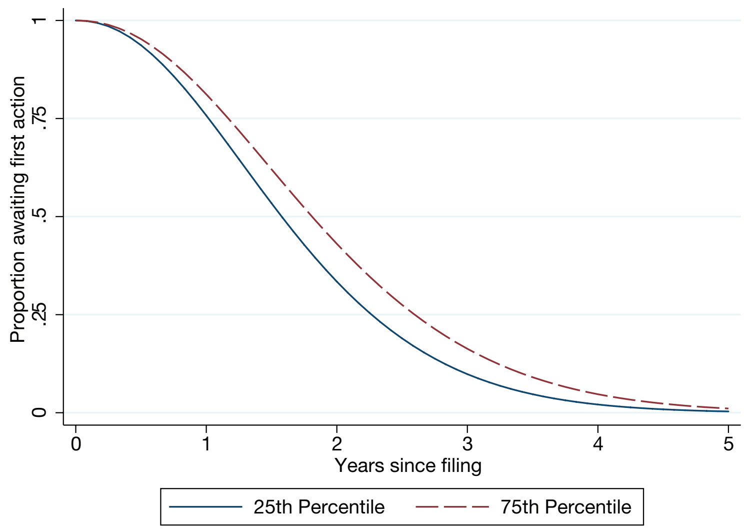
Note: Median survival time increases from 1.38 to 2.12 years. Stock 1 and Stock 2 are increased from the 25th to the 75th percentile.

113 The regression model enables us to predict the survival while varying only one variable.

114 This is equivalent to Stock 1 increasing by approximately 134%, 179%, and 194%, and Stock 2 increasing by approximately 112%, 59%, and 100% for chemical, electrical, and mechanical respectively.

To examine the effect of RCEs, we consider a hypothetical shift in the proportion of S3 applications that are RCEs. That is, we predict the effect of converting some S3a applications into S3b applications, while leaving the total size of S3 unchanged. Like with S1 and S2 above, we consider changing the proportion of RCEs from the 25th percentile (about 14%) to the 75th percentile (about 29%).[115] The result is shown in Figure E3, where the higher intensity of RCEs increases the first-action pendency of incoming applications by about 16%. When more applications file RCEs, resources tend to be spent on those applications rather than on first actions of incoming applications.

Figure E3: Predicted first-action pendency from a shift of Stock 3a applications into Stock 3b

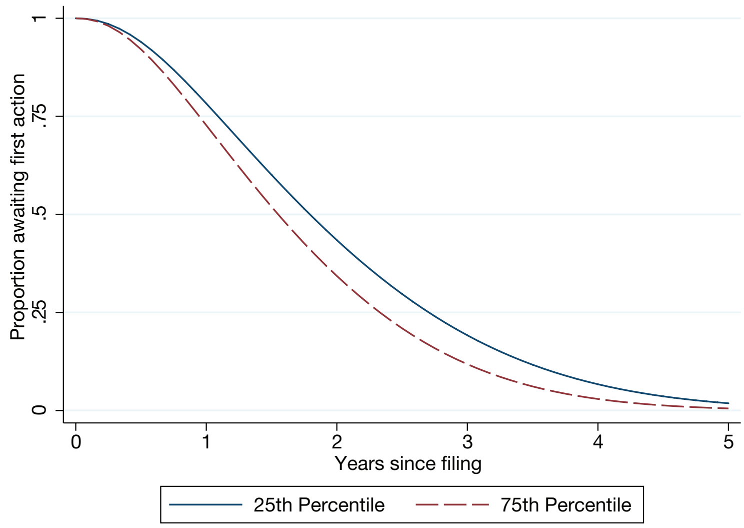


Note: Median survival time decreases from 1.58 to 1.81 years. The proportion of Stock 3 made up of Stock 3b is increased from the 25th to the 75th percentile with a corresponding decrease in Stock 3a.

115 This is equivalent to the proportion increasing from .141 to .285 in chemicals, .158 to .285 in electricals, and .165 to .309 in mechanicals.

Figure E4 shows the effect of increasing the stock of examiners from a relatively low level (the 25th percentile) to a relatively high level (the 75th percentile), which corresponds to an average increase of about 67%.[116] The predicted effect is to decrease the median survival time from 1.79 years to 1.55 years (a 13% decrease).

