BAYESIAN STATISTICAL MODELING FOR ENHANCING PATENT APPLICATION GRANT RATE TIMELINES FOR TEMPORAL PATENT PROSECUTION PREDICTION

by

Turner R. Swanson

An Essay Submitted to the Faculty of the Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Master of Science in Computational Sciences

Milwaukee, Wisconsin

December 2020

TABLE OF CONTENTS

CHAPTER

Bayesian Statistical Modeling for Enhancing Patent Application Grant Rate Timelines for Temporal Patent Prosecution Prediction

By Turner R. Swanson
Department of Mathematical and Statistical Sciences
Marquette University, Milwaukee, Wisconsin, USA
turner.swanson@marquette.edu

Abstract - This paper discusses Bayesian statistical modeling for improving grant rate timelines for making temporal predictions regarding the outcome of patent applications filed at the United States Patent and Trademark Office (USPTO). The model is obtained from data stored for millions of patent applications filed at the USPTO. The focus of the model is on non-provisional patent applications; the most common type of application filed. The model uses all issued and abandoned non-provisional applications filed post-2000 to obtain a probability density function. Probability density functions are obtained from subsets of these applications based on various categorical variables, such as Art Unit, Examiner ID, USPC Class & Subclasses, and Customer Number. Each of these probability density functions are combined using a Bayesian model and aggregated using linear regression to obtain an unweighted aggregate model and weighted linear regression to obtain a weighted aggregated model for making a more accurate grant rate timeline for predicting the outcome of a patent application filing over time.

Keywords - Patent prosecution statistical modeling for, Bayesian statistical modeling, forecasting, grant rate timeline

1. Introduction

Due to the high cost associated with patent prosecution, the decision to pursue or continue pursuing a patent application (application) can be costly. The estimated cost varies from $12,000 to $22,000 to obtain a patent depending on the technological area. This cost includes filing fees from the USPTO, in addition to Attorney fees for performing a patent search of existing related patents; preparing an application's description, drawings and claims; and prosecuting the application [1]. (It should be noted that this only includes the cost associated with filing in the United States and does not consider the additional cost of filing in other countries.) As such, it is important to know if an application is worth filing; or alternatively, since high additional costs can arise with extended prosecution, whether it is worthwhile to continue pursuing an application after an extended prosecution. An example of additional cost for long pending applications is filing an appeal to an Examiner's final rejection of an application. The decision to file or continue pursuing an already pending application can be highly dependent on the counsel's patent strategy and the value and importance of the invention or product to the inventor or company. However, having additional information regarding the probability of an application being granted or becoming abandoned can assist in making more informed decisions. Generally, the counsel (patent attorney or agent) will rely on their personal experience with specific examiners. As discussed more below, some counsel and inventors already utilize data analysis of examiner history to help make decisions during prosecution.

The primary purpose of data analysis of patent applications is to leverage algorithms, machine learning, and data mining to predict the future outcome of a pending application. It allows inventors to pursue applications with the greatest chance of issuing, and reduce spending on the pursuit of applications that have a high chance of being rejected and ultimately abandoned. Even in cases where an application may eventually issue, the cost associated with eventually obtaining a patent may not be worth it [1].

Analysis may also be used to reduce the cost associated with examination (e.g., on office actions, interviews, RCEs, Appeals, etc.) by informing the inventor to adapt their prosecution strategy. For example, deciding to pursue a Request for Continued Examination over an Appeal after a final rejection by an examiner based on data for similar applications or that examiner. Alternatively, by simply abandoning applications that are unlikely to issue earlier. Statistics and data for each of these areas can be gathered and tracked. These statistics include allowance rate, examination timeline, number of pending RCEs, etc. This information can help inform a more successful prosecution strategy.

It is also useful to understand the probability of issuance overtime during prosecution for specific examiners, art units, classes, sub-classes, etc. Similarly, this quantification can also be done for outside law firms (and specific attorneys within the law firm), in-house counsel, or divisions in a company, which can then be used to compare performance and make more informed decisions regarding patent prosecution. This can be done generally or for specific technological areas [2].

Another use of data examination and analysis is in regards to patent portfolios to determine their underlying value. Quantifying the patent portfolio of a company can be otherwise difficult. This is useful for financial analysis in determining how much intellectual capital a company possesses. An example of a way this is performed is by determining the number of forward and backward citations that patents and patent families have in a patent portfolio [3].

2. Background

A patent is a property right granted by the United States Government to the creator of an invention so as “to exclude others from making, using offering for sale, or selling the invention throughout the United States or importing the invention into the United States” for a specified limited time. Currently, this is 20 years from the date of filing of the earliest application which priority is claimed. However, in exchange for this essential monopoly right, the inventor must disclose the details of the invention when the patent is granted - or earlier through the publication of the application [4]. The fundamental idea behind this concept is to encourage scientific and technical innovation through the free exchange of ideas by discouraging inventors from keeping their inventions and ideas secret in exchange for a limited essential monopoly on the invention. There are several different types of patents including design patents, utility patents, and plant patents. Utility patents are the most common type and cover most types of inventions including processes or methods, machines (moving parts or circuitry), manufactured articles (e.g., pencil), and new compositions (e.g., pharmaceuticals). In order to obtain a patent, first, a non-provisional application is filed with the USPTO. The application broadly includes a detailed description of the invention, which almost always includes drawings or diagrams, and one or more claims that describe what the inventor believes to be the bounds defining the claimed invention. In order for a patent to be granted by the USPTO, the claimed invention must (1) be patentable subject matter, (2) have some utility or usefulness, (3) be novel (i.e., new), and (4) not be an obvious variation of what already exists. Due to the complexity of this determination, the USPTO relies on thousands of trained patent examiners with a variety of technical backgrounds. These examiners essentially make the determination of whether the claims of an application meet the above criteria in the form of an office action where the claims are either rejected, allowed, or some combination thereof, if there is more than one claim. If the claims are rejected the inventor's counsel is able to respond with arguments, or amendments to the claims, to sway the examiner to allow the claims through a process called patent prosecution. If the claims are allowed it will eventually be issued to the inventor in the form of a granted patent. However, many applications are unable to overcome this threshold and remain rejected, for various reasons; this typically results in the application being abandoned [4]. In reality, the process is more complex with many different rules and paths, but the details are handled by the inventor's counsel.

Other options after a final office action (i.e., final rejection) from the Examiner include filing a Continuation, Continuation-In-Part, or Divisional applications. Although these can also be filed if a patent is issued on an application [4].

When an application is filed by an inventor or their counsel it will eventually make its way to an examiner having a five digit "Examiner ID." The examiner is generally based on what classification (class) and sub-classification (subclass) that a given application is determined to be categorized by. The classification system is a way of organizing patent documents. Each application is assigned to an examiner and an art unit as part of the Technology Center in the USPTO. For example, art units in Technology Center 1600 include "Biotechnology and Organic fields," art units in Technology Center 2100 include "Computer Architecture Software and Information Security," art units in Technology Center 3700 include "Mechanical Engineering, Manufacturing and Products" [5]. The class and subclass are based on more specific categorization of the invention based on the United States Patent Classification (USPC) system. However, USPC was recently switched to the Cooperative Patent Classification (CPC) system. The CPC system is an effort to harmonize classification between USPTO and the European Patent Office (EPO). For example, class 375 is "pulse or digital communications," having a corresponding subclass 257 of "cable systems and components." It should be noted that other classification systems are used by patent offices of other countries, for example, the International Patent Classification system (IPC). In addition, each application has a "Customer Number," a five digit number based on the company or inventor filing the application [6].

Each of these can be considered a predictor or factor (i.e., a categorical independent variable) known at or soon after filing of the application. The classes and subclasses finally decided on will determine which art unit and ultimately which examiner will receive the application.

3. Related Works

There has been prior work performed in the area of patent prosecution analytics, specifically in the form of tools for attorneys and inventors. Some of the existing patent prosecution analytic tools are Examiner Ninja, PatentBots, Juristat Examiner Reports, and LexisNexisIP. Each of these analytic tools varies in their scope and complexity. The below discussion will primarily focus on Examiner Ninja and PatentBots web tools.

3.1 Examiner Ninja

Examiner Ninja is a free web tool created by Justin Roettger for analysis of examiner specific statistics. Providing specific breakdowns of that examiner's application examination history for making determinations for several different scenarios. Juristat's Examiner Reports provides a very similar data tool on a per examiner basis. These tools can be used to quickly determine tendencies of examiners and update the prosecution strategy, accordingly [7].

The first scenario, looks at the allowance rate (i.e., the percentage of applications that are allowed, aka grant rate) of applications with and without an examiner interview to determine whether to conduct an examiner interview with a particular examiner. For example, if the allowance rate with an examiner interview is higher with an interview than without an interview then it may be wise and worth the extra cost to conduct an examiner interview. The relative benefit of conducting an interview is also provided (i.e., the percent change that conducting an interview provides), as shown in Figure 1. This percent is positive when conducting an interview is beneficial and negative when it is detrimental [7].

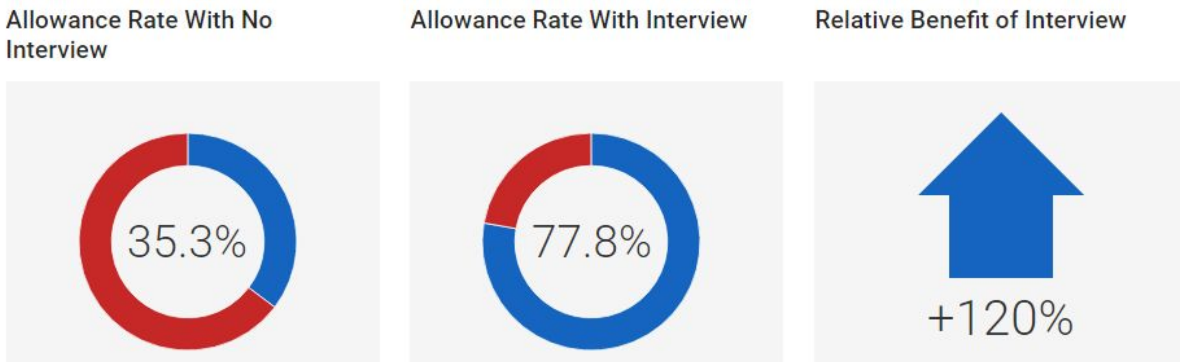

Fig. 1: *A graphical breakdown of allowance rates for a particular examiner with no interview and with an interview, and the relative benefit, as provided by Examiner Ninja.*

The second scenario looks at allowance rates with and without a Request for Continued Examination (RCE), and allowance rates with and without an Appeal for a particular Examiner, in order to better determine how to proceed after a final Office action (rejection) from the Examiner. It also provides the relative benefit of an RCE and Appeal, respectively, as shown in Figures 2 and 3. This relative benefit is the percent change in allowance rate with RCE or Appeal [7].

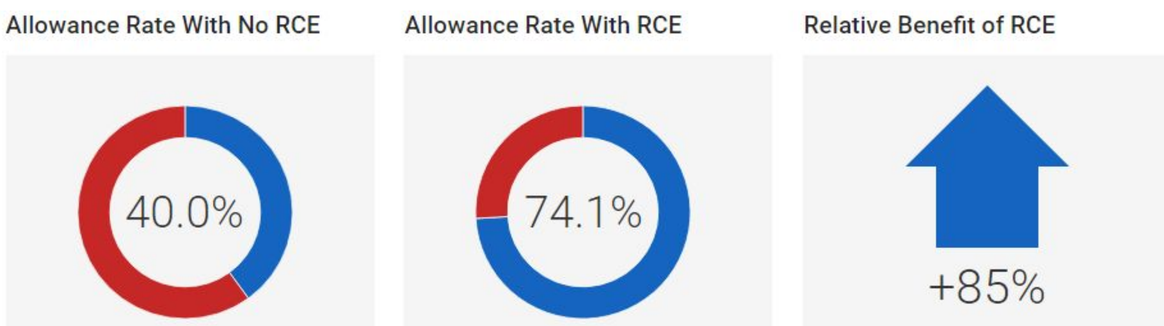

Fig. 2: *A graphical breakdown of allowance rates for a particular examiner with no RCE and with an RCE, and the relative benefit, as provided by Examiner Ninja.*

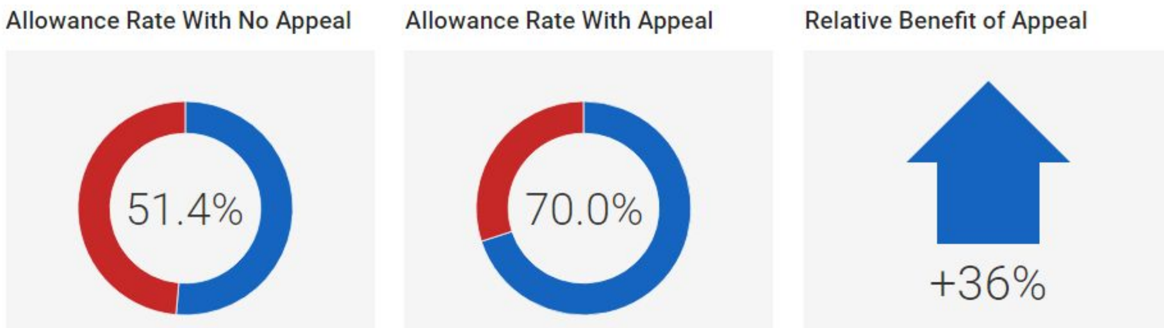

Fig. 3: *A graphical breakdown of allowance rates for a particular examiner with no RCE and with an RCE, and the relative benefit, as provided by Examiner Ninja.*

Filing an RCE or an Appeal are the two main ways to continue pursuing an application after a Final Office action from the examiner. An RCE is a Request for Continued Examination, which simply continues the arguments with the current examiner. An Appeal moves the application to a higher court in an attempt to overrule the examiner's final rejection. Although these are the most common options, they are not the only options. For example, an inventor may request a response after a final rejection, though this may not be granted by the examiner.

3.2 PatentBots

PatentBots is another web tool that provides various statistics; one specifically is the 3-year grant rate (3YGR) of Technical Center, Group, Art Unit, and Examiner. Each of these categories grant rate can be compared relative to each other and the USPTO generally (i.e., all applications). The 3-year grant rate is the percentage of applications that have been granted (i.e., issued) by three years after the first office action [8].

Figure 4 shows a graphical comparison of the 3-year grant rate for example Art Unit 2611, Group 2610, and the USPTO generally [8].

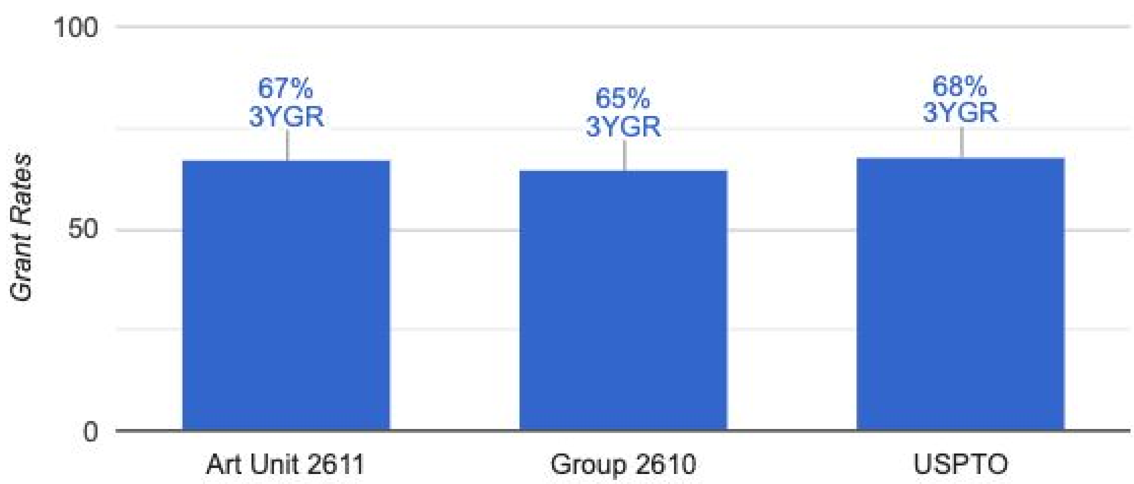

Fig. 4: *A graphical comparison of three year grant rates for Art Unit 2611, Group 2610, and the USPTO generally, as provided by PatentBots.*

Figure 5 shows a grant rate timeline for example Art Unit 2611. The timeline is relative to the date of the first office action [8]. As mentioned, this grant rate timeline can be obtained for any Technical Center, Group, Art Unit, and Examiner. However, this cannot be obtained for an aggregation of these

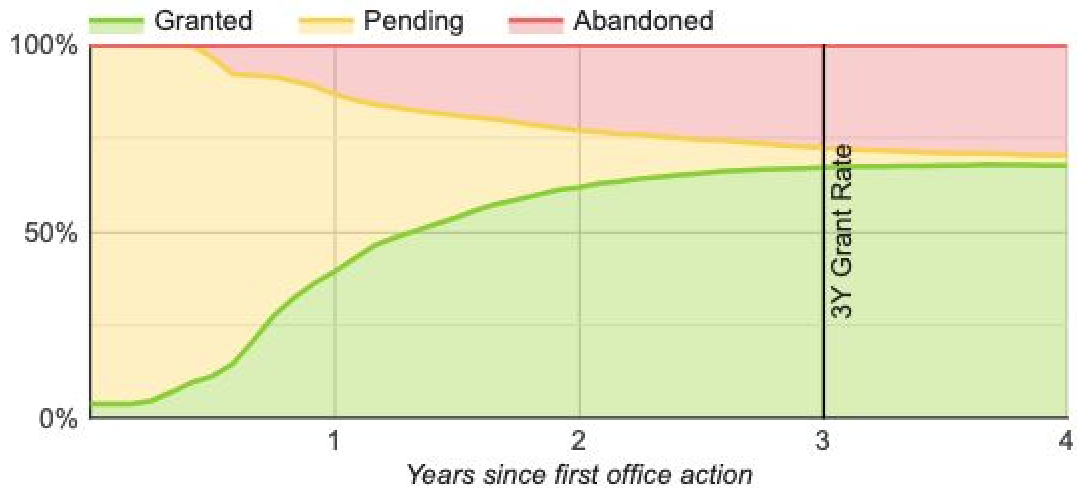

Fig. 5: *An example grant rate timeline for Art Unit 2611, as provided by PatentBots.*

4. Motivation

Currently, the decision to file an application is largely based on a few factors, including: (1) the counsel's belief of whether an invention is patentable over the prior art, based on a patent-search, the counsel's experience in the specific technology area and art-unit, and history with the specific examiner; (2) the inventor's and counsel's knowledge of the specific technology field; and (3) the companies incentive to obtain a patent based on the predicted value of the technology and the competitiveness the technology space. If the counsel does rely on a grant rate timeline for making a decision, currently, they can only obtain a grant rate timeline for one specific factor at a time. Multiple grant rate timelines for specific applications, such as Art Unit and Class, may conflict.

