

A Survival Analysis of the Approval of U.S. Patent Applications

Ying Xie
&
David E. Giles*

*Department of Economics, University of Victoria
Victoria, B.C., Canada V8W 2Y2*

August 2007

Abstract

We model the length of time that it takes for a patent application to be granted by the U.S. Patent and Trademark Office, conditional on the patent actually being awarded eventually. Survival analysis is applied and both the nonparametric Kaplan-Meier and parametric accelerated failure time models are used to analyze the data. We find that the number of claims a patent makes, the number of citations a patent makes, the patent's technological category, and the type of applicant all have significant effects on the duration that a patent is under consideration. A log-normal survival model is the preferred parametric specification, and the results suggest that the hazard function is non-monotonic over time.

Keywords: Patents, research and development, survival analysis, hazard function

JEL Classifications: C16, C29, C41, L10

Author Contact:

David Giles, Dept. of Economics, University of Victoria, P.O. Box 1700, STN CSC, Victoria, B.C., Canada V8W 2Y2; e-mail: dgiles@uvic.ca; Voice: (250) 721-8540; FAX: (250) 721-6214

1. Introduction

In this paper we model the length of time that is taken for patents to be granted by the U.S. Patent and Trademark Office (USPTO), conditional on the patents ultimately being awarded. We are motivated in part by the understanding that the 21st century economy is “one in which knowledge - particularly the technological knowledge that forms the foundation for industrial innovation - is an extremely important economic commodity.” (Jaffe *et al.*, 2002, p.451). The idea of patenting is a kind of “double-edged” sword. On one hand, a patent for an invention grants the inventor a property right to exclude others from making, using, or offering for sale the invention throughout U.S. or from importing the invention into the U.S. for a limited period of time. It provides the patent owner with protection and leverage of an intangible asset. The anticipation of receiving patents increases the motivation for useful invention. On the other hand, patents generate monopoly rights to explore an invention, with associated social costs arising from the barriers to entry on the part of other inventors.

This paper is also motivated by the need to “quantify the extent and impact of knowledge spillovers through utilization patent citations to identify a paper trail that may be associated with knowledge flows between firms.” (Jaffe *et al.*, 2002, p.403). We believe that data relating to patents offer an opportunity to quantify certain aspects of the role of knowledge in the modern economy. We are concerned with how the patent system affects the rate and direction of technological changes, and we are trying to explore question such as, which broad types of technologies obtain patent approval faster than others; whether U.S. inventions get patents granted more quickly than their foreign counterparts, which may contribute to the U.S. leading technology and economy status in the world, *etc.*

Recently, Popp *et al.* (2004) conducted similar research using a slightly earlier version of the NBER patent database that we use in this study, augmented by additional data from the USPTO.¹ Specifically, they used least squares and quantile regression analysis to model the (logarithm of the) length of time that expires between the lodging of a patent application, and its final approval. However, in the context of survival modeling, this log-linear specification amounts to an accelerated failure time model with the very restrictive assumption of a constant hazard. Implicitly, an underlying exponential distribution is being assumed – an assumption that we show in this paper is overwhelmingly rejected by these data. A more comprehensive and flexible analysis of such duration data is required, and apparently no such modelling has been undertaken to date with U.S. or other international patent data.

Survival analysis deals directly with the conditional probability of an event taking place. The question here is: “what is the probability that a patent will be granted at time $t+1$, given that the application process has lasted up to time t ?” The natural way of thinking about that probability that a patent will be granted some time in the future depends on the inherent factors that affect the application process, such as the type of applicant, the number of claims of originality that are made, the number of citations made to other patents, the technological field in which the invention lies, and so on. The characteristics of the patent examiners themselves (such as years of experience) may also affect the duration of patent-awarding process (*e.g.*, Cohen and Merrill, 2003), but we do not have access to such information.

We first use the nonparametric Kaplan-Meier estimator to conduct a preliminary analysis of the life-cycle of the patent-granting process. Then parametric accelerated failure time models, with various distributional assumptions, and covariates are estimated by maximum likelihood. From the latter results, we conclude that the number of claims a patent makes, the number of citations a patent makes, the technological category of the invention, and the type of applicant are all significant factors that affect the duration of time between the lodging of the patent application, and its subsequent award. We also find that duration dependence is positive for the first 3.4 years of the application process, and then negative.

The rest of the paper is organized as follows. Section 2 introduces important concepts for survival analysis. Section 3 presents the estimation methods. Section 4 gives description of the data, and the variables included in the estimation of our models. Section 5 presents the results of our study, and Section 6 offers our conclusions and some suggestions for further related research.

2. Survival Analysis

Duration data measure how long a situation lasts, or how much time elapses before a particular event takes place. The data show when the underlying data-generating process changes from one “state” to another “state”. Examples from outside of economics include how long a machine part will last before breaking down, and how long a patient can survive after a surgical procedure. Economic examples include the length of time a person remains unemployed (*e.g.*, Kiefer *et al.*, 1985; Nickell, 1979; and Sider, 1985); how long a labour dispute lasts (*e.g.*, Kennan, 1985); the time that elapses between changes to official interest rates (*e.g.*, Shih and Giles, 2006); and how long a particular type of exchange rate regime lasts (*e.g.*, Yu, 2005). There are many additional

examples of economic phenomena that involve duration data. Comprehensive discussions that have an economic orientation are provided by Kiefer (1988) and Lancaster (1990), for example.

The purpose of duration (survival) analysis is to model the underlying distribution of the “failure time” variable and to assess the dependence of this variable on covariates. An intrinsic characteristic of duration data is the possibility of observations being censored. Another characteristic of duration data is that the dependent variable cannot be negative. Thus a transformation of the survival time, such as logarithmic transformation, may be appropriate.

We use survival analysis to model the time taken between the application for, and granting of, a patent. The nonnegative random variable T is defined here as the duration (or spell) between the application and grant dates. The unconditional probability that the spell will be shorter than some given value t is given by the cumulative distribution function, written as $F(t) = \Pr.(T < t)$, and the associated probability density function is $f(t) = dF(t)/dt$. Survival analysis makes use of the complement to the cumulative distribution function, referred to as the survival function, which is written as $S(t) = 1 - F(t) = P(T \geq t)$. That is, the survival function measures the probability that the random variable T will equal or exceed the value t . The hazard function captures the conditional probability that the spell will terminate at time $T = t$, given that it has survived until time t . It is given by

$$\lambda(t) = \lim_{dt \rightarrow 0} [\Pr.(t < T < t + dt | T > t) / dt] = f(t) / S(t).$$

Essentially, $\lambda(t)$ is the instantaneous rate at which spells will be completed at duration t , given that they have lasted until t .

