NBER WORKING PAPER SERIES

PATENTS AS OPTIONS: SOME ESTIMATES OF THE VALUE OF HOLDING EUROPEAN PATENT STOCKS

Ariel Pakes

Working Paper No. 1340

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 1984

I have benefited from the comments of many individuals in the course of this study, among them John Bound, Zvi Griliches, Bronwyn Hall, Jerry Hausman, James Heckman, Charles Manski, Daniel McFadden, Andrew Meyers, Dvora Ross, and Mark Schankerman. I am particularly indebted to Charles Manski and Zvi Griliches for a series of discussions which contributed a great deal to the development of this paper; and to James Heckman and the participants in an informal seminar chaired by Charles Manski and Daniel McFadden for comments that facilitated the solution to various problems. This paper is an offshoot of ongoing research with Mark Schankerman. The research was supported by the NSF through grant PRA 81-08635. I am thankful to Andrew Meyers, Dvora Ross and Tom Abbott for superb programming assistance. All errors, of course, remain my responsibility. The research reported here is part of the NBER's research program in Productivity and project in Productivity (World Economy). Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper #1340
April 1984

Patents as Options: Some Estimates of the Value of Holding European Patent Stocks

ABSTRACT

In many countries holders of patents must pay an annual renewal fee in order to keep their patents in force. This paper uses data on the proportion of patents renewed, and the renewal fees faced by, post World War II cohorts of patents in France, the United Kingdom, and Germany, in conjunction with a model of patent holders' renewal decisions, to estimate the returns earned from holding patents in these countries. Since patents are often applied for at an early stage in the innovation process, the model allows agents to be uncertain about the sequence of returns that will be earned if the patent is kept in force. Formally, then, the paper presents and solves a discrete choice optimal stochastic control model, derives the implications of the model on aggregate behaviour, and then estimates the parameters of the model from aggregate data. The estimates enable a detailed description of the evolution of the distribution of returns earned from holding patents over their lifespans, and calculations of both; the annual returns earned from holding the patents still in force (or the patent stocks) in the alternative countries, and the distribution of the discounted value of returns earned from holding the patents in a cohort.

Dr. Ariel Pakes
Department of Economics
Hebrew University of Jerusalem
Mount Scopus Campus
91905 Jerusalem
Israel

In many countries holders of patents must pay an annual renewal fee in order to keep their patents in force. If the renewal fee is not paid in any single year, the patent is permanently cancelled. Assuming that renewal decisions are based on economic criteria, agents will only renew their patents if the value of holding those patents over an additional year exceeds the cost of renewal. Observations on the proportions of different cohorts of patents which are renewed at alternative ages, together with the relevant renewal fee schedules, will, in this case, contain information on the distribution of the values of holding patents, and on the evolution of this distribution function over the lifespan of the patents. Since patent rights are seldom marketed, this is one of the few sources of information on the value of patents available. This paper presents and then estimates a model which recovers the distribution of returns from holding patents at each age over the lifespan of patents from data on the renewal behaviour of, and the renewal fee schedules faced by, post World War II cohorts of patents in each of the United Kingdom, France, and Germany (renewal fees were not instituted in the United States until 1982). The parameters of these distribution functions enable a calculation of; the value, to patent holders, of the proprietary rights created by the patent laws; the distribution of this value among patents; and the process which determines the evolution of the value of patents over their lifespans.

This is not the first time patent renewal data has been used to estimate parameters of the distribution of patent values. In a previous paper (see Pakes and Schankerman, 1978) intercountry differences in the proportion of patents renewed and in renewal fee schedules faced by cohorts of European patents were used to estimate the rate of obsolescence on the returns from holding patents. The earlier paper assumed that cohorts of patents were endowed with a distribution of initial current returns which decayed deterministically thereafter. Methodologically, the major innovation in this paper is that it does not assume that the sequence of returns that will accrue to the patent if it is to be kept in force are known with certainty at the time the patent is applied for. The generalization to an uncertain sequence of returns is to allow for the fact that agents often apply for patents at an early stage in the innovation process, a stage in which the agent is still exploring alternative opportunities for earning returns from use of the information embodied in the patented ideas. In part early patenting arises from the incentive structure created by the patent system, since, if the agent does not patent the information available to him , somebody else might. This incentive is reinforced by the fact that the renewal fees in all countries studied are quite small during the early ages of a patent's life.

A patent holder who pays the renewal fee obtains both; the current returns that accrue to the patent over the coming period, and the option to pay the renewal fee and maintain the patent in force in the following period should he desire to do so. An agent who acts optimally will pay the renewal fee only if the sum of the current returns plus the value of this option exceeds the renewal fee. It will be assumed that the agent values the option at the expected discounted value of future net returns (current returns minus renewal fees), taking account of the fact that an optimal policy will be followed in each future period, and conditional on the information currently at the disposal of the agent. Current decisions, therefore, depend on both current returns, and on the distribution of future returns conditional on current information. An optimal sequential policy for the agent has the form of an optimal renewal (or stopping) rule; a rule determining whether to pay the renewal at each age. The proportion of patents who drop out at age a are the proportion who do not satisfy the renewal criteria at that age, but who did at age a-1. The drop out proportions predicted by the model will be a function of the precise value of the vector of the model's parameters, and of the renewal fee schedules. The data provide the actual proportion of drop outs. Roughly speaking, the estimation problem is to find that value of the vector of the model's parameters which makes the drop out proportions implied by the model as "close" as possible to those we actually observe.

Formally then, this paper presents and solves a dynamic discrete choice optimal stochastic control model, derives the implications of this model on aggregate behaviour, and then estimates the parameters of the model from aggregate data. Dynamic discrete choice models have appeared in the economic literature in several contexts (see for eg. Roberts and Weitzman, 1981); and a small number of them have actually been estimated on micro data. Miller (1982) estimates such a model for the length of job matches, and Wolpin (1982) estimates one for the birth sequences of married women. All three of these models have a range of applications and provide an extremely rich interpretation to the data. They are each, however, based on quite different stochastic specifications (and rightly so given the diversity of the empirical problems they deal with) and, due to the complexity of the estimation problem, it is difficult to determine the robustness of the conclusions to the particular stochastic assumptions chosen. The model used here embeds a Markov assumption, an assumption that the distribution of the next period's return conditional on current information depends only on current returns and the parameters of the problem, in a search model with three types of outcomes. Each year the agents perform experiments to explore alternative ways of best exploiting their patented ideas. One possible outcome of these experiments is that they provide no new information , another is that they determine that the patented ideas can never be profitably exploited, and the third is that the experiments indicate a use which allows the agent to increase the returns which accrue to the patent at subsequent ages. The conditional distribution of beneficial outcomes, should they occur, is not assumed, a priori, to be stationary over ages. This non-stationarity is to allow for the possibility that agents explore their most promising alternatives first; a possibility which is distinctly favored by the data. In addition, since there is a statutory limit to patent lives (an age beyond which the agent cannot keep the patent in force by payment of an annual fee), the model has a finite horizon.

Given our assumptions, it is possible to obtain an explicit solution for the renewal rule as a function of the parameters of the Markov process, the age of the patent, and the renewal fee schedules. However, the model is not as benevolent with respect to the calculation of the aggregate drop out probabilities. To allow for heterogeneity, it is assumed that there is a distribution of initial returns among patents. This distribution is modified over time as agents uncover more profitable ways of exploiting their patented ideas. The distribution of returns at each age does not have, to the best of my knowledge, an analytical form, and, as a consequence, neither do the drop out probabilities. I therefore resort to the simulated frequency approach, suggested by Lerman and Manski (1981), to estimate these probabilities for different values of the parameter vector.

The assumptions of the model, together with the parameter estimates, enable a detailed description of the evolution of the distribution of the current returns earned from holding patents over their lifespans. This information is used to characterize the learning process and to calculate both; the annual returns earned from holding the patents still in force (or the patent stocks) in the alternative countries, and the distribution of the discounted values of returns earned from holding the patents in a cohort. I consider more general aspects of the empirical results, those related to modelling the inventive process and to the private value generated by the patent system, in the final section of the paper.

Section 1 provides an overview of the model, while Section 2 fills in the specific details of its stochastic specification. Section 3 provides a description of how the parameter estimates are actually obtained. The data is described in Section 4, and Section 5 presents and interprets the parameter estimates. Section 6, which closes the paper, provides a brief discussion of some more general implications of the empirical results. An accompanying appendix, which has three parts, deals with different technical points that arise in the course of obtaining the parameter estimates.

Section 1. A Description of the Model

This section provides an overview of the renewal model used in this paper. It begins by considering the decision problem faced by an agent who holds a patent, and ends with the likelihood function implied by our assumptions.

The agent's problem is to decide on whether to pay a renewal fee which will keep the patent in force over the coming year. If the renewal fee is not paid, the patent is permanently cancelled. If the renewal fee is paid and the age of the patent is less than the statutory limit to patent lives, the agent will face a similar problem at the beginning of the next year. If the patent's age equals the statutory limit to patent lives, the current is the last year the agent can keep the patent in force by payment of a renewal fee.

Agents are asssumed to maximize the expected discounted value of the net returns from their actions, and may be uncertain about the sequence of returns that will be earned if the patent is kept in force. This uncertainty allows for the possibility that, at least during the early years of a patent's life, the agent is actively exploring alternative ways to best exploit the ideas embodied in the patent. An implication is that there is a positive probability that the agent will discover a use for the patented ideas which makes future returns to patent protection significantly higher than those being currently earned, and this probability may induce the agent to pay the current renewal fee even if current returns are lower than the cost of renewal.

Let $V_{i,a}$ be the expected discounted value of patent protection to the holder of the i^{th} patent just prior to its a^{th} renewal. If the renewal fee is not paid the patent lapses and $V_{i,a} = 0$. If the renewal fee is paid the agent earns the current return to patent protection and, in addition, maintains the option to renew and keep the patent in force at age $a + 1$. The value of this option to the agent equals the expected discounted value of the patent at age $a + 1$ conditional on the agent's current information set. Formally then,

$$V_{i,a} = \max \{0, r_{i,a} - c_a + \beta E [V_{i,a+1} | \Omega_{i,a}]\} \qquad (1)$$

for all i, and a = 1, ..., L; where L is the statutory limit to patent lives, $r_{i,a}$ is the current returns to patent protection, c_a is the cost of renewal, β is a discount factor, $\Omega_{i,a}$ represents the information set of the agent in the patents a^{th} year, and it is to be understood that zero is an absorbing state in the stochastic process generating $\{V_{i,a}\}_{a=1}^{L}$ (so that if the patent is not renewed at any age it will not be in force thereafter). In equation (1), $r_{i,a} - c_a + \beta E [V_{i,a+1} | \Omega_{i,a}]$ is the value of holding the patent over its a^{th} year. If this expression is negative, the agent will allow the patent to lapse.

To complete the description of the value function we need to specify the conditional distributions of future returns and costs of renewal that are held by the agent. Given these distributions, the solution for the sequence $\{V_{i,a}\}_{a=1}^{L}$ is found by starting with the terminal equation, that is $V_{i,L} = \max \{0, r_{i,L} - c_L\}$, and integrating the system in (1) backwards recursively. Assumptions A1 and A2 provide the general properties of these distribution functions.

A1. $G_a(r_{i,a+1} | \Omega_{i,a}) = G_a(r_{i,a+1} | r_{i,a}, \underset{\sim}{\omega}_g)$ for all i and a=1, ..., L-1; where $G_a(r_{i,a+1} | \Omega_{i,a})$ defines the distribution of $r_{i,a+1}$ conditional on $\Omega_{i,a}$, and $\underset{\sim}{\omega}_g$ is a vector of known parameters.

A2. Agent's hold point expectations on the renewal fees that will be required to keep the patent in force at later ages equal to the current real renewal fees for those ages.

These assumptions simplify the analysis considerably. A2 was motivated by the fact that the renewal fee schedules are published data, and though these schedules are changed periodically, the real renewal fee at any age does not vary much with the year the patent reaches that age. I will assume an exogenously given initial distribution of current returns to patent protection [of $r_{i,1}$]. A1 assumes that the stochastic process generating subsequent sections (i.e. generating $\{r_{i,a}\}_{a=2}^{L}$) is both Markov, and invariant over i. It should be noted that, in addition to A1 and A2, the proofs of the propositions required for our estimation technique use both the empirical fact that all renewal fee schedules are nondecreasing in age (see section 4), and the precise functional form of $\{G_a(\cdot|\cdot)\}_{a=1}^{L-1}$ (see the next section). For an understanding of the general characteristics of the model, however, we need only point out two properties of this sequence of distribution functions. First the probability that the coming year's returns will be greater than a given number is higher the larger are current returns; or if $z_2 > z_2'$, then $G_a\ (z_1 \mid z_2, \underset{\sim}{\omega}_g) <$ $G_a\ (z_1 \mid z_2', \underset{\sim}{\omega}_g)$, for $z_1, z_2, z_2' \in R^+$, and $a=1, \ldots, L-1$. Second, though the solution to the agents decision problem provided in this section does require certain restrictions on the evolution of the sequence of conditional distributions, of $G_a(\cdot \mid \cdot)$, over age (see the next section), it does not require stationarity. This non-stationarity of the stochastic process generating $\{r_{i,a}\}$ turns out to be an important feature of the empirical results.

Note that A1 and A2 imply that the expected value of the option to renew the patent at age a + 1 depends only on current returns ($r_{i,a}$), the parameters of the Markov process generating future returns ($\underset{\sim}{\omega}_g$), β, and the current vector of renewal fees ($\underset{\sim}{c}^a$), that is; $E\ [V_{i,a+1} \mid \Omega_{i,a}] = E\ [V_{i,a+1} \mid r_{i,a}, \underset{\sim}{c}^a, \underset{\sim}{\omega}_g]$, where it is to be understood that this, and subsequent functions, depend also on the discount factor, β. The system in (1) can therefore be rewritten as

$$V\ (a,r;\ \underset{\sim}{c}^a,\ \underset{\sim}{\omega}_g) = \max\ \{0,\ B\ (a,r;\ \underset{\sim}{c}^a,\ \underset{\sim}{\omega}_g) - c_a\} \qquad (2)$$

where $B(a,r; \underset{\sim}{c}^a, \underset{\sim}{\omega}_g) = r_a + \beta E [V_{a+1} \mid r_a, \underset{\sim}{c}^a, \underset{\sim}{\omega}_g]$, and the subscript i is omitted for convenience; for a=1, ..., L-1, and $r \varepsilon R^+$. B(a,r) provides the total benefits from holding the patent for an additional year (the sum of current returns and the value of the option).

The solution to the agents decision problem follows directly from the properties of this benefit function. These properties are provided in the following proposition, and explained immediately thereafter.

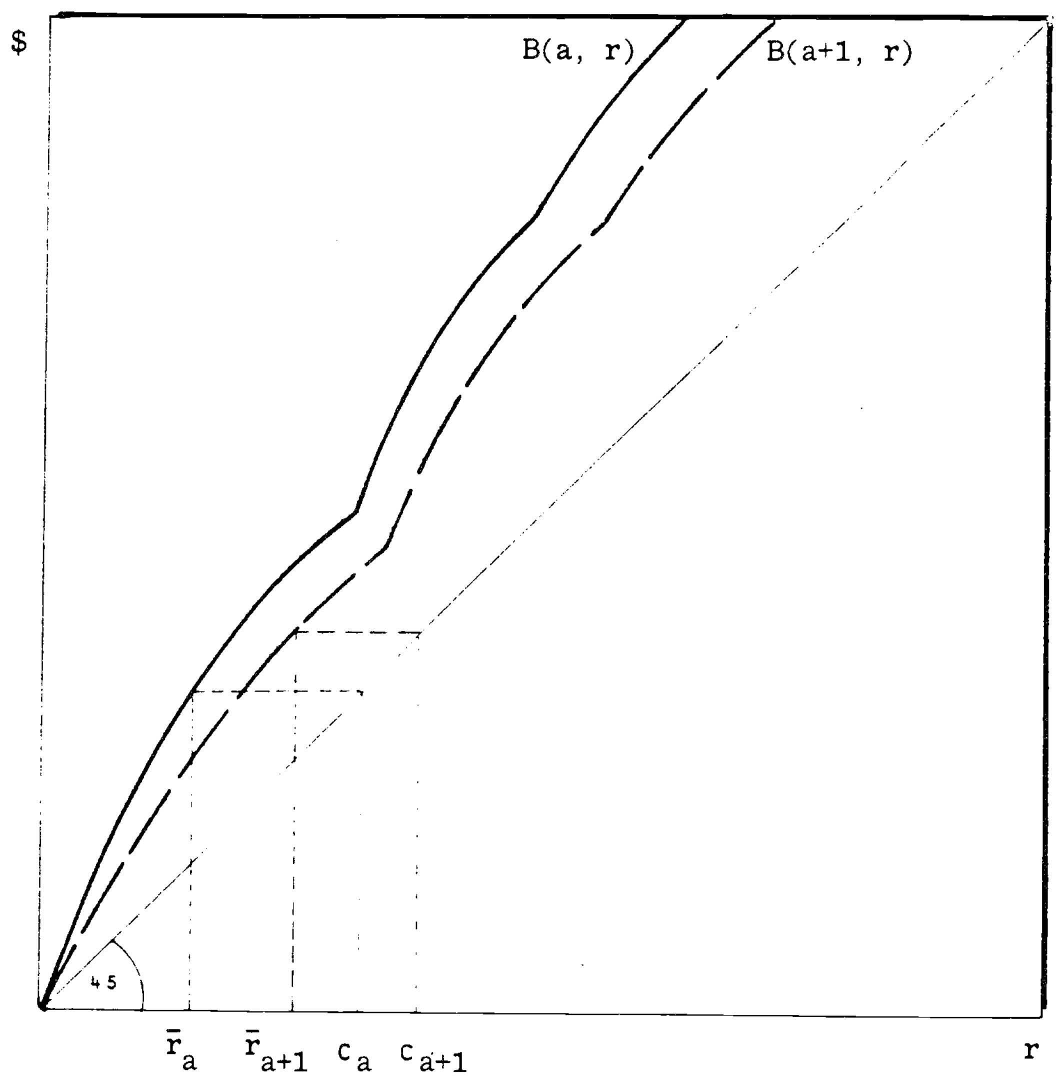
Proposition 1 (proved in Appendix 1). The value of the option, that is B(a,r) - r, is; uniformly continuous and nondecreasing in r, and is nonincreasing in a, for $r \varepsilon R^+$ and a=1,...,L.