The decision making process for continuing to pursue a patent application that is currently pending and has received an adverse Office action is similarly determined based on experience. These decisions are heavily reliant on experience and "feel" without much analytical basis to back them up.

Accordingly, there is a need for methods of analyzing patent application issuance and abandonment data to better inform the inventor, the counsel, and the company making the decision to file an application, or continue pursuing an application. Specifically, a way of aggregating grant rate timeline for multiple factors associated with an application.

5. Research Goals

The goal of this research is to use relevant patent application history data to determine the probability of allowance and abandonment of an application upon initial filing over time. The solution should combine relevant application data based on the art unit, examiner, classification, sub-classification, customer number, and other factors to determine a more specific and precise probability distribution.

The solution would provide for weighting the various factors based on relevance and confidence. This could be in the form of manually adjusting confidence or weights of the various factors. Alternatively, it may take the form of weighting based on importance. The solution would also provide histograms of allowance or abandonment of relevant applications over time for each of the relevant factors. Most importantly the solution would provide a grant rate timeline for issuance and abandonment based on all the relevant factors and compare this to the grant rate timeline for all applications generally.

Additional goals include determining the probability of receiving a second office action, or requiring an appeal, RCE, or other actions in order to obtain an allowance. This would allow for the simulation of prosecution of a specified application a large number of times to determine the length and cost to grant / abandonment. In addition, updating the probability after each relevant action/event occurs during prosecution to improve accuracy.

6. Overview of Proposed Solution

The proposed solution is to quantify the probabilities that an application will issue or abandon at the time of filing and over time after filing in the form of probability density functions. This probability should be based on factors defined for the application at the time of filing, or soon after. These factors include, but are not limited to, the art unit and examiner. In addition, quantifying the change in probability over time. The probability that an application will grant or issue before or after an indicated time, or between two times.

7. Details of Solution

7.1 Characteristics of the Problem

The proposed solution uses bulk static patent application data from the United States Patent and Trademark Office's Bulk Data Storage System initially released October 1, 2015. The data was retrieved from the storage system in February of 2020. The files are updated on a regular ongoing basis. This data comes in the form of a .CSV and includes patent data for applications dating back to February 15, 1910. However, the application data dating back this far is not complete.

The data includes well over 10 million (11,125,755) patent applications including the following static fields: application number, filing date, invention subject matter, application type, examiner last name, examiner first name, examiner middle name, examiner ID, examiner art unit, USPC class, USPC subclass, confirmation number, customer number, attorney docket number, application status code, application status date, file location, file location date, earliest pre-grant publication number, WIPO publication number, WIPO publication date, patent number, patent issue date, abandon date, disposal type, invention title, small entity indicator, and AIA first-to-file.

7.2 Initial Data Cleaning and Analysis

The research focused on regular non-provisional utility applications. Accordingly, applications with the field of "invention subject matter" (i.e., invention_subject_matter) that are "utility" (i.e., UTL) are kept and "design" and "plant" (i.e., DES and PLT) are removed. Further, only applications with an "application type" (i.e., application_type) of "regular" (i.e., REGULAR) are kept and "provisional," "patent cooperation treaty," "reexaminations" and "reissues" (i.e., PROVSNL, PCT, REEXAM, and REISSUE) are removed. This is done since the vast majority of applications are regular nonprovisional utility applications. This reduces the number of applications being analyzed to 9,074,685 utility applications, and then further down to 8,419,611 regular non-provisional utility applications. Conversely, there are 9,025,815 regular applications before filtering out all non-utility applications.

However, prior to 2001 only applications that were issued were actually published. This resulted in selectivity problems from the omission of "non-public" applications since many applications that were abandoned are not generally included in the data. In late 2000 the American Inventors Protection Act (AIPA) was implemented, which includes provisions for the publication of applications prior to grant. In particular, nearly all applications filed are published with a pre-grant publication about 18 months after filing unless the USPTO is specifically instructed otherwise. This publication time varies slightly depending on the art unit. This increases the inclusivity of the Public Patent Application Information Retrieval (PAIR) system for applications received by the Patent Office starting in late November of 2000. This greatly reduces the selectivity issues prior to 2001 [8]. By filtering out applications filed prior to 2001 the number of applications of concern is reduced down to 5,649,154 regular non-provisional utility applications.

Figure 6 and Table 1 shows the total number of applications filed each year post 2000. As expected the number of applications filed increases nearly every year. The only downturn in applications filed each year is 2008 and 2009 following the financial crisis. Since the data was retrieved from the system in early January there are no yet known applications from 2019 and very few known or published applications from 2018. The appearance of a decrease in the total applications filed in 2017 is likely due to late publications of applications filed at the end of 2017. The average number of applications filed from 2001 to 2016 is 332,173. The average number of published applications filed from 2001 to 2018 is 313,842.

Table 1: *Total number of published applications filed for each year 2001 to 2018.*

Years	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**
Apps Filed	269,015	274,671	279,067	295,634	311,754	325,844	335,421
Years	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**
Apps Filed	329,279	304,066	323,902	341,889	364,179	383,800	394,885
Years	**2015**	**2016**	**2017***	**2018***	**Total ('01-'18)* | ('01-'16)**		
Apps Filed	392,711	388,646	302,360	32,031	5,649,154 | 5,314,763		

*Note: incomplete application data for 2017 and 2018.

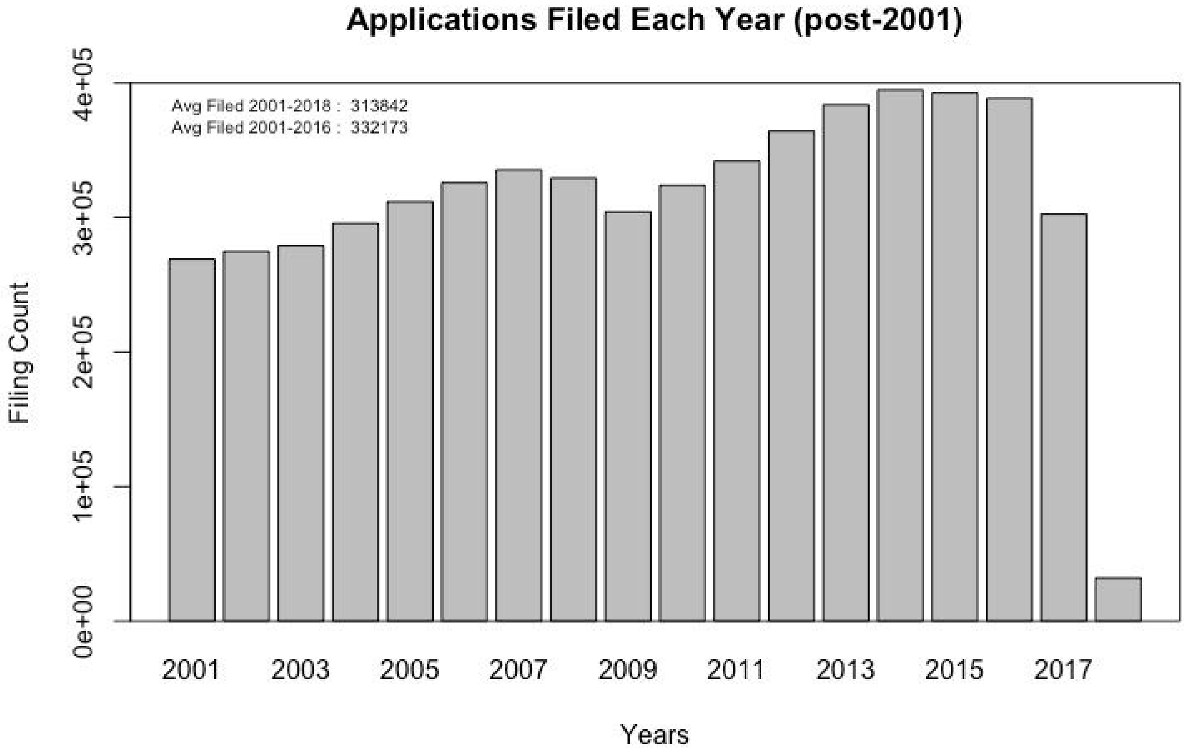

Fig. 6: *Bar plot of total application filing count for each year from 2001 to 2018.*

In order to determine the probability of issuance or abandonment over time, the filing dates are shifted as if they were all filed the same day. This is done by determining the number of days from the date of filing to grant and abandonment. First by separating issued and abandoned applications based on the disposal type (referred to as event status throughout), which is either "ISS," "ABN," and "PEND." For disposal type of "ISS," the time until issue is the difference between the filing date and patent issue date. For disposal type of "ABN," the time until abandon is the difference between the filing date and abandon date. For disposal type of "PEND," the time pending is simply the difference between the filing date and the current date, since as expected there is no disposal date provided. However, most pending applications are from 2014 or later. Over 90% of applications have either issued or been abandoned after 6 years.

Once the applications are separated by disposal type, the percent of applications by disposal type over the total applications can be determined. These percentages are the initial probabilities that an application will eventually be issued or become abandoned.

Figure 7 shows the percentages of applications that will issue or become abandoned for each year. The average percentages of issued, abandoned, and pending applications is 67.29%, 31.64%, and 1.08% for 2001 to 2013. The number of still pending applications becomes non-negligible after 2013.

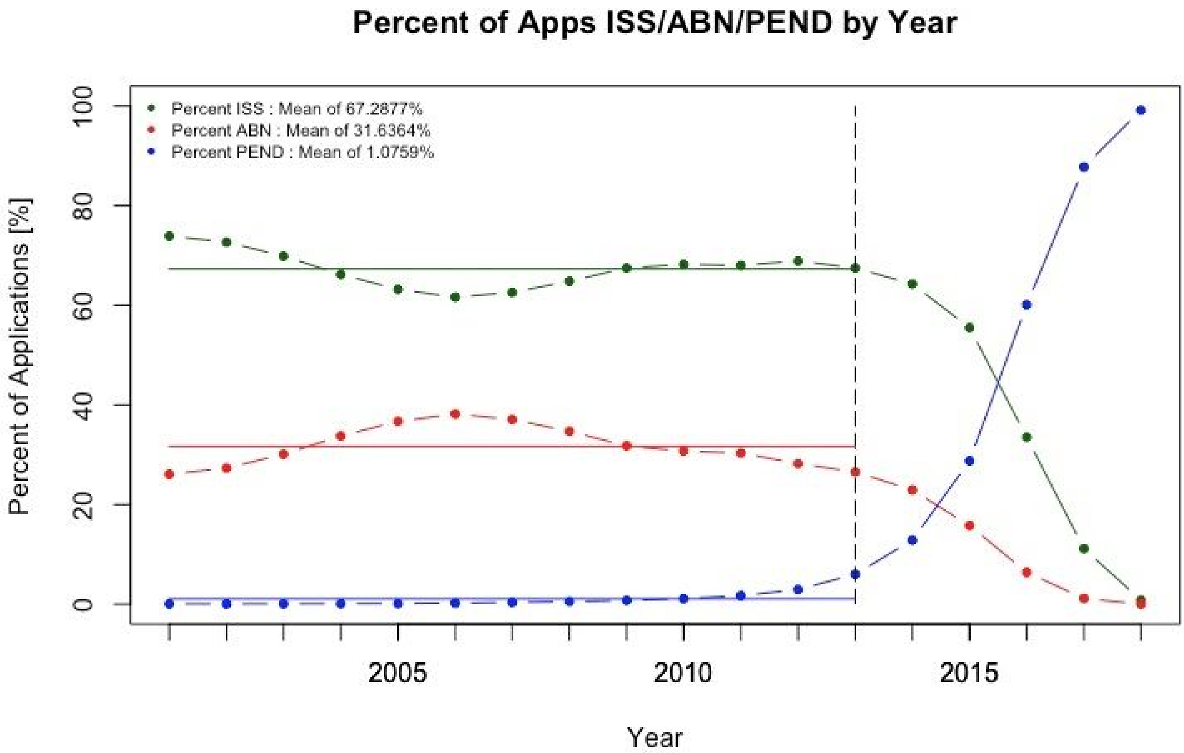

Fig. 7: *Percent of applications that eventually issue, abandon, or are still pending, for each year from 2001 to 2018.*

Figure 8 shows the average times until issue and abandonment for each year. The peak average time to issue is 1344 days in both 2006 and 2007. The peak average time until abandonment is 1310 days in both 2004 and 2005. As can be seen from the plot after peaking the time to issue and abandonment begins decreasing year after year. However, after 2013 the time to issue and abandon continuing to decrease is likely attributed to the application history being too new; since many applications are still pending from more recent years (see Figure 7 above). Although there is significant variation between issue time and abandonment time from 2001 to 2008, the average time to issue and the average time to abandon across 2001 to 2013 are nearly identical. The average time is 1160 days for issued applications and 1164 days for abandoned applications, for 2001 to 2013. In addition, the average time for both abandoned and issued applications follow the same general trend across the years.

Figure 9 shows histograms for issued and abandoned applications over time for each year, using the time to issue and time for abandonment of issued and abandoned applications, respectively. As can be seen from the histograms, both the mean abandonment time and issue time increases slightly each year after 2001. The mean begins to decrease again after 2012, but this is primarily due to a lack of an extended history after that year. There simply has not been enough time for applications filed after 2012. Each of these distributions generally follow an approximate gamma distribution, especially for 2001 to 2013.

The histograms of issued, abandoned, and pending applications for the combined years 2001 to 2018 are shown in Figure 10 below. Figure 11 similarly shows a histogram of issued, abandoned, and pending applications for the combined years 2001 to 2013.