The hazard function provides a convenient definition of duration dependence. Positive (negative) duration dependence exists at the point t^* when $d\lambda(t) / dt > (<) 0$ at $t = t^*$. If $d\lambda(t) / dt = 0$ for every t we have a so-called “memory-less” system. It is possible that the hazard function evolves with time in a non-monotonic fashion alternating between positive and negative dependence. Not all distributions will allow for this possibility. Finally, the integrated (or cumulative) hazard function, defined as $A(t) = \int_0^t \lambda(u) du = -\log S(t)$, is also a useful function for diagnostic purposes in practice. Typically, we look for an integrated hazard function that is an (upward-sloping) straight line as a sign that the model has been correctly specified.

3. Estimation Methods

3.1 Nonparametric Kaplan-Meier Estimator

The Kaplan-Meier estimator is a nonparametric approach. The random spell is written as T^* in the absence of censoring. If there is censoring at time C , then the observed random variable is $T = \min. (T^*, C)$. Suppose that there are k completed spells in our sample of size n , where $k < n$ since some observations are censored, and because of ties. Ties occur when two or more observations have the same duration. We can define a variable δ that takes the value unity if the observation is censored, and zero otherwise. We assume that if $T = t$ and $\delta = 1$, censoring happens immediately after time T . Suppose the completed spells from smallest to largest, $t_1 < t_2 < \dots < t_k$. Let h_j be the number of completed spells of duration t_j , for $j = 1, \dots, K$. In the absence of ties the h_j are all equal to unity in value. Let m_j be the number of observations censored between t_j and t_{j+1} ; m_k is the number of observations with duration greater than t_j , the longest complete spell. Let n_j be the number of spells neither completed nor censored before duration t_j :

$$n_j = \sum_{i \geq j}^K (m_i + h_i)$$

Recall that the hazard, $\lambda(t_j)$, is the rate at which spells are completed at duration t_j , conditional upon the spell having a duration of at least t_j . So, a natural estimator for $\lambda(t_j)$ is:

$$\hat{\lambda}(t_j) = h_j / n_j,$$

i.e., the number of “failures” at duration t_j , divided by the number of spells “at risk” at duration t_j . The corresponding estimator for the survival function is

$$\hat{S}(t_j) = \prod_{i=1}^j (n_i - h_i) / n_i = \prod_{i=1}^j (1 - \hat{\lambda}_i),$$

which is the so-called “product-limit” estimator. The integrated hazard can be estimated by

$$\hat{\Lambda}(t_j) = \sum_{i \leq j} \hat{\lambda}(t_i).$$

The Kaplan-Meier survival, hazard, and integrated hazard functions that are reported below were obtained by writing program code for the SHAZAM econometrics package (Northwest Econometrics, 2004).

3.2 Parametric Models

Another approach is the parametric method which requires an assumption about the underlying distribution of the random variable T . Using a parametric approach also allows us to model the relationship between survival (typically in the form of the hazard function) and one or more explanatory variables (or covariates). While many different parametric models may be used in survival analysis, the most common ones (and the ones that we consider) are based on the Exponential, Weibull, Log-Normal and Log-Logistic distributions.

Table 1

Distribution	Density Function $f(t)$	Survival Function $S(t)$	Hazard Function $\lambda(t)$
Exponential	$\theta \exp(-\theta t)$	$\text{Exp}(-\theta t)$	θ
Weibull	$\theta p(\theta t)^{p-1} \exp(-(\theta t)^p)$	$\text{Exp}[-(\theta t)^p]$	$\theta p(\theta t)^{p-1}$
Log-Normal	$[p/(\theta t)] \phi(-p \log(\theta t))$	$\Phi(-p \log(\theta t))$	$[p/(\theta t)] \phi(-p \log(\theta t)) / \Phi(-p \log(\theta t))$
Log-Logistic	$\theta p(\theta t)^{p-1} / (1 + (\theta t)^p)^2$	$1 / (1 + (\theta t)^p)$	$\theta p(\theta t)^{p-1} / (1 + (\theta t)^p)$

Table 2

Distribution	Density Function $f(w)$	Survival Function $S(w)$
Exponential	$\exp(-\exp(w))$	$\exp(-w)$
Weibull	$\exp(w - \exp(w))$	$\exp(-\exp(w))$
Log-Normal	$\phi(w)$	$\Phi(-w)$
Log-Logistic	$\exp(w)(1 + \exp(w))^{-2}$	$1/(1 + \exp(w))$

The one-parameter Exponential distribution is “memory-less”, in the sense that its hazard function is constant so there is no duration dependence. This very restrictive distribution is a special case of the two-parameter Weibull distribution, which can exhibit either positive or

negative duration dependence - the hazard function increases or decreases monotonically according to whether the shape parameter, p , is larger or smaller than unity. The two-parameter Log-Normal and Log-Logistic distributions can have non-monotonic hazards, so they can capture situations where the duration dependence increases and decreases, over different spell lengths. The density functions $f(t)$, survival functions $S(t)$, and hazard functions $\lambda(t)$ for each of these four models appear in Table 1, where $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and cumulative distribution functions for the standard normal distribution.

When we incorporate a vector of covariates, x , maximum likelihood estimation of the parameters is facilitated by the log-linear transformation, $w = [\log(t) - \beta'x] / \sigma$, where $\theta = \exp(-\beta'x)$, and $p = 1/\sigma$. After this transformation, the covariates enter w linearly. Because the regression coefficients are expressed in terms of the log of the duration times, they convey information about the expected failure times. So, this model is sometimes referred to as a “log expected failure time model” or simply an “accelerated failure time” (AFT) model. With this change of variables, the densities and survival functions for w for the four distributions are presented in Table 2. It is worth noting that only in the case of the Weibull distribution does the AFT model coincide with the well-known proportional hazard model for duration data.

We have used the EViews econometric package (Quantitative Micro Software, 2004), to estimate the various parametric models that are in Table 2. EViews does not incorporate dedicated commands for survival analysis so we constructed LOGL “objects” to formulate and implement the various maximum likelihood estimators.