Figure 1 illustrates the form of B(a,r). Since V(a+1, r) ≥ 0 with probability one, the expected value of the option to renew is nonnegative and B(a,r) ≥ r (the 45 degree line); while the fact that the probability that future returns will be above a given number is larger the higher are current returns implies that the value of the option [the difference between B(a,r) and the 45 degree line] is nondecreasing in r. As the patent ages there are less future years in which the patent can earn returns, and renewal fees rise. Either of these facts is, in general, sufficient to insure that B(a,r) - r decreases in age. Note that though B(a,r) is continuous in r everywhere, there are points at which it is not differentiable in r (see Appendix 3).

Equation (2) implies that it is in the agent's interest to pay the renewal fee if $B(a,r) > c_a$. The following corollary of proposition 1 provides an optimal renewal (or stopping) rule for the agent.

Corollary 1 (proved in Appendix 1 and illustrated in Figure 1). For each age there exists a unique $\bar{r}_a \varepsilon [0, c_a]$, such that it is optimal for the agent to renew the patent if and only if $r_a > \bar{r}_a$. Moreover, the sequence $\{\bar{r}_a\}_{a=1}^{L}$ is nondecreasing in age.

Figure 1. Determining $\{\bar{r}_a\}_{a=1}^{L}$



The first sentence in this corollary provides a simple renewal criteria. The patent ought to be renewed only if current returns are greater than the cutoff, $\bar{r}_a$. Note that $\bar{r}_a < c_a$, so that in general the difference $c_a - \bar{r}_a$ is positive. If $r_a \varepsilon (\bar{r}_a, c_a)$ it is optimal for the agent to take a loss in current net returns ($r_a - c_a < 0$) in order to maintain the option of patent protection in the future. This is one difference between a myopic model, wherein returns decay deterministically over time and an agent would not renew unless $r_a > c_a$, and the stochastic model. It can be shown that the difference between the renewal fee and the cutoff, i.e., $c_a - \bar{r}_a$, is nondecreasing in the current renewal fee (c_a), nonincreasing in the renewal fees for later ages ($c_{a+\tau}$, $\tau > 0$), and, at least in the later ages, nonincreasing in age (since L is the last year the patent can be kept in force $c_L - \bar{r}_L = 0$). The fact that the renewal fees are increasing in age, while B(a,r) is decreasing, implies that the cutoffs are nondecreasing in age. Clearly the cutoffs are functions of; age, $\underset{\sim}{\omega}_g$, and $\underset{\sim}{c}^a$, or $\bar{r}_a = r(a, \underset{\sim}{c}^a, \underset{\sim}{\omega}_g)$, for a = 1, ..., L.

It is now straightforward, at least conceptually, to determine the proportion of patents who drop out, that is who stop paying the renewal fee, at each age. First note that the distribution of initial returns [which we denote by $F_1(r; \underset{\sim}{\omega}_1)$], the stochastic process generating subsequent returns (A1), the renewal fee schedules, and the renewal rule (corollary 1), determine the distribution of returns at each age, say $F_a(r; \underset{\sim}{c}, \underset{\sim}{\omega})$; where $\underset{\sim}{\omega}' = [\underset{\sim}{\omega}_g', \underset{\sim}{\omega}_1']$, (that is $\underset{\sim}{\omega}$ contains the parameters of the Markov process and of the initial distribution of returns), $\underset{\sim}{c}$ is a vector consisting of the renewal fee schedules faced at each age, and, formally

$$1-F_a(r; \underset{\sim}{c}, \underset{\sim}{\omega}) =$$

$$\int_0^\infty \cdots \int_{x_{a-1}=\bar{r}_{a-1}(\underset{\sim}{c}^a, \underset{\sim}{\omega}_g)} \int_{x_a=r} dG_{a-1}(x_a \mid x_{a-1}, \underset{\sim}{\omega}_g) dG_{a-2}(x_{a-1} \mid x_{a-2}, \underset{\sim}{\omega}_g) \cdots dF_1(x_1, \underset{\sim}{\omega}_1)$$

for $r \in R^+$ and $a=1, \ldots, L$. From corollary 1 the proportion of patents who pay the renewal at age a is the proportion with current returns above $\bar{r}_a$ or $1-F_a(\bar{r}_a; \underset{\sim}{c}, \underset{\sim}{\omega})$; while the proportion who drop out at age a, say $\pi(a; \underset{\sim}{c}, \underset{\sim}{\omega})$, is simply the difference between the proportions not paying the renewal fee at age a, and those not paying the renewal fee at age a-1, or

$$\pi(a; \underset{\sim}{c}, \underset{\sim}{\omega}) = F_a\ (\bar{r}_a; \underset{\sim}{c}, \underset{\sim}{\omega}) - F_{a-1}\ (\bar{r}_{a-1}; \underset{\sim}{c}, \underset{\sim}{\omega}), \tag{4}$$

for $a=1, \ldots, L$. Note that $\pi(a; \underset{\sim}{c}, \underset{\sim}{\omega})$ is calculated as the difference of quantiles on two different distribution functions. This is a result of a second difference between the myopic and stochastic models; in the stochastic model the distribution of r changes in a non-trivial manner over age as agents uncover more profitable ways of using their patented ideas.

Equation (4) provides the theoretical probabilities required to calculate the likelihood function implied by the model. In order to formulate this likelihood function explicitly, we require a brief description of the data (section 4 provides more detail on the data set). The data contain information on different cohorts of patents, where a cohort is defined by the year the patent was applied for. For some of these cohorts we do not observe the patents dropping out at later ages, and for some we do not observe those dropping out at earlier ages (there is censoring from both the left and the right). Let the index j distinguish between alternative cohorts, f_j and l_j be the first and last ages at which we observe the number of patents paying the renewal for cohort j, and $A_j = \{f_j, f_j+1, \ldots, l_j, l_j+1\}$, for $j=1, \ldots, J$. Then, for each j, the data contain: i) the sequence $\{n(a,j)\}_{a \in A_j}$ where ; $n(f_j,j)$ denotes the number of patents who did not pay the renewal at f_j, $n(a,j)$ for $f_j < a \leq l_j$ denotes the number of patents who stopped paying the renewal at each subsequent age until (and including) l_j, and $n(l_j+1,j)$ denotes the number of patents which were still in force after l_j; and, (ii) the vector of the renewal fee schedules faced by the cohort, or $\underset{\sim}{c}_j$.

Now consider a patent drawn randomly from a given cohort. It will either drop out by age f_j, drop out at a subsequent age before l_j+1, or still be in force after l_j. Equation (4) implies that for each j the probabilities of these mutually exclusive and exhaustive alternatives are given by:

$\pi\ (f_j;\ \underset{\sim}{c}_j,\ \underset{\sim}{\omega}) = \Sigma_{a=1}^{f_j}\ \pi\ (a;\ \underset{\sim}{c}_j,\ \underset{\sim}{\omega})$; $\pi\ (a;\ \underset{\sim}{c}_j,\ \underset{\sim}{\omega})$ for $a = f_j+1, \ldots, l_j$; and $\pi\ (l_j+1) = 1 - \Sigma_{a=1}^{l_j}\ \pi\ (\underset{\sim}{a};\ \underset{\sim}{c}_j,\ \omega)$, respectively. Given this definition of $\{\pi(a\ \underset{\sim}{c}_j, \underset{\sim}{\omega}_j)\}_{a\varepsilon A_j}$ the (log) likelihood of a particular value of the parameter vector conditional on the observed data, or $\ell(\underset{\sim}{\omega})$, is

$$\ell(\omega) = \sum_{j=1}^{J} \sum_{a\varepsilon A_j} n(a,j) \log \pi\ (a;\ \underset{\sim}{c}_j,\ \underset{\sim}{\omega}) \tag{5}$$

The empirical results presented in section 5 are based on maximizing this likelihood with respect to $\underset{\sim}{\omega}$. Letting n_j be the total number of patents in cohort j and $N = \sum_{j=1}^{J} n_j$, the limiting (as $N\to\infty$) properties of the maximum likelihood estimator are provided in proposition 2.

<u>Proposition 2</u> (proved in Appendix 2). Let $\underset{\sim}{\omega}^*$ be the maximum likelihood estimator of $\underset{\sim}{\omega}$ defined by the equation, $\ell(\underset{\sim}{\omega}^*) = \sup_{\underset{\sim}{\omega}\varepsilon T} \ell(\underset{\sim}{\omega})$; where T is a subset of R^k containing, in its interior, the true value of $\underset{\sim}{\omega}$, say $\underset{\sim}{\omega}^o$. Then, provided an identifiability and invertibility condition are met (see Appendix 2), $\underset{\sim}{\omega}^*$ converges in probability to $\underset{\sim}{\omega}^o$ as $N \to \infty$, holding $\{w_j = n_j/N\}_{j=1}^{J}$ constant, and

$$\sqrt{N}\ (\underset{\sim}{\omega}^* - \underset{\sim}{\omega}^o) \xrightarrow{D} \eta(0,\ [i^o_{r,s}]^{-1}),$$

where $\xrightarrow{D}$ reads converges in distribution, $\eta(.,.)$ denotes the multivariate normal distribution, and $[i_{r,s}]$ denotes the information matrix evaluated in general as

$$i_{r,s} = \sum_{j=1}^{J} w_j \sum_{a\varepsilon Aj} \frac{1}{\pi(a,\underset{\sim}{c}_j,\ \underset{\sim}{\omega})} \frac{\partial\pi(a;\ \underset{\sim}{c}_j,\ \underset{\sim}{\omega})}{\partial\omega_r} \frac{\partial\pi(a;\ \underset{\sim}{c}_j,\ \underset{\sim}{\omega})}{\partial\omega_s}$$

for r,s=1, ..., k; and $[i^o_{r,s}]$ denotes this matrix evaluated at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^o$.

Two points should be noted here. First the dimension in which the properties of $\underset{\sim}{\omega}^*$ approach those provided in proposition 2 is N, the sum of the number of patents in the J cohorts, and as section 4 shows, N is unusually large in our samples. Second, the limiting distribution of $\underset{\sim}{\omega}^*$ follows from a proposition due to Rao (1973, section 5.e.2), and the fact that the functions $\pi(a;\ \underset{\sim}{c}_j,\ \underset{\sim}{\omega})$ $[a\varepsilon A_j,\ j = 1,\ ...,\ J]$ admit first order partials which are continuous at $\underset{\sim}{\omega}^o$ [since the benefit function in equation (2) is not differentiable everywhere, this statement is not immediately obvious]. This same property together with the consistency of the maximum likelihood estimator insure that $[i^*_{r,s}]$, the information matrix when evaluated at $\underset{\sim}{\omega}^*$, is a consistent estimate of $[i^o_{r,s}]$, and, as a result, $[i^*_{r,s}]^{-1}$ is used to estimate the variance-covariance matrix of the parameter estimates.

To complete the specification of the model we require a detailed description of both the Markov process generating the returns from holding a patent, and of the distribution of initial returns. This is provided in the next section. Section 3 explains the procedure used to obtain the maximum likelihood estimates and the information matrix.

Section 2. The Stochastic Process Generating $\{r_a\}_{a=2}^{L}$ and the Distribution of Initial Returns

Equation (6), and the explanation which follows it, describe the Markov process assumed to generate the returns from holding a patent. The conditional distribution of r_{a+1} is defined by

$$r_{a+1}\Big|\, r_a = \begin{cases} 0 & \text{with probability} \quad \exp(-\theta r_a) \\ \max\{\delta r_a, z\} & \text{with probability} \quad 1-\exp(-\theta r_a) \end{cases} \tag{6}$$

where the density of z, $q_a(z)$, is a two parameter exponential, that is

$$q_a(z) = \sigma_a^{-1}\exp[-(\gamma + z)/\sigma_a],$$

and $\sigma_a = \phi^{a-1}\sigma$; for $a=1,\ldots,L-1$.

One advantage of the process specified in (6) is that is permits an explicit solution for the sequence $\{\bar{r}_a\}_{a=1}^{L}$ as a function of the parameter of the model (see the next section). This process also has the following economic interpretation. At each age agents perform experiments designed to enable them to increase the profits from their patented ideas. These experiments can have one of three types of outcomes. First, they may reveal that the patented ideas can never be profitably exploited. This event occurs with probability $\exp(-\theta r_a)$, that is it occurs with smaller probability the larger are the current returns from holding the patent; and if such an outcome does materialize the agent does not pay a renewal fee in the following year (the zero state is an absorbing state in the stochastic process generating current returns, which implies that if it is drawn the agent will let the patent lapse). The second possible outcome is that the absorbing state does not occur, but the experiments do not result in a use for the patented ideas which is more profitable than the current one. In this case current returns decay at the rate $\delta < 1$, as steps forward by other agents in the economy gradually obsolete the returns from the agent's own patent, and the agent must decide whether current returns and/or the possibility of discovering a use which may increase those returns in the future, make it worthwhile to pay the next renewal fee. Finally, the experiments may actually uncover a use for the patented ideas which improves upon the returns which could have been generated with the information of the previous year (the absorbing state does not occur and $z > \delta r_a$). The extent of the improvment depends on the precise realization of z. This random variable has a two parameter exponential distribution, that is; z has probability $\exp(-\gamma/\sigma_a)$ of being greater than zero (experiments do not necessarily lead to outcomes which yield positive returns), and has a density which declines at the constant rate σ_a thereafter. Note that $\sigma_a = \phi^{a-1}\,\sigma$. With $\phi \leq 1$ this allows for the possibility that the probability of uncovering a use which leads to returns greater than a given number declines over age; or for the possibility that agents perform their best experiments first. $\phi \leq 1$ is also a sufficient condition for proposition 1 of the last section.

We have now defined the stochastic process generating the distribution of $(r_2, r_3, \ldots, r_L)$ from the distribution of r_1. Note that this process is a member of a five parameter family, that is $\underset{\sim}{\omega}_g' = (\theta, \gamma, \sigma, \delta, \phi)$. To complete the specification of the model we require also a distribution of initial returns over different patents, that is we require $F_1(r; \underset{\sim}{\omega}_1)$. It is assumed that initial returns distribute lognormally, or

$$\log r_1 \sim \eta\,(\mu, \sigma_R) \qquad (7)$$

This implies that $\underset{\sim}{\omega}_1 = (\mu, \sigma_R)$; so that $\underset{\sim}{\omega}' = (\underset{\sim}{\omega}'_g, \underset{\sim}{\omega}'_1)$ contains seven parameters.

Equations (6) and (7) complete the specification of the model outlined in Section 1. The next section contains a brief description of how the maximum likelihood estimate of $\underset{\sim}{\omega}$, that is $\underset{\sim}{\omega}^*$, was actually obtained, while Section 4 describes the data.

Section 3. Obtaining $\underset{\sim}{\omega}^*$

Three technical problems must be solved before we can obtain $\underset{\sim}{\omega}^*$. First a method must be provided to calculate the cutoffs, or the sequence $\{\bar{r}_a\}_{a=1}^{L}$ as defined in corollary 1, as a function of $\underset{\sim}{c}$ and $\underset{\sim}{\omega}$. Given $\underset{\sim}{\omega}$, these cutoffs determine the drop out probabilities, or the sequence, $\{\pi(a;.)\}^{L}_{a=1}$ as defined in equation (4), which in turn determine the likelihood of $\underset{\sim}{\omega}$ (see equation 5). The second problem, then, is to provide a method which calculates the drop out probabilities correponding to particular values of $\underset{\sim}{\omega}$ and $\{\bar{r}_a\}_{a=1}^{L}$. Finally, a maximization algorithm which finds that value of $\underset{\sim}{\omega}$ that maximizes the likelihood is required. I now consider each of these problems in turn.

Appendix 3 develops a recursive system of analytic equations which solves for the sequence $\{\bar{r}_a = r(a;\underset{\sim}{\omega},\underset{\sim}{c})\}_{a=1}^{L}$. This sytem is obtained by solving for the benefit function in an interval containing $\bar{r}_a$ at each age.[1] The cutoffs corresponding to particular values of $\underset{\sim}{\omega}$ and $\underset{\sim}{c}$ were obtained by simply substituting these values into the system defined in this appendix.

---

[1] Briefly, this problem is first reduced to a more manageable one by expressing B(a,r), for each age, as the sum of L-a component functions. The component functions for age a are definite integrals of the component functions at age a+1 where the limits of integration are determined by the value of r and by the subsequent cutoffs (by $\bar{r}_{a+\tau}$, for $\tau=1,\dots,L-a$). This fact leads to a functional recursion which can be solved using Macsyma (1983; Macsyma is a computer programme designed for symbolic mathematical manipulations) to produce the recursive system of analytic equations for $\{\bar{r}_a\}$. The continuity of the benefit function together with the features of Macsyma enable a check of the Macsyma results for possible programming errors. Finally, the solution can be simplified further by noting that the values of the component functions, evaluated at $\bar{r}_a$, must lie between two simple functions of the parameters of the model. These boundary functions become progressively closer together for the later functions at each age and can, therefore, be used to form an approximation whose error must lie in an easily calculable range. The Macsyma results for this problem were obtained by Andrew Myers and myself.

One cannot, to the best of my knowledge, obtain the drop out probabilities as analytic functions of $\underset{\sim}{\omega}$ and $\{\bar{r}_a\}_{a=1}^{L}$. As a result the simulated frequency approach, suggested by Lerman and Manski (1981), was used to obtain estimates of these probabilities. The simulation estimator of $\{\pi(a;.)\}_{a=1}^{L}$, say $\{\hat{\pi}(a;.)\}_{a=1}^{L}$, is found by taking pseudo random draws from the distribution of initial returns defined by equation (7) and $\underset{\sim}{\omega}_1$, passing each through the stochastic process defined by equation (6) and $\underset{\sim}{\omega}_g$, and calculating the proportion with $r_{a-1} > \bar{r}_{a-1}$ but $\bar{r}_a < r_a$, for a=1,...L [see the definition of $\pi(a;.)$ in equation 4].[2] Let NSIM be the number of pseudo random draws used to evaluate the simulated frequencies. It is well known that $\hat{\pi}(a;.)$ converges almost surely, in NSIM, to $\pi(a;.)$ and has variance equal to $\pi(a;.)[1-\pi(a;.)]/\text{NSIM}$, $(a=1,\ldots,L)$. Define the pseudo likelihood of $\underset{\sim}{\omega}$, say $\hat{\ell}(\underset{\sim}{\omega})$, to equal that value of the likelihood function otained from substituting the simulated for the actual frequencies in equation (5). $\underset{\sim}{\omega}^*$ was obtained by maximizing $\hat{\ell}(\underset{\sim}{\omega})$ with respect to $\underset{\sim}{\omega}$. The information matrix was obtained by perturbing each parameter by one percent from $\underset{\sim}{\omega}^*$, calculating the implied derivatives of the simulated frequencies, and substituting these derivatives into the formula for the information matrix provided in proposition 2. The NSIM used in the final round of the maximization subroutine was twenty thousand (see the next paragraph and section 5); and the change from an NSIM of ten thousand, to an NSIM of twenty thousand, did not have a perceptible effect on the estimates.