Figure E4: Predicted first-action pendency from a hypothetical increase in examiner capacity



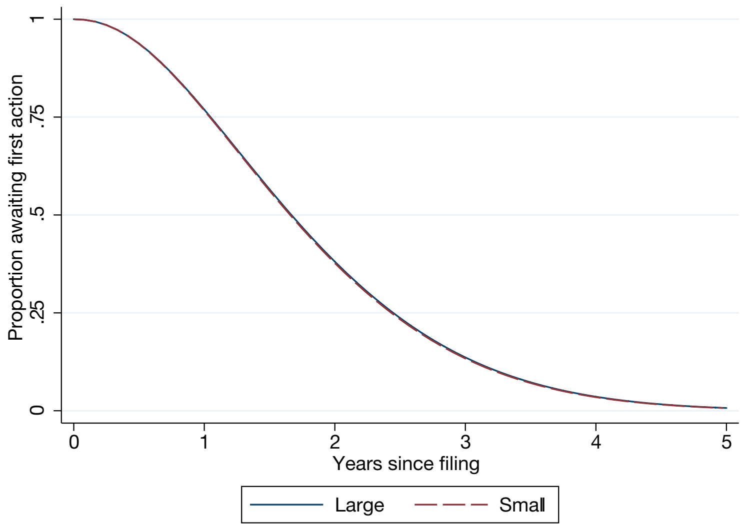
Note: Median survival time decreases from 1.79 to 1.55 years.

116 For junior examiners, the increase is about 38%, 112%, and 75% for chemical, electrical, and mechanical respectively. For senior examiners, the increases are approximately 30%, 97%, and 48% for chemical, electrical, and mechanical respectively.

1.2 Application characteristics

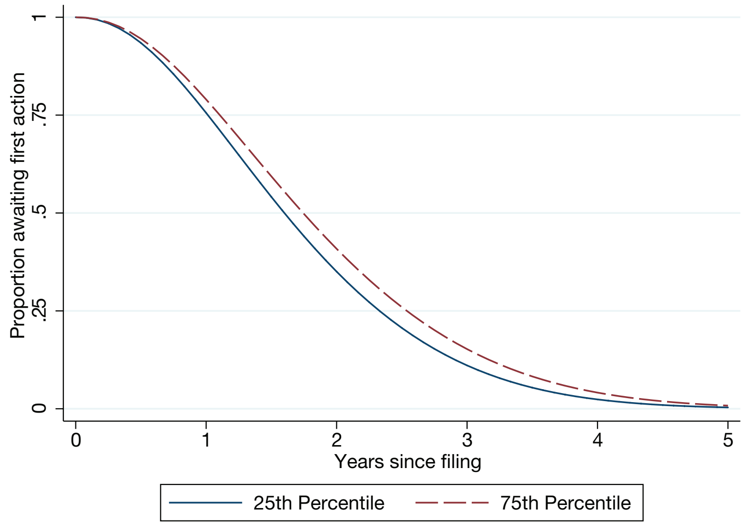
In some sense it should be expected that application characteristics should not have an impact on first-action pendency. If applications are examined on a literal first-in/first-out basis, then this would be so. For instance, Figure E5 shows that there is no discernible difference between the predicted first-action pendency of small entities versus large entities. In fact, it should be a policy goal that small entities and large entities are treated identically.

Figure E5: Predicted first-action pendency by small entity status



The result is not quite the same with respect to the number of claims. Figure E6 shows that a substantial increase in the number of claims (from the 25th percentile to the 75th percentile) is associated with a somewhat longer first-action pendency. Within a given discipline, applications with more claims generally require more processing time. However, this difference would be measured in hours, not in weeks. The remaining difference must be attributable to something else. Because examiners have some discretion over their choice of new work, the difference may be attributable to examiner preferences for applications with fewer claims. However, docket management policies are able to keep this preference from dominating workflow.

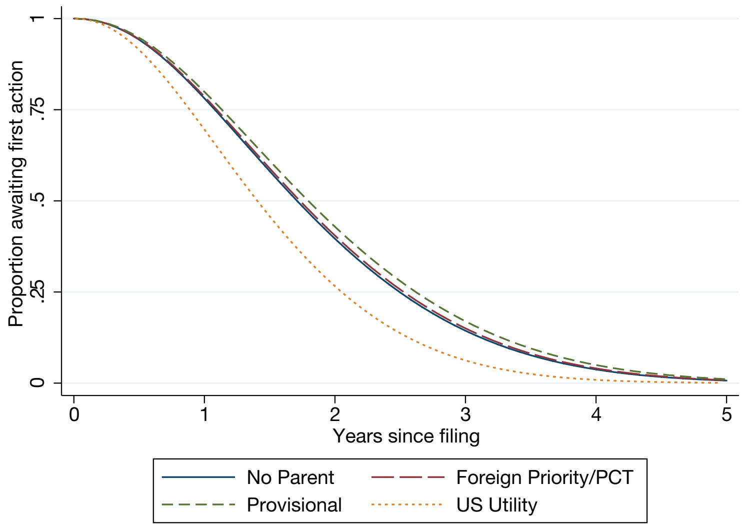
Figure E6: Predicted first-action pendency from a hypothetical increase in the number of claims



Note: Median survival time increases from 1.6 to 1.74 years. Based on the number of claims at filing.

