Econometrica, Vol. 54, No. 4 (July, 1986), 755-784

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

BY ARIEL PAKES[1]

In many countries patentees must pay an annual renewal fee in order to keep their patents in force. This paper presents and then estimates a model which uses observations on the proportion of different cohorts of patents which are renewed at alternative ages, and the relevant renewal fee schedules, to estimate the distribution of the returns earned from holding patents, and the evolution of this distribution function over the lifespan of the patents. Since patents are often applied for at an early exploratory stage of the innovation process, the model allows patentees to be uncertain about the sequence of returns that will be earned if the patent is kept in force. The paper solves the implied optimal stopping problem for the micro units, derives the implications of these solutions on the aggregate proportions renewed, and then estimates the parameters of the model from the aggregate data. Separate estimates are obtained from data on post World War II cohorts of patents in each of France, the United Kingdom, and Germany.

KEYWORDS: optimal stopping, maximum likelihood, simulation estimator, patent rights, renewal fees, option values, the value of patent protection.

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 allows us to recover the distribution of returns from holding patents at each age over the lifespan of patents from information on patent renewals. Separate estimates are obtained from data on 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). These estimates enable calculations

[1] 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, Tom Kurtz, Charles Manski, Daniel McFadden, Andrew Meyers, Dvora Ross, John Rust, Mark Schankerman, three referees, and an editor of this journal. 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, John Rust, 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.

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 have 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 is 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. 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 is 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 those values of the model's parameters which make the drop out proportions implied by the model as "close" as possible to those we actually observe.

Formally then, this paper presents and solves an optimal stopping model, derives the implications of this model on aggregate behavior, and then estimates the parameters of the model from aggregate data. Though optimal stopping models have appeared in the economic literature in several contexts (see, for example, Roberts and Weitzman (1981)), I do not know of another paper which derives an estimator for one's parameters. There have, however, been a small

number of recent studies which estimated alternative types of discrete choice optimal stochastic control models on micro data. In particular, Miller (1984) estimates a job matching model, and Wolpin (1984) estimates a sequential binary choice model for the birth sequences of married women. Though the techniques developed in both of these papers have a range of applications and provide an extremely rich interpretation of the data, they have one troublesome aspect which is shared by this paper. In all these models both the estimation technique and the empirical results depend on the details of the stochastic specification and, because of the complexity of the estimation problem, it is difficult to determine the robustness of the conclusions to the stochastic assumptions chosen (a point which we return to below). 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 nonstationarity 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. This simplifies the estimation problem considerably. On the other hand, 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 analytic 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.

Section 1 provides an overview of the renewal model used in this paper, while Section 2 fills in the details of its stochastic specification. In Section 3, I explain the estimation algorithm. Section 4 describes the data, provides the estimates, and considers their implications. This last section includes a characterization of the process by which the distribution of current returns earned from holding the patents in a cohort evolves over time, and explicit calculations of both the annual flow of returns resulting from the proprietary rights created by the patent laws, and of the distribution of the value of holding the patents in a cohort.

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 assumed 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. An implication of this uncertainty 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 $V(a)$ be the expected discounted value of patent protection to the agent just prior to its ath renewal. If the renewal fee is not paid the patent lapses and $V(a)=0$. 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 equals the expected discounted value of the patent at age $a+1$ conditional on current information. Formally then,

$$V(a)=\max\{0,\ r_a+\beta E[V(a+1)\,|\,\Omega_a]-c_a\} \qquad (a=1,\ldots,L), \tag{1}$$

where L is the statutory limit to patent lives, r_a is the current return to patent protection, Ω_a is the information set of the agent in the patents ath year, c_a is the cost of renewal, and it is understood that zero is an absorbing state in the stochastic process generating $\{V(a)\}_{a=1}^{L}$ (so that if the patent is not renewed at any age it will not be in force thereafter). In equation (1), $r_a+\beta E[V(a+1)\,|\,\Omega_a]$ is the total benefit from holding the patent (the sum of current returns and the discounted value of the option). If this expression is less than c_a, the agent lets the patent lapse.

To complete the description of the value function the conditional distributions of future returns and costs of renewal must be specified. Given these distributions, the solution for the sequence $\{V(a)\}_{a=1}^{L}$ can be obtained by starting with the terminal equation, i.e., $V(L)=\max\{0,\ r_L-c_L\}$, and integrating the system in (1) backwards recursively. Assumptions 1 and 2 provide the general properties of these distribution functions.

ASSUMPTION 1 (A1): $\Pr(z>r_{a+1}\,|\,\Omega_a)=G(z\,|\,r=r_a,\ a,\ \omega_g)$, *where* $\Pr(\cdot\,|\,\cdot)$ *denotes a conditional probability statement, and* ω_g *is a vector of parameters.*

ASSUMPTION 2 (A2): *Agents 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. Moreover the renewal fee schedule in every year is nondecreasing in age.*

These assumptions simplify the analysis considerably. Assumption 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. It is also a fact that all the renewal fee schedules are nondecreasing in age (see Section 4.1). I will assume an exogenously given initial distribution of current returns to patent protection (this differentiates among patents). Assumption A1 assumes that the stochastic process generating subsequent returns (i.e., generating $\{r_a\}_{a=2}^{L}$) is both Markov and the same for all patents.

To characterize the solution to the agent's decision problem I need more detailed assumptions on this Markov process. These additional assumptions are first explained, and then gathered into Assumption 3 below. I assume that the probability that next year's returns are greater than any given number is larger the higher are current returns (A3.3). Second, though the sequence of conditional distributions, i.e. $\{G(\cdot\,|\,\cdot, a)\}_{a=1}^{L}$, need not be stationary over age, they cannot become "better" at too fast a pace. A condition which suffices to rule out this possibility is that, for any given value of current returns, the probability that next year's returns is greater than some number is nonincreasing in age (A3.4). This type of nonstationarity turns out to be an important feature of the empirical results, and is discussed in more detail below. Finally I require regularity conditions that insure the finiteness and continuity of the value function (A3.1 and A3.2).[2]

ASSUMPTION 3: (A3.1) *There exists an* ε *such that* $E[r_a^{1+\varepsilon}|r_1]<\infty$ $(a=2,\ldots,L;\ r_1\in R_+)$. (A3.2) $G(z\,|\,r, a)$ *is continuous in* r *at every* z *except, possibly, at values of* z *at which* $G(z\,|\,r, a)$ *has a discontinuity in* z. (A3.3) $G(z\,|\,r, a)$ *is nonincreasing in* r. (A3.4) $G(z\,|\,r, a)$ *is nondecreasing in* a. [*In* A3.2 *to* A3.4, *it is to be understood that* $a=1,\ldots,L-1$; $(z, r)\in R_+^2$; *and that the conditional distributions are also indexed by the parameter vector,* ω_g.]

Assumptions A1 and A2 imply that the value of the option to renew the patent (i.e., $E[V(a+1)\,|\,\Omega_a]$) depends only on current returns and parameters which do not vary among patents of a given age (β, ω_g, and the current vector of renewal fees). If, for convenience, we omit these latter parameters from the notation, then the value function (equation 1) can be rewritten as

$$V(a, r)=\max\{0,\ r+\beta E[V(a+1)\,|\,r, a]-c_a\} \qquad (a=1,\ldots,L), \tag{2}$$

[2] Note that these conditions do not rule out conditional distributions with mass points, a common characteristic of the Markov processes used in search models.

where $E[V(a+1)\,|\,r, a] \equiv \int_{R_+} V(a+1, z)G(dz\,|\,r, a)$. Clearly in order to characterize the situations in which it is optimal to renew the patent, we require the properties of the function determining the option value. These properties are provided in Lemma 1 and explained immediately thereafter.

LEMMA 1 (proved in Appendix A): *The value of the option, that is* $E[V(a+1)\,|\,r, a]$, *is continuous and nondecreasing in* r, *and nonincreasing in* a ($r \in R_+$, $a = 1, \ldots, L$).