Evaluating the simulated frequencies at a given value of $\underset{\sim}{\omega}$ is a computer time intensive task; the CPU time for a given evaluation being approximately linear in NSIM. A maximization subroutine for a problem involving simulated frequencies should, therefore, conserve on the number of times it evaluates the likelihood function at large NSIM. The subroutine used here varied NSIM within

---

[2] The computer programme to perform the simulation was designed by Bronwyn Hall and myself, and her assistance was, as always, gratefully appreciated.

each run. It was developed by modifying a programme entitled QNMDIF (a quasi Newton method for obtaining the maximum of a function of k variables available from the National Physics Laboratory, 1983; see also Gill, Murray, and Wright, 1981). The j^{th} round of the subroutine was defined by an NSIM, say NSIM(j), and a perturbation vector, say $\Delta\underset{\sim}{\omega}^j = [\Delta\omega_1^j, \ldots, \Delta\omega_k^j]$. The modifications made to QNMDIF directed it to find, with a relatively small number of function evaluations, an $\underset{\sim}{\omega}$, say $\underset{\sim}{\omega}^j$, such that $\hat{\ell}_j(\underset{\sim}{\omega}^j) > \hat{\ell}_j(\omega_1^j, \ldots, \omega_i^j \pm \Delta\omega_i^j, \omega_{i+1}^j, \ldots, \omega_k^j)$, for i=1...k. The J+1 round used $\underset{\sim}{\omega}^j$ as a starting value, an increased NSIM [NSIM(j+1) > NSIM(j)], and a perturbation vector with smaller components ($\Delta\omega_i^{j+1} < \Delta\omega_i^j$; i=1,...,k). The final two rounds used an NSIM of ten and twenty thousand, respectively, and a perturbation vector equal to one percent of the starting value of $\underset{\sim}{\omega}$.[3]

That completes the description of both the model and the estimation technique. The next section describes the data set, while section 5 presents and interprets the parameter estimates.

---

[3] This maximization subroutine was developed by Dvora Ross and myself. Two of the modifications we made to QNMDIF turned out to be particularly important. First to find the gradient vector for each iteration we used the 2k function evaluations obtained from changing each component of the parameter vector by positive and negative values of that component of the perturbation vector. If both perturbations with respect to a parameter resulted in function values less than the starting value for the iteration, the derivative with respect to that parameter was set equal to zero. If not, the derivative was set equal to that implied by the function evaluations. Second the stepsize search was modified so that function values corresponding to small differences in stepsize were not calculated. I am grateful to the staff of the Hebrew University computing center for their help in allocating computer time to us.

Section 4. The Data

The data used in this study were obtained directly from the patent offices of France, Germany, and the United Kingdom (the U.K.) by Mark Schankerman and myself.[4] Table 1 summarizes some of the characteristics of this data.

Row 1 of the table provides the first age at which a renewal fee is due, or f. There is no information on renewals for ages less than f and the renewals at age f reflect events that have occurred over the first f ages. In the U.K. then, the first age at which we have information on the drop outs resulting from events that have occurred over the previous year is a=6. Rows 2,3, and 4 provide, respectively; the last age at which a patent can be kept in force by payment of a mandatory renewal fee (L), the dates of application for the cohorts studied, and the years in which renewals are observed.[5] In all countries, then, we have at least partial information on the renewal behaviour of cohorts applied for in most of the 1950's, throughout the 1960's, and in the early 1970's. The required renewal fee schedules (see assumption 2, or A2, in section 1) were obtained in nominal domestic currency, converted to real domestic currency using the country's own implicit G.N.P. deflator, and then transferred into 1980 U.S. dollars using the official exchange rate in 1980. All monetary values are, therefore, in 1980 U.S. dollars.

Rows 5 and 6 illustrate an important intercountry difference in the characteristics of the data. In France and the U.K. the data include all the patents

---

[4] This data set will be described in more detail in a paper we are currently writing. We are indebted to the respective patent offices for providing us with the data and graciously answering our subsequent queries.

[5] Post world war Germany allowed reapplication of patents previously applied for. By 1952 these were less than 1% of German applications, and this explains the choice of 1952 for the starting cohort for Germany. The French patent office only provided information on renewals between 1970 and 1981. Given the values of f and L in France, this implies that the data contain partial information on the renewal behavior of cohorts applied for between 1951 and 1979 in that country. In light of these facts, I decided to use only post 1950 cohorts for the analysis of the U.K. L was changed to 20 in 1976 in Germany, and in 1980 in the U.K., and this explains the final renewal years for these countries.

Table 1. Characteristics of the Data[*a]

Country Characteristic		France	U.K.	Germany
1.	f	2	5	3
2.	L	20	16	18
3.	Application Dates of Cohorts	1951-79	1950-74	1952-72
4.	First/Last year in which renewals are observed	1970/81	1955/78	1955/74
5.	Patents Studied from Cohort: all patents	Applied for	Applied for	Granted
6.	Estimated Average Ratio of Patents Granted to Patents Applied for [*b]	.93	.83	.35
7.	$\overline{NPAT} = N/J$	36,865	37,286	21,273

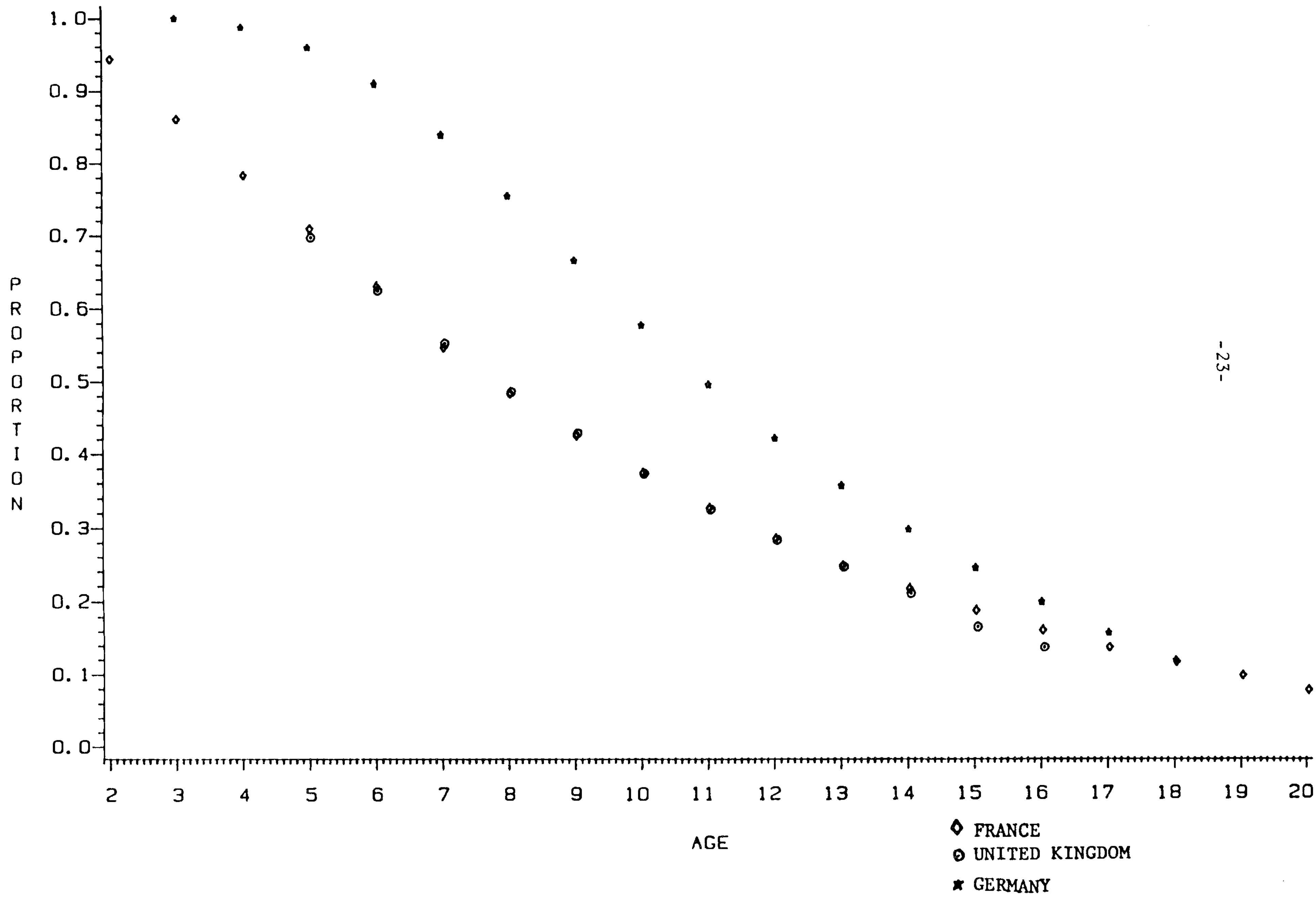
*a Symbols are defined as follows: f = the first age at which a renewal fee is due; L = the last age at which an agent can keep the patent in force by payment of an annual renewal fee; and $\overline{NPAT}$ the average number of patents per cohort.

*b For France and the U.K. these estimates were obtained as follows. Let n_t be the number of patents applied for in year t, and $\tilde{n}_t$ be the number of patents granted. Then the ratio was calculated as $T^{-1} \sum_{t=1}^{T} [(\sum_{\tau=1}^{4} .25\tilde{n}_{t+\tau})/n_t]$.
In Germany the ratio of the patents granted to those applied for from a given cohort was directly available, and these ratios were simply averaged over the cohorts studied.

applied for in the cohorts specified in row 3, but in Germany the data contain only those patents granted. Patents granted by date of application were not available for France and the U.K., though a rough estimate of the ratio of grants to applications in these two countries can be obtained by comparing the number of patents applied for to those granted over time (see the notes to Table 1). This ratio was quite large in France (.93), a bit smaller in the U.K. (.83), but only .35 in Germany (row 6). As a result of the facts that the data contain grants in Germany (in contrast to applications in France and the U.K.), and that the German granting criteria select out only a relatively small portion of the patents applied for, the average number of patents per cohort is smaller in Germany (about 21,000) than in France or in the U.K. (about 37,000; see row 7). Note that rows 3 and 7 imply that the data contain information on about one million patents in each of France and the U.K., and on about half of a million patents in Germany.

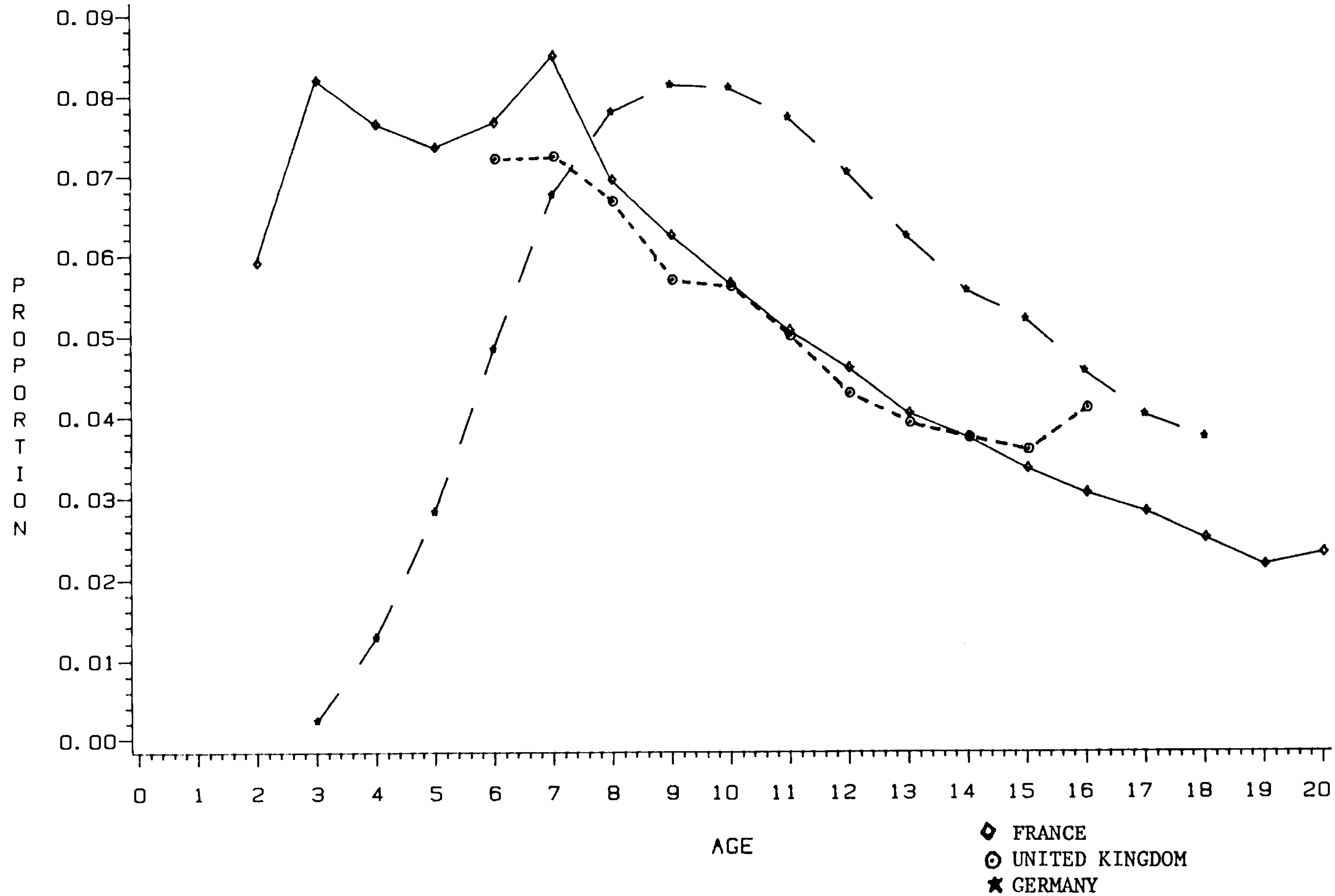
Figures 2 and 3 provide the proportion renewed, and the proportion dropping out, by age, averaged over the cohorts for which these statistics were observed; while figure 4 provides the mean of the renewal fee schedules used in the analysis. Figure 1 makes it clear that there is a distinct difference between the age-path of the proportion renewed in Germany, and those in the other two countries. This difference is magnified in figure 2. In Germany the proportion dropping out is much lower in the early ages, subsequently overtakes and then stays larger than the proportion dropping out in the other two countries. The lower drop out probabilities in the early ages in Germany could reflect the success of the German patent office in weeding out the patents which have high probabilities of not being profitably developed; especially since the renewal fees in the early ages in Germany are relatively small and comparable to those in the other countries (see figure 4). After age five, however, these fees are increasing at a much faster pace in Germany, and this should, all else equal, generate larger drop out probabilities in the later ages in Germany.

FIGURE 2 AVERAGE PROPORTION RENEWED



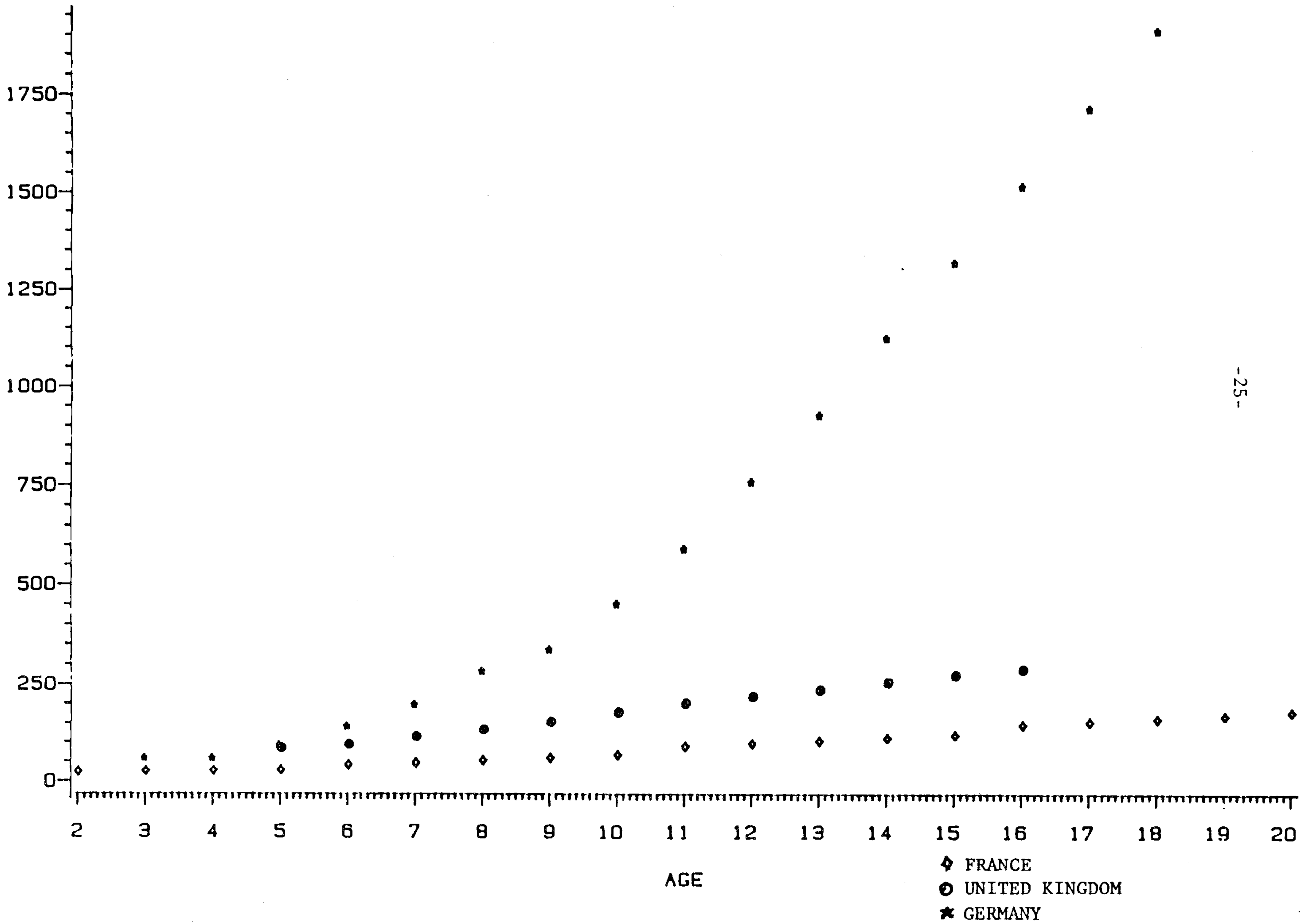
-23-

FIGURE 3: AVERAGE DROP OUT PROPORTION



-24-

FIGURE 4 : AVERAGE RENEWAL FEE SCHEDULES



-25-

Figure 2 also illustrates that there are, in fact, substantial differences in the proportion dropping out both between different ages for a given country, and between countries for a given age (the drop out proportion for age five in the U.K. is not illustrated but equals .305). This understates the total variance in the drop out proportions since there is variance between cohorts at a given age in each country. Most of this latter variance is concentrated in the early ages. Finally, note that in all countries (though to a varying extent) the drop out probabiliites do not decline at as fast a pace in the last few ages as in the ages immediately preceeding them. This is what we would expect from a stochastic model of renewal behaviour, since as the age of the patent approaches L, the option value of holding the patent goes to zero.