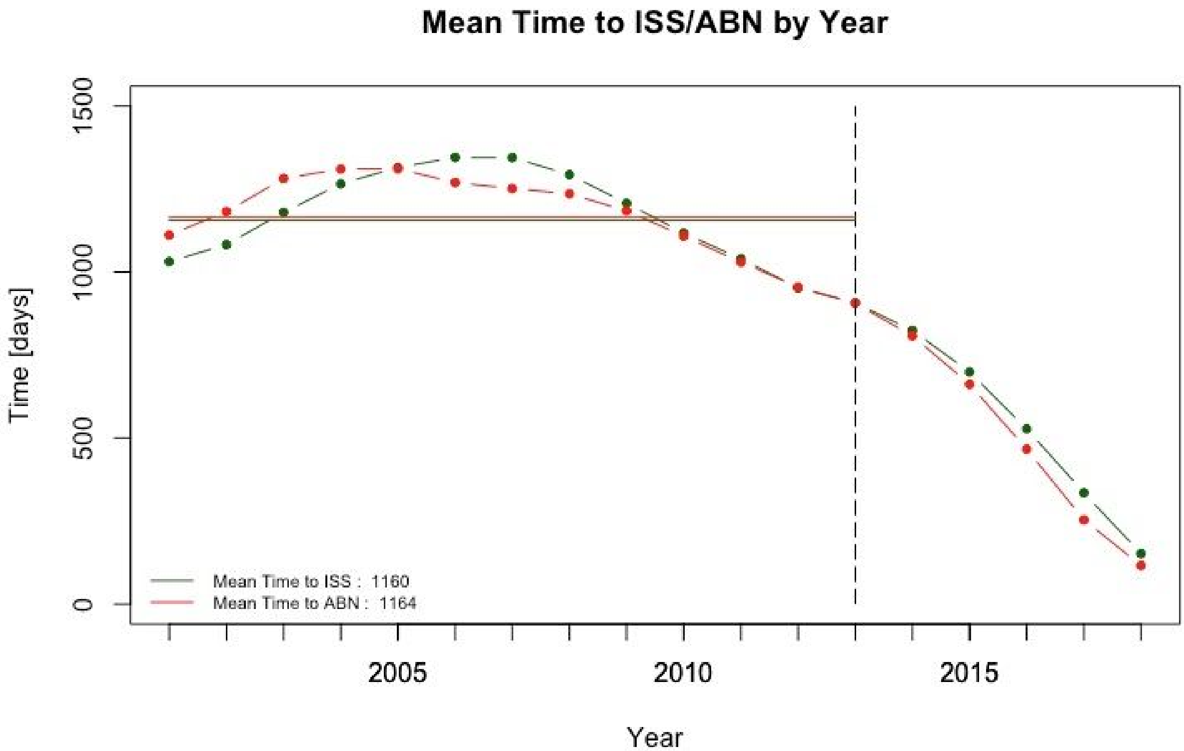

Fig. 8: *Average time until issue or abandonment for each year from 2001 to 2018.*

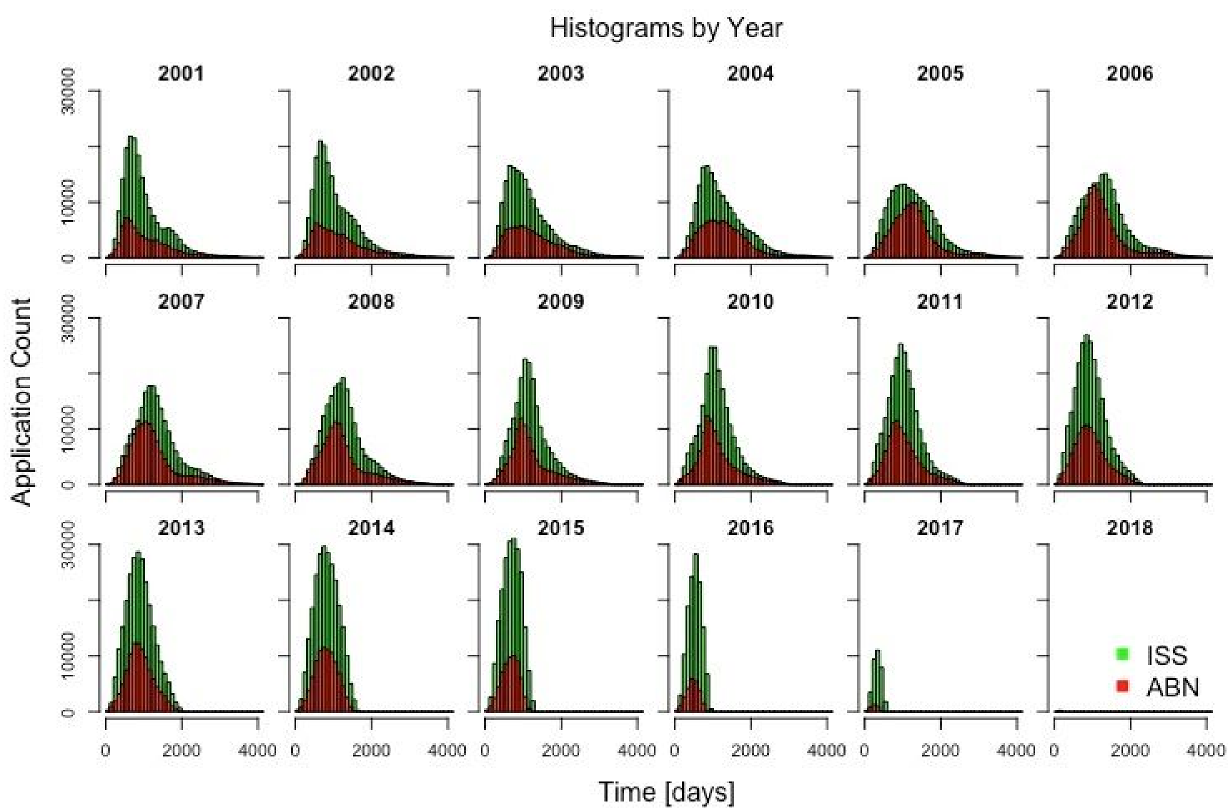

Fig. 9: *Histograms for issued abandoned applications for each year 2001 to 2018.*

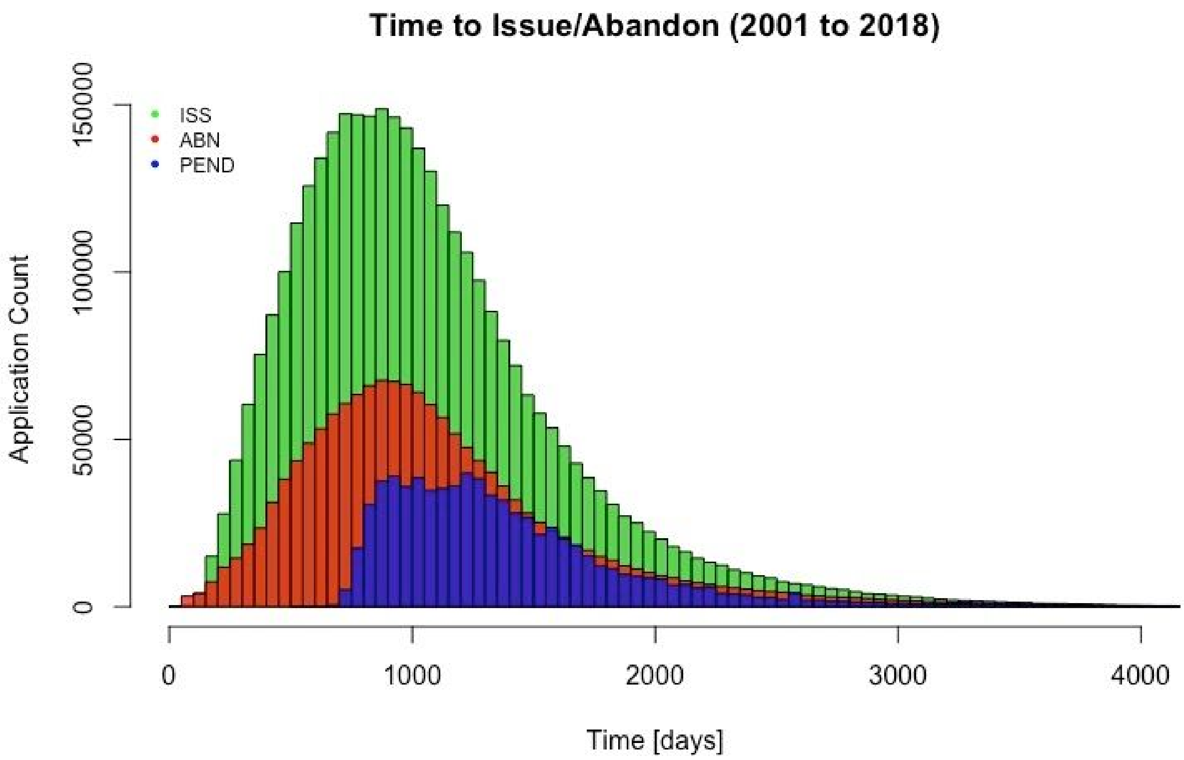

Fig. 10: *Histograms for issued and abandoned applications for all years 2001 to 2018.*

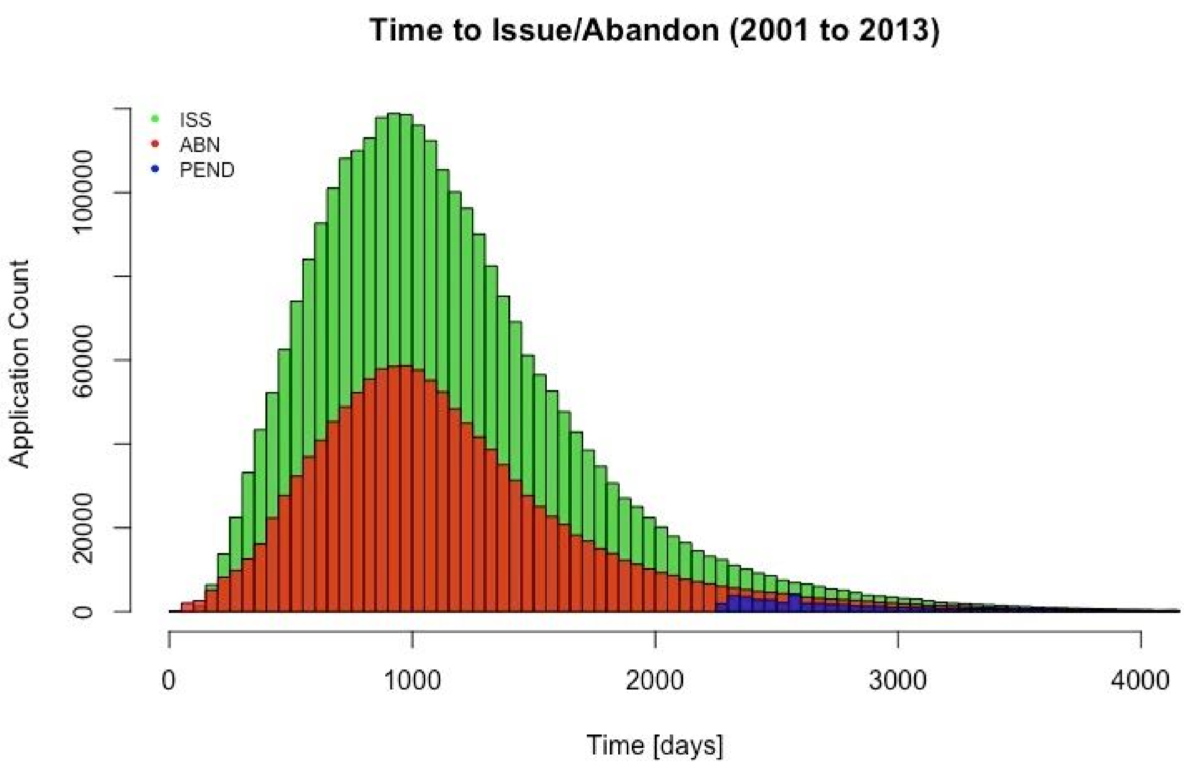

Fig. 11: *Histograms for issued and abandoned applications for all years 2001 to 2013.*

7.3 Survival Functions

A survival function can be obtained for both issued and abandoned applications from the large number of observations of both issued and abandoned applications. The survival function, or complementary cumulative distribution function, $S(t)$, is defined as the probability that an event of interest has not occurred by time, t, see equation (1) [9].

$$S(t) = P(\{T > t\}) \quad (1)$$

For example, the probability that a patient remains alive after a procedure up to time, t. The survival function for issued applications is defined as the probability that an application has not been issued by time, t; and the survival function for abandoned applications is the probability that an application has not been abandoned by time, t .

Specifically, the survival functions were obtained by counting the number of issued or abandoned applications that occur every 30 days from 0 to 5000 days, each 30 day count being indexed. A list of the increasing cumulative sum is obtained for issued and abandoned applications from 0 to 5000 days, each element is subtracted from 1, and then divided by the total issued and abandoned applications to obtain the surviving (i.e., still pending) percentage of eventually issued and abandoned applications for each 30 day bin.

Figure 12 shows a grant rate timeline in the form of survival function curves for each individual year from 2001 to 2018 and the survival function curves for the years cumulatively. However, due to the truncation of years from 2014 to 2018 they lack a complete enough history (e.g., 3000+ days). Accordingly, removing these years results in a significant change as shown in Figure 13 below to the cumulative years survival function, which includes the individual years from 2001 to 2013 and the cumulative of 2001 to 2013. The survival curves for all applications filed from 2001 to 2018 plateaus at 39.6% of applications not issuing (26.4% abandoned and 13.5% still pending), and 60.4% of applications eventually having issued. The survival curves for all applications filed from 2001 to 2013 plateaus at 32.9% of applications not issuing (31.7% abandoned and 1.2% still pending), and 67.1% of applications eventually issuing.

In order to obtain an accurate survival curve it appears only the years from 2001 to 2013 are included. However, it is also possible that the survival curves are truncated to essentially remove the horizontal portion. This truncation would be up to the time that all applications could possibly be pending. For example, for 2018 only about 2 years of the curve would be used for obtaining the cumulative survival curve. This also requires weighting the curves based on the number of applications filed in each year, but would include all applications excluding those still pending to make the survival curve more accurate. Simply excluding pending applications would also result in over weighting of applications which are issued or abandoned earlier in the timeline. A possible solution to this is right censoring of pending applications for both the issued and abandoned survival curves as discussed in more detail later.

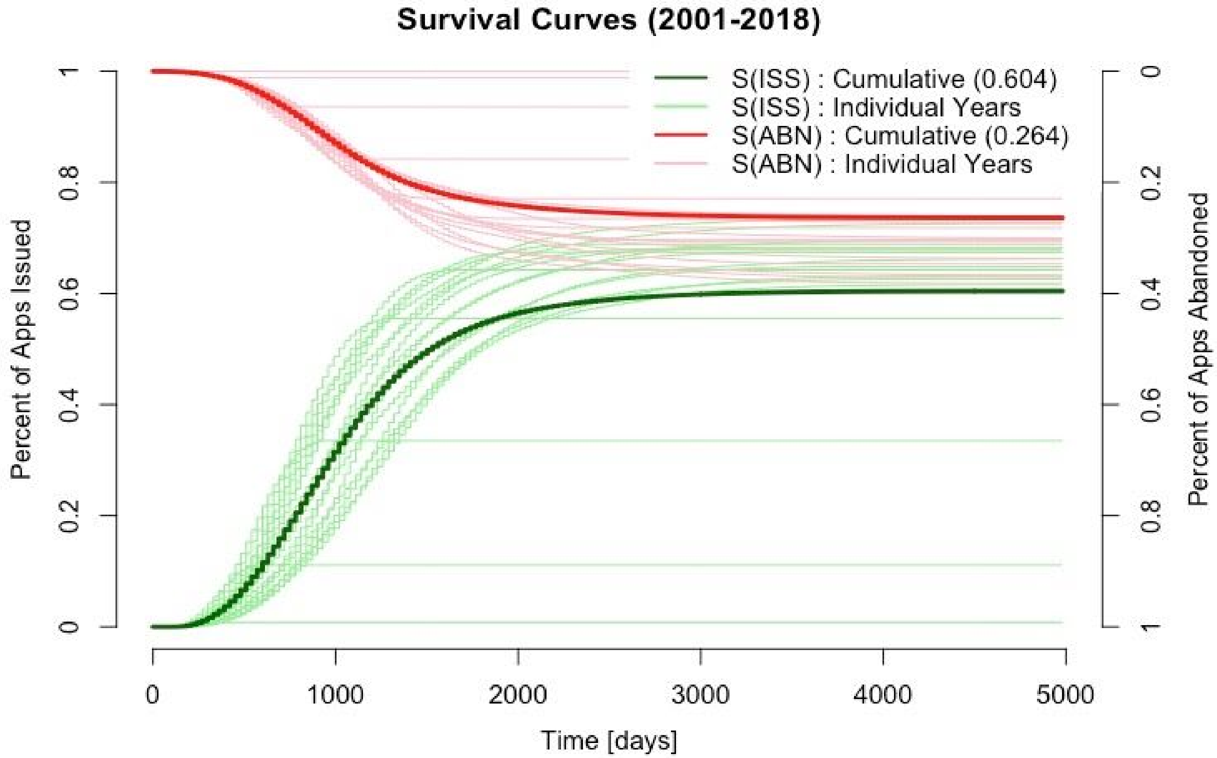

Fig. 12: *Survival curves for issued and abandoned applications for applications filed from 2001 to 2018.*