Figure 1 illustrates the form of the function determining the total benefits from renewing ($r + \beta E[V(a+1)\,|\,r, a]$). Since $V(a+1, z) \geq 0$ with probability one, the value of the option is nonnegative, and the total benefits are greater than r (the 45 degree line). Further, the assumption that the probability that future returns are greater than some number is larger the higher are current returns implies that the value of the option (the difference between the total benefit curve and the 45 degree line) is increasing in r. As the patent ages there are less future years in which the patent can earn returns, renewal fees rise, and the distribution of future returns conditional on current r is not as benevolent (see A3.4). These conditions insure that the option value is decreasing in age for each r. Note that

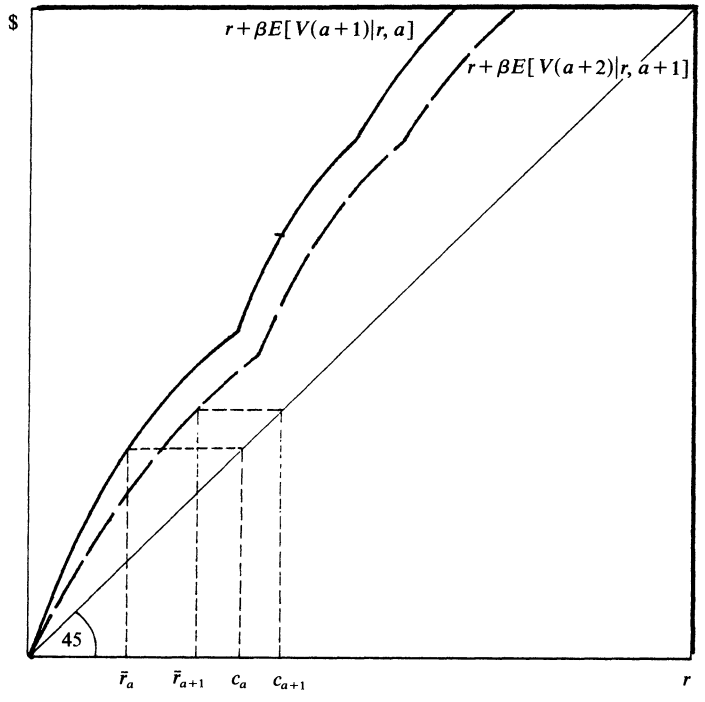


FIGURE 1.—Determining $\{\bar{r}_a\}_{a=1}^{L}$

though the total benefit function is continuous in r everywhere, it need not be differentiable in r.[3]

Recall that the agent renews the patent if the total benefits from renewal exceed c_a. Proposition 1, which provides an optimal renewal rule for the agent, is an immediate consequence of Lemma 1.

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

Figure 1 explains this proposition. The 45 degree line enables us to find the point at which the vertical axis equals c_a. Comparing the values of the total benefit curve to this point it is clear that $r+\beta E[V(a+1)|r, a] \gtreqless c_a$, according as $r \gtreqless \bar{r}_a$. This, in turn, implies the simple renewal criteria provided in the proposition—renew if and only if current returns, r_a, are greater than the cutoff, $\bar{r}_a$. Note that $\bar{r}_a \leq c_a$, so that in general the difference $c_a - \bar{r}_a$ is positive. If $r_a \in (\bar{r}_a, c_a)$ 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 the option value is decreasing, implies that the cutoffs are increasing in age. This result is used in the derivation of the properties of our estimator, and, in addition, enables us to simplify the form of the function of age, ω_g, and the renewal fee schedules which determines the precise values of the cutoffs (see Section 3).

It is now straightforward, at least conceptually, to determine the proportion of patents holders who drop out, that is who stop paying the renewal fee, at each age. First note that the distribution of initial returns (which I denote by $F(r, 1; \omega_1)$, where ω_1 is a vector of parameters), the stochastic process generating subsequent returns, the renewal fee schedules, and the renewal rule, determine the (unconditional) distribution of returns at each age, say $F(r, a)$, where

$$1 - F(r, a) = \Pr\{r_a \geq r, r_{a-1} \geq \bar{r}_{a-1}, \ldots, r_2 \geq \bar{r}_2, r_1 \geq \bar{r}_1\} \quad (a = 2, \ldots, L, r \in R_+). \tag{3}$$

Here it is understood that $F(\cdot)$ depends on $\omega = [\omega_g, \omega_1]$, and on the vector of renewal fee schedules faced by the cohort, say c. Note also that the definition in (3) insures that if the patent is not renewed in any period there are no returns to patent protection thereafter (i.e. subsequent returns are less than any $r \in R_+$).

[3] This is a result of the twin facts that the value function is calculated as the maximum of two other functions, and that the conditional distributions of returns may have mass points. The value function is not differentiable everywhere for the particular special case we estimate.

Proposition 1 implies that the proportion of patent holders who pay the renewal at age a is the proportion with current returns above $\bar{r}_a$, or $1-F(\bar{r}_a, a)$. Since the proportion who drop out at age a, say $\pi(a)$, is simply the difference between the proportions not paying the renewal fee at age a and those not paying the renewal at age $a-1$,

$$\pi(a) = F(\bar{r}_a, a) - F(\bar{r}_{a-1}, a-1) \qquad (a=2,\ldots,L), \tag{4}$$

where it is understood that $\pi(\cdot)$ depends also on ω and c, and that $F(\bar{r}_1, 1)=0$ (there is no renewal fee required for the initial year of patent protection).[4]

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 some characteristics 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, let f_j and l_j be the first and last ages at which we observe the number of patentees paying the renewal for cohort j, and let $A_j=\{f_j, f_j+1, \ldots, l_j, l_j+1\}$, for $j=1,\ldots,J$. The set A_j indexes the distinct cells in which a patent of cohort j could be observed. The first cell corresponds to patents which dropped out before (or at) age f_j, the next l_j-f_j-1 cells correspond to patents which drop out at each subsequent age until (and including) l_j, and the final cell corresponds to patents which were still being renewed at l_j. The data include, for each cohort, the number of patents observed in each of these cells, or the sequence $\{n(a,j)\}_{a=f_j}^{l_j+1}$, and the vector of renewal fee schedules faced by the cohort, or 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 (or including) l_j, or still be paying the renewal fee at l_j. Equation (4) implies that the probabilities of these mutually exclusive and exhaustive alternatives are given by

$$\tilde{\pi}(a,j) = \begin{cases} \sum_{a=1}^{f_j} \pi(a, c_j) & \text{for } a=f_j, \\ \pi(a, c_j) & \text{for } f_j < a \leq l_j, \\ 1-\sum_{a=1}^{l_j} \pi(a, c_j) & \text{for } a=l_j+1, \end{cases} \qquad (j=1,\ldots,J). \tag{5}$$

With these definitions, the (log) likelihood of a particular value of the parameter vector conditional on the observed data, or $l(\omega)$, is

$$l(\omega) = \sum_{j=1}^{J} \sum_{a\in A_j} n(a,j) \log \tilde{\pi}(a,j;\omega). \tag{6}$$

[4] For the particular stochastic specification introduced in the next section this is insured by setting $c_1=0$, which will imply both that $\bar{r}_1=0$, and that the presence of $\bar{r}_1$ in equations (3) and (4) does not affect their values.

Our estimator of ω maximizes the likelihood in equation (6), and properties of this maximum likelihood estimator are provided in Proposition 2 below. First, however, an explanation is in order. The asymptotic distribution provided in this proposition follows from a theorem due to Rao (1973, Section 5.e.2). This theorem requires that the functions $\pi(a, j; \omega)$ $[a \in A_j, j = 1, \ldots, J]$ admit continuous first order partials with respect to ω at ω equal to the true value of ω, say ω^0. Since Assumptions 1 to 3 do not insure the differentiability of either the value function, or of the conditional Markov distributions, they do not insure the required condition. As a result, though it is convenient to introduce Proposition 2 here, its proof depends on the precise specification of both the distribution of initial returns and of the Markov process generating subsequent returns—neither of which are introduced until the next section.[5]

PROPOSITION 2 (proved in Appendix B): *Let n_j be the total number of patents in cohort j, let $N = \sum_{j=1}^{J} n_j$, let n_j/N converge in probability to w_j as $N \to \infty$ $(j = 1, \ldots, J)$, and let ω_N^* be the maximum likelihood estimator defined by the equation $l_n(\omega_N^*) = \sup_{\omega \varepsilon Y} l_N(\omega)$, where Y is a subset of R^k containing ω^0, the true value of ω, in its interior. Then, provided identifiability and invertibility conditions are satisfied (see Appendix B), ω_N^* converges in probability to ω^0, and*

$$\sqrt{N}(\omega_N^* - \omega^0) \xrightarrow{D} \eta(0, [i_{r,s}^0]^{-1}),$$

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

$$i_{r,s} = \sum_{j=1}^{J} w_j \sum_{a \in A_j} \frac{1}{\pi(a, j)} \frac{\partial \pi(a, j)}{\partial \omega_r} \frac{\partial \pi(a, j)}{\partial \omega_s}$$

for $r, s = 1, \ldots, k$, and $[i_{r,s}^0]$ denotes this matrix evaluated at $\omega = \omega^0$.