Turning to figure 3 note that the average cost of renewal schedules are nondecreasing in age. This is also true for the renewal fee schedules of each year and underlies the form of the solution to the agents decision problem provided in Corollary 1. The renewal fees are quite small in all countries in the early years, and increase significantly faster in Germany thereafter.

Section 5. The Empirical Results

Table 2 provides the parameter estimates, different dimensions of the data, and some summary statistics, for each country. It was decided at the outset to set the discount factor (β) equal to .9 in all runs; and the results presented in the table are conditional on β = .9.[6]

The parameter estimates in Germany and France are all positive and highly significant. Recall that the dimension in which parameter estimates converge to their true values is the total number of patents or NPAT. The extremely large values of NPAT (row B.2) explain the relatively low estimated standard errors in France and Germany. On the other hand the estimated information matrix for the U.K. was singular (see footnote b to the table). As will become clear presently, this occurs because the estimates imply that in order to distinguish between different possible values of the parameter vector we require independent information on events which occur during the early ages; and in the U.K. we do not have such information until age 6.

To get an indication of the fit of the model the difference between the estimated and acutual π's was squared and averaged over the NCHRTAGE (row B.4) distinct cohort-age cells for which these proportions are observed. The resulting numbers appear as MSE[π] in row C.1 of the table. Comparing them to the variance in the actual π's (i.e., to V[π;data] in row C.3), it is clear that in France and Germany only a small fraction of the variance in the acutal π's is not accounted for by the model (1.4% in France, and .6% in Germany), while in the U.K. this fraction is somewhat larger (6.4 %). To see whether there was any indication of cohort specific differences in the fit of the model, the differences between the estimated and actual π's were also used to calculate a pseudo

---

[6] This decision, and the decision to make only one run for each country, were taken, in part, to minimize on computer time. The CPU time for each run increases more than proportionately to the number of parameters estimated.

Table 2. Parameter Estimates[a]

	Country		
	France	U.K.[b]	Germany
A. Parameter			
σ	5689 (8.24)	5467(6.09)	7460 (19.72)
γ	9162 (13.67)	6919(10.29)	8687 (17.09)
ϕ	.5084 (5.66×10^{-4})	.4383(2.17×10^{-3})	.4896 (1.16×10^{-3})
δ	.8475 (2.62×10^{-4})	.8102(1.81×10^{-3})	.8861 (2.48×10^{-4})
σ_R	1.579 (2.92×10^{-3})	1.525(3.04×10^{-3})	1.158 (2.36×10^{-3})
μ	4.705 (2.75×10^{-3})	5.425(2.55×10^{-3})	6.718 (3.70×10^{-3})
θ	.0990 (6.36×10^{-4})	.36[b]	.0855 (2.46×10^{-3})
B. Dimension[c]			
B.1 NPAT	1,069,095	983,471	446,741
B.2 NSIM	20,000	20,000	20,000
B.3 Age:f/L	2/20	5/16	3/18
B.4 NCHRT	29	26	21
B.5 NCHRTAGE	238	272	237
C. Summary Statistic[d]			
C.1 MSE[π]	5.42×10^{-4}	6.91×10^{-4}	1.48×10^{-4}
C.2 PDW[π]	1.65	2.24	1.85
C.3 V[π; data]	3.90×10^{-2}	1.07×10^{-2}	2.65×10^{-2}

Notes to Table 2

[a]Patents are assigned to cohorts by year of application. Numbers in parenthesis beside parameter estimates are their estimated standard errors.

[b]Letting $[i^*_{r,s}]$ be the estimated information matrix, then, for the U.K., $i^*_{\theta\theta} = 0$. The standard errors of this column were obtained by inverting a six by six matrix consisting of $i^*_{r,s}$ for $r,s \neq \theta$. They are, therefore, conditional on $\theta = \theta^*$.

[c]See also the notes to Table 1. NPAT = the total number of patents covered by the data. NSIM = the number of random draws used to evaluate the simulated frequencies in the final iteration of the maximization subroutine and in the estimation of the information matrix (see Section 4). NCHRT = number of cohorts covered by the data. NCHRTAGE = the number of cohort-age cells covered by the data.

[d]Let $e_{a,j}$ be the difference between the estimated and the actual $\pi(a,j)$ for $a \varepsilon A_j$, $j=1,\ldots,J$. Then MSE $[\pi] = (\text{NCHRTAGE})^{-1} \sum_{j=1}^{J} \sum_{a \varepsilon A_j} e^2_{a,j}$, and

$$\text{PDW}(\pi) = [\sum_{j=1}^{J} \sum_{a=f_j}^{l_j-1} (e_{a+1,j} - e_{a,j})^2 / \sum_{j=1}^{J} (l_j - f_j)] / [\sum_{j=1}^{J} \sum_{a=f_j}^{l_j} e^2_{a,j} / \sum_{j=1}^{J} (l_j - f_j + 1)].$$

$V[\pi;\ \text{data}]$ is the sample variance of $\pi(a,j)$.

Durbin-Watson statistic for each country (see note d to Table 2). These are provided in row C.2 of the table, and seem to distribute about two. I return to further comments on the fit of the model after a brief description of some of the implications of the parameter estimates; particularly those relating to the characteristics of the learning process. Since it is these characteristics that the data in the U.K. are not rich enough to determine, I shall concentrate on the estimates for France and Germany.

The parameters whose estimates exhibit substantial intercountry differences are μ, σ_R, and σ. The estimates of μ and σ_R imply that a substantial fraction of the patents in the French data started out with low, almost negligible, initial returns; while the higher mean and the lower coefficient of variation in Germany imply that this phenomena was not nearly as pronounced among German patents (the mode of the estimated distribution of initial returns is under ten dollars in France but is over two hundred dollars in Germany; and the parameter estimates indicated that about thirty percent of the French patents had initial returns under fifty dollars, while under one percent of the German patents do). The larger σ in Germany implies that, on average, the holders of the patents included in the German data had a higher probability of discovering uses which increased the returns to their patented ideas. Recall that the German data includes only patents granted while the French data includes all patents applied for; and that the granting criteria seem to be particularly stringent in Germany (Table 1). It seems, then, that the German patent office was, on the whole, successful in weeding out patents with low initial returns and a smaller probability of increasing those returns over time.

The estimates of θ, δ, ϕ, and γ do not vary much between the two countries. The low estimates of ϕ (about .5) implies that the learning process is concentrated in the early ages. Table 3 illustrates this point. The descriptive

Table 3. The Evolution of Implicit Revenues in the Early Ages*[a]

	Country	
	France	Germany
Characteristic		
$E(r_1)$ $[r_1 \mid r_1 > 0]$	380.43	1608.57
Pr (Downside); Pr (Upside)	.0637; .1807	.0004; .2705
π (2; $\underset{\sim}{c}$, $\underset{\sim}{\omega}$)	.0637	(no required renewal)
$E(r_2)$ $[r_2 \mid r_2 > 0]$	1414.72	3400.98
Pr (Downside); Pr (Upside)	.0387; .0331	.0006; .0584
π (3; $\underset{\sim}{c}$, $\underset{\sim}{\omega}$)	.0907	.0013
$E(r_3)$ $[r_3 \mid r_3 > 0]$	1432.24	3224.56
Pr (Downside); Pr (Upside)	.0118; .0012	.0005; .0039
π (4; $\underset{\sim}{c}$, $\underset{\sim}{\omega}$)	.0792	.0121
$E(r_4)$ $[r_4 \mid r_4 > 0]$	1339.05	2899.41
Pr (Downside); Pr (Upside)	.0048; 0.00	.0003; 0.0
π (5; $\underset{\sim}{c}$, $\underset{\sim}{\omega}$)	.0381	.0277
$E(r_5)$ $[r_5 \mid r_5 > 0]$	1192.70	2641.40
$\overline{\text{NPAT}}$	36,865	21,273

NOTES TO TABLE 3

[1]The estimates were based on a simulation run with 20,000 draws using the estimates of $\underset{\sim}{\omega}$ given in Table 2 and the mean of the renewal fee schedules. Pr (Downside) is the average probability of discovering that the patented ideas will never by profitably exploited (of drawing the absorbing state); averaged over the patents still in force. Pr (Upside) is the average probability of discovering a use which enables the agent to increase returns in the following year (of $z > \delta r_a$); averaged over the patents still in force.

$E_r[r \mid r > 0]$ = the mean of r for patents still in force.

$\pi(a; \underset{\sim}{c}, \underset{\sim}{\omega})$ = the proportion of patents who drop out at the a^{th} renewal.

statistics provided in this and in subsequent tables were obtained from a simulation run of 20,000 draws based on the mean of the renewal fee schedules and the parameter estimates of Table 2. Consider first the column of figures for France. The mean of the initial distribution of returns was 380 dollars. During the initial year just under twenty percent of the French patent holders discovered a use which enabled them to increase subsequent returns, while over six percent discovered that their patented ideas could never be profitably exploited. These six percent were the only patents whose renewal fees were not paid in the second year. The holders of the remaining patents paid the renewal fee and maintained the option of patent protection on the results of the second year's experiments. The substantial learning that occurred over the first year caused a sharp increase in the average returns of the patents still in force in the second year. During the second year much less learning occured than occurred during the first year. An additional nine percent of the patent holders stopped paying the renewal fee at the third age. Of these, about five percent were owned by agents who, after doing experiments for two years, had decided that it was not worthwhile to pay the renewal fee in order to have the option of patent protection on the results of subsequent experiments. Average learning probabilities decreased further over the next two ages. They were just about sufficient to keep the mean of the current returns earned on the patents still in force constant. There was essentially no learning after the fifth age, and the effect of the obsolescence process clearly dominates the learning processes when comparing the means of the patents still in force in the fifth, to those still in force in the fourth, ages. The major qualitative difference between the German and the French columns in this table arises from the fact, noted earlier, that the German parameter estimates imply that a much smaller proportion of the patents in the German data started out with negligible returns. As a result most of the patents included in the German data were known to be worth something at the outset, and more of the German patent holders who did not discover a more profitable use over time had current returns which induced them to pay the renewal fee until the ages in which those fees started rising sharply (which was after age five, see figure 4).[7]

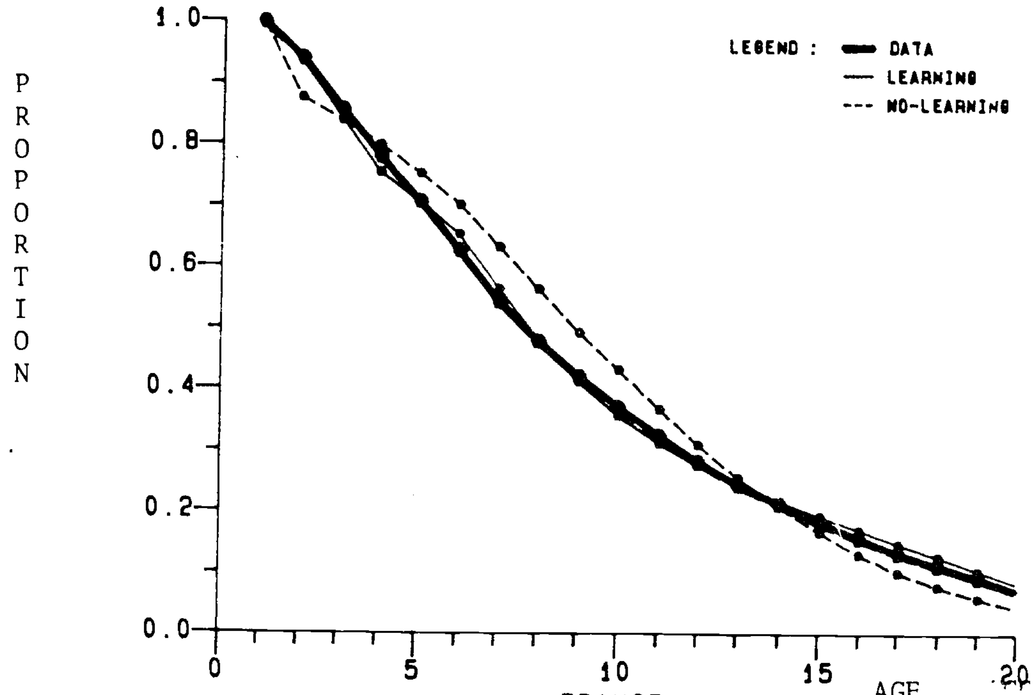
I now return briefly to the issue of the fit of the model. Figure 5 provides the proportion renewed, by age, averaged over the cohorts for which this proportion was observed. The thick lines provide the proportions in the data, the thin lines those estimated by our model, and, for comparison, we also provide the proportions estimated from a model which does not allow for learning, (the broken lines). The no learning model is a model in which patents are endowed with an initial distribution of returns which decay deterministically thereafter. It is obtained by changing the probability statement in equation (6) to read; $r_{a+1} = \delta r_a$ with probability one.[8] In this figure it is hard to distinguish the curve estimated by the model with learning, from the data. On the other hand the model without learning predicts too few renewals in the early ages (i.e., too many drop outs), too many renewals in the middle ages, and too few again in the later ages. Recall that the renewal fees are close to constant over the initial ages. As a result, the model without learning cannot accommodate both the small number of drop outs in the initial age, and the sharp increase in the number of drop outs over the next few ages. This point is magnified

---

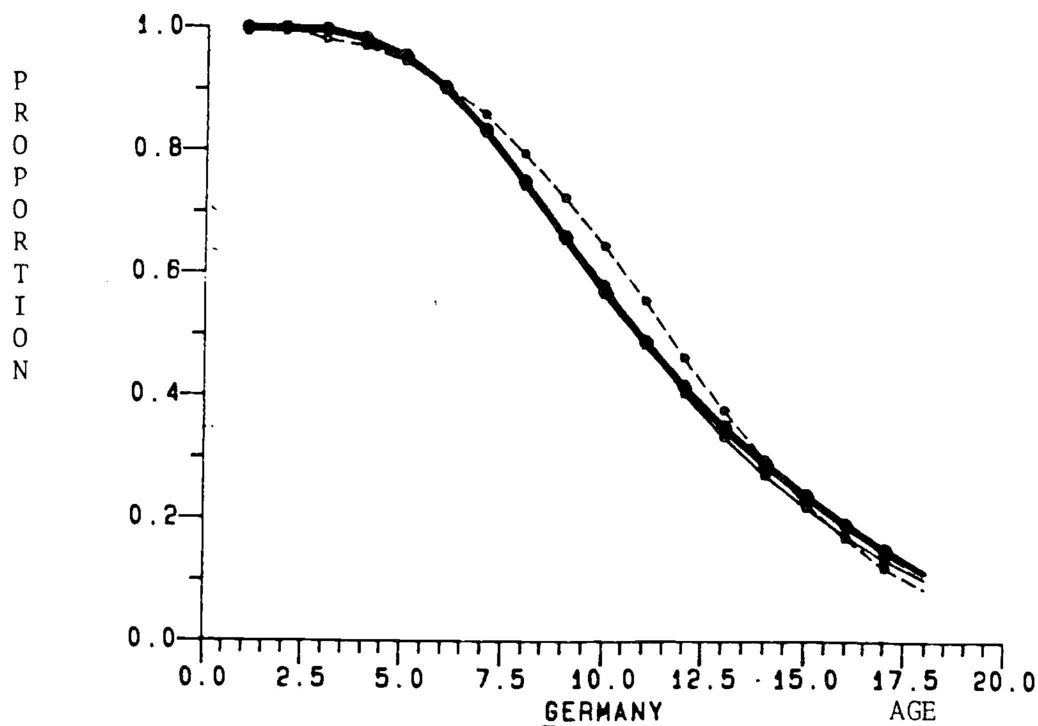
[7] The parameter estimates for the U.K. presented in Table 2 imply a learning process which is similar to those described for France and Germany. The problem with the U.K. estimates was that without independent information on the drop outs over the first few ages, the likelihood function could not distinguish between different values for the learning parameters, particularly θ and ϕ. As a result of this fact the maximization algorithm also had much more difficulty in finding the estimates for the U.K.

[8] As one would expect from the large size of our samples (NPAT) the likelihood ratio test statistic for the null hypothesis that there was no learning was inordinately large; over 20,000 for Germany and over 60,000 for France.