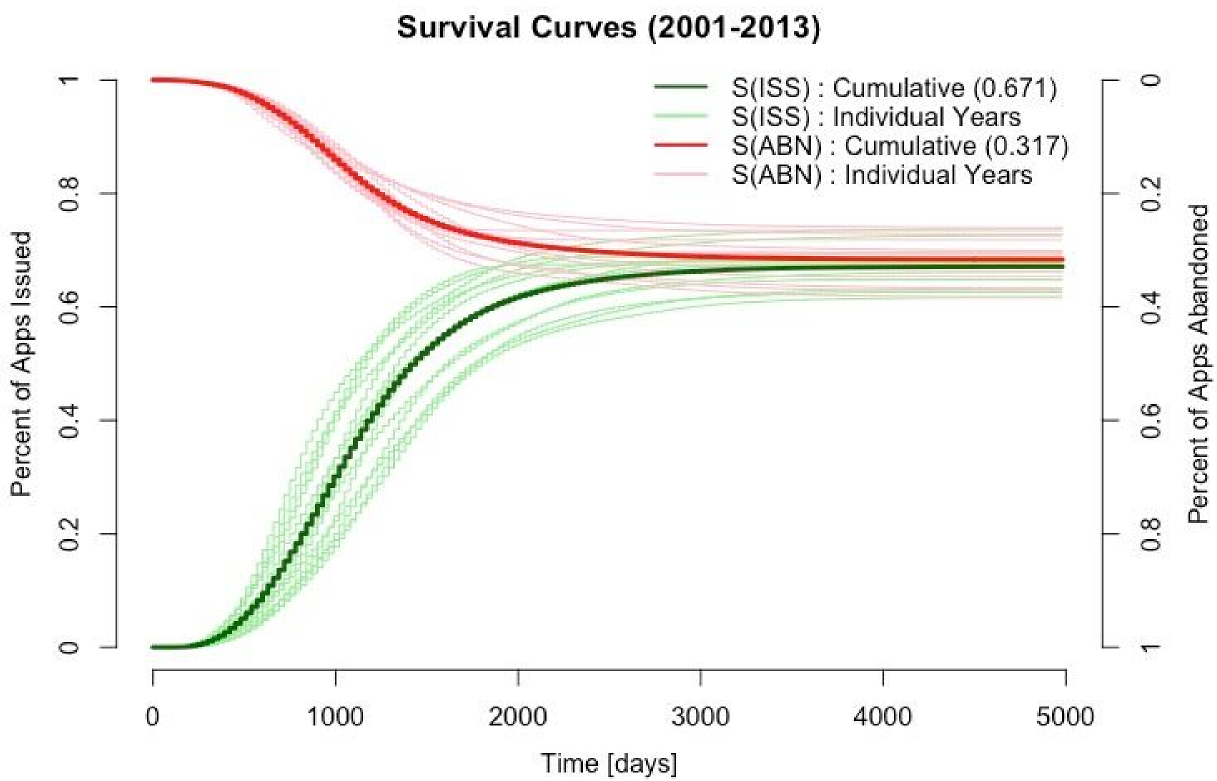

Fig. 13: *Survival curves for issued and abandoned applications for applications filed from 2001 to 2013.*

The complement to the survival function is the cumulative distribution function (CDF), $F(t)$. In our case, the CDF for an issued application is simply defined as the probability that an application has been issued by time, t; and the CDF for abandoned applications is the probability that an application has been abandoned by time, t. The survival function is defined in terms of CDF and can be obtained through integration of the probability density function (PDF), $f(x)$, as shown in equation (2) [9].

$$S(t) = P(\{T > t\}) = 1 - F(t) = \int_t^{\infty} f(x)dx \quad (2)$$

The CDF curves for the cumulative of applications from 2001 to 2018 and 2001 to 2013, can be obtained by subtracting the each value in the curve from 1. Simply by rearranging the above equation to solve for $F(t)$, as shown in equation (3).

$$F(t) = 1 - S(t) = P(\{T < t\}) \quad (3)$$

The survival function, $S(t)$, curve and the cumulative distribution function, $F(t)$, curve for issued applications filed from 2001 to 2018 are shown in Figure 14 below. Similarly, Figure 15 also shows the survival function curve and CDF curve for abandoned applications filed from 2001 to 2013.

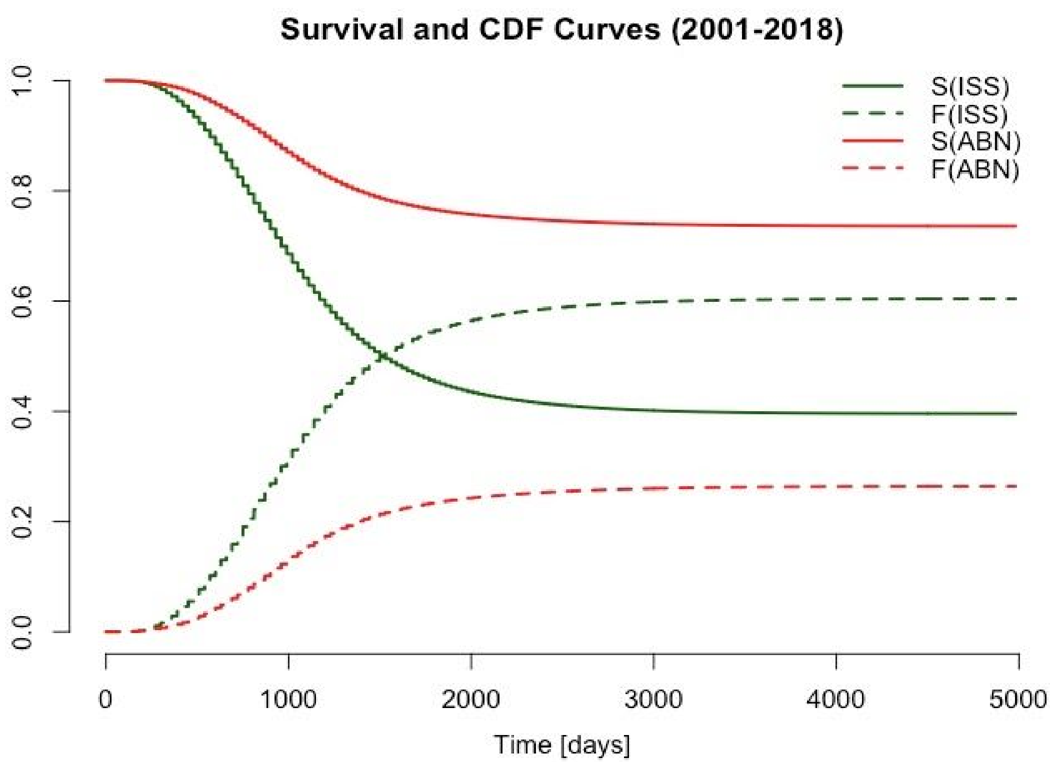

Fig. 14: *Survival function and CDF curves for issued and abandoned applications from 2001 to 2018.*

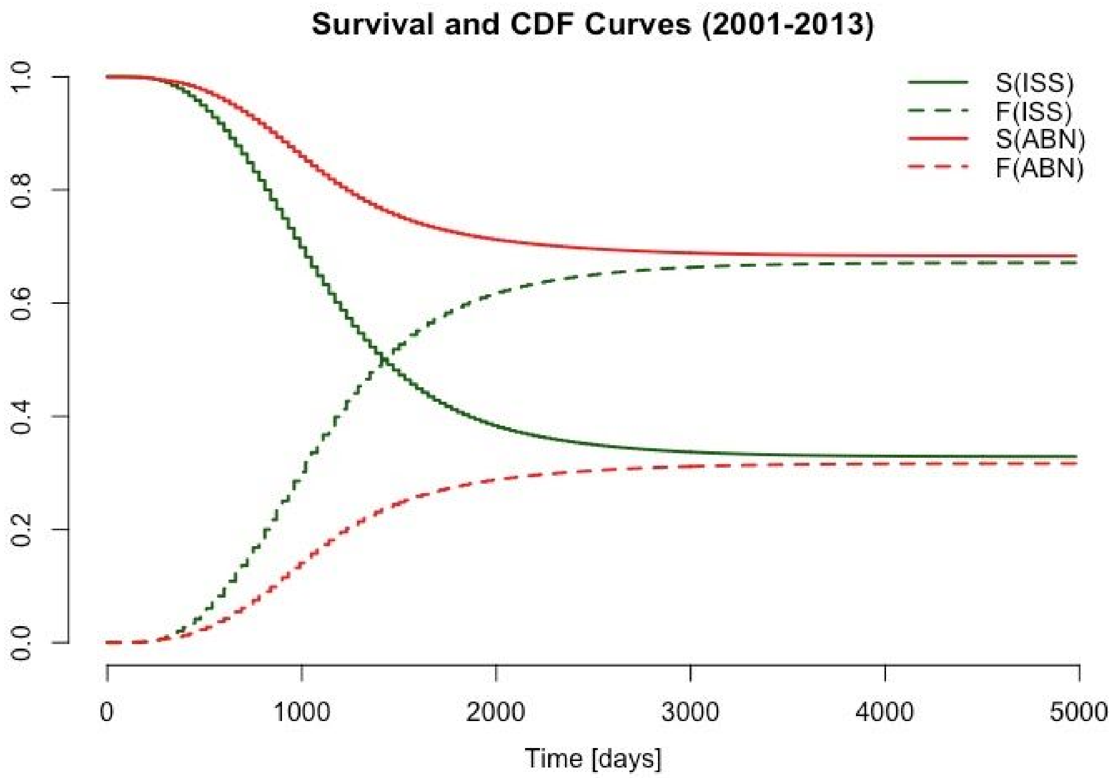

Fig. 15: *Survival function and CDF curves for issued and abandoned applications from 2001 to 2013.*

In order to obtain the PDF, $f(t)$, the derivative is taken of the survival function from equation (3) above, to obtain equation (4) .

(4) $$f(t) \ = \ \frac{d}{dt}S(t) = -\frac{d}{dt}(1 - F(t)) = \frac{d}{dt}\left(\int_t^{\infty} f(x)dx\right)$$

This is only true if the survival curve and CDF has support of 0 to 1. This ensures that PDF integrates to 1, which is required by definition [9]. As shown and discussed above, the event that an application issues is not certain to occur; an application has two possible outcomes of being issued or abandoned, or may remain pending for a very long time. Accordingly, $S(\infty)=0$ is not true for the above survival curves.

One way to deal with this problem is to obtain survival curves for issued and abandoned applications independently, thus guaranteeing $S(\infty)=0$. This results in survival curves that go from 1 to 0, which can then be used to obtain PDFs that integrate to 1. Then Bayes theorem can be used to account for other additional information, as discussed in more detail below.

7.4 Bayesian Statistical Model

In order to improve the survival function based on different factors defined for a filed application, individual survival curves need to be obtained for each factor. As an example case the following application factors are used, Art Unit 2611, Examiner ID 57219, USPC Class 375, USPC Subclass 257000, and Customer Number 74365.

Figure 16 shows in the upper right most plot a histogram obtained by sorting by event status including issued, abandoned, and still pending. Counting for each event status the number of applications that occur for each 30 day time interval from the time of filing (i.e., day 0) to 5000 days. The remaining plots are histograms obtained by pulling out just the applications for each factor and sorting by event status.

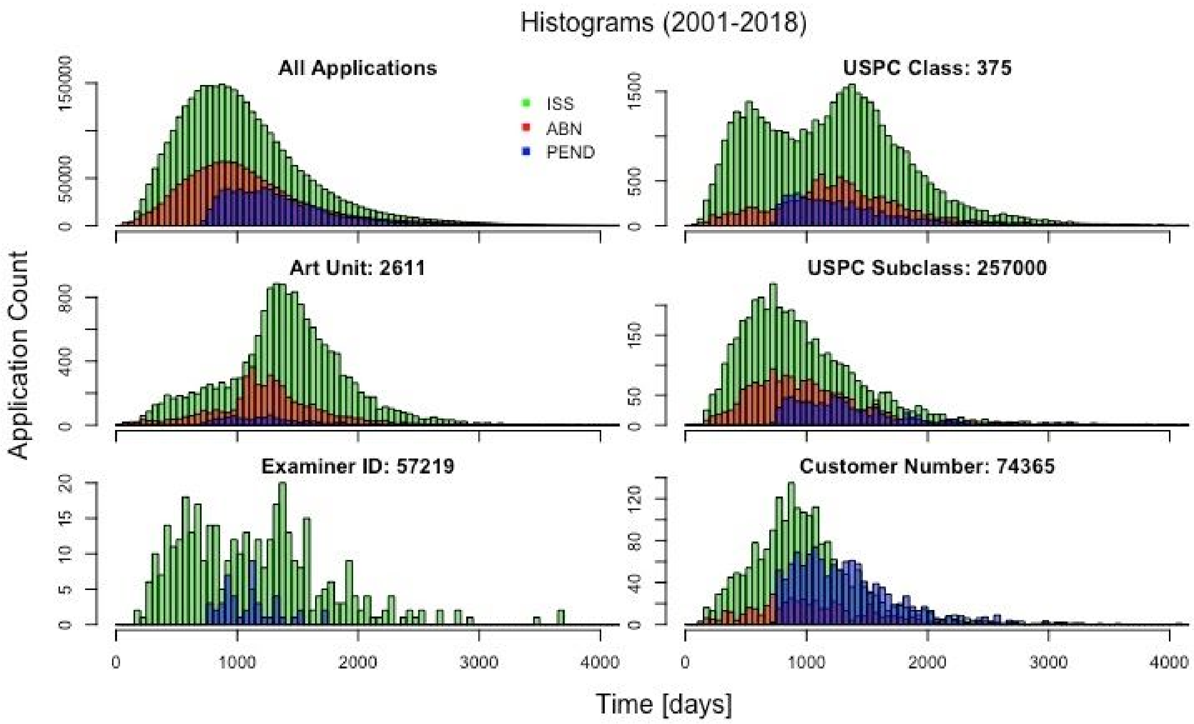

Fig. 16: *Histograms of issued, abandoned, and pending applications for all applications and applications with various factors.*

The survival curves are obtained numerically first for all applications, then for each factor. These are obtained by simply counting the number of applications that have issued, abandoned, or are still pending before each 30 day time interval inclusively, $d_{event}[i]$, from the time of filing (i.e., day 0) to 5000 days. Then, dividing the count of each interval by the total number of applications for the factor and status, N_{status}, and subtracting this value from 1; as shown in equation (5) below. Note *i* indicates is the index of the time interval.

(5)
$$S_{event}[i] = 1 - \frac{d_{event}[i]}{N_{event}}$$

Figure 17 shows, in the upper right most plot, the general survival curves for all applications for each of the possible event statuses of issued, abandoned, and still pending (i.e., $S(ISS)$, $S(ABN)$, and $S(PEND)$). The remaining plots show the conditional survival curves for each specified factor given each of the possible status events of issued, abandoned, and still pending (i.e. $S(\theta|ISS)$, $S(\theta|ABN)$, and $S(\theta|PEND)$). The survival curves for each factor regardless of status event, $S(\theta)$, is also shown in each plot. This survival curve basically is a survival curve for a specific factor where an event is issuance, abandonment, or still pending.