Two comments are in order here. First the dimension in which ω^* converges to the true value of ω(i.e., ω^0) is N, the sum of the number of patents in the J cohorts, and as Section 3 shows, N is unusually large in our samples. Second, the fact that the $\pi(\cdot)$ functions admit first order partials which are continuous at ω^0, together with the consistency of the maximum likelihood estimator, insures that $[i_{r,s}^*]$, the information matrix when evaluated at $\omega = \omega^*$, is a consistent estimator of $[i_{r,s}^0]$. As a result, $[i_{r,s}^*]^{-1}$ is used to estimate the variance-covariance matrix of 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 estimates.

[5] Indeed the proof of this proposition consists entirely of showing that the $\tilde{\pi}(\cdot)$ functions which are implicitly defined by the distribution of initial returns, the Markov process, and Proposition 1, satisfy this differentiability condition.

2. DETAILS OF THE STOCHASTIC SPECIFICATION

Equation (7), 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

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

(7) 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 (7) is that it permits an explicit solution for the sequence $\{\bar{r}_a\}_{a=1}^{L}$ as a function of the parameters of the model (see below). 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 make obsolete the returns from the agent's own patent, and the agent must decide whether current returns, and 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 improvement 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. $\phi \leq 1$ also suffices for assumption A3.4 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 $\omega_g = (\theta, \gamma, \sigma, \delta, \phi)$. To complete the specification

of the model we require also a distribution of initial returns over different patents. It is assumed that initial returns distribute lognormally, or

$$\log r_1 \sim \eta(\mu, \sigma_R). \tag{8}$$

This implies that $\omega_1 = (\mu, \sigma_R)$; so that $\omega = (\omega_g, \omega_1)$ contains seven parameters.

Equations (7) and (8) complete the specification of the model outlined in the first section. Section 3 contains a brief description of how the maximum likelihood estimate of ω, that is ω^*, was actually obtained. The reader who is not interested in the details of the estimation procedure is advised to omit this section and go directly to Section 4, which first describes the data, and then analyzes the empirical results.

3. OBTAINING ω^*

Three technical problems must be solved before we can obtain ω^*. First a method must be provided to calculate the cutoffs, or the sequence $\{\bar{r}_a\}_{a=2}^{L}$ as defined in Proposition 1, as a function of c and ω. Given ω, these cutoffs determine the drop out probabilities, or the sequence $\{\pi(a)\}_{a=2}^{L}$ as defined in equation (4), which in turn determine the likelihood of ω (see equation (6)). The second problem, then, is to provide a method which calculates the drop out probabilities corresponding to particular values of ω and $\{\bar{r}_a\}_{a=1}^{L}$. Finally, a maximization algorithm which finds that value of ω that maximizes the likelihood is required. I now consider each of these problems in turn.

It is possible to develop a recursive system of analytic equations which solves for the sequence $\{\bar{r}_a = r(a; \omega, c)\}_{a=2}^{L}$ for our problem. The system is obtained by solving for the benefit function in an interval containing $\bar{r}_a$ at each age.[6] The cutoffs corresponding to particular values of ω and c were obtained by simply substituting those values into this system of equations.

One cannot, to the best of my knowledge, obtain the drop out probabilities as analytic functions of ω 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 $\{\pi(a)\}_{a=2}^{L}$, say $\{\hat{\pi}(a)\}_{a=2}^{L}$, is found by

[6] Briefly, this problem is first reduced to a more manageable one by expressing the function determining the benefits from renewal, at each age, as the sum of $L-1$ 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 $\tau = 1, \ldots, L-a$). This fact leads to a functional recursion which can be solved using Macsyma (1983), a computer program 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 functional recursion and the boundary functions are developed in Pakes (1984, Appendix 3). The Macsyma results were obtained by Andrew Myers and myself.

taking pseudo random draws from the distribution of initial returns defined by equation (8) and ω_1, passing each through the stochastic process defined by equation (7) and ω_g, and calculating the proportion with $r_{a-1} \geq \bar{r}_{a-1}$ but $r_a < \bar{r}_a$, for $a = 2, \ldots L$ (see the definition of $\pi(a)$ in equation 4).[7] 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 $\pi(a)[1-\pi(a)]/NSIM$ $(a = 2, \ldots, L)$. Define the pseudo likelihood of ω, say $\hat{l}(\omega)$, to equal that value of the likelihood function obtained from substituting the simulated for the actual frequencies in equation (6). ω^* was obtained by maximizing $\hat{l}(\omega)$ with respect to ω. The information matrix was obtained by perturbing each parameter by one per cent from ω^*, 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 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 ω is a computer time intensive task; the CPU time for a given evaluation is 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 each run. It was developed by modifying a program 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 jth round of the subroutine was defined by an *NSIM*, say $NSIM(j)$, and a perturbation vector, say $\Delta\omega^j = [\Delta\omega_1^j, \ldots, \Delta\omega_k^j]$. Modifications were made to QNMDIF (to both the procedure for finding the gradient vector, and to the stepsize search; see footnote 8 for details) which directed it to find, with a relatively small number of function evaluations, an ω, say ω^j, such that $\hat{l}_j(\omega^j) \geq \hat{l}_j(\omega_1^j, \ldots, \omega_i^j \pm \Delta\omega_i^j, \omega_{i+1}^j, \ldots, \omega_k^j)$, for $i = 1 \ldots k$. The $j+1$ round used ω^j as a starting value, an increased *NSIM* $[NSIM(j+1) > NSIM(j)]$, and a perturbation vector with smaller components $(\Delta\omega_i^{j+1} < \Delta\omega_i^j;\ i = 1, \ldots, k)$. The final two rounds used an *NSIM* of ten and twenty thousand, respectively, and a perturbation vector equal to one per cent of the starting value of ω.[8]

That completes the description of both the model and the estimation algorithm.

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

[8] This maximization subroutine was developed by Dvora Ross and myself. There were two modifications made to QNMDIF which 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.

The next section describes the data set and then provides the empirical results.

4. THE EMPIRICAL ANALYSIS

4.1. *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.[9] Table I summarizes some of the characteristics of this data.

Row 1 of the table provides the first age for which a renewal fee is due, or f. There is no renewal requirement for ages less than f and the renewals at age f reflect events that have occurred over the first f ages.[10] 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.[11] In all countries, then, we have at least partial information on the renewal behavior of cohorts applied for in most

TABLE I

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 is the first age for which a renewal fee is due; L is the last age at which an agent can keep the patent in force by payment of an annual renewal fee; and $\overline{NPAT}$ is 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.

[9] 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.

[10] In terms of the model, we have, for $a<f$, $c_a=0$, which implies that $\bar{r}_a=0$, and that there will not be any patents with $r_a<\bar{r}_a$ in these ages. On the other hand there may be patents that draw the zero state before $a=f$. Since the zero state is absorbing, these patents will not be renewed at $a=f$ (see equation (7)).

[11] Post World War Germany allowed reapplication of patents previously applied for. By 1952 these were less than 1 per cent 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.

of the 1950's, throughout the 1960's, and in the early 1970's. The renewal fee schedules 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 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 note b to Table I). This ratio was quite large in France (.93), a bit smaller in the U.K. (.83), but only .35 in Germany (row 6). Two implications of these facts should be noted. First, when interpreting the estimates for France and the U.K. one should keep in mind that one event that would lead to a draw of the zero (or absorbing) state in these countries is a patentee who is told that his application is not granted (the model will then correctly insure that there will be no subsequent returns to patent protection for that agent). Second, the twin facts that the data include only patents granted in Germany, and that the proportion granted is small in that country, imply that the average number of patents per cohort is smaller in Germany (about 21,000) than in France or the U.K. (about 37,000). Note also 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.