FIGURE 5. AVERAGE PROPORTION RENEWED



FRANCE



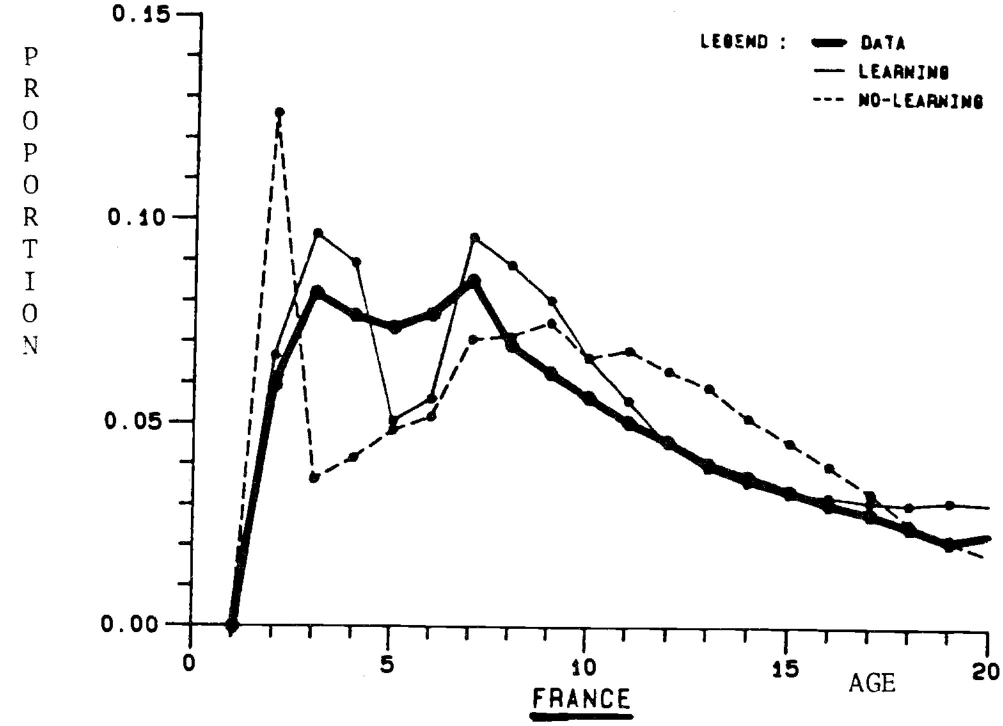
GERMANY

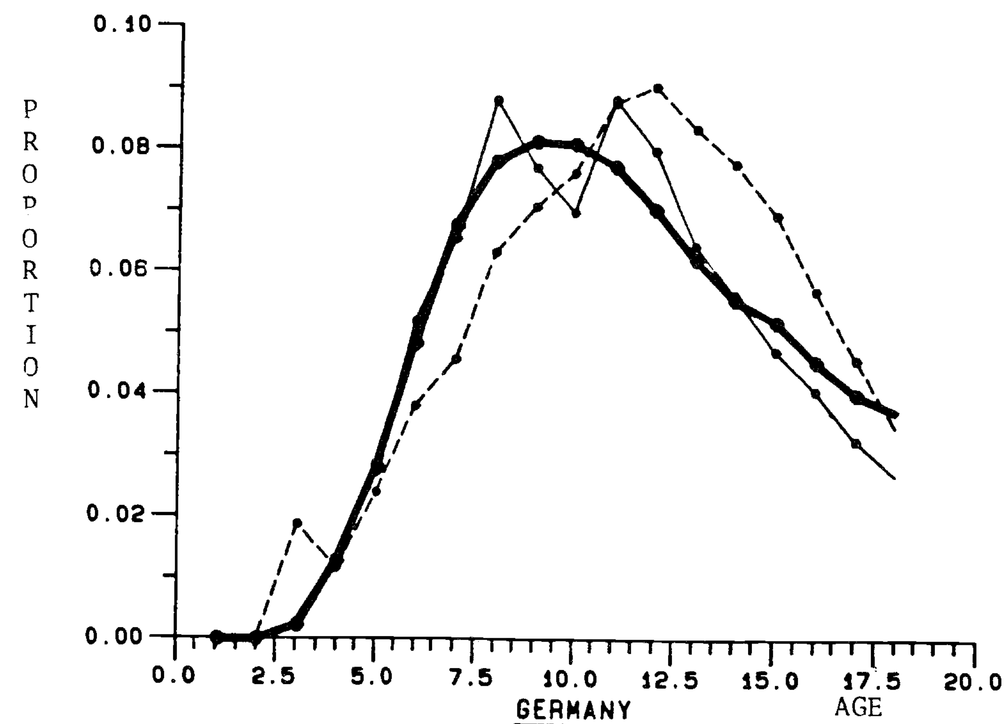
in figure 6 which provides the proportion dropping out, by age, averaged over the cohorts for which this proportion is observed. The model with learning accounts for the combination of the low initial drop outs and the increase in the number of drop outs over the next few ages by estimates which imply that the option value of patents which start out with low returns is initially high, but then declines rather rapidly. As will be shown presently, this model accounts for the spread of those who do drop out over the later ages by a somewhat skewed distribution of initial returns, and, more importantly, by a learning process which increases the skew in the distribution of returns substantially over the next few ages.

In figure 6 we can actually see the differences between the estimates from the model with learning, and the data. These differences are concentrated in the middle ages. The age-specific average drop out probabilities in the French data have two local maxima (at ages three and seven). The estimates from the model for France also have two local maxima (and at the same ages), but the model's estimates of these maxima are somewhat too high, and its estimate of the trough between them is too low. In Germany the data provide a rather flat age distribution of average drop out probabilities between ages eight and eleven. The model's estimates replace this with two local maxima and a minimum; though neither the maxima nor the minimum are nearly as pronounced as those estimated for the earlier ages in France. In addition, the model's estimate of the average drop out probabilities in the later ages are a bit too high in France, and a bit too low in Germany.

The proportions which actually enter the likelihood function (the π's) are a combination of the renewal proportions (for the last age for each cohort, and the first where there was left censoring) and the drop out proportions (for all other cohort-age combinations). These are also the proportions which underly

FIGURE 6. AVERAGE DROP OUT PROPORTIONS





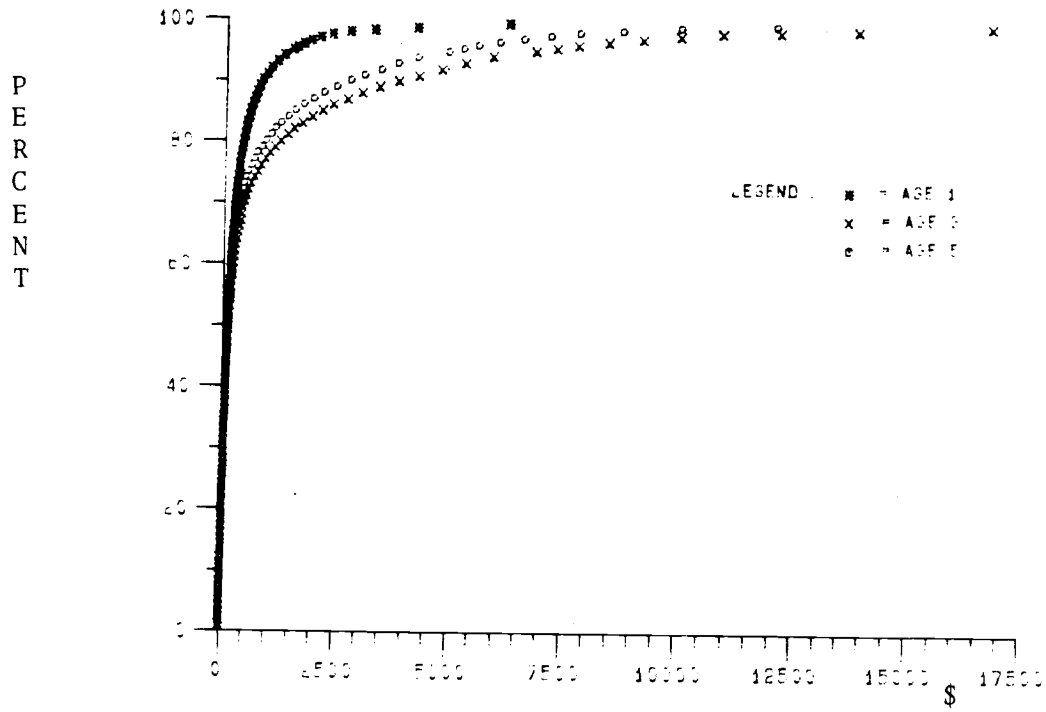
the mean square error measures provided in row C.2 of Table 2. Though figures 5 and 6 indicate why the mean square error measures are low relative to the variance of π in the data; they also point out that there are still some aspects of the data that the model does not account for, and this should be kept in mind when considering the implications of the parameter estimates.[9]

Figures 7 and 8 provide the distribution of current returns at age one, age three, and age five, in France, and in Germany, respectively. Part A of these figures contain the first 99.5 percentiles of these distributions, while part B narrows in on the lowest 75 percentiles. Two points come out clearly from the figures. First the distributions, particularly those for the later ages, are extremely skewed. As a result it is hard to distinguish any interage differences in the lower 75 percentiles of the distribution from part A of the figures. Second, there is a distinct pattern to the evolution of these distribution functions over age. Between the first and the third ages there is a substantial increase in the dispersion of the distribution functions. This is easiest to see in part B of the figures. It occurs because of two implications of the parameter estimates. First between these ages the experiments of a

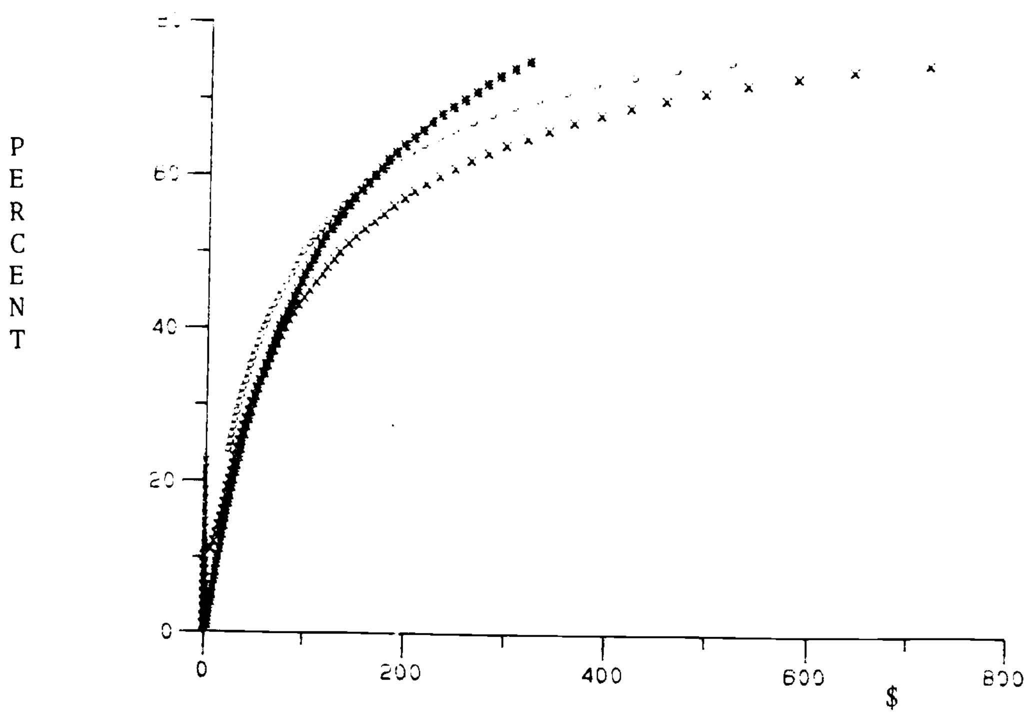
---

[9] It is worth noting that the intercohort differences in the drop out proportions for a given age and country were fit quite well by the data. As noted in section 4 most of this variance was concentrated in the early ages. This was also what the model predicted since the last age at which the option value of patents with low returns induced payment of renewal fees depended on the precise age at which the renewal fees began to rise. Even given this point, however, it is still undoubtedly the case that the extremely large values of NPAT and NSIM in table 2 imply that the MSE(π) statistics provided in that table are too large to be rationalized in terms of binomial sampling error in the empirical and estimated frequencies. Though this problem (which is called the problem of extra binomial sampling variance by Williams, 1982; see also the review by Haseman and Kupper, 1979) occurs frequently when analyzing the determinants of proportions, I do not know of any logically consistent way of accounting for it when the model has a sequential dimension. It is also worth restating here the related problem noted in the introduction; that is, the complexity of the estimation procedure in discrete choice optimal stochastic control problems such as ours makes it difficult to determine the robustness of the conclusions to the particular distributional assumptions made (for a discussion of related issues, see Heckman and Singer, forthcoming). This is one reason for examining (in some detail) the consistency of our empirical results with known intercountry differences in patenting procedures.

FIGURE 7. FRENCH RETURNS

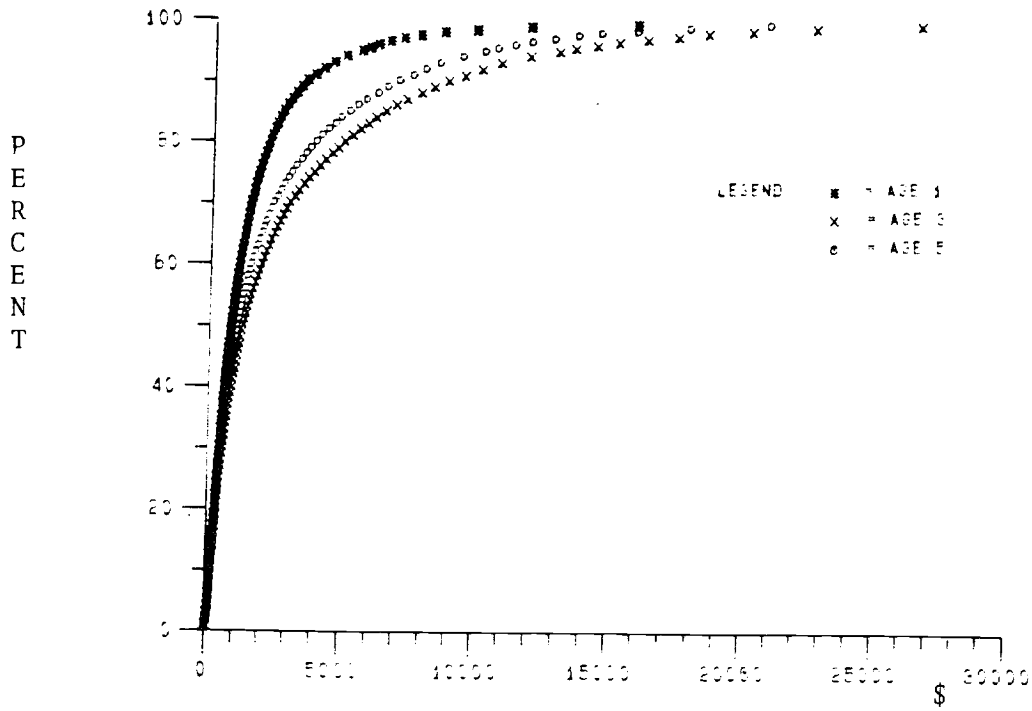


A. LOWEST 99.5 PERCENTILES

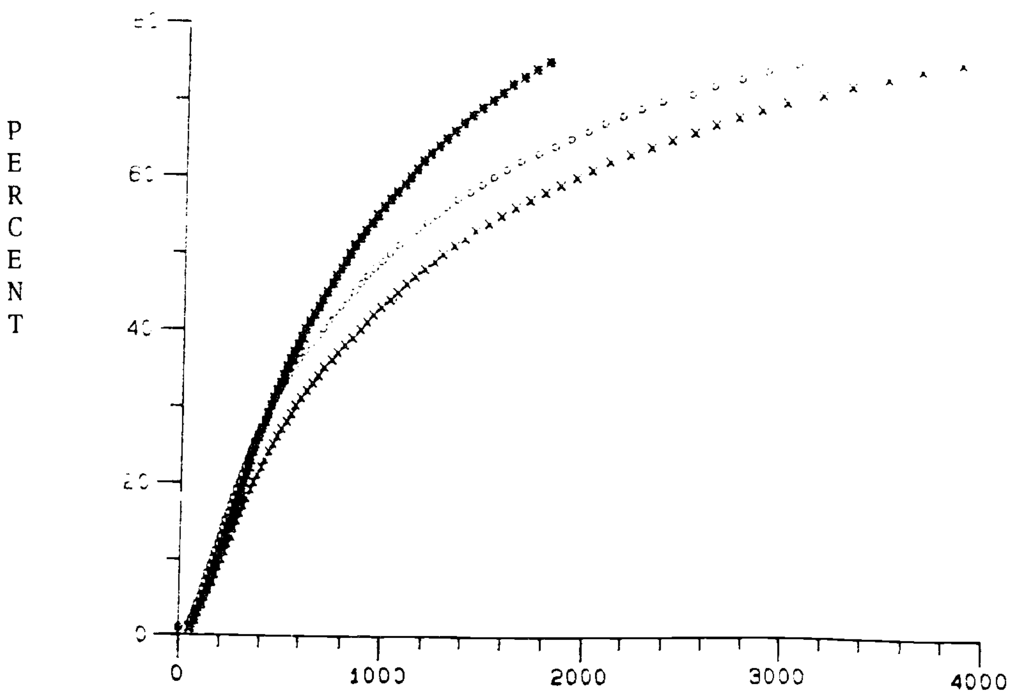


B LOWEST 75 PERCENTILES

FIGURE 8. GERMAN RETURNS



A. LOWEST 99.5 PERCENTILES



B. LOWEST 75 PERCENTILES

large portion of the patent holders had enabled them to increase their returns, and the effect of this was to thicken the tail of the distribution function and push it to the right. On the other hand, many of the holders of patents which had negligible initial returns, and recall that such patents were a much larger fraction of the French data, had discovered by age three that their patented ideas were not likely to enable them to increase their returns in the future, and have consequently dropped out. By age three, then, the distribution developed a mass at zero. A comparison of the percentiles for age five to those of age three reveals the onset of the obsolescence process; that is the percentiles from the age five distribution are below (strictly speaking, never above) the same percentiles from the distribution at age three.

The skew in the distribution of initial returns combined with the substantial increase in this skew over the early ages, lead to a highly skewed distribution of realized patent values. Table 4 provides percentiles and lorenz curve coefficients from the distribution of realized patent values; where the realized value of a patent is defined as the discounted sum of net returns (current returns minus renewal fees) from age one to the last age the given patent is kept in force. Again I begin by considering the column of figures for France. Twenty-five percent of the patents in the French data had realized values of seventy-five dollars or less.[10] These patents contributed about a half of one percent to the total value of the patents in a cohort, while the patents in the lower half of the distribution contributed less than two percent of the total value of a cohort. The median of the distribution of realized values ($534) was

---

[10] Of course some of these patents had negative (though small in absolute value) realized values, as they were patents who paid early renewals for options which did not materialize. If, for example, we had defined the realized values as the discounted sum of net returns from age two, rather than from age one (as in the table), the lorenz curve coefficient corresponding to p = .25 would have been negative, though close to zero.