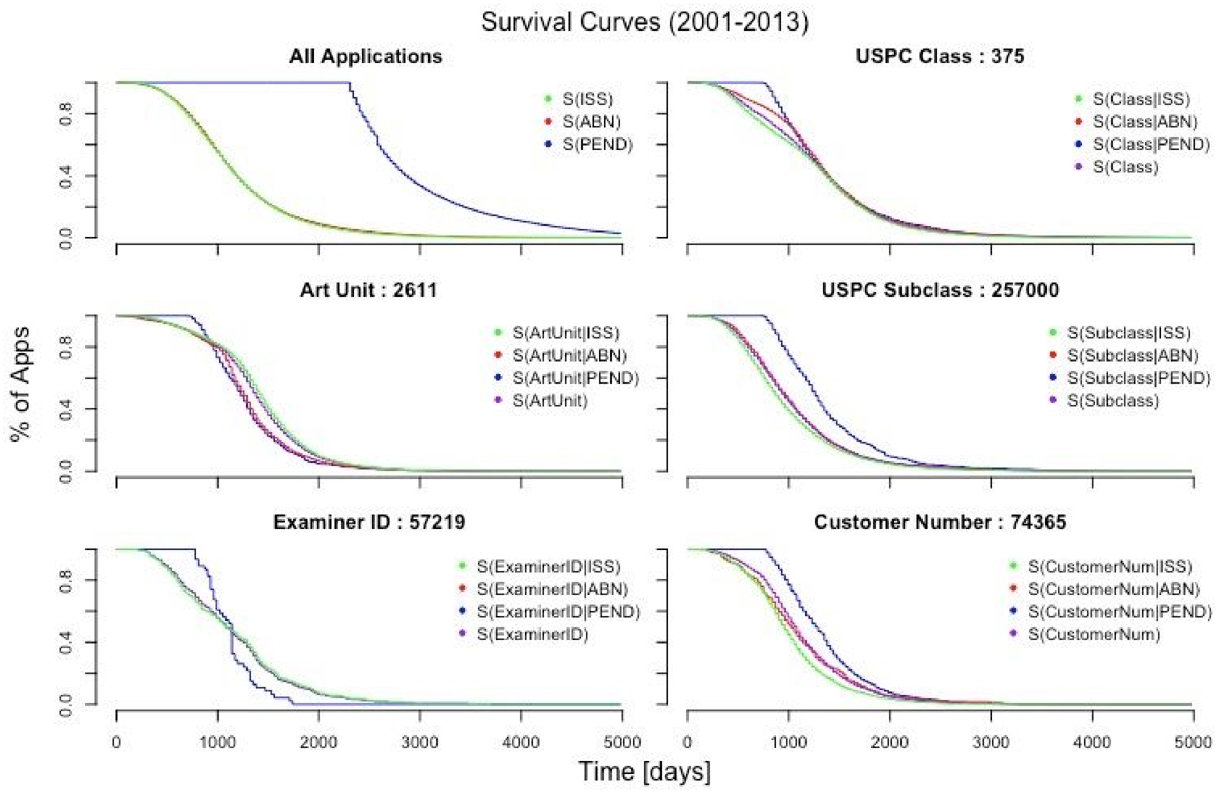

Fig. 17: *Survival curves for each factor for issued, abandoned, and pending applications.*

The PDF corresponding to each survival curve from the above are obtained through numerical differentiation of each survival curve. These PDFs are the conditional probabilities, or likelihoods, of observing an application with a specified factor, θ, given each possible event status is true, $P(\theta|ISS)$, $P(\theta|ABN)$, and $P(\theta|PEND)$. The evidential probability of observing an application with any of the possible event status (i.e., ISS, ABN, or PEND) with a specific factor is the PDF, $P(\theta)$, which is obtained from the corresponding survival curve. The quotient $P(\theta|ISS)/P(\theta)$ is the support that the specified factor, θ, provides the event status. Bayes's theorem can be used to obtain the posterior PDF, which is the probability that an application will issue or become abandoned given a specified factor, as shown in the equation (6) [10].

(6) $$P(ISS|\theta) = \frac{P(\theta|ISS)P(ISS)}{P(\theta)}$$

Note that ISS may be replaced with any event status. The priors $P(ISS)$, $P(ABN)$, and $P(PEND)$ are obtained from the survival curves for all applications through differentiation. $P(\theta)$ can also be expanded and rewritten as the following equation (7.1).

(6.1) $$P(\theta) = P(\theta|ISS)P(ISS)+P(\theta|ABN)P(ABN)+P(\theta|PEND)P(PEND)$$

Figures 18 and 19 show the probability density functions needed to compute the probability of issuance and abandonment given a specified factor is true, $P(ISS|\theta)$ and $P(ABN|\theta)$, are shown for each factor in below.

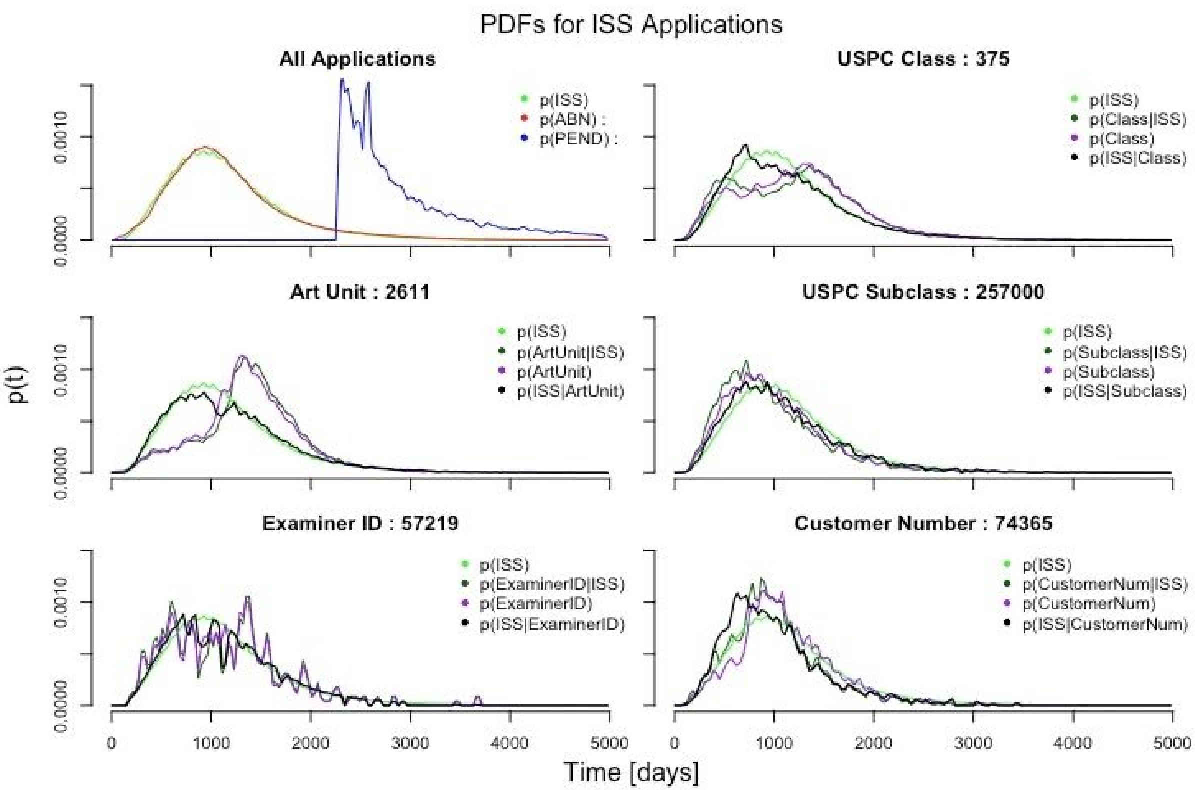

Fig. 18: Plots of *PDF curves for calculation of the posteriors,* $P(ISS|\theta)$ *, for each factor.*

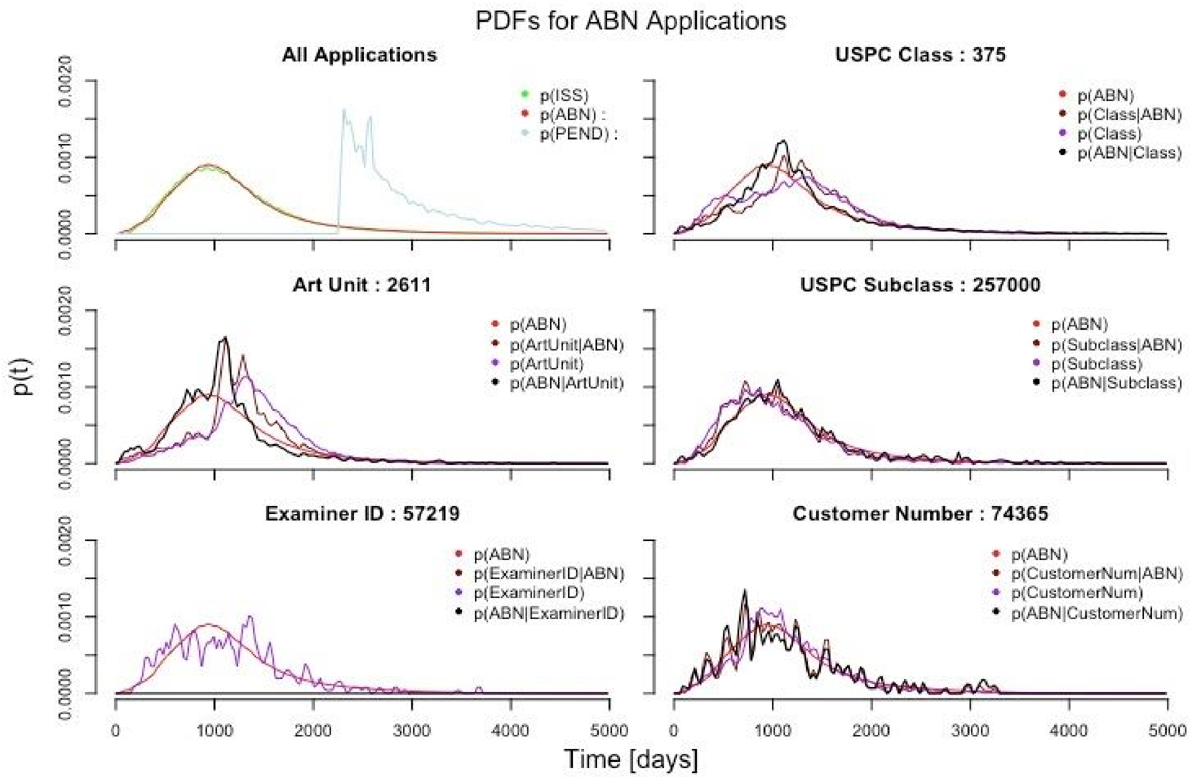

Fig. 19: Plots of *PDF curves for calculation of the posteriors,* $p(ABN|\theta)$*, for each factor.*

Figures 20 and 21 show the computed PDFs for issued and abandoned event statuses, respectively, for each of the factors above, $P(Event|\theta)$.

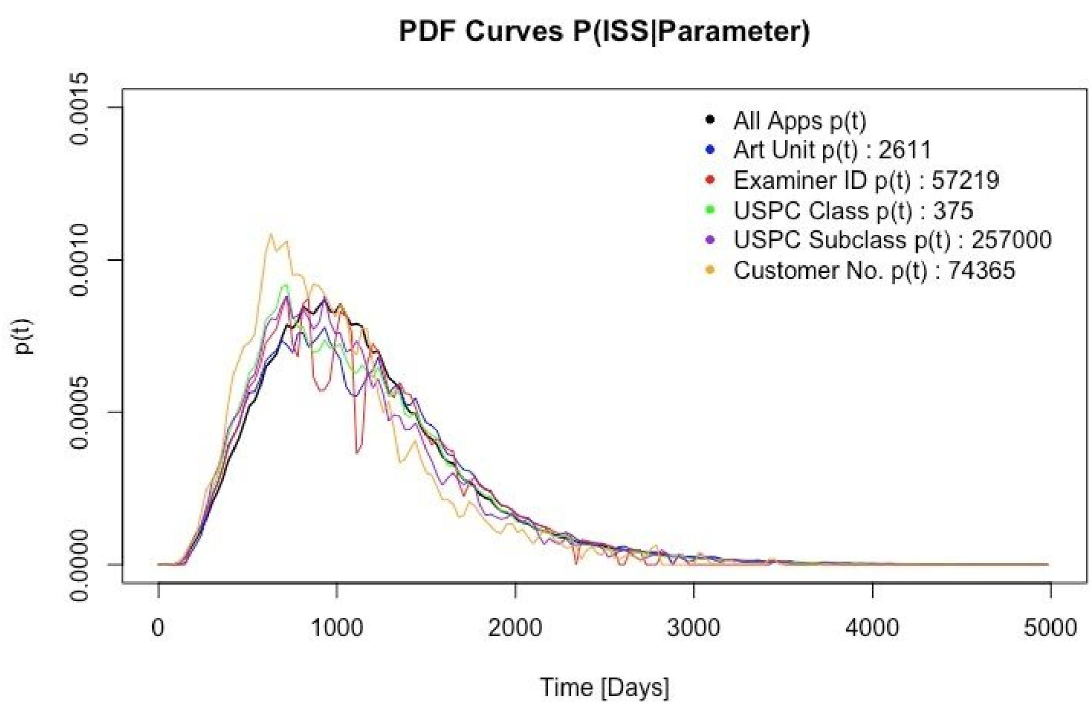

Fig. 20: Plot of *PDF curves for probability of being issued given each factor.*

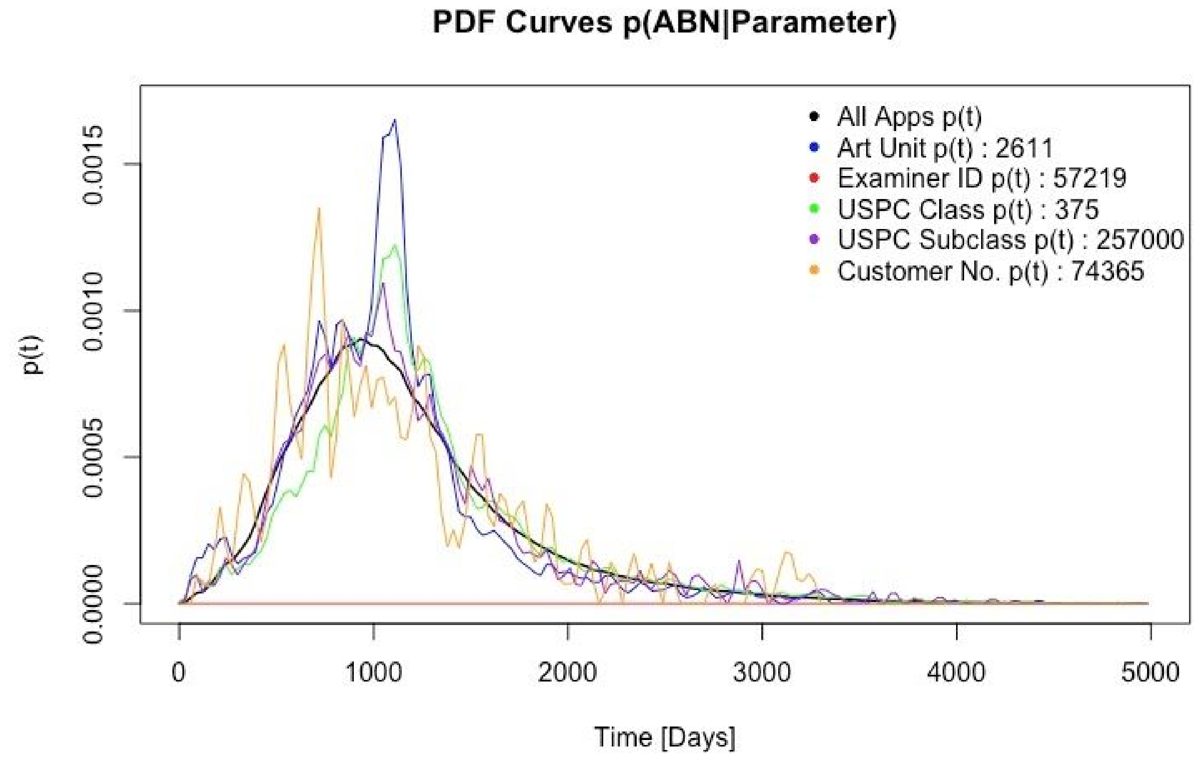

Fig. 21: Plot of *PDF curves for probability of being abandoned given each factor.*

Each of these probability density functions are numerically integrated to ensure that each of them integrates to approximately 1. The error is within 10%; the largest being for $P(ABN|ArtUnit)$.

Each of these PDFs appears to generally follow an approximate gamma distribution, see the equation (8) below, where k is a shape parameter, and θ is a scale factor, where $\Gamma(k)$ is the gamma function.

(7) $$f(t) = \frac{1}{\Gamma(k)\theta^k} t^{k-1} e^{-\frac{t}{\theta}}$$

The mean and variance for the gamma distribution is $\mu = k\theta$ and $\sigma^2 = 1/k$.

A gamma distribution can be fitted using regression to both the issued and abandoned PDFs for each factor as shown in Figures 22 and 23 below. The shape and scale parameter determined for each fitted gamma distribution is included in each figure.

The parameters are determined using Maximum likelihood estimation. The likelihood function for N independent identically distributed observations $(t_1, ..., t_N)$ in equation (8).

(8) $$L(k, \theta) = \prod_{i=1}^{N} f(t_i; k, \theta)$$

The log likelihood function is obtained from equation (8) to obtain equation (9).

(9) $$LL(k,\theta) = (k-1)\sum_{i=1}^{N} ln(t_i) - \sum_{i=1}^{N} \frac{x_i}{\theta} - Nkln(\theta) - Nln(\Gamma(k))$$

Taking the derivative with respect to θ and setting equal to zero, then solving for θ yields the maximum likelihood estimator of the scale parameter, θ, see equation (10).

(10) $$\hat{\theta} = \frac{1}{kN}\sum_{i=1}^{N} t_i$$

Substituting equation (10) into equation (9).

(11) $$LL(k) = (k-1)\sum_{i=1}^{N} ln(t_i) - Nk - Nkln(\frac{\sum t_i}{kN} - Nln(\Gamma(k))$$

And, similarly taking the derivative with respect to *k* and setting equal to zero yields the following equation.