Figure 2 provides the proportion dropping out at each age averaged over the cohorts for which the proportion is observed. For "a" not equal to the first or last observed cell for the cohort ($a \neq f_j$ or l_j+1), it is these proportions, disaggregated by cohort, that enter the likelihood function. That is, the estimation procedure compares the disaggregated proportions to the drop out probabilities implied by different values of the model's parameter vector. For $a=f_j$ or l_j+1, the estimation procedure tries to match the total proportion renewed, and the averages of these proportions are provided in Figure 4 below. Figure 3 provides the mean of the renewal fee schedules used in the analysis.

Figure 2 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. 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 proportions 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 3). 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 proportions in the later ages in Germany.

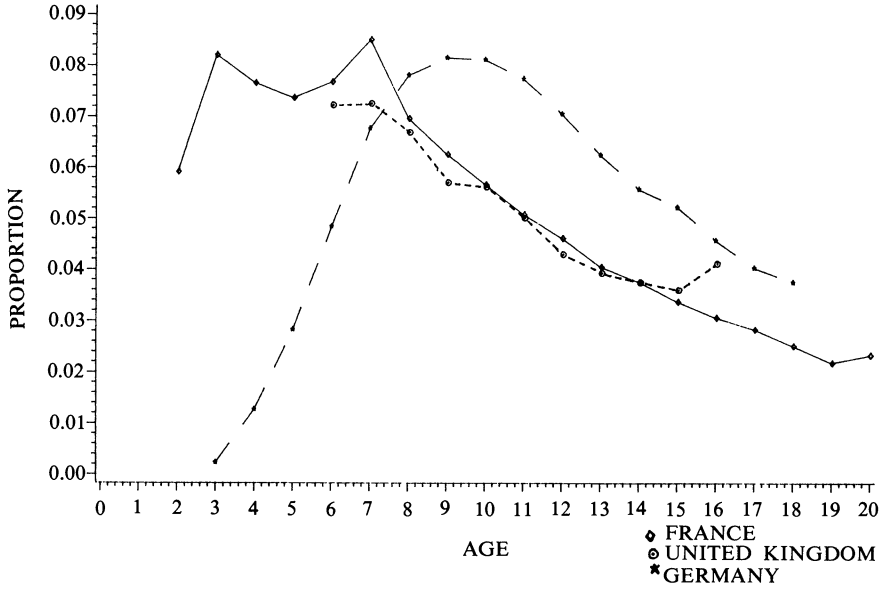


FIGURE 2.—Average drop out proportions.[a]

[a] In terms of the notation of Section 1, for $a \neq f$, the graphs are of the average of $n(a, j)/n(j)$ over all j for which $a \in A_j$, but $a \neq f_j$ and $a \neq l_{j+1}$.

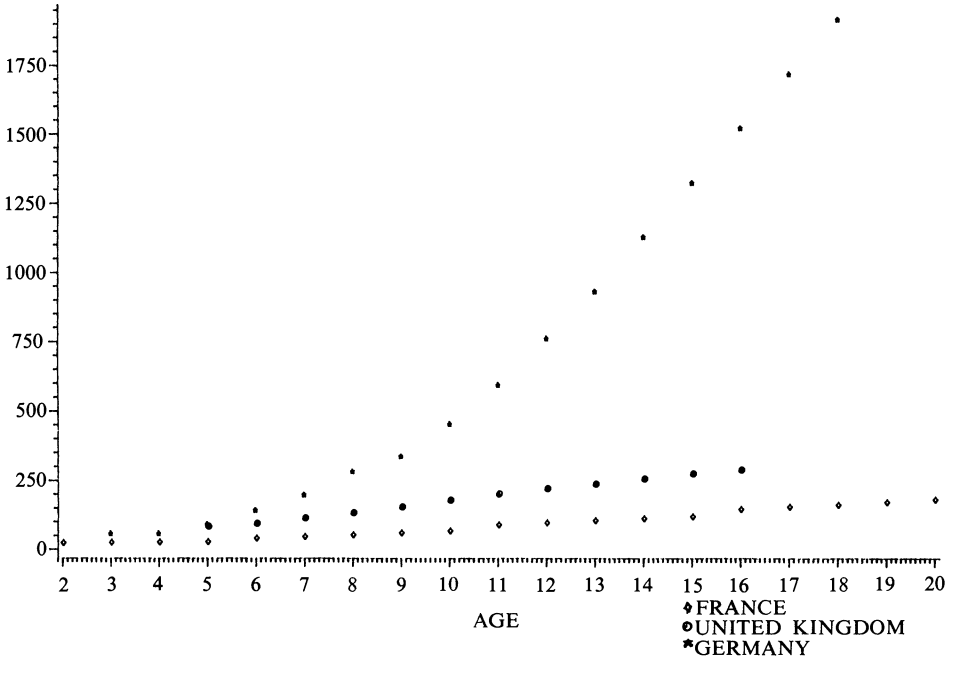


FIGURE 3.—Average of renewal fee schedules.

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 proportions do not decline at as fast a pace in the last few ages as in the ages immediately preceding them. This is what we would expect from a stochastic model of renewal behavior, 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 (which justifies the last statement in Assumption 2, Section 1). The renewal fees are quite small in all countries in the early years, and increase significantly faster in Germany thereafter.

4.2. *The Empirical Results*

Table II 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 $\beta = .9$.[12]

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 note b to Table II). 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. As a result I pay little attention to the U.K. estimates in what follows.[13]

To get an indication of the fit of the model, the difference between the estimated and actual $\tilde{\pi}$'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[$\tilde{\pi}$] in row C.1 of the table. Comparing them to the variance in the actual $\tilde{\pi}$'s (i.e., to $V[\tilde{\pi}$; data] in row C.3), it is clear that in France

[12] The decision to fix β, and a related decision to do only one run per country, served to save on computer time. The CPU time for each run increases more than proportionately to the number of parameters estimated.

[13] On the whole the estimates presented in the table for the U.K. have implications which are very similar to those that will be described for the French estimates. It should also be noted that the maximization algorithm had much more difficulty in converging for the run on the U.K. data.

and Germany only a small fraction of the variance in the actual $\tilde{\pi}$'s is not accounted for by the model (1.4 per cent in France, and .6 per cent in Germany; in the U.K. this fraction is a somewhat larger 6.4 per cent). To see whether there was any indication of cohort specific differences in the fit of the model, the differences between the estimated and actual $\tilde{\pi}$'s were also used to calculate a pseudo Durbin-Watson statistic for each country (see note d to Table II). 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 related to the characteristics of the learning process.

The parameters whose estimates exhibit large 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;

TABLE II

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[$\tilde{\pi}$]	5.42×10^{-4}	6.91×10^{-4}	1.48×10^{-4}
C.2. PDW[$\tilde{\pi}$]	1.65	2.24	1.85
C.3. V[$\tilde{\pi}$; data]	3.90×10^{-2}	1.07×10^{-2}	2.65×10^{-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 I. *NPAT* is the total number of patents covered by the data. *NSIM* is 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 3). *NCHRT* is the number of cohorts covered by the data. *NCHRTAGE* is the number of cohort-age cells covered by the data.