Table 4. Percentiles (pl) and Lorenz Curve Coefficients (lc) From the Distribution of Realized Patent Values[a]

Percent p	France pl($)	France lc (%)	U.K. pl($)	U.K. lc (%)	Germany pl($)	Germany lc (%)
.25	75.23	.544	355.55	.554	1,999.60	2.249
.50	533.96	1.833	1,516.84	3.247	6,252.93	7.341
.75	3,731.35	8.087	7,947.55	16.369	19,576.26	25.288
.85	10,292.06	19.575	15,357.09	31.721	32,428.14	41.001
.90	17,423.11	31.261	22,206.21	44.257	44,241.87	52.672
.95	31,609.59	52.461	34,740.07	62.960	65,753.61	69.223
.97	42,905.78	65.514	43,889.95	73.640	78,299.01	78.348
.98	51,215.84	73.729	51,277.22	80.072	94,842.63	83.800
.99	66,515.40	84.011	65,075.08	87.858	118,354.78	90.330
maximum	259,829.27	-	374,028.70	-	419,217.55	-
mean	5,631.03	-	7,357.05		16,169.48	-
$\overline{\text{NPAT}}$	36,865		37,826		21,273	

[a] The realized value for patent i is $\sum_{\tau=1}^{\tau_i^*} \beta^{(\tau-1)} (r_{i,\tau} - c_\tau)$, where τ_i^* = the last age at which patent i was kept in force. The distribution of values was estimated from a simulation run of 20,000 draws using the estimates of $\underset{\sim}{\omega}$ provided in Table 2 and the mean of the renewal fee schedules.

less than one tenth its mean ($5,631); and the five percent of the distribution with the highest realized values contribute about half of the total value of a cohort. The German distribution of realized values was somewhat less skewed than the distribution in France; though even the German distribution was extremely skewed. The difference between the two distributions was, as might have been expected from the fact that in Germany the data refers to grants rather than applications, most pronounced at the lowest percentiles. In Germany these percentiles were non-negligible, albeit, quite small. Still only about 7 percent of the patents in Germany had realized values in excess of $50,000; in France only two and a half percent had values this large. Given the size of the cohorts this implies that, on average, about a thousand patents which had realized values in excess of $50,000 were applied for annually in France, and about fifteen hundred such patents were granted annually in Germany.

One other point is worthy of note here. The estimate of the ratio of the average realized value in a cohort of patents applied for in France, to that value in a cohort of patents granted in Germany, is .35, which is just equal to the average of the ratios of grants to applications in the German cohorts (see table 1). The estimates seem to imply, then, that the mean of the realized values of the patents applied for in the two countries was similar. On the other hand, there were a significantly larger number of patents applied for per year in Germany than in France [about 60,780 in Germany, versus 36,865 in France]; so that, on average, the total value of a cohort of patents in Germany was larger than the value of a French cohort.

The patent stock held in a country at a given point in time consists of the patents from the cohorts applied for over the previous L years which are still in force at that time. Table 6 provides the annual net returns earned by the patents of each age contained in a stock which is constructed by assuming that each of the previous L cohorts began with the average number of patents per cohorts and faced the mean of the renewal fee schedules. The entries in this table are in thousands of 1980 U.S. dollars. The net annual returns earned from holding the patent stocks in these countries is estimated at; .315 billion dollars in France, .386 billion dollars in the U.K., and .512 billion dollars in Germany. The next section considers likely implications of this, and the other results provided in this section.

Table 5. Estimates of the Annual Flow of Returns from Holding the Patent Stocks of European Countries
(Entries in Thousands of 1980 U.S. Dollars)[a]

	Country		
	France	U.K.	Germany
Age			
1	13,948.7	27,067.4	34,031.0
2	47,998.5	75,287.4	72,321.4
3	43,886.0	64,487.3	67,332.2
4	37,116.1	52,290.0	59,696.3
5	31,258.2	39,832.3	52,045.4
6	26,060.7	31,633.7	44,938.5
7	21,814.2	24,771.9	38,452.2
8	18,223.8	19,283.3	32,321.6
9	15,169.9	14,842.0	27,348.7
10	12,595.6	11,210.4	22,283.2
11	10,220.6	8,335.2	17,709.1
12	8,318.6	6,112.9	13,647.8
13	6,902.4	4,094.4	10,275.4
14	5,599.6	3,013.6	7,432.6
15	4,523.1	2,027.4	5,159.3
16	3,475.7	1,261.7	3,469.6
Total to Age 16	307,111.8	385,730.9	508,464.3
17	2,735.4	-	2,204.3
18	2,113.9	-	1,351.5
Total to Age 18	311,961.1	385,730.9	512,020.1
19	1,606.5	-	-
20	1,194.3	-	-
Total to Age 20	314,761.9	385,730.9	512,020.1

---

[a] The estimates assume that all cohorts currently in force began with $\overline{NPAT}$ patents, and faced the mean of the renewal fee schedules. The entry for age a is calculated as $\sum (r_{i,a} - c_a)$, where the summation extends over all patents still in force at that age. The estimates of the distribution of $r_{i,a}$ (a = 1, ..., L) are obtained from a simulation run of 20,000 draws using the parameter estimates of Table 2.

Section 6. More General Aspects of the Empirical Results

Many of the detailed implications of the parameter estimates were presented in the last section. This final section provides a brief discussion of their relationship to a few, more general, issues in the economics of technological change. I focus primarily on; the estimated characteristics of the learning process (Table 3 and Figures 5 and 6), the estimated skew in the distribution of the value of holding patents (Table 4), and on the estimates of the annual returns earned from holding the patent stocks of the alternative countries (the last row of Table 5).

To get an indication of the importance of the incentives created by the patent laws we would like to compare the estimates provided in Table 5 to either, the total returns that accrued to the patented ideas, or to the expenditures that went into developing them. Neither of these two figures are available, but the OECD (1975; Tables III and IV) does provide estimates of the R & D expenditures funded by the business enterprises in these countries in 1963 (which is the midcohort in our data). The estimates of the annual returns from holding the patent stocks were respectively, 15.56%, 11.03%, and 13.83% of the R & D expenditures of the business enterprises in France, the U.K. and Germany; and the sum of these returns across countries was 13.14% of the sum of their R & D expenditures. Since there may be returns earned as a result of patenting per se, regardless of whether the patents were ever renewed, and since our estimates only pertain to the returns earned by renewing (or holding) patents already in force, the numerator of this ratio may slightly understate the annual monetary value of the incentives created by the patent system. Moreover, the ratio suffers from the fact that we have not netted out various balance of trade effects (business enterprises in these countries also own patents in force elsewhere, and foreign business enterprises own patents in force in these countries; while not all the business sector's R & D expenditures are directed towards patentable innovations, and not all patentees are business enterprises). The ratio does, however, suggest that the proprietary rights resulting from the patent laws create annual returns which are non-negligible in comparison to privately funded R & D activity.

The returns earned from holding patents may, of course, be only a small fraction of the returns that accrue to patented ideas. Nevertheless the general similarity between the shape of the estimated distributions of the value of holding patents on the one hand (see Table 4), and currently available evidence on the distribution of the values of patented ideas on the other, is quite striking. In particular the evidence available from disaggregated case studies indicates an extremely skewed distribution of the values of patented ideas (see Sanders, Rossman, and Harris, 1958; and Gabrowski and Vernon, 1983). Scherer (1965, p1098), for example, notes that the data provided in Sanders, Rossman, and Harris (1958) suggests a Paraeto-Levy distribution with an infinite mean for the distribution of profits from patented ideas; while Garbrowski and Vernon (1983) summarize their studies on the profitability of new pharmaceutical entities with the statement,

> "In effect, these results indicate that pharmaceutical firms are heavily dependent on obtaining an occasional 'big winner' to cover their R & D costs and generate profits" [Gabrowski and Vernon, 1983, p.11].

Larger sample econometric studies have focused on the relationship between the number of patents applied for and alternative measures of the outputs and the inputs into inventive activity [see the articles in Griliches (ed.), forthcoming]. Pakes, (1981) provides a detailed time-series cross-section analysis of the reduced form relationship between patent applications, R & D expenditures, and changes in the stock market value of firms, that allows for dynamic error components to intercede between these variables. That article concludes that changes in the number of patents applied for by firms are a very noisy measure of the changes in stock market value of their R & D related output, but that, on average, increases in patent applications are associated with large increases in the firm's value; just what we would expect from a highly skewed distribution of the value of patented ideas. In addition, a strong simultaneous relationship between the factors driving R & D expenditures and those driving patents was found; suggesting that a significant search for uses and improvements to the patented ideas continues at least during the early years of a patent's life.

There is an explanation of the patenting process which is at least consistent with both the empirical results found in this paper, and with those cited above. Patents are applied for at an early stage in the inventive process, a stage in which there is still substantial uncertainty concerning both the returns that will be earned from holding the patents, and the returns that will accrue to the patented ideas. Gradually the patentors uncover the true value of their patents. Most turn out to be of little value, but the rare "winner" justifies the investments that were made in developing them. If this explanation captures the nature of the patenting process we would not expect to find a very stable relationship between profits and current and past patents, or between profits and the current and past R & D expenditures which lead to them; except possibly for very large aggregates. For individual economic units we would expect most increases in patents not to lead to any increase in profits, and for there to be an occasional jump in profits which is not necessarily preceeded by any increase in patenting. Similar statements can be made concerning the relationship between profits and the R & D expenditures that lead to the development of the patented ideas. Growth through discovery will occur in spurts, and these spurts will be probabilistically related to the investments which preceeded them. Traditional production function approaches to obtaining estimates of either the rate of return to the investments which produced the patents, or the determinants of the quantity of resources invested in their development, are not likely to be very precise. Nor will they provide much evidence on the characteristics of the distribution of possible outcomes; features of the problem that are likely to be particularly important in analyzing the rich set of issues determining the evolution of firm and industry structure. An alternative, pointed out by Nelson and Winter (1982), and Telser (1982), is to be more careful in the econometric modelling of the inventive process itself; employing, perhaps, controlled search processes in which investment expenditures affect the distribution of possible outcomes.[11]

Disaggregated patent renewal data from over fifty national and regional patent offices containing both the technical field of the patent, and the patentor, with coverage, in most cases, dating back at least to 1973, is currently available from INPADOC (see references). A more disaggregate patent renewal study which estimates the return to patent protection by technical field, and by nationality and type of patentor (e.g., individuals, small business enterprises, large corporate entities) is likely to prove extremely valuable. Issues related to which sectors in a particular country, and which countries, derive disproportionate returns from the patent laws lie at the heart of much of the discussion of the social costs and benefits of alternative patent systems (see Scherer 1979 chapter 16, and the literature cited there). Moreover, a more disagrregated study would provide information on both, differences in the characteristics of the learning process and in the distribution of possible outcomes, and differences in the relationship between patent statistics and alternative measures of inventive inputs and outputs, between different industries. Policy, as well as empirical and theoretical analysis would benefit considerably from such information.

---

[11] A step in this direction has been made by Ericson and Pakes, 1983.

REFERENCES

Ericson, Richard, and Ariel Pakes, 1983. "The Rise and Fall of R & D Labs," paper presented at the NBER conference on Technical Change and R & D Incentives, Cambridge, Massachusetts, August 10-12th, 1983.

Garbrowski, Henry G., and John M. Vernon, 1983. "Studies on Drug Substitution, Patent Policy and Innovation in the Pharmaceutical Industry",final report to the NSF, Duke University.

Gill, Phillip E., Walter Murray and Margeret H. Wright, 1981. Pratical Optimization, Academic Press.

Griliches, Zvi (ed.) forthcoming. Patents, R & D and Productivity, the University of Chicago Press, Chicago.

Haseman, J.K., and L.L. Kupper, 1979. "Analysis of Dichotomous Response Data from Certain Toxological Experiments", Biometrics, Vol. 35 (March), pp 281-293.

Heckman, J. and B. Singer, forthcoming. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", Econometrica.

INPADOC, International Patent Documentation Center, Vienna, Austria.

Lerman, Steven R., and Charles F. Manski 1981. "On the Use of Simulated Frequencies to Approximate Choice Probabilities", pp. 305-315 in C.F. Manski and D. McFadden (eds.) Structural Analysis of Discrete Data with Econometric Applications, M.I.T. Press, Cambridge.

Macsyma, 1983. Macsyma Reference Manual, version 10 the Mathlab Group, Labaratory for Computer Science, M.I.T.

Miller, Robert A., 1982. "Job Specific Capital and Labor Turnover", unpublished Ph.D. dissertation, the University of Chicago, Chicago.

National Physics Laboratory, 1983. ONMDIF, Numerical Analysis Group, Downer's Grove, Illinois.

Nelson, Richard R., and Sidney Winter, 1982. An Evolutionary Theory of Economic Change, Harvard University Press, Cambridge.

Organization for Economic Cooperation and Development, 1975. Patterns of Resources Devoted to Research and Experimental Development in the OECD Areas, 1963-1971, OECD, Paris.

Pakes, Ariel, 1981. "Patents, R & D, and the Stock Market Rate of Return," National Bureau of Economic Research Working Paper No. 786, Cambridge.

Pakes, Ariel and Mark Schankerman, 1978. "The Rate of Obsolescence of Knowledge, Research Gestation Lags, and the Private Rate of Return to Research Resources", Harvard Institute of Economic Research Discussion Paper No. 659, also forthcoming in Zvi Griliches (ed.) op. cit.

Rao, C. Radhakrishna, 1973. Linear Statistical Inference and its Applications, John Wiley and Sons, New York.

Roberts, Kevin and Martin L. Weitzman, 1981. "Funding Criteria for Research, Develepment and Exploration Projects", Econometrica, Vol. 49. No. 5, pp 1261-1288

Sanders, J., J. Rossman and L. Harris, 1958. "The Economic Impact of Patents", the Patent, Trademark and Copyright Journal of Research and Education, II (September), pp 340-363.

Scherer, Frederic M., 1965. "Firm Size, Market Structure, Opportunity, and the Output of Patented Innovations", American Economic Review LV (December), pp 1097-125.

Scherer, Frederic M., 1980. Industrial Market Structure and Economic Performance. 2nd edition, Rand McNally College Publishing Company, Chicago.

Telser, Lester G., 1982. "A Theory of Innovation and Its Effects", the Bell Journal of Economics, Vol. 13 pp 69-92.

Williams, D.A., 1982. "Extra-binomial Variation in Logistic Linear Models", Applied Statistics, Vol. 31, pp 144-148.

Wolpin, Kenneth I., 1982. "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality", Economic Growth Center Discussion Paper No. 424, Yale University, New Haven.

March 1984

Accompanying Appendix to: "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks"

by Ariel Pakes

Appendix 1. Proposition 1 and Corollary 1

Proposition 1. $B(a,r) - r$ is : (i) uniformly continouous and nondecreasing in r; and, (ii) nonincreasing in a; for $r \in R^+$ and $a = 1,..,L$.

Proof: The proof of both (i) and (ii) is obtained by induction on a. I begin with part (i) and show that the assumption that $B(a + 1,r) - r$ is uniformly continuous and nondecreasing in r, implies that $B(a,r) - r$ is also. Note that equations (2) and (7) in the text imply that

$$B(a,r) - r = \beta[1 - \exp(-\theta r)]D(a,r), \qquad (A1.1)$$

where $D(a,r) = Q_{a+1}(\delta r)V(a+1,\delta r) + \int_{\delta r}^{\infty} V(a+1,y)dQ_{a+1}(y)$. Now recall that $V(a+1,r) = \max\{0, B(a+1,r) - c_{a+1}\}$, so that the assumed continuity and monotonicity of $B(a+1,r) - r$ implies that $V(a+1,r)$ is uniformly continuous and nondecreasing in r. Clearly then for $r' > r$

$$- V(a+1,\delta r')[Q_{a+1}(\delta r') - Q_{a+1}(\delta r)] \leq -\int_{\delta r}^{\delta r'} V(a+1,y)dQ_{a+1}(y)$$

$$\leq - V(a+1,\delta r)\,[Q_{a+1}(\delta r') - Q_{a+1}(\delta r)]. \qquad (A1.2)$$

The monoticity of $B(a,r) - r$ follows directly from (A1.1) and the first inequality in (A1.2) as they imply that

$$D(a,r') - D(a,r) = Q_{a+1}(\delta r')V(a+1,\delta r') - Q_{a+1}(\delta r)V(a+1,\delta r) - \int_{\delta r}^{\delta r'} V(a+1,y)dQ_{a+1}(y)$$

$$\geq Q_{a+1}(\delta r)[V(a+1,\delta r') - V(a+1,\delta r)] \geq 0, \qquad (A1.3).$$

To prove uniform continuity of $D(a,r)$ [which implies the uniform continuity of $B(a,r) - r$, see (A1.1)] we must show that for every $\varepsilon > 0$ there exists an $h(\varepsilon)$ such that

$$\left| D(a,\tilde{r}) - D(a,r) \right| < \varepsilon \quad \text{provided} \quad \left| \tilde{r} - r \right| < h(\varepsilon), \qquad \text{(A1.4).}$$

From the definition of $D(a,r)$ the second inequality in (A1.2) and the fact that $Q_{a+1}(y) < 1$

$$\left| D(a,r) - D(a,\tilde{r}) \right| < \left| V(a+1,\delta r) - V(a+1,\delta \tilde{r}) \right| , \qquad \text{(A1.5).}$$

Since the hypothesis of the inductive argument implies that $V(a+1,r)$ is uniformly continuous in r, there exists an $h^{*}(\varepsilon)$ such that

$$\left| V(a+1,\delta r) - V(a+1,\delta\tilde{r}) \right| < \varepsilon \quad \text{provided} \quad \left| \delta \right| \left| r - \tilde{r} \right| < h^{*}(\varepsilon), \qquad \text{(A1.6).}$$

(A1.5) and (A1.6) imply (A1.4) with $h(\varepsilon) = \left| \delta \right|^{-1} h^{*}(\varepsilon)$.

We have shown that $B(a,r) - r$ is uniformly continuous and nondecreasing in r provided $B(a+1,r) - r$ is. To complete the inductive argument we need only note that $B(L,r) - r = 0$, which is clearly uniformly continuous and nondecreasing in r.

To prove part (ii) first assume that $B(a+1,r) - r > B(a+2,r) - r$ and note that this implies that

$$\begin{aligned} B(a,r) - r &= \beta \int_{R^+} V(a+1,y)dG_{a+1}(y|r) > \beta \int_{R^+} V(a+2,y)dG_{a+1}(y|r) \\ &> \beta \int_{R^+} V(a+2,y)dG_{a+2}(y \mid r) = B(a+1,r) - r, \end{aligned}$$

where; the first inequality follows from the fact that, since $c_{a+2} > c_{a+1}$, $B(a+2,r) < B(a+1,r)$ implies $V(a+2,r) < V(a+1,r)$; and the second inequality follows from the monoticity of $V(a+1,r)$ in r and the fact that $G_{a+2}(y_1 \mid y_2) > G_{a+1}(y_1 \mid y_2)$ for all $y_1, y_2 \in R^+$ [see equation (7)]. We have shown that if $B(a+1,r) - r > B(a+2,r) - r$ then $B(a,r) - r > B(a+1,r) - r$. To complete the inductive argument we need only note that

$$B(L-1,r) - r = \beta \int_{c_L}^{\infty} (y - c_L)dG_L(y \mid r) > 0 = B(L,r) - r.$$

Corollary 1: For each age there exists a unique $\bar{r}_a \in [0,c_a]$ such that it is optimal to renew the patent if and only if $r_a > \bar{r}_a$. Moreover, the sequence $\{\bar{r}_a\}_{a=1}^{L}$ is nondecreasing in a.