(12) $$ln(k) - \psi(k) = ln(\frac{1}{N}\sum_{i=1}^{N} t_i) - \frac{1}{N}\sum_{i=1}^{N} ln(t_i)$$

Where $\Psi(k)$ is the digamma function. Solving for k cannot be done with a closed-form, thus a numerical method such as Newton's method is required [11].

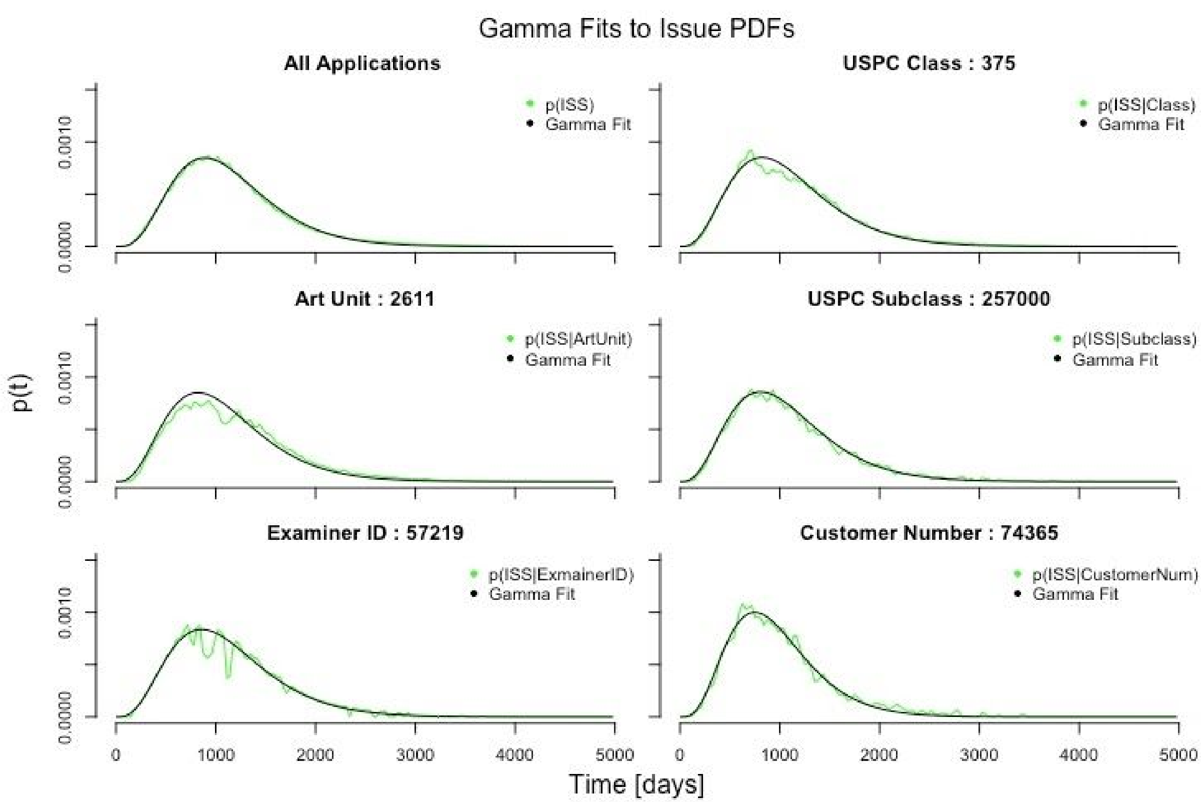

Fig. 22: Plots of fitted *gamma distributions to corresponding PDFs for issued applications given each factor.*

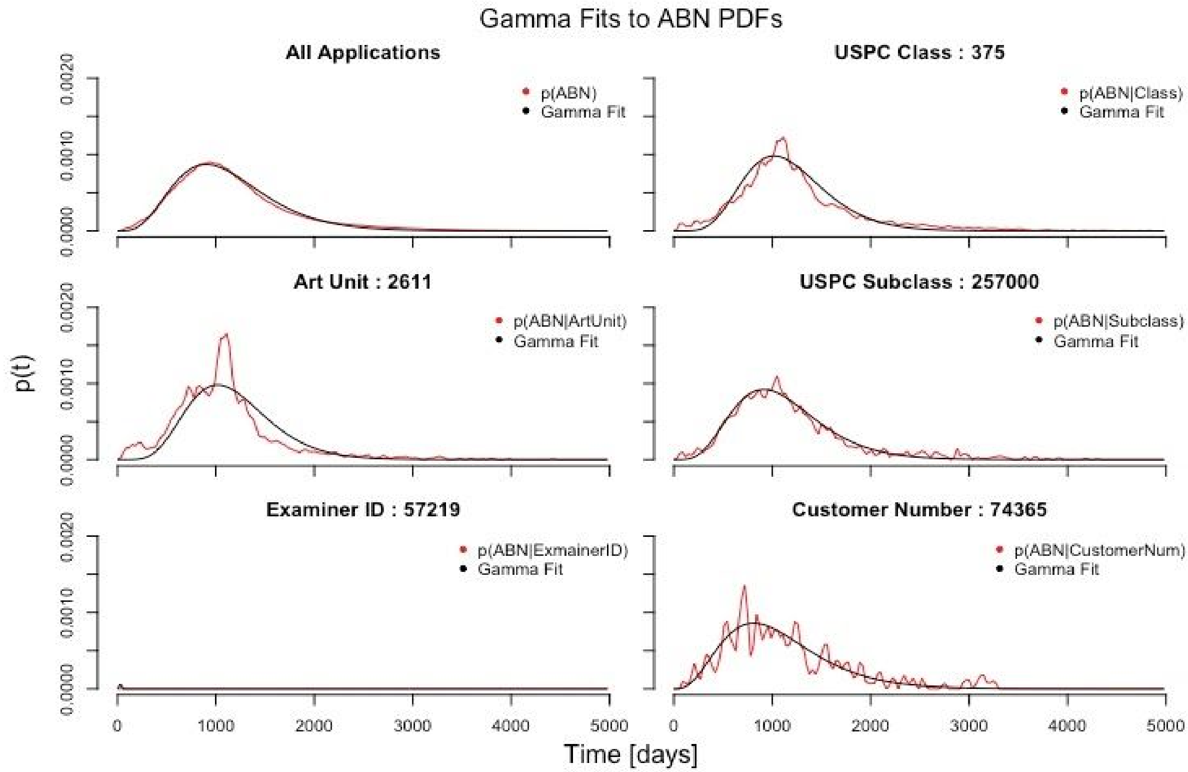

Fig. 23: Plots of fitted *gamma distribution to corresponding PDFs for abandoned applications given each factor.*

The individual PDFs for each factor can be combined. This is done using a regression method similar to what is discussed above to fit a gamma distribution, where the gamma distribution is the aggregated distribution for all factors, which is discussed in more detail below. The fit for issued and abandoned applications uses all of the factor specific distributions from Figures 22 and 23 above, respectively.

The distributions for specific factors can all be weighted equally or they can be weighted based on where they fall in the hierarchy. The hierarchy of the factors includes all applications (post-2001) at the top, underneath these are the factors of “Attorney/Firm,” “Small Entity Indicator,” “AIA/Pre-AIA,” “Customer Number,” “Art Unit,” “Examiner ID,” “USPC Class,” and “USPC Subclass.” Only the “Examiner ID” and the “USPC Subclass” are directly dependent on the “Art Unit” and the “USPC Class,” respectively. Although it is possible that an examiner may switch “Art Units” while working at the USPTO so this is not necessarily true, but likely. For the other factors they are not directly dependent. The Small Entity Indicator is simply a “0,” if not considered a small entity, or a “1” if considered a small entity. The AIA/Pre-AIA is similarly a “0,” if considered under AIA rules (filed after March 16, 2013), or a “1” if considered under Pre-AIA (filed before March 16, 2013).

There are no absolute direct dependencies for the above factors. However, “Technical Center” and “Group” may optionally also be included. The factor of “Art Unit” depends directly from “Group” which depends directly from “Technical Center.”

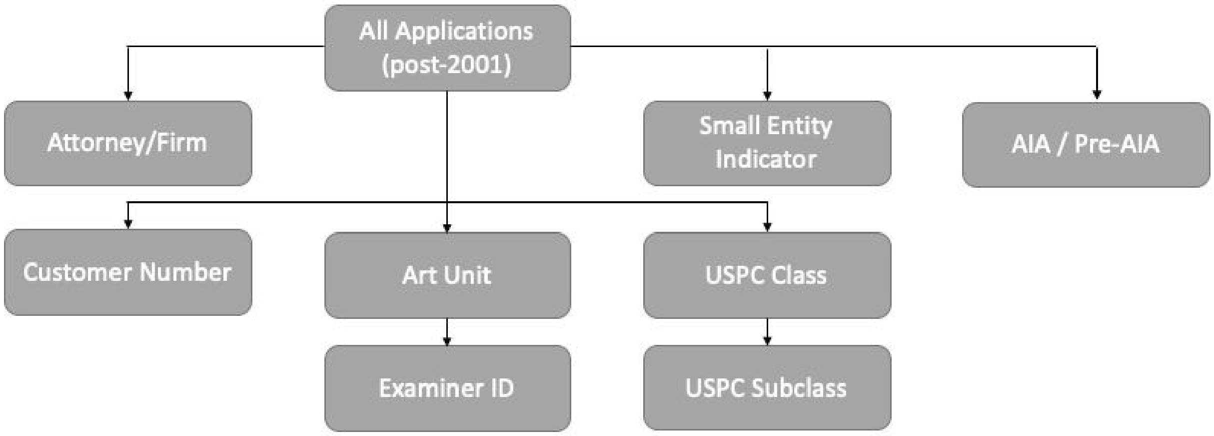

Fig. 24: *Hierarchy of application factors.*

Each of the factors also has a different number of applications associated with it. This can be based on the number of applications, or the scale of the histogram. The scale of "All Applications" is 100,000. The scale for "Art Unit" is 1,000. The scale for "Examiner ID" is 10-30. The scale for "USPC Class" is 1000-2000. The scale for "USPC Subclass" is around 100-200. The scale for "Customer Number" can vary widely depending on the size of the company or inventor, but for medium to large companies a scale of 50-100 is reasonable. Note other factors can also vary significantly, in particular "Examiner ID" can vary significantly depending on how long the examiner has been working at the USPTO.

The weights were assigned based on an inverse relationship with the scale. Accordingly, "All Applications" have the smallest weight, then "Art Unit," "USPC Class," "USPC Subclass," "Customer Number," and "Examiner ID." "Examiner ID" and "Customer ID" will be the closest to the actual application. The table below shows a list of possible weights. These weights initially do not add to one, but are multiplied by a scaling factor of 20.08, in order to sum to one.

Table 2: *Weights for each application factor.*

Weights	All Apps	Art Unit	Examiner ID	USPC Class	USPC Subclass	Customer Number
Unscaled	10^-5	0.001	0.0333	0.0005	0.005	0.01
Scaled	0.0002	0.02	0.666	0.01	0.1	0.2

Alternatively, the scale can be used to directly determine the weights for aggregating the individual PDFs based on each of the factors. This would allow the factoring in varying scales. This is useful when there is a new Examiner or Customer with a short history thus few applications.

The weights are used in weighted least squares, or weighted linear regression, for fitting an aggregated gamma distribution to the various factor-based PDF distributions. It should be noted for the example set of application factors that for the aggregated abandoned application PDF curve there are no abandoned applications for the specified Examiner ID. As such, the unscaled weight is set to zero, and the scaled weights adjusted accordingly by a different scaling factor.

The fit of the aggregated model is based on the residual, r_{ij}, which is defined as the difference between the actual value from each of the PDFs, $p_j(t_i)$, and the value predicted by the model, $f(t_i;\theta,k)$, which is the gamma function from equation (7), as shown in equation (13).

$$r_{ij} = p_j(t_i) - f(t_i;\theta,k) \tag{13}$$

The score function is the sum of the weighted squared residuals, shown in equation (14).

$$S = \sum_{j=1}^{n} \sum_{i=1}^{N} w_j r_{ij}^2 \tag{14}$$

Where *n* is the number of posterior PDFs being used for the aggregate model. The list *w* is the weights for the individual posterior PDFs, the list *w* is of length *n*. The score function is then minimized by an optimization technique to determine the optimal scale and shape parameters θ and k.

Figures 25 and 26 show plots of the weighted aggregated PDF from issued and abandoned PDFs for each factor. The mean and mode is 990 and 825 days, respectively, for the weighted aggregate of issued PDFs. The mean and mode is 1020 and 855 days, respectively, for the weighted aggregate PDF for abandoned. The mean is determined through numerical integration of the weighted aggregated PDF up to the time, in days, where 50% of the area is to the left and 50% of the area is to the right. This is where 50% of applications that will issue, have issued, and 50% of applications that will become abandoned, are abandoned. The mode is determined through numerical differentiation of the PDF curve and is the time where this curve peaks and the derivative is approximately zero.

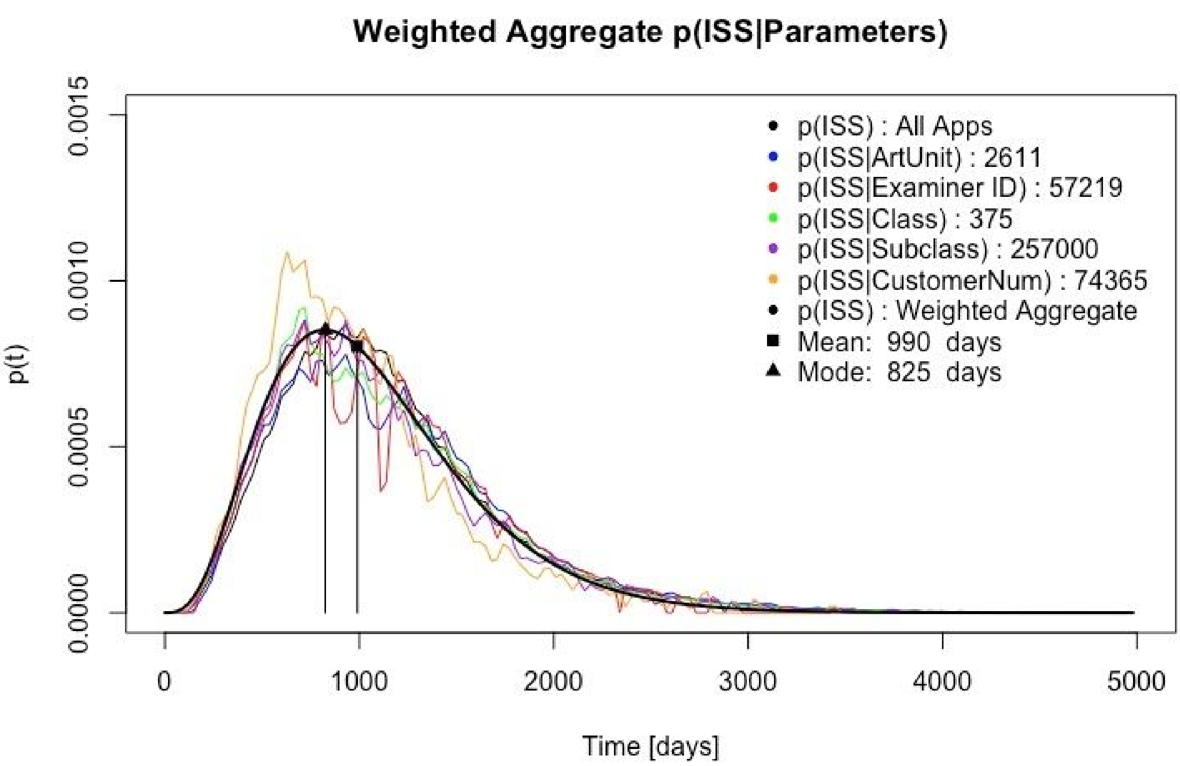

Fig. 25: *Aggregated issuance PDF curve using weighted regression of multiple factor specific PDF curves.*

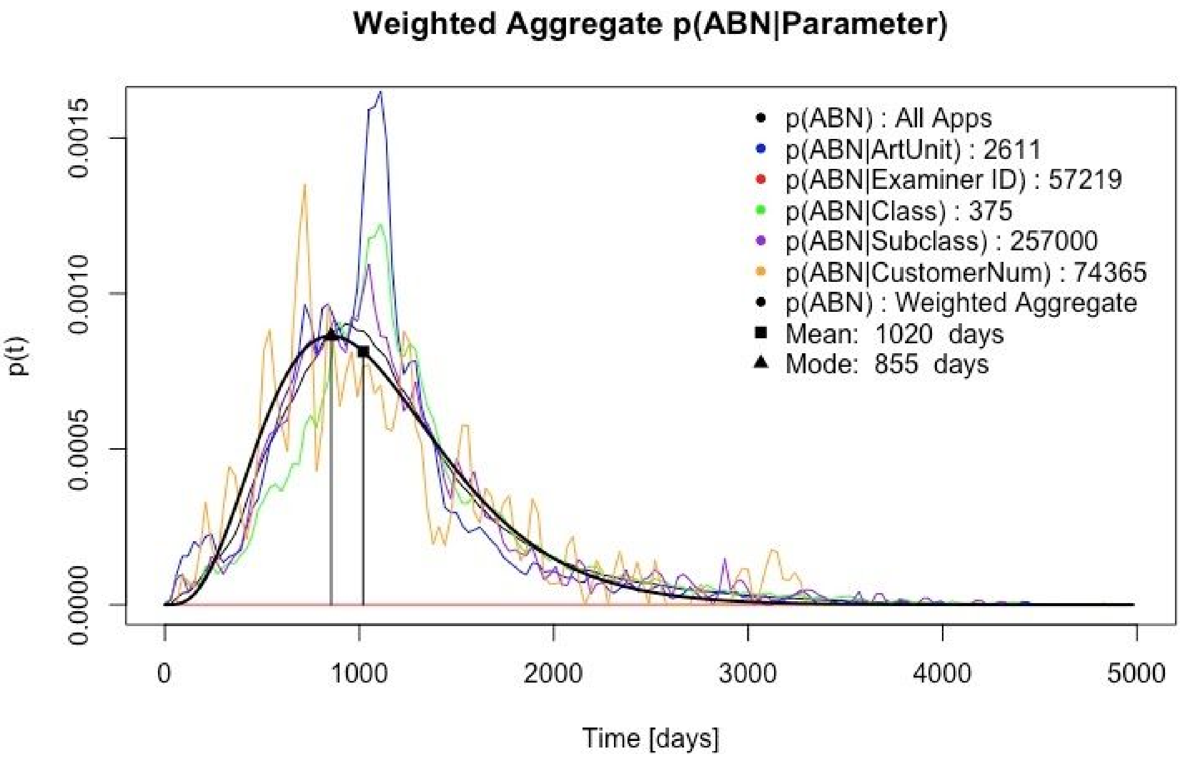

Fig. 26: Aggregated abandonment PDF curve using weighted regression of multiple factor specific PDF curves.