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

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

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

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 per cent of the French patents had initial returns under fifty dollars, while under one per cent 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 include only patents granted while the French data include all patents applied for; and that the granting criteria seem to be particularly stringent in Germany (Table I). 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) imply that the learning process is concentrated in the early ages. Table III illustrates this point. The descriptive 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 II. 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 per cent of the French patent holders discovered a use which enabled

TABLE III

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
$\pi(2)$	.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
$\pi(3)$	.0907	.0013
$E_{(r_3)}[r_3 \mid r_3 > 0]$	1432.24	3224.56
Pr (Downside); Pr (Upside)	.0118; .0012	.0005; .0039
$\pi(4)$	.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
$\pi(5)$	.0381	.0277
$E_{(r_5)}[r_5 \mid r_5 > 0]$	1192.70	2641.40
$\overline{NPAT}$	36,865	21,273

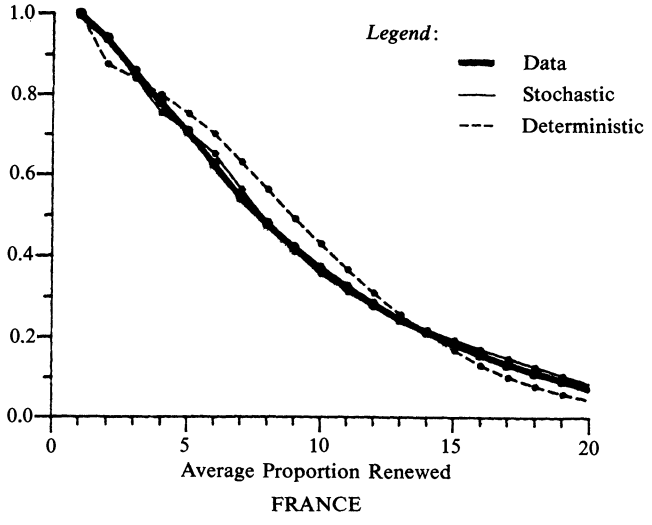
[a] The estimates in this and the following tables were obtained from a simulation run of 20.000 draws using the estimates of ω given in Table II and the mean of the renewal fee schedules. Pr (Downside) is the average probability of discovering that the patented ideas will never by profitably exploited (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. $E_r[r \mid r > 0]$ is the mean of r for patents still in force. $\pi(a)$ is the proportion of patents which drop out at the ath renewal.

them to increase subsequent returns, while over six per cent discovered that their patented ideas could never be profitably exploited. These six per cent 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 large learning probabilities in 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 occurred than occurred during the first year. An additional nine per cent of the patent holders stopped paying the renewal fee at the third age. Of these, about five per cent 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 3).

I now return briefly to the issue of the fit of the model. Figure 4 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 deterministic model (the broken lines). The deterministic 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 (7) to read: $r_{a+1} = \delta r_a$ with probability one.[14] In this figure it is hard to distinguish the curve estimated by the stochastic model from the data. On the other hand the deterministic model 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 deterministic model 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

[14] As one would expect from the large size of our samples (*NPAT*) the likelihood ratio test statistic for the null hypothesis that the model was deterministic was inordinately large (over 20,000 for Germany and over 60,000 for France; see also note 15). More informative is the fact that the MSE($\tilde{\pi}$) statistics for the deterministic model were about twice their values for the stochastic model.

magnified in Figure 5 which provides the proportion dropping out, by age, averaged over the cohorts for which this proportion is observed. The stochastic model 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



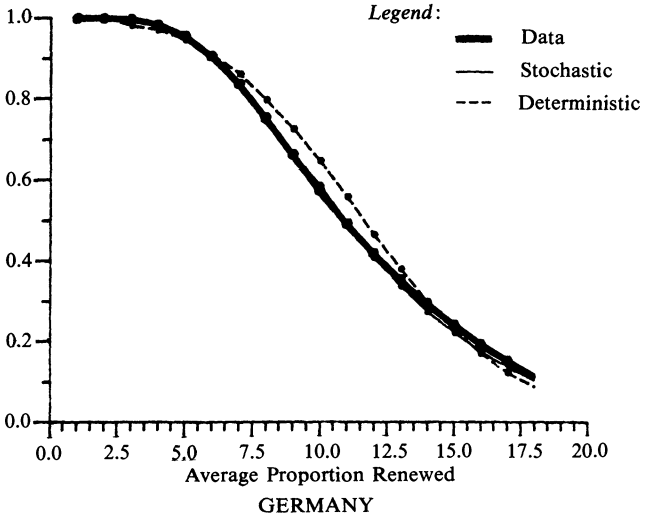


FIGURE 4—Comparisons of average proportion renewed.

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 5 we can actually see the differences between the estimates from the stochastic model and the data. These differences are concentrated in the middle ages. The age-specific average drop out proportions in the French data have two local maxima (at ages three and seven). The estimates from the model for France

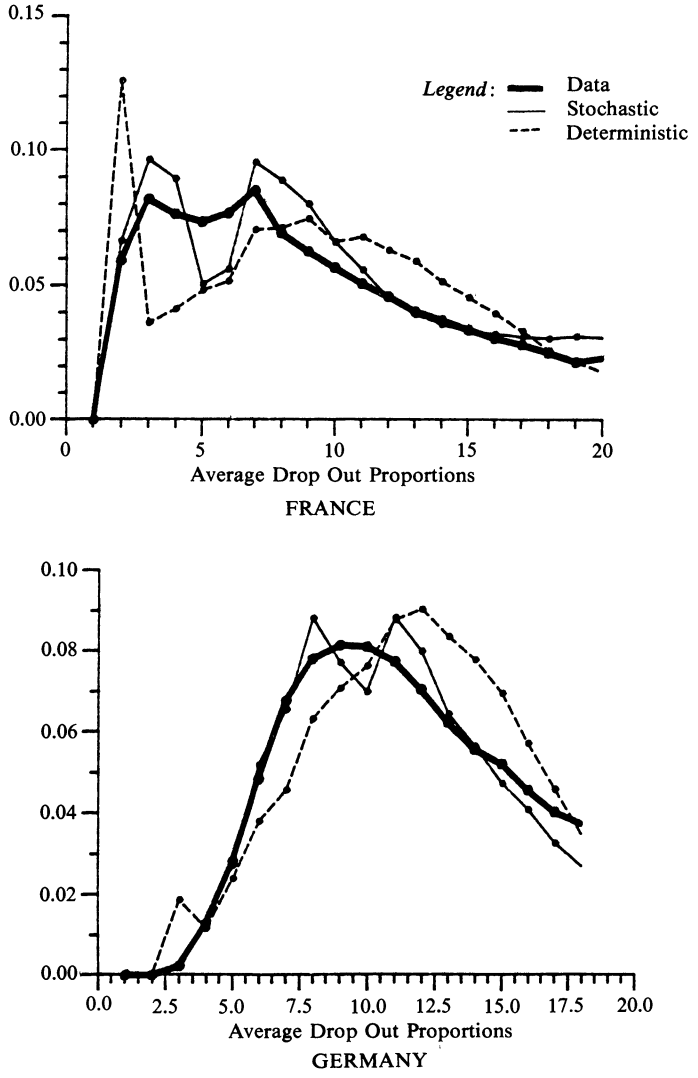


FIGURE 5—Comparisons of average drop out proportions.

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 distributon of average drop out proportions 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 estimates of the average drop out proportions in the later ages are a bit too high in France, and a bit too low in Germany. In sum, though Figures 4 and 5 indicate why the mean square error of the differences between the observed and estimated proportions are small relative to $V(\tilde{\pi}; \text{data})$, they also indicate that the model is not perfectly specified, and this should be kept in mind when considering the implications of the parameter estimates.[15]

Table IV provides a summary of the distribution of returns at ages one, three, and five respectively. Two implications of this table are of interest. First there is a distinct pattern to the evolution of these distribution functions over age. Between ages one and three the upper tail of the distribution becomes thicker and is pushed to the right. That is, a substantial fraction of the patentees who had the "upside draws" in Table III uncovered uses for their patented ideas which increased the returns earned from holding their patents by large amounts. A comparison of the quantiles for age five to those of age three reveals the onset of the obsolescence process; that is, the quantiles from the age five distribution are always below the same quantiles from the distribution at age three. The second point to note is that there is a skew in the initial distribution of returns,

TABLE IV[a]

THE DISTRIBUTION OF RETURNS IN THE EARLY AGES [$F(r, a)$]

Country	France			Germany		
r/a	1	3	5	1	3	5
0	0	.155	.270	0	.001	.04
50	.31	.315	.375	.01	.01	.04
150	.580	.525	.585	.07	.065	.095
500	.830	.710	.745	.34	.27	.325
2,500	.975	.86	.895	.83	.655	.705
5,000	.990	.925	.950	.940	.800	.845
15,000	.995	.990	.990	.990	.95	.97

[a] See the note to Table III.