Similar results can be shown in Figure E7 with respect to the origin of applications. Again, a literal first-in/first-out policy should lead to no distinction between parent types. For the most part, this is the case—with the exception of applications with a US parent (traditional continuations, continuations-in-part, or divisionals). While the child applications of US parent applications may not be docketed to the same examiner, having a US parent may indicate that an examiner can—in part—rely on the work of the previous examiner. Thus, they may be considered "low cost," and may be favoured by examiners over other applications. Again, docket management policies may be able to balance the results of this preference to some degree; however, from a policy perspective, we may want to allow examiners some discretion in terms of a triage of new applications. In some cases it may be socially beneficial to push through the easy applications over "harder" applications, because total wait time can be reduced.

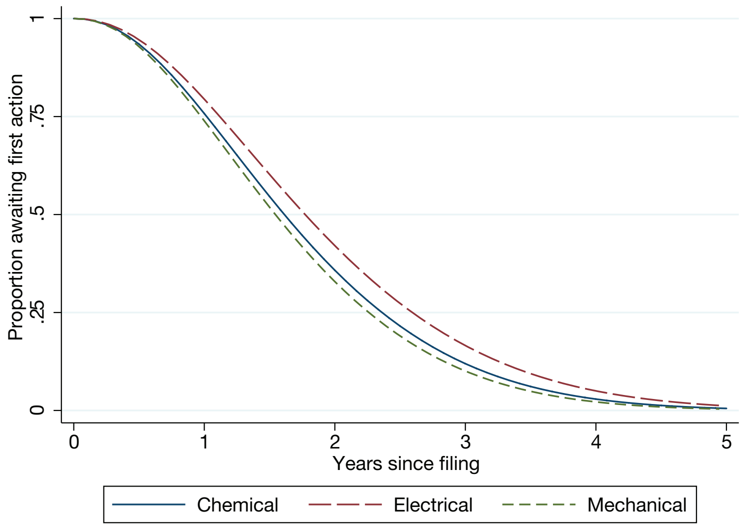
Figure E7: Predicted first-action pendency by parent type



Note: Median survival time ranges from 1.41 to 1.8 years.

Like parent type, the technological discipline of an incoming application has an effect on its predicted first-action pendency. Figure E8 shows that mechanical patent applications have the shortest first-action pendency, followed by chemical, and then electrical (roughly in order of complexity). We should expect the technological disciplines to have a noticeable effect on first-action pendency because neither incoming stocks nor examiners can switch between disciplines. Thus if one discipline is "under-staffed" relative to another discipline, it will take time to adjust hiring rates in order to balance the disciplines.

Figure E8: Predicted first-action pendency by discipline



2. Post-first-action pendency

Columns two and three of Table E1 show the results of the two post-first-action pendency regressions. Column two estimates the pendency from first action to patenting, for those applications that are patented. And column three estimates the pendency from first action to abandonment, for those applications that are abandoned. Together, the two models can be used to estimate overall disposal, as well as the propensity to patent versus abandon. As discussed above, patent issuance and abandonment are called ***competing risks***, because patent issuance precludes the observation of when the application would have abandoned, had it not patented. Similarly, abandonment precludes the observation of when the application would have patented, had it not abandoned.[117]