4. Data Characteristics

Our data-set relates to all of the *utility* patents granted in the U.S. between January 1 1963 and December 30, 1999 (37 years).² The total number of observations (patents) is 2,923,922. The data are provided in the “NBER Patent-Citation Data File”, which is supplied in electronic form by Jaffe and Trajtenberg (2002). A detailed discussion of the history and characteristics of this data-file is provided by Hall *et al.* (2002). The patent data themselves were procured from the USPTO, except for the citations from patents granted in 1999, which come from MicroPatent. (Hall *et al.*, 2002, p.408.)

Our study conditions on a patent being awarded, as no other information is available. The earliest grant year available is 1963, but no application year information is available for patents granted

before 1967, so we restricted our sample to patents with a grant year no earlier than 1967. An inspection of the data revealed that there are some patents with exceptionally long time-lags between application (all prior to 1963) and the award of the patent.³ These outlier applications were apparently “militarily sensitive”, and were eliminated by setting the sample to have an application year later than or equal to 1963. This yielded a sample with 2,693,365 observations. Our sample is free of “obvious errors” reported by Popp *et al.* (2004, p.14), such as grant dates that precede application dates. Our parametric modelling includes the covariate “number of claims”. As no claims information is available for granted patents until 1975, we constrained the sample to application years no earlier than 1963 and non-zero number of claims. This reduces the sample to 1,983,420 observations. By way of comparison the non-parametric Kaplan-Meier analysis is conducted with both the full and reduced samples. There is no censoring problem in our data because the study conditions on the awarded patents and the selected sample includes patents for which both application year and grant year information available.

Our data contains the grant *date* for all patents in the file (*i.e.*, of all utility patents granted since 1963) and the application *year* for patents granted since 1967. The duration for each patent’s consideration is calculated in years. For patents whose grant year equals the application year, we have coded the duration to 0.5 years. The summary statistics for the duration data, and the covariates used in the parametric modeling are shown in Table 3.

A key data item in the patent database is the number of citations (“*CITES*”) made in an application. Patent citations serve an important legal function, as they determine the scope of the property rights awarded by the patent. The applicant has a legal duty to disclose any knowledge of the “prior art”.⁴ The number of claims (“*CLAIMS*”) made in a patent application specify in detail the distinct “components,” or building blocks, of the patented invention, and hence their number may be indicative of the scope of the invention (Jaffe *et al.*, 2002, p.43), and also the scope of protection afforded by the patent. Positive relationships are expected between the duration variable and both the number of citations and number of claims made in the application, as increases in each will complicate the task facing the patent examiner.

The USPTO has developed a highly elaborate classification system for the technologies to which the patented inventions belong, consisting of about 400 main (3-digit) patent classes, and over 120,000 patent subclasses. For the vast majority of uses one is likely to resort to only the original

3 – digit patent class, and serving as controls in regressions, six main categories are aggregated here in the database. We code the technological categories by dummy variables, CAT_i ($i = 1, \dots, 6$) as follows: Chemical (excluding Drugs); Computers and Communications; Drugs and Medical; Electrical and Electronics; Mechanical; and Others. For modeling purposes, and to avoid the “dummy variable trap”, we normalize on CAT_6 , and we have no prior expectations regarding the signs of the marginal effects of these variables.

The USPTO also classifies patents by type of “assignees” (applicants), and we have coded dummy variables for the main types of assignees (ASS_i ; $i = 1, \dots, 7$): Unassigned; U.S. Non-Government organization (mostly corporations); Non-U.S. non-Government organizations (mostly corporations); U.S. individuals; Non-U.S. individuals; U.S. Federal Government; Non-U.S. Governments. For regression purposes we normalize on the first category.⁵ Our only prior expectations are that the U.S assignees will have smaller marginal effects than their non-U.S. counterparts.

Other variables that may be important include the year in which the application is lodged or the patent is granted. These were found to be insignificant after controlling for other covariates and are not discussed further.⁶ It has also been suggested that observable characteristics among the patent examiners, including their experience at the USPTO, their degree of technological specialization may have an important impact on patent approval times (Cohen and Merrill, 2003, p.22). The additional data procured from the USPTO by Popp *et al.* (2004) and used in their study included information regarding the experience of the patent examiners, and the number of drawings and sheets for each application. Interestingly, the last of these variables entered their regressions with an unanticipated negative coefficient, though it was statistically insignificant; and although examiner experience reduced the examination duration, as expected, this variable was barely significant (taking account of the very large sample size).⁷

Table 3
Summary Statistics

Observations											
<i>N</i> = 1,983,420											
<i>(N</i> = 2,693,365)											
	Application Year	Grant Year	Duration	CLAIMS	CITES	CAT₁	CAT₂	CAT₃	CAT₄	CAT₅	CAT₆
Mean	1986	1988	1.8908 (1.9980)	12.0839	7.4934	0.2041	0.1091	0.0829	0.1709	0.2222	0.2107
Median	1987	1989	2.0000 (2.0000)	10.0000	6.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Maximum	1998	1998	34.0000 (35.0000)	868.0000	770.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Minimum	1963	1975	0.5000 (0.5000)	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sum						404859	216452	164482	338877	440765	417985
Percentage						20.41%	10.91%	8.29%	17.09%	22.22%	21.07%
	ASS₁	ASS₂	ASS₃	ASS₄	ASS₅	ASS₆	ASS₇				
Mean	0.1750	0.4374	0.3597	0.0068	0.0033	0.0137	0.0041				
Median	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000				
Maximum	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000				
Minimum	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000				
Sum	347064	867586	713419	13414	6557	27156	8224				
Percentage	17.50%	43.74%	35.97%	0.68%	0.33%	1.37%	0.41%				

Note: Figures in parentheses correspond to a sample size with *N* = 2,693,365.

Using the variables described in Table 3, we define w_t as:

$$w_t = [\log(t) - \beta_1 - \beta_2 CLAIMS - \beta_3 CITES - \sum_{i=2}^7 \beta_{i+2} ASS_i - \sum_{i=1}^5 \beta_{i+9} CAT_i - \sum_{i=2}^7 \beta_{i+13} (CLAIMS * ASS_i) - \sum_{i=2}^7 \beta_{i+19} (CITES * ASS_i) - \sum_{i=1}^5 \beta_{i+26} (CITES * CAT_i) - \sum_{i=1}^5 \beta_{i+31} (CLAIMS * CAT_i)] / \sigma,$$

where σ is a scale parameter, and the choice of dummy variables reflects the normalizations on the dummy variables CAT_6 and ASS_1 .