[15] Given the values of *NSIM* and *NPAT* for our problem (see Table I) the binomial sampling error in both the empirical and estimated frequencies have variances very close to zero. As a result even our, relatively small, sample values of MSE($\tilde{\pi}$) are too large for sampling variance to be the only source of error in the model. Though this problem, which is called the problem of extra-binomial sampling variance by Williams (1982) (see also the review in Haseman and Kupper (1979) and the discussion in Heckman and Singer (1984)), occurs frequently in models designed to analyze proportions when the underlying sample size is large, I do not know of any consistent way of accounting for it when the model has a sequential dimension.

and that this skew increases substantially over the first few ages. This fact leads to a highly skewed distribution of realized patent values.

Table V 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 per cent of the patents in the French data had realized values of seventy-five dollars or less.[16] These patents contributed about a half of one per cent to the total value of the patents in a cohort, while the patents in the lower half of the distribution contributed less than two per cent of the total value of a cohort. The median of the distribution of realized values ($534) was less than one tenth its mean ($5,631); and the five per cent 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 refer to grants rather than applications, most pronounced at the lowest percentiles. In Germany these percentiles were nonnegligible, albeit, quite small. Still only about 7 per cent of the patents in Germany had realized values in excess of $50,000; in France only

TABLE V

PERCENTILES (p1) AND LORENZ CURVE COEFFICIENTS (1c) FROM THE DISTRIBUTION OF REALIZED PATENT VALUES[a]

	Country					
	France		U.K.		Germany	
Per cent p	p1 ($)	1c per cent	p1 ($)	1c per cent	p1 ($)	1c per cent
.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	—
$\overline{NPAT}$	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^* is the last age at which patent i was kept in force. See also the note to Table III.

[16] Of course some of these patents had negative (though small in absolute value) realized values, as they were patents on which early renewals were paid 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.

two and a half per cent 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 I). 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). On average, then, the total value of a cohort of patents in Germany was larger than the value of a French cohort.

4.3. *The Value of Patent Protection and the Characteristics of the Patenting Process*

A word of caution is in order before proceeding. Though it may well be the case that the patent renewal data are the most extensive and detailed information source on the value of patent protection available, they, in themselves, contain only a limited amount of information: the age path of the proportion of patents in different groups paying a renewal fee and the renewal fee schedules. Mixing this information with additional assumptions has lead to a set of quite detailed conclusions, but it should be clear that these may depend on the additional assumptions chosen (both behavioral and stochastic). The only exogenous check of these conclusions I have considered is a broad check of the implications of the parameter estimates against known intercountry differences in the data. In this section I consider more general implications of the parameter estimates. Though here it will be possible to provide rough checks for the consistency of some of the conclusions we derive with alternative sources of information, it should be kept in mind that there may be many models that do as well as ours in all these respects (as well as in fit), but differ substantially in others.

To get an indication of the annual returns earned from holding the patent stock in a country, we must account for the fact that the patent stock held 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. Assuming that each of the previous L cohorts began with the average number of patents per cohort and faced the mean of the renewal fee schedules, and using the parameter estimates of Table II, we find that the net annual flow of returns from holding the patent stocks in France, the U.K., and Germany were .315, .385, and .512 billion dollars, respectively. To consider whether these figures imply large gains from patenting we would like to compare them to either the total returns that accrued to the patented ideas, or to the expenditures that went into developing them. Neither of these two numbers are available, but the OECD (1975; Tables III and IV) does provide estimates of the R&D expenditures funded by the business enter-

prises 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 per cent, 11.03 per cent, and 13.83 per cent 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 per cent 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.[17] Still, the ratio does suggest that the proprietary rights resulting from the patent laws create annual returns which are nonnegligible 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 V), 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 (1958, p. 1098), for example, notes that the data provided in Sanders, Rossman, and Harris suggest a Pareto-Levy distribution with an infinite mean for the distribution of profits from patented ideas; while Garbrowski and Vernon 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 (1984)). Pakes (1985) 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 during the early years of a patent's life.

[17] Business enterprises in these countries also own patents in force elsewhere, and foreign business enterprises own patents in force in these countries. Moreover, not all the business sector's R&D expenditures are directed towards patentable innovations, and not all patentees are business enterprises.

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 preceded by any increase in patenting. Growth through discovery will occur in spurts, and these spurts will be probabilistically related to the investments which preceded 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.[18]

One final point: Disaggregated patent renewal data, data which enable an investigation of the returns to patent protection by technical field of the patent and by nationality and type of patentor (e.g. individuals, small business enterprises, large corporate entities), is gathered by INPADOC (International Patent Documentation Center, Vienna, Austria). These data should prove extremely valuable. Issues related to which sectors of a particular economy, and which economies, derive disproportionate benefits from the patent laws lie at the heart of most discussions of the cost and benefits of alternative patent systems (see Scherer (1965, Chapter 16), and the literature cited there). Moreover, intersectoral differences in the patenting and R&D processes are central to the literature on market structures, industrial policy, and technical progress.

National Bureau of Economic Research, Cambridge, MA 02138, U.S.A

Manuscript received May, 1984, revision received November, 1985.

APPENDIX A

This appendix proves the following lemma.

LEMMA 1: *The value of the option, that is,* $E[V(a+1)|r, a]$, *is*: (*i*) *continuous and nondecreasing in* r, *and* (*ii*) *nonincreasing in* a ($r \in R_+$, $a = 1, \ldots, L$).

[18] A step in this direction has been made by Ericson and Pakes (1983).

PROOF: The proofs of both (i) and (ii) are obtained by backward induction on "a", and use the following lemma (for a proof, see Ross (1983, p. 154)). Let $f(z)$ be nondecreasing in z and $\Pr\{z \le k \mid z_1\} \le \Pr\{z \le k \mid z_1'\}$ for all $k \in R_+$. Then, provided $E[f(z) \mid z_1] < \infty$, $E[f(z) \mid z_1] \ge E[f(z) \mid z_1']$ (for nonincreasing $f(\cdot)$, $E[f(z) \mid z_1] \le E[f(z) \mid z_1']$).

Part (*i*). Since $E[V(L+1) \mid r] = 0$ for all r, the initial condition of the inductive argument is satisfied trivially, and it suffices to show that if the proposition is true for $a = \tau+1$, it is also true for $a = \tau$. Recall that $V(\tau+1, z) = \max\{0, z - c_{\tau+1} + \beta E[V(\tau+2) \mid z, \tau+1]\}$, and note that since the hypothesis of the inductive argument implies that $E[V(\tau+2) \mid z, \tau+1]$ is continuous and nondecreasing in z, $V(\tau+1, z)$ is also.

To establish continuity take any $r \in R_+$. $E[V(\tau+1) \mid r, \tau]$ will be continuous at r if for every sequence $\{r_n\}$ such that $\lim r_n = r$ (or $r_n \to r$), $E[V(\tau+1) \mid r_n, \tau] \to E[V(\tau+1) \mid r, \tau]$ (Royden (1968, p. 48). For any such sequence, let $V_n(\tau+1)$ be the random variable $V(\tau+1, z)$ with distribution $G(z \mid r_n, \tau)$ ($V(\tau+1)$ has distribution $G(z \mid r, \tau)$). Since A3.2 implies that $G(z \mid r_n, \tau)$ converges in distribution to $G(z \mid r, \tau)$, and $V(\tau+1, z)$ is continuous in z, the distribution of $V_n(\tau+1)$ converges to that of $V(\tau+1)$ (Billingsley (1979, Theorem 25.7)). This fact will insure that $E[V(\tau+1) \mid r_n, \tau] \to E[V(\tau+1) \mid r, \tau]$, if there exists an $\varepsilon > 0$ for which $E[V(\tau+1)]^{1+\varepsilon} \mid r_n, \tau] < \infty$ for all n (Billingsley (1979, Theorem 25.12 and its corollary)). Note also that since r was arbitrary, if we show that $E[V(\tau+1)]^{1+\varepsilon} \mid r_n, \tau] < \infty$, then $E[V(a+1) \mid r, a]$ is continuous in r for all $r \in R_+ (a = 1, \ldots, L)$.