Figures 27 and 28 show plots of the unweighted aggregate PDFs from issued and abandoned PDFs for each factor. The mean and mode is 990 and 855 days, respectively, for the weighted aggregate of issued PDFs. The mean and mode is 1050 and 945 days, respectively, for the weighted aggregate of abandoned PDFs. As discussed above, the mean is determined through numerical integration of the weighted aggregated PDF up to the time, in days, where 50% of the area is to the left and 50% of the area is to the right. This is where 50% of applications that will issue, have issued, and 50% of applications that will become abandoned, are abandoned. The mode is determined through numerical differentiation of the PDF curve and is the time where this curve peaks and the derivative is approximately zero.

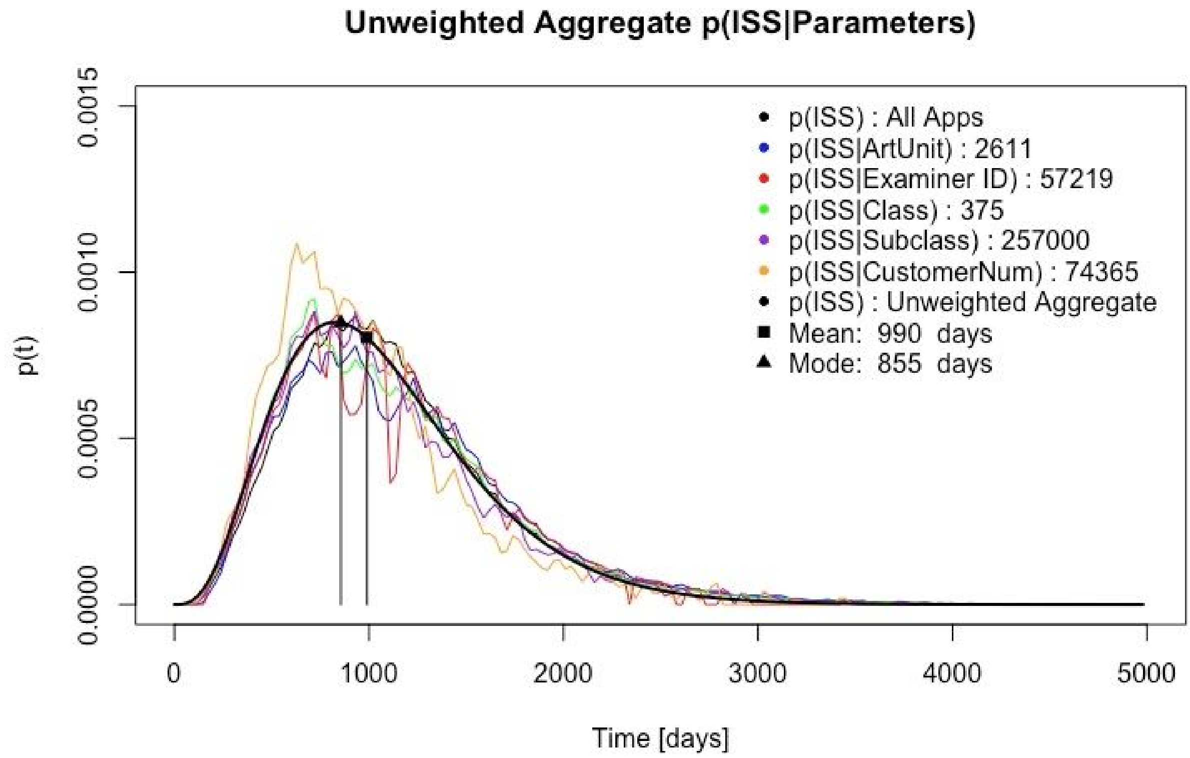

Fig. 27: *Aggregated Issuance PDF curve using unweighted regression of multiple factor specific PDF curves.*

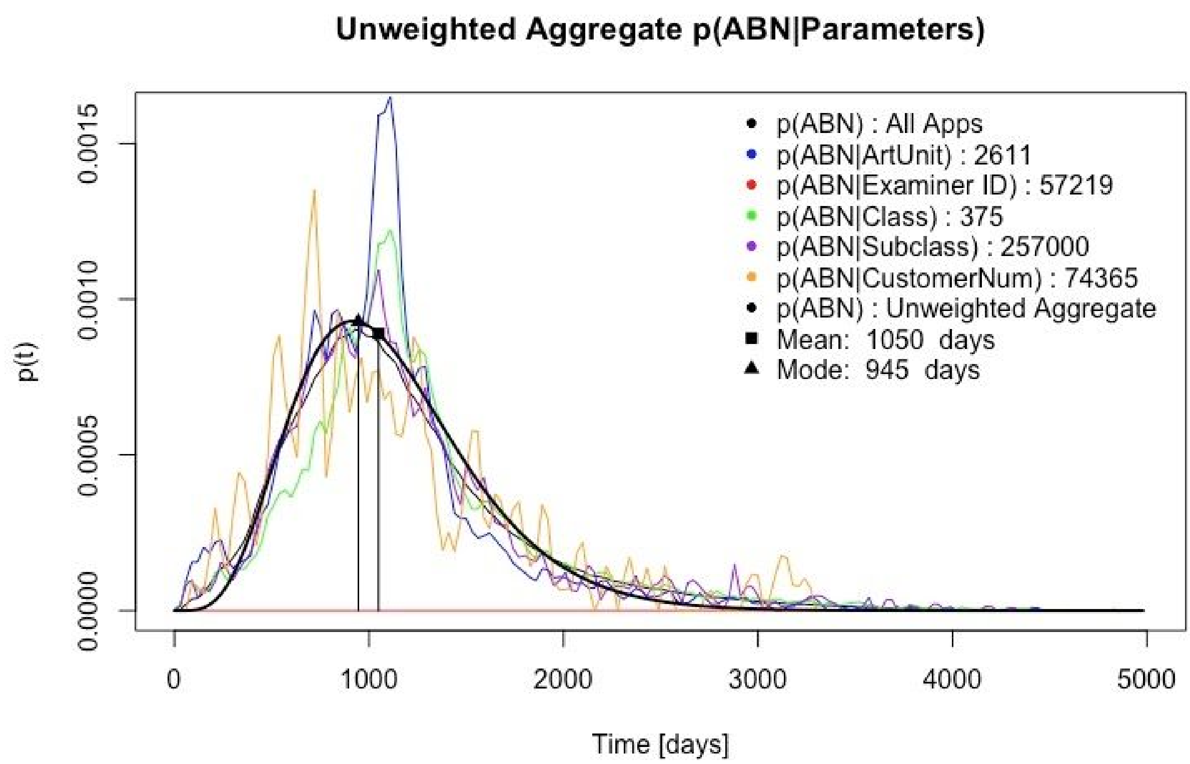

Fig. 28: *Aggregated Abandonment PDF curve using unweighted regression of multiple factor specific PDF curves.*

From the unweighted and weighted aggregated PDF curves for issued and abandoned applications the percentiles can be obtained. The area under the PDF curve can be determined through numerical integration between two time values. Integrating the issued PDF curve from the time of filing, time 0 days, forward under the curve and tracking the percentage of area under the curve, the time at which a desired percentage (e.g., 25%, 50%, 90% 99%) of applications that will issue, have issued can be determined. This can be similarly done using the abandonment PDF curve to determine the percentiles for applications that have abandoned from those that will eventually be abandoned.

Clearly, at this point issued and abandoned applications are separate. The PDFs for issued and abandoned applications integrate to unity. As mentioned above, for the issued PDF, integrating gives the percentages of applications that have been issued from those that will eventually issue. For the abandoned PDF curve, integrating gives the percentage of applications that have abandoned from those that will eventually be abandoned. However, these percentages and times for issued and abandoned applications need to reflect the percentage of the total applications filed including those that issue, become abandoned, and remain pending.

To do this, the probabilities that an application will eventually issue, become abandoned, or remain pending for any given specified factor are determined. These are determined through the following set of simple equation (15):

(15) $$p(Event|\theta) = N_{Event,\theta}/N_{\theta}$$

(15.1) $$p(ISS|\theta) = N_{ISS,\theta}/N_{\theta}$$

(15.2) $$p(ABN|\theta) = N_{ABN,\theta}/N_{\theta}$$

(15.3) $$p(PEND|\theta) = N_{PEND,\theta}/N_{\theta}$$

Where $N_{ISS,\theta}$ is the total number of applications that have an event status of being issued with the specified factor (e.g., Art Unit 2611), and N_{θ} is the total number of applications that meet the specified factor including all event statuses. Similarly, $N_{ABN,\theta}$ is the total number of applications that have an event status of being abandoned with the specified factor. Similarly, $N_{PEND,\theta}$ is the total number of applications that have an event status of still pending with the specified factor.

Table 3 shows the probability that an application will eventually be issued or become abandoned given specific factors including “Art Unit,” “Examiner ID,” “USPC Class,” “USPC Subclass,” and “Customer Number” which are computed using the above equations.

Table 3: *Probabilities that an application will eventually issue or become abandoned given example specified factors (2001-2013).*

Factor, θ	$p(ISS\|\theta)$	$p(ABN\|\theta)$	$p(PEND\|\theta)$
All Applications	0.6713	0.3166	0.0121
Art Unit: 2611	0.7790	0.2199	0.0011
Examiner ID: 57219	1.0000	0.0000	0.0000
USPC Class: 375	0.7654	0.2240	0.0105
USPC Subclass: 257000	0.6854	.3064	0.0118
Customer Number: 74365	0.7551	0.2140	0.0308
Unweighted Aggregate	0.7760	0.2134	0.0111
Weighted Aggregate	0.9089	0.0801	0.0071

Each of the probabilities from Table 3 above are multiplied by the corresponding weights from Table 2 above and then added together to obtain the weighted aggregated probability that the application will issue or become abandoned eventually. This approach to aggregation of probabilities is a linear opinion pool as defined by equation (10) [12].

(10) $$p(Event|\theta_i) = \sum_{i=1}^{n} w_i \, p_i(Event|\theta_i)$$

Where n is the number of factors, i, being aggregated, w_i is a weight corresponding to each factor where the weights sum to 1, $p_i(Event|\theta_i)$ represents factor i's probability, and $p\,(Event|\theta_i)$ is the combined or aggregated probability. Alternatively, a logarithmic opinion pool may be used.

The weighted aggregated probability that the application with the above specific factors will eventually issue is 0.9089, become abandoned is 0.0801, or remain pending is 0.0071. The unweighted aggregate probabilities that an application will issue or become abandoned eventually simply is an average of the issued probabilities and abandoned probabilities above. The unweighted aggregate probabilities that the application with the above specified factors will eventually issue is 0.7760 become abandoned is 0.2134 and remain pending is 0.0111.

The weighted aggregated probabilities for issuance and abandonment is multiplied by the area under the corresponding PDF curve to determine the percentage out of the total applications that will issue or become abandoned. This can then be converted back to a survival curve (or CDF) through numerical integration. These survival curves can then be used to determine the probability that an application will issue or become abandoned before or after a specified time. The probability of issuance or abandonment between two specified can similarly be determined.

Figure 29 shows grant rate timeline or the survival curves for issued and abandoned applications using the aggregated Bayesian model and the survival curves for issued and abandoned applications generally. As can be seen from the figure below, the aggregate issuance survival curve plateaus at 90.89% of applications; this is an increase of 24.18%. The aggregate abandonment survival curve plateaus at only 8.01% of applications; this is a decrease by -23.16%.

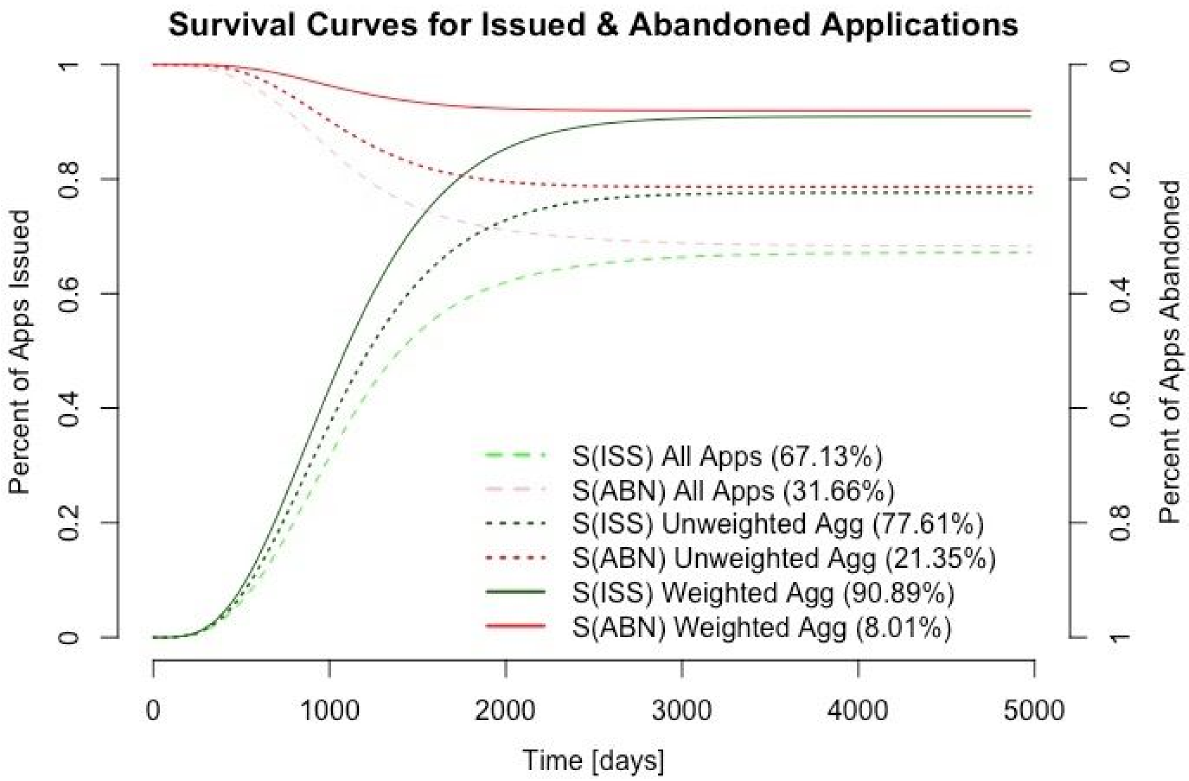

Fig. 29: *Survival curves for issued and abandoned applications for the unweighted and weighted aggregate models, and the general model.*

Figure 30 shows survival curves starting at the percentage of applications that will eventually issue or be abandoned (i.e., initial probability) for the aggregated model and general all application model, decreasing toward zero over time. This helps show the decrease in probability of issuance or abandonment over time.