5. Results

5.1 Nonparametric Estimation

Figures 1-3 plot the Kaplan-Meier survival functions, hazard functions and integrated hazard functions with sample sizes of $N = 1,983,420$ and $N = 2,693,365$. The estimated hazard function shows a non-monotonic pattern of duration dependence. It starts by increasing, then it is decreasing, and finally it follows an increasing pattern. Therefore, duration dependence cannot be qualified as being either positive, or negative: it depends upon the survival time. More precisely, the probability of a patent getting granted rises for the first four years or so, then falls, and rises again around thirty years or later. Of course, patents that are under assessment for twenty years or more represent a very extreme part of the sample, and the earlier parts of the hazard and cumulative hazard functions are really of much greater interest. This point is relevant to the preferred parametric model discussed below.

5.2 Parametric Modelling

To determine the role of key covariates on the application-grant duration, we estimate accelerated failure time (AFT) models, with the covariates noted earlier. First, to choose between the (nested) Weibull and Exponential distributions, we test $H_0: \sigma = 1$, which is equivalent to testing $p = 1$. From Table 4, an asymptotically valid (Wald) test of this restriction is based on $z = [(\hat{\sigma} - 1) / a.s.e.(\hat{\sigma})] = -4956.667$. So, the null hypothesis is very strongly rejected, and the Weibull distribution is preferred to the Exponential model. Recalling that the the log-linear specifications adopted by Popp *et al.* (2004) amount to the use of the Exponential model, this result has adverse consequences for their results, as the latter are based on a mis-specified model and so the parameter estimates are inconstant.