Now

$$E[V_n(\tau+1)^{1+\varepsilon}] \le E\left[\left(\sum_{j=1}^{L-\tau} r_{\tau+j}\right)^{1+\varepsilon} \,\middle|\, r_\tau = r_n\right] \le 2^\varepsilon \sum_{j=1}^{L-\tau} E[r_{\tau+j}^{1+\varepsilon} \mid r_\tau = r_n]$$

(Rao, 1973, p. 149). Since Assumption A3.1 insures that there exists an $\varepsilon > 0$ such that $E[r_{1+j}^{1+\varepsilon} \mid r = r_n] < \infty$, it will suffice to show that $E[r_{\tau+j}^{1+\varepsilon} \mid r_\tau = r_n] \le E[r_{1+j}^{1+\varepsilon} \mid r_1 = r_n]$, for all n and $j = 1, \ldots, L-1$.

For this we require only that

$$G_{\tau,j}(k \mid r_n) \equiv \Pr\{k \ge r_{\tau+j} \mid r_\tau = r_n\} \ge \Pr\{k \ge r_{1+j} \mid r_1 = r_n\} \equiv G_{1,j}(k \mid r_n)$$

for all $k, j = 1, \ldots, L-\tau$, and $r_n \in R_+$. A second inductive argument proves this point. Since Assumption A3.4 insures the inequality for $j = 1$, it will suffice to show that if the inequality is true for $j = j'$, it is also true for $j = j'+1$. Now

$$\begin{aligned}
&\Pr\{k \ge r_{\tau+j'+1} \mid r_\tau = r_n\} \\
&\quad = \int G(k \mid z, \tau+j') G_{\tau,j'}(dz \mid r_n) \\
&\quad \ge \int G(k \mid z, 1+j') G_{\tau,j'}(dz \mid r_n) \\
&\quad \ge \int G(k \mid z, 1+j') G_{1,j'}(dz \mid r_n) = \Pr\{k \ge r_{2+j'} \mid r_1 = r_n\}
\end{aligned}$$

where the first inequality follows from A3.4, and the second from the hypothesis of the inductive argument and the Lemma since $G(k \mid z, 1+j')$ is nonincreasing in z (from A3.3) and bounded by 1.

To establish that $E[V(\tau+1) \mid r, \tau]$ is nondecreasing in r apply the Lemma directly and note that: $V(\tau+1, z)$ is nondecreasing in z (from the hypothesis of the inductive argument); $G(z \mid r, \tau) \le G(z \mid r', \tau)$ whenever $r \ge r'$ (from A3.3); and $V(\tau+1, z)$ is integrable with respect to $G(z \mid r, \tau)$ (from the argument given above). *Q.E.D.*

Part (*ii*). For the first step of the inductive argument I assume that $E[V(a+2) \mid r, a+1] \le E[V(a+1) \mid r, a]$ and show that this implies that $E[V(a+1) \mid r, a] \le E[V(a) \mid r, a-1]$ for $r \in R_+$. Recall that $V(a+1, z) = \max\{0, z - c_{a+1} + \beta E[V(a+2) \mid z, a+1]\} \le \max\{0, z - c_a + \beta E[V(a+1) \mid z, a]\} = V(a, z)$, where the inequality follows from the hypothesis of the inductive argument and the fact that $c_{a+1} \ge c_a$ (see A2). Therefore, for any $r \in R_+$,

$$\begin{aligned}
E[V(a+1) \mid r, a] &= \int V(a+1, z) G(dz \mid r, a) \le \int V(a, z) G(dz \mid r, a) \\
&\le \int V(a, z) G(dz \mid r, a-1) = E[V(a) \mid r, a-1],
\end{aligned}$$

where the last inequality follows from the Lemma, since $V(a, z)$ is nondecreasing in z and integrable with respect to $G(z \mid r, a-1)$ (see above), and $G(z \mid r, a) \ge G(z \mid r, a-1)$ from Assumption A3.4. To establish the initial condition for the inductive argument it suffices to note that $E[V(L+1) \mid r, L] = 0 \le \int_{c_L} (z - c_L) G(dz \mid r, L-1) = E[V(L) \mid r, L-1]$. *Q.E.D.*

APPENDIX B. PROPOSITION 2

Proposition 2, which provides properties of the maximum likelihood estimator, follows directly from a theorem due to Rao (1973, Section 5.e.2) provided the following regulatory conditions are satisfied:

(i) The functions $\pi(a, j; \omega)(a=2, \ldots, L; j=1, \ldots, J)$ admit first order partials which are continuous at $\omega=\omega^0$.

(ii) For every $\omega \in Y$ such that $\omega \neq \omega^0$, $\pi(a, j; \omega) \neq \pi(a, j; \omega^0)$ for at least one couple (a, j) $(a=2, \ldots, L; j=1, \ldots, J)$.

(iii) The information matrix, $[i_{r,s}]$, is nonsingular at $\omega=\omega^0$.

Since neither the conditional Markov distributions nor the function determining the benefits from renewing are differentiable everywhere, it is not immediately obvious that condition (i) is satisfied in our problem, and a formal proof of this condition is given below.[19] Given this proof, I simply assume (ii) and (iii). They will be satisfied provided there is sufficient variation in the cost schedules and ages covered by the data.

PROOF OF CONDITION (i): Omitting the index j for simplicity we have, from equation (5) in the text,

$$\pi(a; \omega) = F(\bar{r}_a(\omega), a; \omega) - F(\bar{r}_{a-1}(\omega), a-1; \omega),$$

where $F(\cdot, a)$ and $\bar{r}_a(\cdot)$ provide the distribution of returns and the cutoff at age a respectively $(a=2, \ldots, L)$. Lemma B.1 below shows that $\bar{r}_a(\omega)$ is continuously differentiable in ω at $\omega=\omega^0(a=2, \ldots, L)$. It therefore suffices to prove that $F(r, a)$ is both continuously differentiable with respect to ω at $\omega=\omega^0$ in an interval containing $r=\bar{r}_a(\omega^0)$, and has a density which is continuous at $r=\bar{r}_a(\omega^0)(a=1, \ldots, L$; recall that $F(\bar{r}_1, 1) \equiv 0)$. Lemma B.2 proves a condition which suffices for this point.

Q.E.D.

LEMMA B.1 *Each element of the sequence of functions* $\{\bar{r}_a(\omega)\}_{a=1}^{L}$ *admits first partials which are continuous at* $\omega=\omega^0$.

PROOF: The proof is by 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 $\bar{r}_a(\omega)$ admits continuous first partials with respect to ω at $\omega=\omega^0$ provided the $\bar{r}_{a+\tau}(\omega)[\tau=1, 2, \ldots, L-a]$ do. Proposition 1 and equation (7) imply that $\bar{r}_a(\cdot)$ is defined by the implicit function

$$\mu(\bar{r}_a, \omega) = \bar{r}_a + \beta[1-\exp(-\theta\bar{r}_a)]\int_{\bar{r}_{a+1}}^{\infty} V(a+1, z; \omega)Q(dz, a; \omega) - c_a = 0.$$

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 ω at $\omega=\omega^0$. The hypothesis of the inductive argument implies that $\bar{r}_{a+1}(\omega)$ has continuous first partials; and $Q(z, a; \omega)$ is an exponential distribution which has a density which possesses continuous first partials with respect to ω everywhere for $z \in R_+$. It will, therefore, suffice to show that $V(a+1, z; \omega)$ has continuous first partials with respect to ω at $\omega=\omega^0$ for every $z \in (\bar{r}_{a+1}, \infty)$, except possibly a set of z of Lebesgue measure zero, provided that $\bar{r}_{a+\tau}(\omega)$ [for $\tau=1, \ldots L-a$] have continuous first partials at $\omega=\omega^0$. A second inductive argument suffices to prove this point.