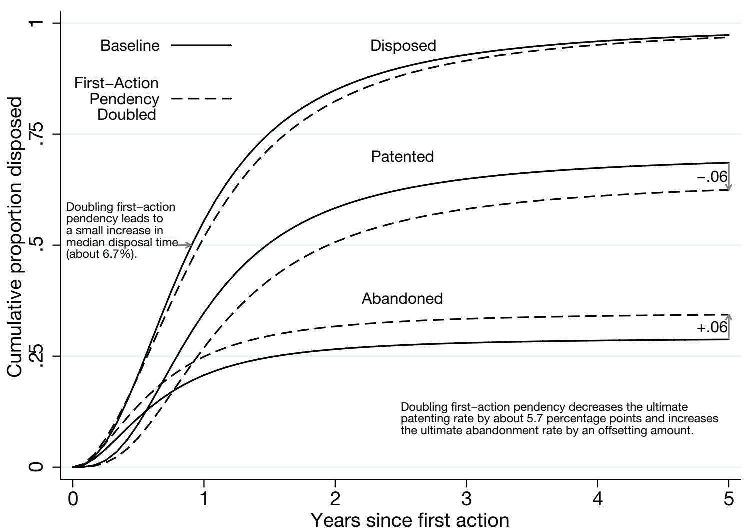
There are three slight changes in the post-first-action pendency regressions relative to first-action pendency. First, we include a variable measuring the first-action pendency time for the application in question. Our interest in this variable is whether an application's first-action pendency is correlated with a longer duration after first action. Second, the stocks of the explanatory variables are measured as of the date of the first action, because this is the point at which they enter S3 and are at risk for disposal. For the same reason, S1 and new arrivals are combined because we do not expect the time to disposal to be sensitive to new arrivals as distinct from S1. Last, the year fixed effects indicate the date of first action rather than the date of filing.

117 Technically, survival time models assume that each event type would eventually occur, over a long enough time horizon. We employ a method of calculating an adjusted CIF that accounts for the competing risk for each of the disposal types. The adjusted CIFs can be used to assess the net effect on ultimate issuance rates. See Putter, H., M. Fiocco, R. B . Geskus. 2007. *Tutorial in Biostatistics: Competing Risks and Multi-State Models.* Statistics in Medicine, 26: 2389-2430.

2.1 Effect of first-action pendency

Because all the subject applications have already received a first action, we include the time until first action for that application as an explanatory variable for its post-first-action pendency. We find that first-action pendency does have a significant effect on both patenting and abandonment. First-action pendency tends to slow down patenting and speed up abandonment. As far as the total disposal rate is concerned, the effects somewhat cancel out. However, the overall abandonment rate will be higher for applications with higher first-action pendency.

Figure E9: Predicted disposal after first action if first-action pendency is doubled



The result can be seen graphically in Figure E9. The figure considers the cumulative incidence functions for a hypothetical application where we hold all factors constant, but where we increase the observed *first-action pendency* to see the effect on *post-first-action pendency.* By combining the CIFs for patenting and abandonment, we can observe the total disposal CIF (which is the sum of the patenting and abandonment CIFs).

We consider a doubling of first-action pendency (e.g., from one year to two years). The result is that patenting is slower (and less likely) and abandonment is faster (and more likely). The ultimate probability of patenting falls by six percentage points; similarly the ultimate probability of abandonment increases by the same six percentage points. Note that the CIF for overall disposal does not change significantly. This is important because it shows that slower first-action pendency leads to slower disposal. But, because some applications "fall out" of the pending stock through a higher abandonment rate, the total disposal rate shown in the CIF understates the impact of higher first-action pendency.

We focus first on the first-action impact on post-first-action pendency, because it creates a separate indirect impact that compounds the effect of the stocks and application characteristics. That is, according to Figure 5.17, each explanatory variable has a direct impact on post-first-action pendency. In addition, the explanatory variables influence first-action pendency, which in turn influences post-first-action pendency. We call this effect the indirect effect.

2.2 Stocks

Table E2 presents a summary table that describes the results of the competing risk model for post-first-action pendency. Two regression results are used here, using a competing risk methodology: post-first-action pendency for granted patents and for abandonments, as detailed in Tables 5.3 and E1. Those results jointly determine the combined CIF (as shown in Figure E9). The CIF for patented applications and the CIF for abandoned applications are then constructed by scaling each individual CIF down to account for the alternative risk. The net result is summarized in Table E2.