Figure 1 : Kaplan-Meier Survival Functions

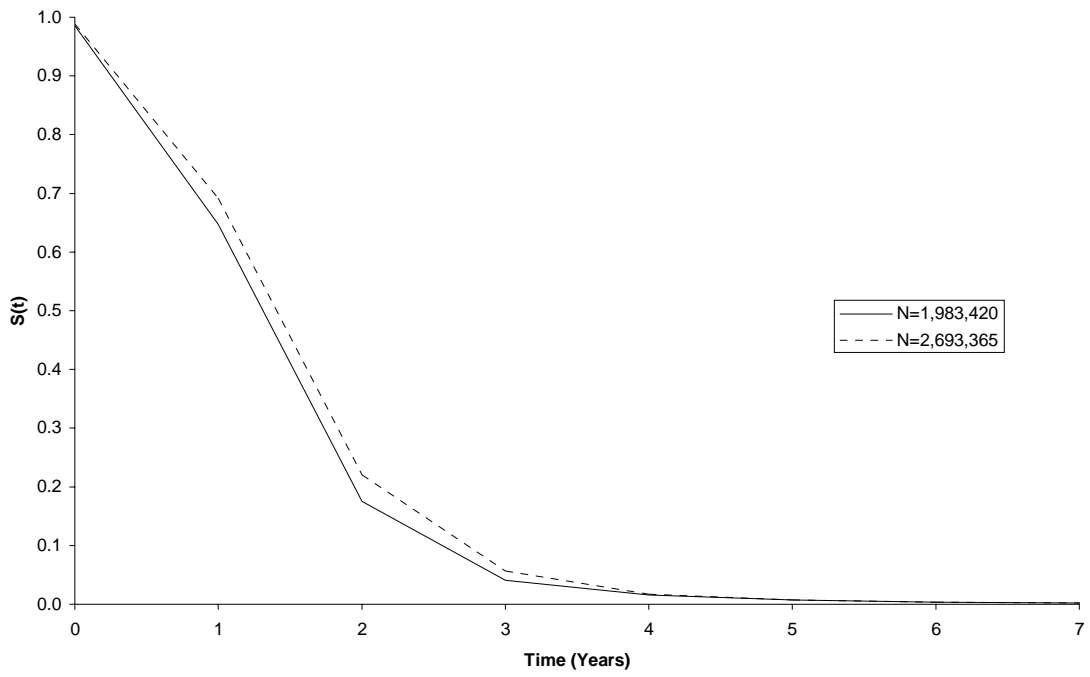


Figure 2: Kaplan- Meier Hazard Functions

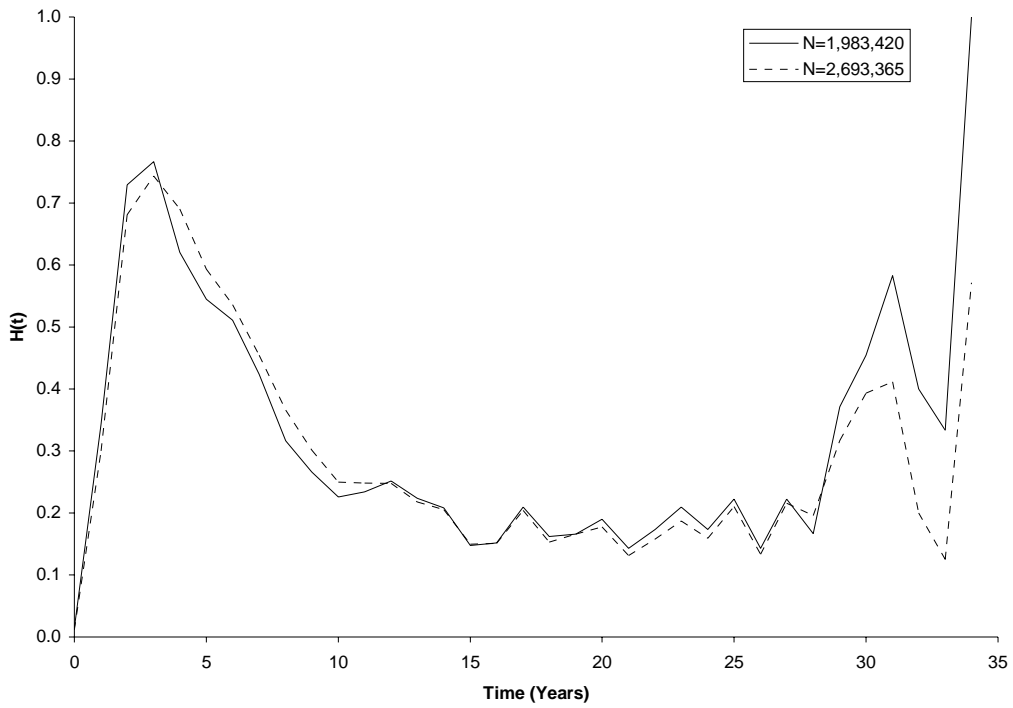


Figure 3: Kaplan-Meier Cumulative Hazards

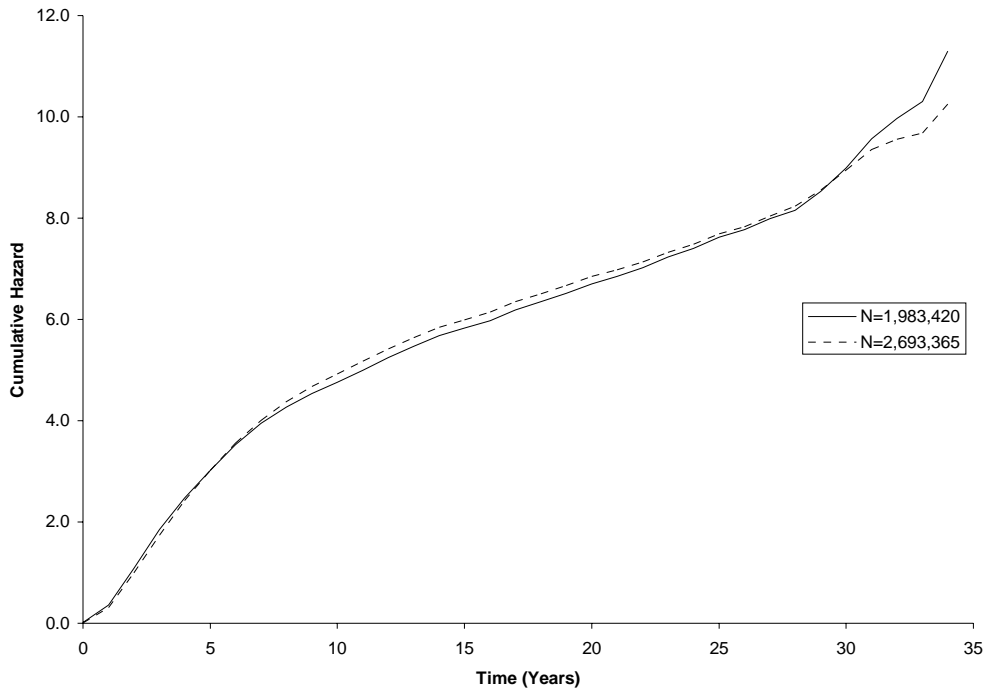


Figure 4: Weibull, Log-Normal, Log-Logistic Cumulative Hazards

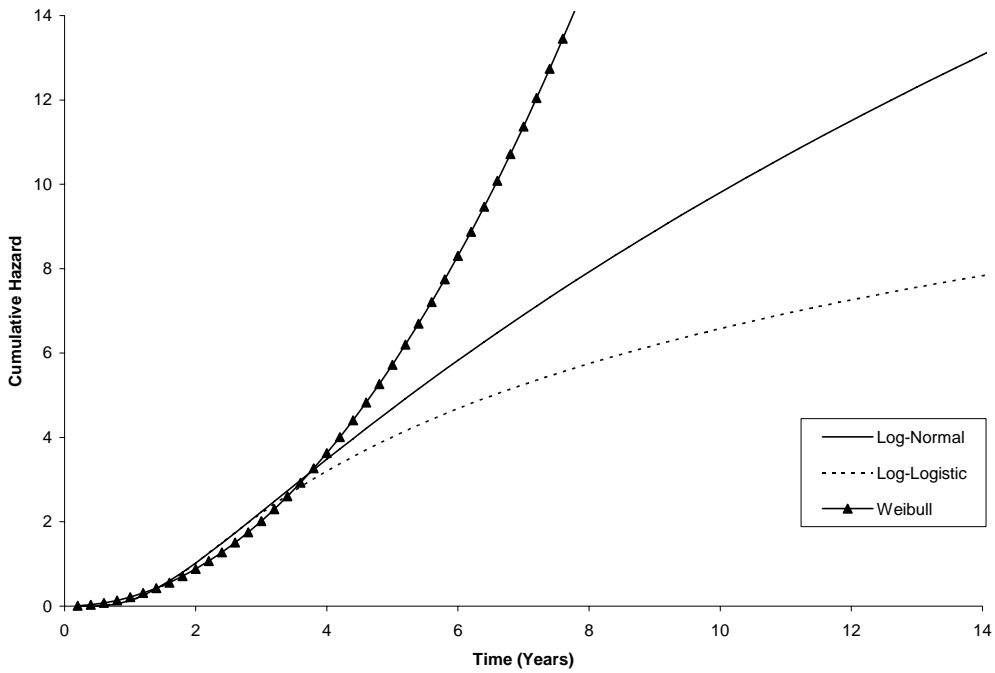


Table 4**Maximum Likelihood Estimation Results**

	<u>Weibull</u>	<u>Log-Normal</u>	<u>Log-Logistic</u>
β_1	0.6205 (672.22)	0.3986 (294.38)	0.3986 (306.68)
β_2	0.0016 (30.08)	0.0022 (27.62)	0.0026 (33.26)
β_3	0.0021 (49.37)	0.0021 (25.31)	0.0023 (27.57)
β_4	0.0289 (54.41)	0.0122 (12.74)	0.0107 (11.60)
β_5	-0.0147 (-16.29)	0.0250 (17.96)	0.0336 (24.68)
β_6	0.0411 (15.84)	0.0272 (6.84)	0.0267 (6.89)
β_7		0.0539 (5.36)	0.0637 (6.59)
β_8	0.4973 (617.94)	0.0600 (25.77)	0.0106 (3.97)
β_9	0.1156 (23.50)	0.0833 (9.57)	0.0879 (10.02)
β_{10}	0.1157 (114.37)	0.0695 (41.26)	0.0691 (41.78)
β_{11}	0.2344 (224.85)	0.2025 (99.64)	0.2175 (104.82)
β_{12}	0.1688 (105.90)	0.1561 (73.63)	0.1685 (78.93)
β_{13}	0.0517 (46.26)	0.0547 (29.01)	0.0638 (34.78)
β_{14}	0.0397 (41.25)	0.0350 (19.96)	0.0400 (23.64)
β_{16}	0.0008 (16.24)	0.0005 (6.73)	0.0004 (5.32)

Table 4 (continued)