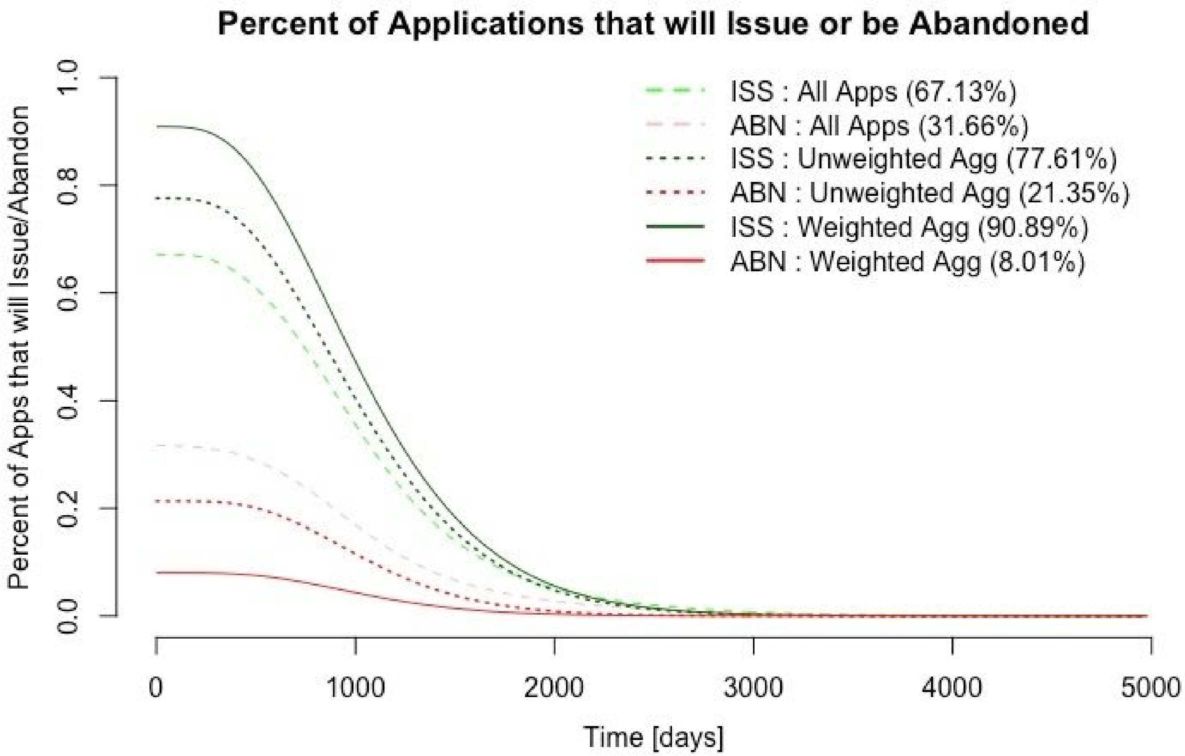

Fig. 30: *Percent of applications remaining to be issued or abandoned over time for unweighted and weighted aggregated models.*

7.5 Kaplan-Meier Survival Model

The Kaplan-Meier survival estimator is a non-parametric statistic used to estimate the survival function of lifetime data. Kaplan-Meier survival estimator method can be used for estimating the survival function of issued and abandoned applications with right censoring of abandoned and pending applications obtains the survival curve. The survival estimator of the survival function, $S(t)$, is given with equation (16) below.

$$\widehat{S}(t) = \prod_{i:\, t_i \leq t} \left(1 - \frac{d_i}{n_i}\right) \quad (16)$$

where d_i is the number of events (issued applications) that occurred within time increment, n_i is the number of individuals (applications) known to have survived (not issued, or still pending or not abandoned after time increment inclusive). Still pending applications that occur within an increment are right censored. Right censorship simply removes that the applications from the count of pending (surviving) applications, n_i, but are not added to the number of events that occurred, d_i [9].

Applying this method to each of the years 2001 to 2018 and these years cumulatively, the survival curves shown in Figure 31 below are obtained.

As can be seen some of the later years (e.g., 2014 to 2018) do not have long enough history. As expected, these years need to be removed to achieve more consistency and remove front loading of the survival curve. Repeating the above Kaplan-Meier for each of the years 2001 to 2013 and these years cumulatively the survival curves shown in Figure 32 below are obtained.

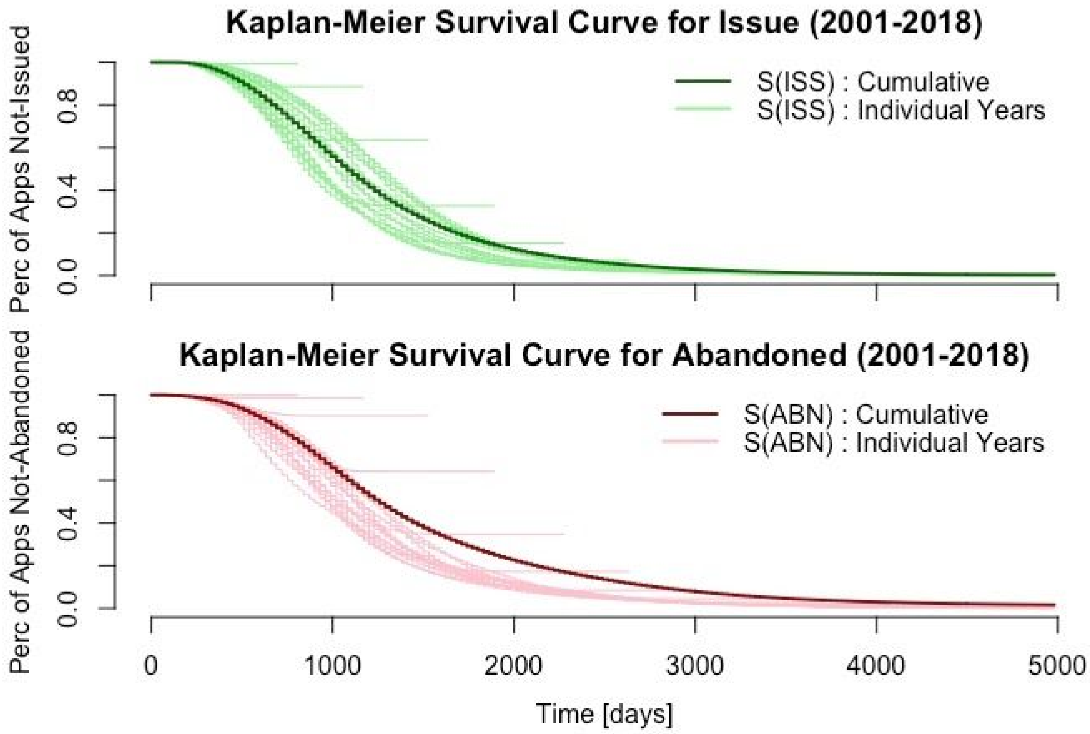

Fig. 31: *Kaplan-Meier estimator issued and abandoned survival curves for the years 2001 to 2018 individually and cumulatively.*

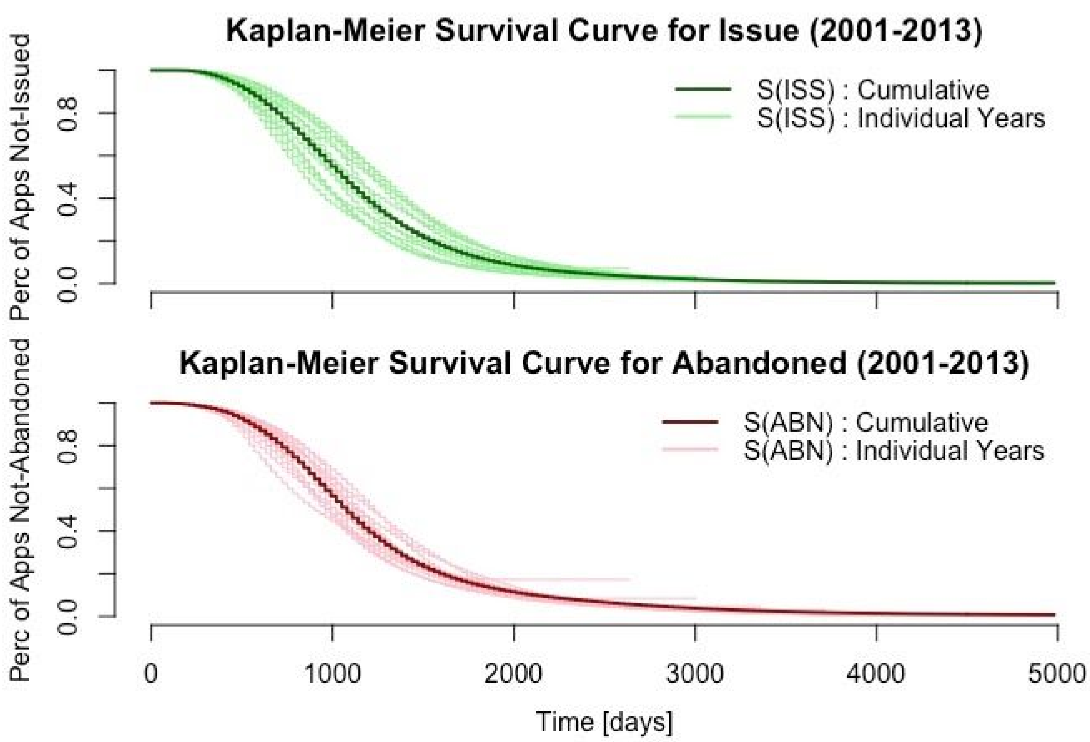

Fig. 32: *Kaplan-Meier estimator issued and abandoned survival curves for the years 2001 to 2013 individually and cumulatively.*

Survival curves can be independently obtained for only issued applications and only abandoned applications for each of the factors discussed above. These factor specific survival curves can be used to obtain PDFs and proceed with the process as discussed above.

8. Evaluate Solution and Improvements

In the determined solution, the data is combined by using Bayes' theorem. This allowed for the full accounting of applications for the probability density functions of both issuance and abandonment curves. It allows for multilevel modelling of the probability density functions such as a Bayesian hierarchical model, and could be expanded on in future iterations. However, this method shifts the probability density mean towards earlier times if the data for all years 2001 to 2018 are included in the model. This is due to the years after about 2013 not having data with long enough history, with many applications still pending and the ones that have been abandoned or been issued occurring before 1500 days. This leads to over representation of data with earlier issuance and abandonment dates, which can skew the data incorrectly toward earlier times. This was solved by simply excluded data post 2013.

This problem could be solved in a better way by incorporating all data from 2001 to 2018. The solution requires the use of regression to combine the survival curves of each year 2001 to 2018, weighting each year's survival curve by the number of applications filed that year. Years are then only included up to their difference from the current date. For example, 2001 would span the entire x-axis of the regression model, while 2018 would only extend approximately 2 years, or 700 days, any days after 700 would not be used for fitting.

Another solution to this problem is using Kaplan-Meier survival estimator discussed above to obtain survival functions for events of issued, abandoned, and either event using all known data. This requires right censoring of still pending applications when obtaining the survival curves for issued and abandoned applications. This will compensate for the skew towards earlier issuance and abandonment times and shift the curves to the right.

The factors, or categorical independent variables, used to improve the aggregated model included Art Unit, Examiner ID, Class, Subclass, and Customer Number. However, Small Entity Indicator status and AIA status could also be factored in. These are fairly high level as they are simply a 1 or 0 in the data, either they are a small entity or not, or they are pre-AIA or AIA. These were left out of the model since from initial analysis Small Entity Indicator only slightly varied from all applications, and it assumed that any information captured by this small entity indicator was likely captured by the Customer Number data. AIA status was also left out because the data for AIA only includes application filed post March 16, 2013. Thus, the data is not as complete as pre-AIA application data, and would thus likely be unreliable. As the AIA date becomes further in the past then the data for AIA will become more complete and more reliable, and thus more useful. In addition, certain factors that are not tracked (at least accurately tracked) by the USPTO database include specific attorney, and specific firm, working on the application prosecution case. These would likely need to be tracked externally and then included into the aggregated model if complete enough data is available.

In addition, Technical Centers and Groups, which are categories above Art Units in the hierarchy of factors, can also be factored into the aggregated model. These can be incorporated and weighted less than Art Unit, or can be included as part of a Bayesian hierarchical model.

In addition, an application can possibly have multiple classes and subclasses. These could be incorporated by including them in the model and weighting them equally in the class and subclass categories. Alternatively, they may be weighted based on the inverse of the number of applications and treated as separate factors with their own individual weight for the weighted aggregate model.

The aggregated model is obtained from regression fitting a gamma distribution to the various curves. Although a gamma distribution very closely fits with the example factors, it may be beneficial to use a kernel density function or another non-parametric way of obtaining the probability density function. This would be able to account for more variation in the model and in particular multimodal distributions.

There are also certain event based factors such as first and second office action / rejections, appeals, RCE, etc., that can have an effect on the probability of issuance or abandonment during prosecution of an application. These were not included in this initial model.

The aggregated model could be improved by using a multiple stage Bayesian Hierarchical model in which the PDFs of class and subclass, and art unit and examiner ID are chained together in a 2-stage chain. In addition, or alternatively, each factor can be given a second stage that is a beta distribution that specifies the degree of membership to that factor.

9. Impact of Results and Discussion

The results of this research show that a model of the probability of a filed patent application issuing or being abandoned can be generated with multiple factors in the form of an aggregated model grant rate timeline. Previously including multiple factors would not have enough data to generate a grant rate timeline, since very few applications if any will match all factors or any several of the factors. This model is a significant improvement over simply looking at the issuance and abandonment probabilities and survival curves (i.e., grant rate timeline) for each factor independently, in comparison to the timeline for all applications. However, there is still room to improve upon the model by including more factors, and building improving the model with various techniques such as using Kaplan-Meier or using a Bayesian Hierarchical model to combine the PDFs as discussed above.

References

[1] Quinn, Gene. *The Cost of Obtaining a Patent in the US*. IPWatchdog. Published April 4, 2015. Retrieved from <https://www.ipwatchdog.com/2015/04/04/the-cost-of-obtaining-a-patent-in-the-us/>.

[2] *How Predictive Analytics is Reshaping Patent Prosecution*. LexisNexisIP. June 6, 2016. Retrieved from <https://www.lexisnexisip.com/knowledge-center/how-predictive-analytics-is-reshaping-patent-prosecution/>.

[3] Trippe, Anthony. *Data Mining Lessons Applied to Analyzing Patent Documents.* IPWatchdog. March 12, 2013. Retrieved from <https://www.ipwatchdog.com/2013/03/12/data-mining-lessons-applied-to-analyzing-patent-documents-2/id=37137/>.

[4] Schox Esq., Jeffrey. *Not So Obvious: An Introduction to Patent Law and Strategy*. Third Edition. September 2013.

[5] "Patent Technology Centers Management." USPTO.gov. Published on: Jul 20, 2018 09:49 AM EDT. Retrieved from <https://www.uspto.gov/patent/contact-patents/patent-technology-centers-management>.

[6] "Patent Classification." USPTO.gov. Published on: Jan 29, 2019 10:55 AM EST. Retrieved from <https://www.uspto.gov/patents-application-process/patent-search/classification-standards-and-development>.

[7] Obissi, Adnan M. *Patent Prosecution Data Analytics*. MBHB. Spring 2017. Retrieved from <https://www.mbhb.com/intelligence/snippets/patent-prosecution-data-analytics>.

[8] Graham, Stuart et al. *The USPTO Patent Examination Research Dataset: A Window on the Process of Patent Examination.* USPTO Economic Working Paper No. 2015-4. Office of Chief Economist November 2015. Retrieved from <http://www.uspto.gov/economics>.

[9] Rodríguez, German. *Lecture Notes on Generalized Linear Models.* Chapter 7: Survival Models. Princeton. (2007). Revised September 2010. Retrieved from <http://data.princeton.edu/wws509/notes/>

[10] Congdon, Peter. *Bayesian Models for Categorical Data*. John Wiley & Sons. 2005.

[11] Choi, S. C.; Wette, R. *Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias*. Technometrics. 1969. 11 (4): 683–690.

[12] Clemen, Robert T. et al. Aggregating Probability Distributions. *Advances in Decision Analysis: From Foundations to Applications*. Cambridge: Cambridge University Press. 2007. 154-176.

Additional References

Campbell, Richard S. Patent Trends as a Technological Forecasting Tool. World Patent Information. Volume 5, Issue 3, Pages 137-143. June, 1982.

Cox, D. and D. Oakes. Analysis of Survival Data. London: Chapman-Hall. 1984.

Lemley, Mark A. and Sampat, Bhaven. *Examining Patent Examination.* Retrieved from <http://ssm.com/abstract=1485011>.