Table E2 Results of competing risk regressions for post-first-action pendency

	Predicted response in post-first-action pendency		Predicted response in ultimate abandonment rate	
	Direct effect	**Total effect**	**Direct effect**	**Total effect**
Explanatory variable	**(1)**	**(2)**	**(3)**	**(4)**
	(% change)		(percentage point change)	
Stocks				
(impact of doubling)				
Stock 1	-1.9%	-0.1%	1.7	3.2
Stock 2	4.3%	6.0%	4.3	5.8
Stock 3a (non-RCEs)	10.0%	6.4%	0.5	-2.1
Stock 3b (RCEs)	2.5%	3.1%	-0.2	0.3
Junior examiners	-9.4%	-10.3%	-2.1	-2.9
Senior examiners	-14.7%	-15.5%	-3.2	-3.9
Application characteristics				
(impact of doubling)				
First-action pendency	6.7%	6.7%	5.7	5.7
Claims	8.7%	9.7%	-2.3	-1.7
(impact of discrete change)				
Small entity status	-14.0%	-14.1%	12.7	12.6
Electrical (relative to Chemical)	1.0%	2.1%	-10.9	-10.1
Mechanical (relative to Chemical)	-8.6%	-9.0%	-6.1	-6.5
International parent	0.4%	0.6%	-0.3	-0.2
US provisional parent	12.2%	12.7%	2	2.4
US parent	8.7%	6.5%	0.2	-1.3

Notes:

Baseline post-first-action pendency is 10.9 months, with 29.2% abandonment.

Impact of a discrete change represents the effect of that indicator relative to the alternative (omitted) category.

Total effect includes the direct effect of the explanatory variable plus the indirect effect of the explanatory variable operating through its impact on first-action pendency (see Figure 5.17). Because first-action pendency itself has no indirect effect, the total effect is equal to the direct effect.

Columns 1 and 2 of Table E2 show the effect of the explanatory variable on post-first-action pendency. Column 1 shows the direct effect, which takes into account only the change in the explanatory variable. The total effect in Column 2 combines the direct effect of the explanatory variable, and adds the indirect effect that comes from the explanatory variables impact on first-action pendency, which then affects post-first-action pendency (see Figure 5.17). For instance, Stock 2 has a positive direct effect on post-first-action pendency (doubling S2 increases post-first-action pendency by 4.3%). However, S2 also has an effect on first-action pendency (it increases it, as shown in Tables 5.3 and E1), and higher first-action pendency has an effect on post-first-action pendency.

The regression results (Table 5.3 and E1) show that increased first-action pendency is associated with slower patenting and faster abandonment. But, it slows down patenting ***by more than*** it speeds up abandonment. The net effect is shown in Column 1 of Table E2: doubling the first-action pendency increases the post-first-action pendency by 6.7%. The indirect effect is combined with the direct effect to generate the total effect in Column 2. The total effect for first-action pendency is equal to the direct effect, because the indirect effect of the other explanatory variables operates ***through*** the first-action pendency variable.

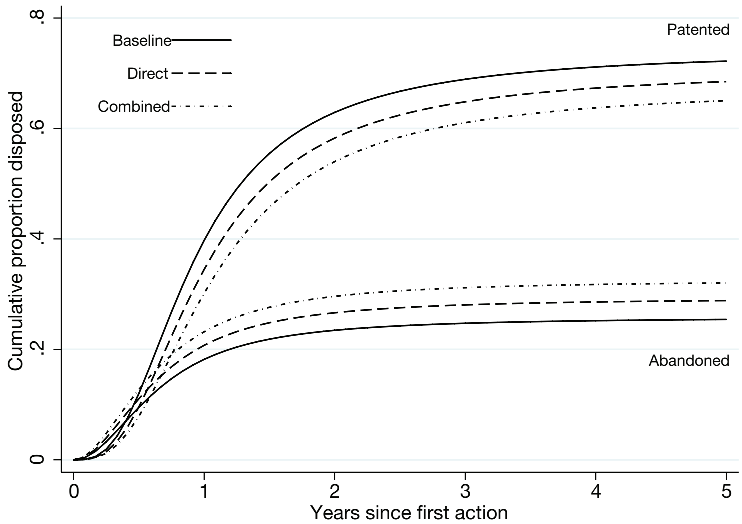
A further consequence of the differential impacts on patenting and abandonment is that the overall abandonment rate will be higher for applications with higher first-action pendency (by 5.7 percentage points, given a doubling in first-action pendency). Thus, Columns 3 and 4 of Table E2 show the impact of the explanatory variable on the predicted abandonment rate of newly examined applications, where Column 4 incorporates the direct effect of Column 3 as well as the indirect effect.

In considering the effects of the stocks on post-first-action pendency, we can see from Table E1 that S1 and S2 are correlated with longer post-first-action pendency for patented applications, and shorter pendency for abandoned applications. Both of these effects are interesting because applications in S1 and S2 are—in a sense—***behind*** the applications that are the subject of the regressions. This difference between the effects of the application stocks on patenting and abandonment provide interesting inferences. S1 and S2 may provide competing pressures on examiners, who have to choose between continuing prosecution and working on first actions. The higher the stock of S1 and S2, the greater the pressure to work on first actions, to the exclusion of applications in S3. Empirically, this competition seems to slow down patenting and speed up abandonment for applications that have already received a first action.