Maximum Likelihood Estimation Results

	Weibull	Log-Normal	Log-Logistic
β_{22}	0.0053 (51.95)	0.0036 (31.89)	0.0038 (32.25)
β_{24}	0.0046 (6.66)	0.0023 (1.71)	0.0023 (1.76)
β_{26}	0.0027 (3.07)	0.0033 (2.47)	0.0033 (2.48)
β_{27}	-0.0019 (-25.42)	-0.0007 (-6.53)	-0.0005 (-4.80)
β_{28}		0.0009 (6.94)	0.0010 (7.33)
β_{29}	-0.0031 (-34.17)	-0.0022 (-16.42)	-0.0022 (-16.34)
β_{30}	0.0008 (8.84)	0.0011 (7.82)	0.0011 (8.04)
β_{31}		0.0004 (3.22)	0.0005 (4.07)
β_{32}	-0.0012 (-18.52)	-0.0010 (-9.83)	-0.0011 (-11.13)
β_{33}	-0.0010 (-14.46)	-0.0005 (-4.05)	-0.0007 (-5.92)
β_{34}	0.0009 (9.97)	0.0008 (7.17)	0.0006 (5.01)
β_{35}	-0.0003 (-4.71)	-0.0003 (-3.19)	-0.0005 (-4.27)
β_{36}	-0.0006 (-9.58)	-0.0008 (-7.33)	-0.0009 (-8.87)
σ	0.4796 (4584.31)	0.4587 (1965.07)	0.2678 (1323.18)
<i>LogL</i>	-1467971	-1268554	-1306554
<i>AIC</i>	1.4803	1.2792	1.3175
<i>SIC</i>	1.4804	1.2794	1.3177

Note: Asymptotic “z-statistics” appear in parentheses.

Second, we choose the Log-Normal Distribution as the preferred model over the (non-nested) Weibull and Log-Logistic specifications on the basis of the Schwarz and Akaike Information Criteria (SIC, AIC) values and the shapes of the estimated cumulative hazard functions. The AIC and SIC values are included in Table 4, and the cumulative hazards, including the (relatively) linear one for the Log-Normal specification, appear in Figure 4. Table 4 presents the main estimation results, with only significant regressors retained. In these AFT models a positively (negatively) signed coefficient implies that the expected duration increases (decreases) for changes in the values of the covariate.

Figures 5 and 6 plot the survival and hazard and functions for a sample size of $N = 1,983,420$ for the preferred Log-Normal model.⁸ The scale parameter $\sigma (= 0.4587)$ is less than unity, implying that the hazard rate rises and then begins to decline at some point, as we see in Figure 6. The percentiles of the survival times for the Log-Normal model may be calculated by $t(\text{percentile}) = [\theta^{-1} \exp(-\Phi^{-1} - m)]/p$, where m corresponds to the fractional representation of the percentile of interest and Φ^{-1} is the standard normal percentile distribution. For the median duration, $m = 0.5$ and so $t(50) = \theta^{-1}$. (See Box-Steffensmeier and Jones, 2004, p.34.)

From this preferred parametric estimation result, first, we observe that after 1.7 years (the median estimated survival time) there is a 50% probability that the patents will still be in the application-grant process.⁹ After two years, there is a 36% probability that a patent will still be in the process. Then the survival function falls to close to the horizontal axis after five years, which indicates that for (ultimately successful) patents that have been in the application-grant process for nearly five years, it is highly likely that they will be granted very soon. The hazard function reaches its mode after 3.4 years, with a modal value of 2.13. It then starts decreasing, which is quite similar to the Kaplan-Meier result and confirms our preliminary non-parametric analysis. In terms of the hazard rate, we can see the probability of a patent getting granted in the next month ($dt = 1/12$), conditional on it having been in the examining process for 3.4 years is 17.7% ($= 2.13 \times 1/12$), and the probability of getting granted in the next month, conditional on it having been in the examining process for 5 years is 16.6% ($= 1.994236 \times 1/12$), *etc.*

Figure 5: Log-Normal Survival Function

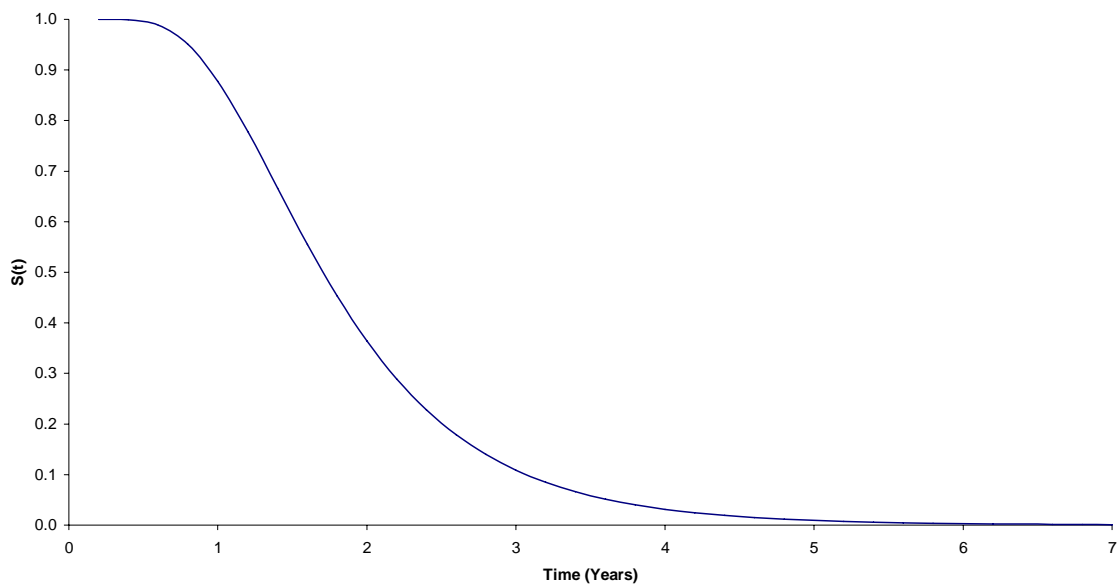
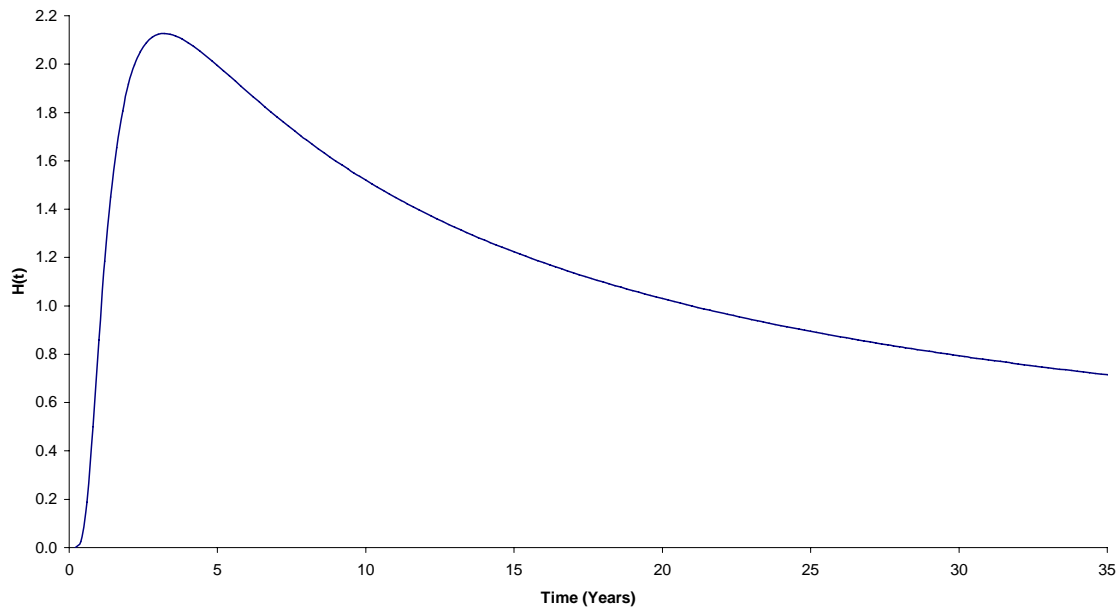