Proof: Recall that $V(a,r) = \max\{0, B(a,r) - c_a\}$, and note that $B(a,0) = 0$, while $B(a,c_a) > c_a$. The fact that $B(a,r) - c_a$ has a unique zero at an $\bar{r}_a \in [o,c_a]$ now follows directly from part (i) of proposition 1 as it implies that $B(a,r)$ is continuous and strictly increasing in r. Since $B(\cdot)$ is nonincreasing in a, $B(a,\bar{r}_{a+1}) > B(a+1,\bar{r}_{a+1}) = c_{a+1} > c_a = B(a,\bar{r}_a)$, for all a. The second statement in the corollary follows from this inequality and the fact the $B(a,r)$ is nondecreasing in r, for $r \in R^+$ and $a=1...L$.

Appendix 2. Properties of the Maximum Likelihood Estimator

The properties of the maximum likelihood estimator used in the text are stated in proposition 2. This proposition follows directly from a theorem due to Rao [1973, section 5.e.2] provided the following regularity conditions are satisfied:

A. The functions $\pi(a; \underset{\sim}{c}_j, \underset{\sim}{\omega})$ $[a \in A_j;\ j = 1, \ldots, J]$ admit first order partials which are continuous at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^o$.

B. For every $\underset{\sim}{\omega} \in T$, such that $\underset{\sim}{\omega} \neq \underset{\sim}{\omega}^o$, $\pi(a; \underset{\sim}{c}_j, \underset{\sim}{\omega}) \neq \pi(a; \underset{\sim}{c}_j, \underset{\sim}{\omega}^o)$ for at least one couple (a, j) $[a \in A_j;\ j = 1, \ldots, J]$.

C. The information matrix, $[i_{rs}]$, is non-singular at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^o$.

As noted in the text, the benefit function from the model of section one [equation (2)] has points with discontinuous first partials. As a result it is not immediately obvious that condition A is satisfied, and a formal proof of this condition is provided below.[1] Given this proof, I simply assume conditions B and C. They will be satisfied provided there is sufficient variation in the cost schedules and the ages covered by the data.

Proof of Condition A. From the definition of π_a [equation 5], and Bayes Law

$$\pi_a = \Pr\{r_a < \bar{r}_a,\ r_{a-1} > \bar{r}_{a-1}\} = \Pr\{r_a < \bar{r}_a \mid r_{a-1} > \bar{r}_{a-1}\}[1 - \sum_{i<a} \pi_{a-i}] \tag{A2.1}$$

for $a=1, \ldots, L$; where the index j has been omitted for convenience, and it is understood both that; $\bar{r}_0 = \sum_{i<1} \pi_{1-i} = 0$, so that $\pi_1 = \Pr\{r_1 < \bar{r}_1 \mid r_0 > 0\}$; and that all functions depend also on $\underset{\sim}{\omega}$. The proof of the continuity of the first

---

[1] To see why condition A can be satisfied despite the fact that $B(a,r)$ is not differentiable everywhere, note that the sequence $\{\pi(a;\underset{\sim}{c},\underset{\sim}{\omega})\}_{a=1}^{L}$ depends on the benefit function only through the cutoffs (see equation 5); i.e. through the solution $\bar{r}_a = r(a,\underset{\sim}{c}^a,\underset{\sim}{\omega})$ to the equations $B(a,\bar{r}_a;\underset{\sim}{c}^a,\underset{\sim}{\omega}) - c_a = 0$ for $a=1,\ldots,L$. Though not differentiable in r everywhere, $B(a,r)$ is differentiable in r in an interval containing $\bar{r}_a$; see equations A3.2 and A3.3 of Appendix 3 (this is also illustrated in Figure 1 of section 1).

partials of $\pi_a(\underset{\sim}{\omega})$ is by induction on a. The first step of the inductive argument is to show that $\pi_a(\underset{\sim}{\omega})$ has continuous first partials with respect to $\underset{\sim}{\omega}$ at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^o$ provided $\pi_{a-1}(\underset{\sim}{\omega}), \ldots, \pi_1(\underset{\sim}{\omega})$ do. The second step is to show that $\pi_1(\underset{\sim}{\omega})$ has continuous first partials at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^o$. Equation (A2.1) makes it clear that if we establish that each element of the sequence $\{\Pr\,(r_a < \bar{r}_a \mid r_{a-1} > \bar{r}_{a-1})\}_{a=1}^{L}$ has continuous first partials with respect to $\underset{\sim}{\omega}$ at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^o$, we will have completed both steps of the inductive argument.

Let $\psi_1(z) = \Pr\,\{r_a = 0 \mid r_{a-1} = z\}$, and $\psi_{2,a}(z) = \Pr\,\{0 < r_a < \bar{r}_a \mid r_{a-1} = z\}$, for a = 1, ..., L. Then, from the definition of $F_{a-1}(\cdot)$ in equation (3), and equation (6)

$$\Pr\,\{r_a < \bar{r}_a \mid r_{a-1} > \bar{r}_{a-1}\} = \int_{\bar{r}_{a-1}}^{\infty} \psi_1(z)dF_{a-1}(z) + \int_{\bar{r}_{a-1}}^{\infty} \psi_{2,a}(z)dF_{a-1}(z) \qquad \text{(A2.2)}$$

where it is understood that the point $\bar{r}_{a-1}$ is excluded from the limit of integration, for a=1, ..., L. Lemma 1 below shows that $\{\bar{r}_a(\underset{\sim}{\omega})\}_{a=1}^{L}$ admit continuous first partials at $\underset{\sim}{\omega}^o$; while lemma 2 shows that $F_{a-1}(z)$ has a density which is both continuous in z at $z = \bar{r}_{a-1}(\underset{\sim}{\omega}^o)$, and admits continuous first partials with respect to $\underset{\sim}{\omega}$ at $\underset{\sim}{\omega}^o$ everywhere for $z \in (\bar{r}_{a-1}, \infty)$. Thus, to prove the desired result, it suffices to show that $\psi_1(\cdot)$ and $\psi_{2,a}(\cdot)$ are both continuous in z at $z = \bar{r}_{a-1}(\underset{\sim}{\omega}^o)$, and have continuous first partials with respect to $\underset{\sim}{\omega}$ at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^o$ almost everywhere with respect to the Lebesgue measure (a.e.) for $z \in (\bar{r}_{a-1}, \infty)$. Since $\psi_1(z) = \exp\,(-\theta z)$, it obviously satisfies these conditions. Equation (6) implies that

$$\psi_{2,a}(z) = \begin{cases} 0 & \text{if } z > \delta^{-1}\bar{r}_a \\ [1 - \exp\,(-\theta z)]Q_{a-1}(\bar{r}_a) & \text{if } 0 < z < \delta^{-1}\,\bar{r}_a, \end{cases}$$

for a=2, ..., L. Note that $\bar{r}_{a-1}(\underset{\sim}{\omega}^o) < \bar{r}_a(\underset{\sim}{\omega}^o)/\delta^o$ (from corollary 1 and the fact that $\delta<1$). Clearly then $\psi_{2,a}(z)$ is both continuous in z at $z = \bar{r}_{a-1}\,(\underset{\sim}{\omega}^o)$, and has continuous first partials with respect to $\underset{\sim}{\omega}$ at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^o$ everywhere except at the point $z = \bar{r}_a\,(\underset{\sim}{\omega}^o)\,/\,\delta^o$, which is a set of Lebesgue measure zero (for a = 1, ..., L).

Lemma 1. Each element of the sequence of functions $\{\bar{r}_a(\underset{\sim}{\omega})\}_{a=1}^{L}$ admits first partials which are continuous at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^o$.

Proof. The proof is by (this time backwards) induction on a. Since $\bar{r}_L = c_L$, the initial condition of the inductive argument is satisfied trivially, and it suffices to show that $\bar{r}_a(\underset{\sim}{\omega})$ admits continuous first partials with respect to $\underset{\sim}{\omega}$ at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^o$ provided $\bar{r}_{a+\tau}(\underset{\sim}{\omega})$ [τ=1, 2, ..., L-a] do. Corollary 1 and equation (6) imply that $\bar{r}_a(\cdot)$ is defined by the implicit function

$$\mu(\bar{r}_a, \underset{\sim}{\omega}) = \bar{r}_a + \beta[1-\exp(-\theta\bar{r}_a)]\int_{\bar{r}_{a+1}}^{\infty} V_{a+1}(z)dQ_a(z) - c_a = 0. \tag{A2.3}$$

Clearly $\mu(\cdot)$ possesses a continuous, strictly positive, partial derivative with respect to $\bar{r}_a$. The implicit function theorem therefore implies the lemma provided $\mu(\cdot)$ admits continuous first partials with respect to $\underset{\sim}{\omega}$ at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^o$. The hypothesis of the inductive argument implies that $\bar{r}_{a+1}(\underset{\sim}{\omega})$ has continuous first partials; and $Q_a(z)$ is an exponential distribution which has a density which possesses continuous first partials with respect to $\underset{\sim}{\omega}$ everywhere for $z \in R^+$. It will, therefore, suffice to show that $V_{a+1}(z)$ has continuous first partials with respect to $\underset{\sim}{\omega}$ at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^o$ a.e. for $z \in (\bar{r}_{a+1}, \infty)$, provided that $\bar{r}_{a+\tau}(\underset{\sim}{\omega})$ [for τ=1, ... L-a] has continuous first partials at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^o$. A second inductive argument suffices to prove this point.

First, assume $V_{a+2}(z)$ has continuous first partials with respect to $\underset{\sim}{\omega}$ a.e. for $z \in (\bar{r}_{a+2}, \infty)$ provided that $\bar{r}_{a+\tau}(\omega)$[τ=2, ..., L-a] have continuous first partials at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^o$. To see that this implies that $V_{a+1}(\cdot)$ has the required properties, note that

$$V_{a+1}(z) = \begin{cases} v^1_{a+1}(z) = z - c_{a+1} + \beta[1 - \exp(-\theta z)]\int_{\bar{r}_{a+2}}^{\infty} V_{a+2}(s)dQ_{a+1}(s), \\ \qquad\qquad \text{if } z \in [\bar{r}_{a+1}, \delta^{-1}\bar{r}_{a+2}]; \\ v^2_{a+1}(z) = z-c_{a+1} + \beta[1-\exp(-\theta z)]\{Q_{a+1}(\delta z)V_{a+2}(\delta z) + \int_{\delta z}^{\infty} V_{a+2}(s)dQ_{a+1}(s)\}, \\ \qquad\qquad \text{if } z \in (\delta^{-1}\bar{r}_{a+2}, \infty). \end{cases} \tag{A2.4}$$

Given the hypothesis of the inductive argument, it is clear that $v^1_{a+1}(z)$ has continuous first partials with respect to $\underset{\sim}{\omega}$ at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^o$ for every $z \in [\bar{r}_{a+1}, \infty)$. For $z \in (\delta^{-1} r_{a+2}, \infty)$ the points of discontinuity of the first partials of $v^2_{a+1}(z)$ are the points of discontinuity of the first partials $V_{a+2}(\delta^{-1}z)$. Thus, if $S_{a+1}[S_{a+2}]$ is the set of points in $(\bar{r}_{a+1}, \infty)$ $[(\bar{r}_{a+2}, \infty)]$ where $V_{a+1}(z)$ $[V_{a+2}(z)]$ has discontinuous first partials with respect to $\underset{\sim}{\omega}$ at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^o$, then

$$m(S_{a+1}) \leq m(S_{a+2}) + m(\bar{r}_{a+2}\delta^{-1} \mid_{\underset{\sim}{\omega} = \underset{\sim}{\omega}^o}) = m(S_{a+2}) = 0,$$

where $m(\cdot)$ provide the Lebesgue measure of alternative sets, and the last equality follows from the hypothesis of the inductive argument.

Thus $V_{a+1}(z)$ has continuous first partials with respect to $\underset{\sim}{\omega}$ at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^o$ a.e. for $z \in (\bar{r}_{a+1}, \infty)$ provided $V_{a+2}(\cdot)$ has the required property. To complete the inductive argument then, we need only note that $V_L(z) = \max\{0, z - c_L\}$ a function that has continuous first partials at $\omega = \omega^o$ everywhere for $z \in (\bar{r}_L = c_L, \infty)$.

<u>Lemma 2</u>. $F_a(z)$ has a density which is both continuous in z at $z = \bar{r}_a(\underset{\sim}{\omega}^o)$ and admits continuous first partials with respect to $\underset{\sim}{\omega}$ at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^o$ everywhere for $z \in (r_a, \infty)$ (and $a = 1, \ldots, L$).

<u>Proof</u>. The proof is by forward induction on a. First assume $F_{a-1}(\cdot)$ has a density with the required properties and denote that density by $f_{a-1}(\cdot)$. Then equation (6) and corollary 1 imply that

$$\Pr\{z \geq r_a \geq \bar{r}_a\} = \int_{\bar{r}_{a-1}}^{\infty} \Pr\{z \geq r_a \geq \bar{r}_a \mid s\}\, f_{a-1}(s)ds, \tag{A2.5}$$

where

$$\Pr\{z \geq r_a \geq \bar{r}_a \mid s\} = \begin{cases} [1 - \exp(-\theta s)]Q_{a-1}(z), & \text{if } \delta^{-1}z \geq s > \bar{r}_{a-1} \\ 0 & , \text{if } s > \delta^{-1}z, \end{cases} \tag{A2.6}$$

and $Q_{a-1}(z)$ denotes the exponential distribution evaluated at z. Substituting (A2.6) into (A2.5) we have

$$\Pr\{z > r_a > \bar{r}_a\} = Q_{a-1}(z) \int_{\bar{r}_{a-1}}^{\delta^{-1}z} [1 - \exp(-\theta s)] f_{a-1}(s) ds. \qquad \text{(A2.7)}$$

The density, $f_a(z)$ for $z \in (\bar{r}_a, \infty)$, can be derived directly by differentiation of (A2.7). The fact that it is continuous in z at $z = \bar{r}_a(\omega^o)$ and possesses continuous first partials with respect to $\underset{\sim}{\omega}$ at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^o$ everywhere for $z \epsilon (\bar{r}_a, \infty)$ follows from the same properties of; the exponential distribution and its density, of $f_{a-1}(z)$ everywhere for $z \in [\bar{r}_a / \delta^o, \infty)$ [which follows from the hypothesis of the inductive argument since $\bar{r}_a / \delta^o > \bar{r}_{a-1}$ from Corollary 1], and from the continuity of the first partial of $\bar{r}_{a-1}(\underset{\sim}{\omega})$ at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^o$ [lemma 1].

To complete the inductive argument then we need only show that $F_2(z)$ has a density with continuous first partials at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^o$ everywhere for $z \in (\bar{r}_2, \infty)$. This follows from the same argument used above (substituting a = 2) and the fact that $f_1(z)$ is the density of the lognormal distribution [equation (7)] which clearly has continuous first partials with respect to $\underset{\sim}{\omega}$ at $\underset{\sim}{\omega} = \underset{\sim}{\omega}^o$ everywhere for $z \in R^+$.

Appendix 3. A Solution for the Sequence $\{\bar{r}_a\}_{a=1}^{L}$

I begin by outlining the form of the solution used here. To find $\bar{r}_{a-1} = r(a-1;\, \underset{\sim}{c},\, \underset{\sim}{\omega})$ we require properties of the function $E[V_a \| \Omega_{a-1}]$. The model [equations (2) and (6) and assumptions A1 and A2] implies that the distribution of V_a conditional on Ω_{a-1} can be written as

$$V_a \Big| \Omega_{a-1} = \begin{cases} 0 & \text{with probability } \exp(-\theta r_{a-1}) \\ \max\{0,\, \max(\delta r_{a-1},\, z) - c_a & \\ \quad + \beta\, g_a\, [\max(\delta r_{a-1},\, z)]\} & \text{with probability } [1-\exp(-\theta r_{a-1})] \end{cases} \tag{A3.1}$$

where $\Pr\{z < y\} = Q_a(y)$, and $g_a(y) = E[V_{a+1} \mid y]$.

It follows from corollary 1 that if $r_a < \bar{r}_a$, the agent will let the patent lapse, and $V_a = 0$. Now consider a patent with $\delta r_{a-1} < \bar{r}_a$. With probability $\exp(-\theta r_{a-1})$, $r_a = 0$. With probability $[1-\exp(-\theta r_{a-1})]$, the current return the patent would earn were it to be renewed is $\max(\delta r_{a-1},\, z)$. If $z < \bar{r}_a$, then $\max(\delta r_{a-1},\, z) < \bar{r}_a$, and $V_a = 0$. If $z > \bar{r}_a$, then $\max(\delta r_{a-1},\, z) = z$ and $V_a = z - c_a + \beta g_a(z)$. Formally then, A3.1 and corollary 1 imply that

$$\begin{aligned} E[V_a \mid r_{a-1},\, r_{a-1} < \delta^{-1}\bar{r}_a] &= [1 - \exp(-\theta r_{a-1})]\, \Big\{\int_{\bar{r}_a}^{\infty} [z - c_a + \beta g_a(z)]\, dQ_a(z)\Big\} \\ &\equiv [1-\exp(-\theta r_{a-1})]\, h^0_{a-1} \end{aligned} \tag{A3.2}$$

where h^0_{a-1} is independent of r_{a-1}.

Substituting (A3.2) into equation (2),

$$V_{a-1}(r_{a-1} \mid r_{a-1},\, r_{a-1} < \delta^{-1}\bar{r}_a) = \max\{0,\, r_{a-1} - c_{a-1} + \beta[1-\exp(-\theta r_{a-1})]h^0_{a-1}\} \tag{A3.3}$$

Now note from corollary 1 that $\bar{r}_{a-1} < \bar{r}_a < \delta^{-1}\bar{r}_a$. Clearly then (A3.3) implies that $\bar{r}_{a-1}$ is the unique solution to,

$$\bar{r}_{a-1} - c_{a-1} + \beta[1-\exp(-\theta\bar{r}_{a-1})]h^0_{a-1} = 0 \tag{A3.4}$$

for a = 1, ..., L. Equation A3.4 provides a solution for $\{\bar{r}_a\}_{a=1}^{L}$ in terms of $\{h_a^0\}_{a=1}^{L}$. Below I find functions $B_{a+1}^{v}(\cdot)$ and $U_{a+1}^{v}(\cdot)$ such that

$$h_a^0 = \sum_{v=0}^{L-a-1} B_{a+1}^{v}(\bar{r}_{a+1+v}, \underset{\sim}{\omega}) + \sum_{v=0}^{L-a-2} U_{a+1}^{v}(\bar{r}_{a+1+v}, \underset{\sim}{\omega}) h_{a+1+v}^0 \qquad (A3.5)$$

for a=1, ..., L-1 ($h_L^0 = 0$). Together (A3.4) and (A3.5) provide a system of 2L-1 equations which can be solved recursively for $\{\bar{r}_a\}_{a=1}^{L}$.