The net result (the total effect in Column 2 of Table E2) is that S2 slows down post-first-action pendency, and both stocks increase the level of abandonment (Column 4). The Stocks 1 and 2 on post-first-action pendency can be seen graphically in Figure E10. We consider a hypothetical increase in S1 and S2 from a low level to a high level (25^{th} percentile to the 75^{th} percentile), holding all other factors constant. The baseline is where S1 and S2 are at their 25^{th} percentile levels. When S1 and S2 are increased to their 75^{th} percentile levels, there is a direct effect in that this serves to slow down patenting (lowering the patenting CIF) and speed up abandonment (raising the abandonment CIF). This serves to also increase the ultimate abandonment proportion. Further, increasing S1 and S2 also tends to increase first-action pendency. This yields a second (indirect) effect as described above, which magnifies the direct effect. That is, patenting becomes even less likely and abandonment even more likely. While overall post-first-action pendency for disposals will not change much, an increase in S1 and S2 will have two important effects: it will increase overall pendency (because it will increase first-action pendency), and it will increase abandonments relative to patents.[118]

118 The impact on overall pendency is easy to see in Figure 5.18.

Figure E10: Predicted disposal after first action if Stock 1 and Stock 2 are increased



Note: Stock 1 and Stock 2 are increased from the 25th percentile to the 75th percentile.

For Stock 3, the post-first-action pendency results differ distinctly from those of the first-action pendency. With respect to stocks, the primary difference is that both S3a and S3b serve to increase the post-first-action pendency of patents. Recall that S3a appears to be a signal of the carrying capacity of examiners *at the time of filing*. However, the size of S3a has the expected congestion effect on pendency by the time the application actually reaches S3a. Thus, at the time of first action other S3a applications compete for resources and tend to increase the pendency.[119] Stock 3b has similar effects to those of S3a, but smaller in magnitude. Overall, the largest impacts from Stock 3 derive from the impacts on first-action pendency (see Figure 5.19)

More examiners tend to shorten the duration for patenting (Table E1), without changing abandonment rates. In fact, senior examiners have a stronger effect than junior examiners. Thus, more examiners are correlated with shorter post-first-action pendency and a lower ultimate abandonment proportion. This fact by itself is good: more examiners do not encourage or discourage applicants with respect to abandonments. They merely speed up patenting, and in doing so, they increase the ultimate issuance rate. Thus, the combined effects in Table E2 show examiners correlated with a significantly lower post-first-action pendency and lower abandonment rate.

119 The size of S3a does not significantly impact the abandonment rate directly. The total effect is due to the fact that S3a speeds up first actions, which decrease abandonment.

2.3 Application characteristics

Several other factors are found to have a strong correlation with post-first-action pendency and abandonment rates. First, the number of claims—to the extent that it measures complexity—tends to increase post-first-action pendency but reduce abandonment, as shown in Table E2. This result may seem counter-intuitive. However, the regression results in Table E1 show that this pattern is due to the fact that higher claims are associated higher post-first-action pendency of both abandonments and patents. The important feature is that it slows down abandonment much more than it slows down patenting. Thus, in net it leads to less overall abandonment.

Small entities tend to abandon much faster than large entities, according to the regression results in Table E1. Thus, the small entity indicator is correlated with shorter post-first-action pendency and higher abandonment in Table E2.

Relative to chemical applications, both electronics and chemical applications have shorter post-first-action pendency for patented applications, but much longer pendency for abandonments. This leads to a lower abandonment rate for both disciplines, relative to chemicals. However, the relative magnitudes work out such that mechanical applications have a lower overall post-first-action pendency and electronics applications have a higher overall post-first-action pendency.

Patent applications that claim priority to provisional applications tend to have longer post-first-action pendency (due mostly to their longer pendency for patented applications). The same is true for applications with parents that are regular US applications (i.e., continuations, continuations-in-part, and divisionals). However, recall that those with regular US parents have a much shorter first-action pendency, on average. Thus, the indirect effect works in different directions for the two parent types, and applications with a provisional parent end up with a higher level of abandonment.

Concept House
Cardiff Road
Newport
NP10 8QQ

Tel: 0300 300 2000
Minicom: 0300 0200 015
Fax: 01633 817 777

For copies in alternative formats please contact our Information Centre.

When you no longer need this booklet, please recycle it.

DPS/IP Research-06/13