Figure 6: Log-Normal Hazard Function



Overall, we observe that the coefficients on these variables carry the expected signs, where these are known. For example, the greater the number of claims or citations a patent makes, the longer it will take to get approval, *ceteris paribus*. In the former case this is essentially independent of assignee type. Also, non-government organizations (mostly corporations) take a shorter period of time to get patents granted compared to individuals and governments in both the U.S. and non-U.S. cases. This seems reasonable because corporations usually have stronger legal capabilities and better resources dedicated to patent applications than do individuals, and corporations also have stronger business motivations and operational efficiency than do government organizations. Furthermore, patents of U.S. applicants get granted faster than non-U.S. applicants in general, which reflects the reality that U.S. applicants may have certain advantages over their non-U.S. counterparts. For example, a foreign applicant may make more citations to foreign inventions than does a U.S. applicant, and such applications may involve patent lawyers in more than one country, and these factors may slow down the examining and patent approval process. Interestingly, Popp *et al.* (2004, p.26) find that U.S. applications suffer a *longer* duration in the examining process than most non-U.S. applications. In part, this may be a consequence of a bias associated with implicitly using an Exponential specification for the hazard model.

We also observe that once we control for other significant factors, patents in the Mechanical category take the shortest time in getting granted among all the technical fields, followed by patents in the Electrical & Electronics and Chemical (excluding drugs) categories. Patents in the Computers & Communications category take the longest time for the granting of patents. Interacting these category effects with the number of claims or citations can reduce or increase the predicted duration relative to the control group, depending on the category in question.

6. Conclusions and Implications

In order to explore some of the mechanisms that actually facilitate technology innovation, this paper conducts a survival analysis of patent applications in the U.S. Patent and Trademark Office. We use both nonparametric and parametric techniques to obtain estimators of the hazard functions associated with the length of time that it takes for a (successful) patent to be approved. The application of the nonparametric Kaplan-Meier estimator indicates significant non-monotonic patterns of duration dependence and gives us some preliminary insights, without controlling for covariate effects. We also use Accelerated Failure Time framework to estimate parametric hazard functions based on the Exponential, Weibull, Log-Logistic and Log-Normal distributions. The Log-Normal model is preferred among all these distributions when we conduct

a formal model-selection. Having controlled for the effects of various covariates, the hazard function of a patent application remains non-monotonic, with a pattern that increases initially and then decreases.

We also conclude that the number of claims a patent makes, the number of citations a patent makes, the industry category, and the type of the assignee will all significantly affect the length of time that it takes for a patent to get granted. We also note that patents in the “Mechanical” category have the shortest duration, while those in the “Computers & Communications” category take the longest time to get granted among all of the technological fields, other things being equal. These results may have significant implications for policy makers who wish to better allocate federal funding in terms of the technology-transfer process if they are interested in stimulating the development of computers and communications, increasing the rate of patenting and promoting the associated spillovers. Therefore it may have a significant impact on innovation and estimates of rates of return in this field. It could also be helpful in terms of determining resource allocation within the USPTO.

We also find that U.S. patent applications are awarded more quickly than non-U.S. applications, once we control for other attributes. This provides some reassurance in response to the observation that: “Obviously, the economic benefits to the U.S. economy of domestic research depend on the fruits of that research being more easily or more quickly harvested by domestic firms than by foreign firms.” (Jaffe and Trajtenberg, 2002, p.179).

The survival times of patent applications deserve further investigation. It would be interesting to examine the impact of measures of patent generality and of patent originality (*e.g.*, Jaffe *et al.*, 2002). Differences in USPTO practices across time or across technological areas may produce differences in citation intensities that are unrelated to the “true” impacts of the patents. For example, we cannot tell if the rise in the number of citations made by the late 1990’s is simply a response to the introduction of computerized patent files at the USPTO during the 1980’s. Thus, we do not know whether the associated durations reflect a “real” phenomenon (*e.g.*, fields with patents citing fewer earlier patents are truly more original), or if different citation practices that are somewhat artificial. (*e.g.*, Jaffe *et al.*, 2002). Finally, the increasing role of government-owned laboratories in the technology commercialization process in the U.S. has had a significant impact on technology transfer (*e.g.*, Jaffe and Lerner, 2002), but has yet to be examined in the context of patent application survival models.