Since $V(L, z) = \max\{0, z-c_L\}$ the initial condition for the inductive argument is satisfied trivially and it will suffice to show that $V(a+1, z)$ has the required property provided that $V(a+2, z)$ does. Equations (7) and (2) imply that

[19] The following two points help to explain why condition (i) can be satisfied despite the nondifferentiability of $r+\beta E[V(a+1)|r, a]$ and $G(z|r, a)$. First, the direct dependence of $\pi(a)$ on the benefit function is through the fact that the cutoff, $\bar{r}_a$, is defined as the unique solution to $r+\beta E[V(a+1)|r, a]=c_a$, and for our problem it is possible to show that $r+\beta E[V(a+1)|r, a]$ is differentiable in r in a region about $\bar{r}_a(a=2, \ldots, L)$. Second, though conditional on any value of r, $G(z|r, a)$ has points at which it is not differentiable in z; there does not exist a value of $z \in R_+$ which is a discontinuity point for a set of r of positive Lebesgue measure; as a result, the unconditional distributions of returns, that is $F(\cdot, a)$, is differentiable (see below).

$$V(a+1, z)=\begin{cases} v^1(a+1, z)=z-c_{a+1}+\beta[1-\exp(-\theta z)]\displaystyle\int_{\bar r_{a+2}}^{\infty} V(a+2, s)Q(ds, a+1), \\ \qquad\qquad \text{if } z\in[\bar r_{a+1}, \delta^{-1}\bar r_{a+2}]; \\ v^2(a+1, z)=z-c_{a+1}+\beta[1-\exp(-\theta z)]\{Q(\delta z, a+1)V(a+2, z) \\ \qquad\qquad +\displaystyle\int_{\delta z}^{\infty} V(a+2, s)Q(ds, a+1)\}, \quad \text{if } z\in(\delta^{-1}\bar r_{a+2}, \infty). \end{cases}$$

The argument of the last paragraph together with the hypothesis of the inductive argument implies that $v^1(a+1, z)$ has continuous first partials with respect to ω at $\omega=\omega^0$ for every $z\in[\bar r_{a+1}(\omega^0), \infty)$. For $z\in(\delta^{-1}\bar r_{a+2}, \infty)$ the values of z at which $v^2(a+1, z)$ has discontinuous first partials are the values of z at which $V(a+2, \delta z)$ has the same property. Now, for $\tau=1$ or 2, let $S(a+\tau)$ be the set of $z\in(\bar r_{a+\tau}(\omega^0), \infty)$ at which $V(a+\tau, z)$ has discontinuous first partials with respect to ω at $\omega=\omega^0$. Then

$$m[S(a+1)]\leq m[S(a+2)]+m[\bar r_{a+2}\delta^{-1}]=m[S(a+2)]=0,$$

where $m[\cdot]$ provides the Lebesgue measure of alternative sets, and the last equality follows from the hypothesis of the inductive argument. *Q.E.D.*

LEMMA B.2: *$F(r, a)$ has a density which is both continuous in r and admits continuous first partials with respect to ω at $\omega=\omega^0$ for every $r\in[\bar r_a(\omega^0)-\varepsilon, \infty)$ and some $\varepsilon>0$ ($a=2, \ldots, L$).*

PROOF: The proof is by forward induction on "a". First assume $F(\cdot, a-1)$ has a density with the required properties and denote that density by $f(\cdot, a-1)$. Choosing $0<\varepsilon<\bar r_a(\omega^0)-\delta^0\bar r_{a-1}(\omega^0)$ (that such an ε exists follows from the facts that $\bar r_a\geq\bar r_{a-1}$ and $\delta^0<1$), Proposition 1 and equation (7) imply that for any $r\in[\bar r_a(\omega^0)-\varepsilon, \infty)$

$$\text{(L2.1)}\qquad \Pr\{r>r_a\geq\bar r_a-\varepsilon\}=\int_{\bar r_{a-1}}^{\infty}\Pr\{r>r_a\geq\bar r_a-\varepsilon\,|\,z\}f(z, a-1)\,dz$$

where

$$\text{(L2.2)}\qquad \Pr\{r>r_a\geq\bar r_a-\varepsilon\,|\,z\}=\begin{cases}[1-\exp(-\theta z)]Q(r, a-1) & \text{if } \delta^{-1}r\geq z>\bar r_{a-1}, \\ 0 & \text{if } z>\delta^{-1}r;\end{cases}$$

and $Q(\cdot, a-1)$ denotes an exponential distribution. Substituting we have

$$\text{(L2.3)}\qquad \Pr\{r>r_a\geq\bar r_a-\varepsilon\}=Q(r, a-1)\int_{\bar r_{a-1}}^{\delta^{-1}r}[1-\exp(-\theta z)]f(z, a-1)\,dz.$$

For any r in the required interval the density, $f(r, a)$, can be derived by direct differentiation of (L2.3). The fact that it is continuous in r and possesses continuous first partials with respect to ω at $\omega=\omega^0$ follows from the same properties of: the exponential distribution and its density, of $f(z, a-1)$ for all $z\in[\bar r_{a-1}, \infty)$ (which follows from the hypothesis of the inductive argument), and from the continuity of the first partials of $\bar r_{a-1}$ with respect to ω at $\omega=\omega^0$ (Lemma B.1).

To complete the inductive argument we need only show that $F(\cdot, 2)$ has a density with the required properties. This can be shown by substituting $a=2$ in the argument given above and noting that $F(r, 1)$ is the lognormal distribution which has both a continuous density and continuous first partials with respect to ω for all $r\in R_+$. *Q.E.D.*

REFERENCES

BILLINGSLEY, PATRICK (1979): *Probability and Measure.* New York: John Wiley and Sons.

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): *Practical Optimization.* New York: Academic Press.
GRILICHES, ZVI (1984): *Patents, R&D and Productivity.* Chicago: The University of Chicago Press.
HASEMAN, J. K., AND L. L. KUPPER (1979): "Analysis of Dichotomous Response Data from Certain Toxological Experiments," *Biometrics*, 35, 281-293.
HECKMAN, J., AND B. SINGER (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52, 271-320.
LERMAN, STEVEN R., AND CHARLES F. MANSKI (1981): "On the Use of Simulated Frequencies to Approximate Choice Probabilities," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. F. Manski and D. McFadden. Cambridge: M.I.T. Press, 305-315.
MACSYMA (1983): *Macsyma Reference Manual*, Version 10 the Mathlab Group, Labaratory for Computer Science, M.I.T.
MILLER, ROBERT A. (1984): "Job Matching and Occupational Choice," *Journal of Political Economy*, 92, 1086-1120.
NATIONAL PHYSICS LABORATORY (1983): QNMDIF, Numerical Analysis Group, Downer's Grove, Illinois.
NELSON, RICHARD R., AND SIDNEY WINTER (1982): *An Evolutionary Theory of Economic Change.* Cambridge: Harvard University Press.
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 (1985): "Patents, R&D, and the Stock Market Rate of Return," *Journal of Political Economy*, 93, 390-409.
——— (1984): "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," National Bureau of Economic Research Working Paper No. 1340.
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 in Zvi Griliches (ed.) op. cit.
RAO, C. RADHAKRISHNA (1973): *Linear Statistical Inference and its Applications.* New York: John Wiley and Sons.
ROBERTS, KEVIN, AND MARTIN L. WEITZMAN (1981): "Funding Criteria for Research, Development and Exploration Projects," *Econometrica*, 49, 1261-1288.
ROSS, SHELDON M. (1983): *Introduction to Stochastic Dynamic Programming.* New York and London: Academic Press.
ROYDEN, H. L. (1968): *Real Analysis.* New York and London: MacMillan Publishing Company.
SANDERS, J., J. ROSSMAN, AND L. HARRIS (1958): "The Economic Impact of Patents," in the *Patent, Trademark and Copyright Journal of Research and Education*, II (September), pp. 340-363.
SCHERER, FREDERIC M. (1958): "Firm Size, Market Structure, Opportunity, and the Output of Patented Innovations," *American Economic Review*, 55, 1097-1125.
——— (1965): *Industrial Market Structure and Economic Performance*, 2nd edition. Chicago: Rand McNally College Publishing Company.
TELSER, LESTER G. (1982): "A Theory of Innovation and Its Effects," *Bell Journal of Economics*, 13, 69-92.
WILLIAMS, D. A. (1982): "Extra-binomial Variation in Logistic Linear Models," *Applied Statistics*, 31, 144-148.
WOLPIN, KENNETH I. (1984): "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality," *Journal of Political Economy*, 92, 852-874.