To solve for h_a^0 we require $E[V_{a+1} \mid r_a]$ for $r_a \varepsilon R^+$. We now introduce a sequence of elementary functions which are used to construct $E[V_{a+1} \mid r_a]$. To begin partition R^+ into the L-a+1 intervals, $\{I_a^{\,p}\}_{p=0}^{L-a}$; where $I_a^0 = (0, \delta^{-1}\bar{r}_{a+1}]$; $I_a^p = (\delta^{-p}\bar{r}_{a+p}, \delta^{-(p+1)}\bar{r}_{a+p+1}]$, for p=1, ..., L-a-1; and $I_a^{L-a} = (\delta^{-(L-a)}\bar{r}_L, \infty)$. From corollary 1, $E[V_{a+1} \mid r_a, r_a < \bar{r}_a] = 0$; while from equation (A3.2), $E[V_{a+1} \mid r_a, \bar{r}_a < r_a \varepsilon I_a^0] = [1-\exp(-\theta r_a)]h_a^0$. To complete the specification of the function $E[V_{a+1} \mid r_a]$, define

$$h_a^v(r_a) = [1-\exp(-\theta r_a)]^{-1}\{E[V_{a+1} \mid r_a, r_a \varepsilon I_a^v] - E[V_{a+1} \mid r_a, r_a \varepsilon I_a^{v-1}]\},$$

for v-1,..., L-a; so that

$$E[V_{a+1} \mid r_a, r_a \varepsilon I_a^p] = [1-\exp(-\theta r_a)][\sum_{v=0}^{p} h_a^v(r_a)] \qquad (A3.6)$$

for p=1, ..., L-a, and a=1, ..., L-1. These equations and the definition of $V_a(r_a)$ [equation (2)] imply the graph of $V_a(r_a)$ provided in Figure A3. This graph is now used to illustrate how the sequence $\{h_{a-1}^v(\cdot)\}_{v=0}^{L-a+1}$ can be derived from the sequence $\{h_a^v(\cdot)\}_{v=0}^{L-a}$. The functional recursion that results from this derivation uniquely determines the sequences $\{h_a^v(\cdot)\}_{v=0}^{L-1}$ for a=1, ..., L-1. The first element of each of these sequences is them used in conjunction with (A3.4) to solve for $\{\bar{r}_a\}_{a=1}^{L}$.

Figure A3: Graph of $V_a(r_a)$

	$r_a - c_a$	$r_a - c_a$	...	$r_a - c_a$	$r_a - c_a$	...
	+	+		+	+	
	$\beta[1-\exp(-\theta r_a)]$	$\beta[1-\exp(-\theta r_a)]$	...	$\beta[1-\exp(-\theta r_a)]$	$\beta[1-\exp(-\theta r_a)]$	...
0	•	•		•	•	
	h_a^0	$[h_a^0 + h_a^1(r_a)]$	...	$\sum_{v=0}^{q-1} [h_a^v(r_a)]$	$\sum_{v=0}^{q} [h_a^v(r_a)]$	...
		$(\leftarrow\text{-----}I_a^1\text{-----}\rightarrow]$	...	$(\leftarrow\text{-----}I_a^{q-1}\text{--}\rightarrow]$	$(\leftarrow\text{-----}I_a^q\text{-----}\rightarrow]$	
$\bar{r}_a$	$\delta^{-1}\bar{r}_{a+1}$	$\delta^{-2}\bar{r}_{a+2}$	$\delta^{-(q-1)}\bar{r}_{a+(q-1)}$	$\delta^{-q}\bar{r}_{a+q}$	$\delta^{-(q+1)}\bar{r}_{a+q+1}$ $[r_a]$	

Recall that $V_{a-1}(r_{a-1}) = \max\{0,\ r_{a-1} - c_{a-1} + \beta\ E[V_a \mid r_{a-1}]\}$. Consider first the case in which $\bar{r}_{a-1} < r_{a-1} \in I^0_{a-1}$. In this case the agent will let the patent lapse in the next period if either the absorbing state is drawn or if $z < \bar{r}_a$. Thus the only states in which $V_a > 0$ are those in which the absorbing state does not occur and $z > \bar{r}_a$. For each such state one obtains $V_a(r_a{=}z)$ by substituting z for r_a in Figure A3. To obtain $E[V_a \mid r_{a-1},\ \bar{r}_{a-1} < r_{a-1} \in I^0_{a-1}]$ we simply intergrate these values over their measure, or

$$\begin{aligned} E[V_a \mid r_{a-1},\ \bar{r}_{a-1} < r_{a-1} \in I^0_{a-1}] &= [1-\exp(-\theta r_{a-1})]\ [\int_{\bar{r}_a}^{\infty} (z-c_a)dQ_a(z) \\ &\quad + \beta \sum_{v=0}^{L-a} \int_{\delta^{-v}\bar{r}_{a+v}}^{\infty} [1-\exp(-\theta z)]h^v_a(z)dQ_a(z)] \qquad \text{(A3.7)} \\ &\equiv [1-\exp(-\theta r_{a-1})]h^0_{a-1}, \end{aligned}$$

which defines h^0_{a-1}. To find $h^1_{a-1}(\cdot)$ in terms of $\{h^v_a(\cdot)\}_{v=0}^{L-a}$ consider $E[V_a \mid r_{a-1},\ r_{a-1} \in I^1_{a-1}]$, and recall that if $r_{a-1} \in I^1_{a-1}$, then $\delta r_{a-1} \in (\bar{r}_a, \delta^{-1}\bar{r}_{a+1}]$. If the absorbing state does not occur and $z < \delta r_{a-1}$ then $V_a(r_a) = V_a(r_a = \delta r_{a-1} \in I^0_a)$; while if the absorbing state does not occur and $z > \delta r_{a-1}$, $V_a(r_a) = V_a(r_a{=}z)$. Taking these values of $V_a(\cdot)$ from the graph and integrating them over their measure

$$\begin{aligned} E[V_a|r_{a-1},\ r_{a-1} \in I^1_{a-1}] &= [1-\exp(-\theta r_{a-1})]\ \{Q_a(\delta r_{a-1})[\delta r_{a-1}-c_a+\beta h^0_a[1-\exp(-\theta\delta r_{a-1})] \\ &\quad + \int_{\delta r_{a-1}}^{\infty} [z-c_a+\beta h^0_a[1-\exp(-\theta z)]]dQ_a(z) + \beta \sum_{v=1}^{L-a} H^v_a(\infty)\} \end{aligned}$$

where here, and in the discussion below, $H^v_a(x) = \int_{\delta^{-v}\bar{r}_{a+v}}^{x} \{[1-\exp(-\theta z)]h^v_a(z)\} dQ_a(z)$. This equation together with (A3.6) and (A3.7) imply that

$$\begin{aligned} h^1_{a-1}(r_{a-1}) &= Q_a(\delta r_{a-1})\{\delta r_{a-1}-c_a+\beta h^0_a[1-\exp(-\theta\delta r_{a-1}]\} \\ &\quad - \int_{\bar{r}_a}^{\delta r_{a-1}} \{z-c_a+\beta h^0_a\ [1-\exp(-\theta z)]\} dQ_a(z). \qquad \text{(A3.8)} \end{aligned}$$

Finally, following an analogous procedure for v=1, ..., L-a-1, we have

$$h_{a-1}^{v+1}(r_{a-1}) = \beta Q_a(\delta r_{a-1})[1-\exp(-\theta\delta r_{a-1})]h_a^v(\delta r_{a-1}) - \beta H_a^v(\delta r_{a-1}). \qquad (A3.9)$$

Noting that $\int_y zdQ_a(z) = [y + \sigma_a][1-Q_a(y)]$, and that $\bar{r}_a - c_a = -\beta h_a^0[1-\exp(-\theta\bar{r}_a)]$, the system defined by (A3.7), (A3.8), and (A3.9) can be simplified to read

$$\begin{aligned} h_a^0 &= B_{a+1}^0 + \beta\, U_{a+1}^0\, h_{a+1}^0 + \beta \sum_{v=1}^{L-a-1} H_{a+1}^v(\infty) && \text{(a)} \\ h_{a+1}^1(z) &= b_{a+1}^1(z) + \beta\, u_{a+1}^1(z)\, h_{a+2}^0 && \text{(b)} \qquad (A3.10) \\ \text{and}\quad h_{a+1}^v(z) &= \beta Q_{a+2}(\delta z)[1-\exp(-\theta\delta z)]h_{a+2}^{v-1}(\delta z) - \beta H_{a+2}^{v-1}(\delta z), && \text{(c)} \end{aligned}$$

where $B_{a+1}^0 = \sigma_{a+1}[1-Q_{a+1}(\bar{r}_{a+1})]$; $U_{a+1}^0 = B_{a+1}^0\theta\exp(-\theta\bar{r}_{a+1})/(1+\sigma_{a+1}\theta)$; $b_{a+1}^1(z) = \delta z - \bar{r}_{a+2} - B_{a+2}^0\{1-\exp[-\sigma_{a+2}^{-1}(\delta z-\bar{r}_{a+2})]\}$; and $u_{a+1}^1(z) = \exp(-\theta\bar{r}_{a+2})\{1-\exp[-\theta(\delta z-\bar{r}_{a+2})]\} - U_{a+2}^0\{1-\exp[-(\theta+\sigma_{a+2}^{-1})(\delta z-\bar{r}_{a+2})]\}$; for v=2, ..., L-a-1; and a=1, ..., L-1.

Direct substitution shows that the system in (A3.10) provides a solution for the sequence $\{h_a^0\}_{a=1}^L$ of the form given in (A3.5). Even for moderate L, however, solving for the sequences $\{B_a^v\}_{v=0}^{L-a-1}$ (for a=1, ..., L-1) and $\{U_a^v\}_{v=0}^{L-a-2}$ (for a=1, ..., L-2) manually would be both a painstaking and an error-prone task. One advantage of the form given by (A3.10) is that it can be programmed into Macsyma (1983; Macsyma is a computer program designed for symbolic mathematical manipulations) which will (with some prodding) produce the exact form of the required coefficients.[1] The properties of the model [in particular the continuity of the value function, see proposition 1] together with the features of

---

[1] The macsyma solution was obtained by Andrew Myers and myself. We are grateful to the Mathlab Group at the MIT Laboratory for Comuputer Science for providing us with free access to Macsyma, and for guiding us through out initial queries. The Mathlab Group is supported, in part, by the U.S. Energy Research and Development Administration under Contract No. E(11-1) 3070, and by the National Aeronautics and Space Administration under Grant No. NSG 1323.

Macsyma [its ability to distinguish an argzero of a function] allow a check of the Macsyma solution for possible programming errors. Finally, for large L, it is not necessary to solve for all of the coefficients $\{B_a^v\}$ and $\{U_a^v\}$. This is the second advantage of writing the solution in the form of (A3.10). It follows from the fact that for each (v, a) there exist easily calculable functions of the parameters of the model $(\underline{B}_a^v, \bar{B}_a^v)$ and $(\underline{U}_a^v, \bar{U}_a^v)$, such that $\underline{B}_a^v < B_a^v < \bar{B}_a^v$ and $\underline{U}_a^v < U_a < \bar{U}_a^v$. These boundary functions allow one to form the approximations $\hat{B}_a^v = 1/2(\bar{B}_a^v + \underline{B}_a^v)$ and $\hat{U}_a^v = 1/2\ (\bar{U}_a^v + \underline{U}_a^v)$; each of which have a maximum possible approximation error equal to the value of the approximation itself. Both these functions decline monotonically (and rather rapidly) in v; and have zero limits as $v\to\infty$ (which implies that $L\to\infty$). I now derive the boundary functions, and consider their limiting rates of convergence. The researcher can use the exact approximations to calculate where the approximation error is within tolerable limits for the problem at hand.

Note first that (A3.10) and the fact that $h^1_{a+v+1}(z)$ is nonnegative and nondecreasing in z for $z\varepsilon[\delta^{-v}\bar{r}_{a+1+v}, \infty)$, implies that $h^{v-j}_{a+j+1}(\delta^j z)$ is nonnegative and nondecreasing in z for $z\varepsilon[\delta^{-v}\bar{r}_{a+1+v}, \infty)$. This fact allows us to show that $\beta H^v_{a+1}(\infty)$ is bounded by two simple functionals $h^1_{a+v}(\cdot)$. The upper bound is found by noting that (A3.10c) implies that

$$h^v_{a+1}(z) \leq \beta h^{v-1}_{a+2}\ (\delta z) \leq \dots \leq \beta^{v-1} h^1_{a+v}(\delta^{v-1} z)$$

for $z\varepsilon[\delta^{-v}\bar{r}_{a+1+v},\infty)$. Substituting this inequality into the definition of $\beta H^v_{a+1}(\infty)$,

$$\beta H^v_{a+1}(\infty) \leq \beta^v \int_{\delta^{-v}\bar{r}_{a+1+v}}^{\infty} h^1_{a+v}\ (\delta^{v-1} z)dQ_{a+1}(z), \qquad A(3.11).$$

To obtain the lower bound note that the monotonicity of $h^{v-1}_{a+2}(\cdot)$ and the definition of $H^{v-1}_{a+2}(\cdot)$ imply that

$$\beta H^{v-1}_{a+2}(\delta z) \leq \beta[1-\exp(-\delta\theta z)]h^{v-1}_{a+2}(\delta z)\{Q_{a+2}(\delta z) - Q_{a+2}(\delta^{-v+1}\bar{r}_{a+1+v})\}$$

for $z\varepsilon[\delta^{-v}\bar{r}_{a+v+1}, \infty)$. Substituting this inequality into (A3.10c), noting that for $z\varepsilon[\delta^{-v}\bar{r}_{a+v+1}, \infty)$, $[1 - \exp(-\theta\delta z)] \geq [1 - \exp(\theta\delta^{-v+1}\bar{r}_{a+v+1}]$, and repeating these operations recursively,

$$h^{v}_{a+1}(z) \geq \beta^{v-1}\kappa'_{v,a}h^{1}_{a+v}(\delta^{v-1}z) \tag{A3.12}$$

where $\kappa'_{v,a} = \prod_{j=1}^{v-1} [1-\exp(-\theta\delta^{-v+j}\bar{r}_{a+1})]\, Q_{a+1+j}(\delta^{-v+j}\bar{r}_{a+1+j})$, for $v=1,\ldots L-a-1$, and $z\varepsilon[\delta^{-v}\bar{r}_{a+1+v},\infty)$. Substituting (A3.12) into the definition of $\beta H^{v}_{a+1}(\infty)$ we have,

$$\beta H^{v}_{a+1}(\infty) \geq \beta^{v}\kappa_{v,a} \int_{\delta^{-v}\bar{r}_{a+1+v}}^{\infty} h^{1}_{a+v}(\delta^{v-1}z)dQ_{a+1}(z), \tag{A3.13}$$

where $\kappa_{v,a} = [1-\exp(-\theta\delta^{-v}\bar{r}_{a+1+v})]\kappa'_{v,a}$. From (A3.10b) then

$$\underline{A}^{v}_{a+1} + \underline{U}^{v}_{a+1}h^{0}_{a+v+1} \leq \beta^{v}H^{v}_{a+1}(\infty) = A^{v}_{a+1} + U^{v}_{a+1}\, h^{0}_{a+v+1} \leq \bar{A}^{v}_{a+1} + \bar{U}^{v}_{a+1}\, h^{0}_{a+1+v} \tag{A3.14}$$

where;

$\bar{A}^{v}_{a+1} = \beta^{v}\delta^{v}\sigma_{a+1}[1-Q_{a+1}(\delta^{-v}\bar{r}_{a+v+1})][(\delta/\phi)^{v} + Q_{a+v+1}(\bar{r}_{a+v+1})]/[1 + (\delta/\phi)^{v}]$;

$\underline{A}^{v}_{a+1} = \kappa_{v,a}\bar{A}^{v}_{a+1}$; $\underline{U}^{v}_{a+1} = \kappa_{v,a}\bar{U}^{v}_{a+1}$; and

$\bar{U}^{v}_{a+1} = \beta^{v+1}\delta^{v}\sigma_{a+1}[1-Q_{a+1}(\delta^{-v}\bar{r}_{a+v+1})]\theta\exp(-\theta\bar{r}_{a+v+1})\{[(\delta/\phi)^{v}$

$+(1+\sigma_{a+1}\delta^{v}\theta)Q_{a+v+1}(\bar{r}_{a+v+1})]/(1+\sigma_{a+1}\delta^{v}\theta)(1+\theta\delta^{v}\sigma_{a+1} + (\delta/\phi)^{v}\}$,

and $\phi = \sigma_{a+1}/\sigma_{a}$, for $v=1,\ldots,L-a-1$; and $a=1,\ldots,L-1$.

Recall that we are considering approximations for the sequences $\{A^{v}_{a+1}\}$ and $\{U^{v}_{a+1}\}$ of the form $\{\hat{A}^{v}_{a+1} = 1/2\ (\bar{A}^{v}_{a+1} + \underline{A}^{v}_{a+1})\}$ and $\{U^{v}_{a+1} = 1/2(\bar{U}^{v}_{a+1} + \underline{U}^{v}_{a+1})\}$.

Clearly the resulting approximation errors will be less than the values of the sequences themselves and these can be obtained directly from (A3.14). To consider their limiting properties as $v \to \infty$ (which implies $L \to \infty$) recall that $\beta, \delta, \phi, < 1$ and assume that the nondecreasing sequence $\{c_a\}_{a=1}^{\infty}$ has limit equal to $\bar{c}$ [this implies that $\lim(a \to \infty)\ \bar{r}_a = \bar{c}$]. Then A3.14 implies that, for each age, both sequences $\{\bar{A}_{a+1}^{v}\}_{v=0}^{\infty}$ and $\{\bar{U}_{a+1}^{v}\}_{v=0}^{\infty}$ converge to zero at the limiting rate $\beta^{v}\delta^{v}\exp(-\delta^{-v}\ \bar{c}/\sigma_{a+1})$. Since the lower bounds, $\underline{A}_{a+1}^{v}$ and $\underline{U}_{a+1}^{v}$ decline faster then the upper bounds, so does the approximation error.

The following procedure was used to obtain the results reported in this paper. For each (relevant) age an exact calculation of the first four elements for both sequences $\{A_{a+1}^{v}\}_{v=0}^{L-a-1}$ and $\{U_{a+1}^{v}\}_{v=0}^{L-a-2}$ was made using (A3.10). After checking that using the approximation for the fourth element instead of its exact value did not have a perceptible effect on the cutoffs, that and subsequent elements were replaced by the approximations for them.