References

- Box-Steffensmeier, J. M. and B. S. Jones (2004), *Event History Modeling, A Guide for Social Scientists*, Cambridge University Press, Cambridge.
- Cohen, W. M. and S. A. Merrill (2003), *Patents in the Knowledge Based Economy*, National Academies Press, Washington DC.
- Hall, B. H., A. B. Jaffe and M. Trajtenberg (2002), “The NBER Patent-Citations Data File: Lessons, Insights and Methodological Tools”, Chapter 13 in A. B. Jaffe and M. Trajtenberg (eds.), *Patents, Citations and Innovations: A Window on the Knowledge Economy*, MIT Press, Cambridge MA, 404-451.
- Granger, C. W. J. (2003), “Some Methodological Questions Arising From Large Data Sets”, in D. E. A. Giles (ed.), *Computer-Aided Econometrics*, Marcel Dekker, New York, 1-9.
- Grossman, G. and E. Helpman (1991), *Innovation and Growth in the Global Economy*, MIT Press, Cambridge MA.
- Jaffe, A. B. and J. Lerner (2002), “Reinventing Public R&D: Patent Policy and the Commercialization of National Laboratory Technologies”, Chapter 10 in A. B. Jaffe and M. Trajtenberg (eds.), *Patents, Citations & Innovations: A Window on the Knowledge Economy*, MIT Press, Cambridge MA, 287-335.
- Jaffe, A. B. and M. Trajtenberg, eds., (2002), *Patents, Citations & Innovations: A Window on the Knowledge Economy*, MIT Press, Cambridge MA
- Jaffe, A. B. and M. Trajtenberg (2002), “Flows of Knowledge from Universities and Federal Laboratories: Modeling the Flow of Patent Citations over Time and across Institutional and Geographic Boundaries”, Chapter 6 in A. B. Jaffe and M. Trajtenberg (eds.), *Patents, Citations & Innovations: A Window on the Knowledge Economy*, MIT Press, Cambridge MA, 179-197.
- Jaffe, A. B., M. Trajtenberg and M. S. Fogarty (2002), “The Meaning of Patent Citations: Report on the NBER/Case-Western Reserve Survey of Patentees”, Chapter 12 in A. B. Jaffe and M. Trajtenberg (eds.), *Patents, Citations & Innovations: A Window on the Knowledge Economy*, MIT Press, Cambridge MA, 380-401.
- Jaffe, A. B., M. Trajtenberg and R. Henderson (2002), “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations”, Chapter 5 in A. B. Jaffe and M. Trajtenberg (eds.), *Patents, Citations & Innovations: A Window on the Knowledge Economy*, MIT Press, Cambridge MA, 155-177.
- Kaplan, E. L. and P. Meier (1958), “Nonparametric Estimation From Incomplete Observations”, *Journal of the American Statistical Association*, 53, 457-481.

- Kennan, J. (1985), "The Duration of Contract Strikes in U.S. Manufacturing", *Journal of Econometrics*, 28, 5-28.
- Kiefer, N. M. (1988), "Economic Duration Data and Hazard Functions", *Journal of Economic Literature*, XXVI, 646-679.
- Kiefer, N. M., S. Lundberg and G. R. Neumann (1985), "How Long is a Spell of Unemployment", *Journal of Business and Economic Statistics*, 3, 118-128.
- Lancaster, T. (1990), *The Econometric Analysis of Transition Data*, Cambridge University Press, Cambridge.
- Nickell, S. J. (1979), "Estimating the Probability of Leaving Unemployment", *Econometrica*, 47, 1249-1266.
- Northwest Econometrics (2004), *SHAZAM Econometrics Software. User's Reference Manual, Version 10*, Northwest Econometrics, Vancouver BC.
- Popp, D. T. Juhl and D. K. N. Johnson (2004), "Time in Purgatory: Examining the Grant Lag for U.S. Patent Applications", *B.E. Journal of Economic Analysis and Policy*, 4 (Topics), Article 29.
- Quantitative Micro Software (2004), *EViews 5 User's Guide*, Quantitative Micro Software, Irvine CA.
- Shih, R. and D. E. Giles (2006), "Modelling the Duration of Interest Rate Spells Under Inflation Targeting in Canada", *Econometrics Working Paper EWP0605*, Department of Economics, University of Victoria.
- Sider, H. (1985), "Unemployment Duration and Incidence: 1968-82", *American Economic Review*, 75, 461-472.
- U.S. Patent and Trademark Office (2006), www.uspto.gov.
- Yu, Y. Y. (2005), *Modelling Exchange Rate Regime Durations*, M.A. Extended Essay, Department of Economics, University of Victoria.

Footnotes

* We are grateful to Nilanjana Roy and David Scoones for their helpful comments on an earlier version of this work.

1. We were unaware of this research until our own work was completed, and are grateful to Nilanjana Roy for bringing it to our attention. Popp *et al.* (2004, p.14) report that their data-set had to be cleaned to remove obviously erroneous data, resulting in a sample of 1,653,854 observations.

2. According to the U.S. Patent and Trademark Office (2006), there are three types of patents that are distinguished in the U.S.: “Utility patents may be granted to anyone who invents or discovers any new and useful process, machine, article of manufacture, or composition of matter, or any new and useful improvement thereof; Design patents may be granted to anyone who invents a new, original, and ornamental design for an article manufacture; and Plant patents may be granted to anyone who invents or discovers and asexually reproduces any distinct and new variety of plant.”

3. Using the individual patent numbers, we looked into these long duration patents on the U.S. Patent and trademark Office (2006) website and examined the individual file information. There were 243 items with application process time greater than 20 years, but all of these applications started before 1963, and we randomly selected 85% of these items and noticed that all of these patents were related to the US military, either directly or indirectly. Since the US-Cuba crisis occurred in 1962, we thus presume these outlier observations were due to the concern of the sensitive US-Soviet military confrontation in the 1960s-1970s, and hence all of these patents were kept “under wraps” for an extended period of time for security reasons. Current evidence indicates that there are special examination procedures for particular inventions. For example, according to the examination of applications for countering terrorism, in view of the importance of developing technologies for countering terrorism and the desirability of prompt disclosure of advances made in these fields, the USPTO will accord “special” status to patent applications for inventions which potentially materially contribute to countering terrorism. Thus, a similar situation was presumed to be applied in these earlier military-related inventions.

4. Citations made may constitute a paper-trail for spillovers; and citations received may indicate the importance of the cited patents. The number of times a patent document has been cited may be a measure of its technological significance. (Jaffe *et al.*, 2002, p.418.)
5. “Unassigned” patents are those for which the inventors have not yet granted the rights to the invention to a legal entity such as a corporation, university, or government agency, or to other individuals.
6. In contrast, Popp *et al.* (2004) found dummy variables for the grant year to be statistically significant, but this may be the result of their use of a log-linear model and least squares estimation.
7. With sample sizes in the millions of observations (as we also have in our own analysis), true significance is associated with much smaller p-values than we traditionally adopt – almost all covariates will be “significant” by the latter standards. (Granger, 2003).
8. These graphs and the median and model survival times noted below use $\hat{\theta} = \exp(-\hat{\beta}'x)$, evaluated at the sample means of the covariates.
9. This estimated median duration compares well with the sample median of 2.0 years. (The sample standard deviation is 0.981 years.)