Pakes (1986): "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks"[*]

Spring 2009

Main question: How much are patents worth? Answering this question is important, because it helps inform the debate about optimal patent length and design. For example, are patents good tools for rewarding innovation?

- Q_a: value of a patent at age a

- Goal of paper is to estimate Q_a using data on their renewal. Q_a is inferred from patent renewal process via a structural model of optimal patent renewal behavior.

1 Behavioral Model

- Treat patent renewal system as exogenous – only looking at the European system

- For $a = 1, \ldots, L$, a patent can be renewed by paying the fee c_a

- Timing

  - At age $a = 1$ patent holder obtains period revenue r_1 from patent

  - Decides whether or not to renew. If renew then pay c_1 and proceed to age $a = 2$

- If don't renew, lose patent and get 0.

  - At age $a = 2$ patent holder obtain period revenue r_2 from patent

  - Decides whether or not to renew. If renew then pay c_2 and proceed to age $a = 3$

  - And so on ….

- Let V_a denote the value of a patent at age a

$$V_a \equiv \max_{t \in [a,L]} \sum_{a'=1}^{L-a} \beta^{a'} R(a + a')$$ (1)

where

$$R(a) = \begin{cases} r_a - c_a & \text{if } t \geq a \text{ (when you hold onto the patent)} \\ 0 & \text{if } t < a \text{ (after you allow the patent to expire)} \end{cases}$$ (2)

and t is the age at which the agent allows the patent to expire. Hence $R(a)$ denote the profits from a patent during the a-th year.

- The sequence $R(1), R(2), \ldots$ is a controlled stochastic process – it is inherently random, but also affected by the agents actions (i.e. renewing the patent).

- This type of problem is called an "optimal stopping" problem. Unlike Rust's bus engine paper, this is not a regenerative optimal stopping problem.

- Since the maximal age is finite, L, this is a finite-horizon problem. Most dynamic problems are either (a) infinite-horizon, stationary problems or (b) finite-horizon, non-stationary problems

- Stationarity means that the value functions and decision rules are time-invariant functions of the state variables. Only get dependence on time through the values of the state variables (e.g. mileage in Rust's bus engine paper).

- State variable in this paper: r_a – the single period revenue.

- Finite-horizon problems are solved via *backward recursion*. Start with the last period of the problem and work backwards. Value function is

$$V_a(r_a) = \max\left\{0, r_a + \beta E[V_{a+1}(r_{a+1})|\Omega_a] - c_a\right\} \qquad (3)$$

  where the value of a patent $Q_a = r_a + \beta E[V_{a+1}(r_{a+1})|\Omega_a]$ and you choose to renew if $Q_a > c_a$.

- Ω_a is the history of revenue up to age a, or $\{r_1, r_2, \ldots, r_a\}$.

- Expectation is over $r_{a+1}|\Omega_a$. The sequence of conditional distributions, $G_a \equiv F(r_{a+1}|\Omega_a), a = 1, 2, \ldots, L$ is an important component of the model specification.

- Pakes assumes

$$r_{a+1} = \begin{cases} 0 & \text{with prob. } \exp(-\theta r_a) \\ \max\{\delta r_a, z\} & \text{with prob. } 1 - \exp(-\theta r_a) \end{cases} \qquad (4)$$

  where density of z is $q_a = \frac{1}{\sigma_a}\exp(-(\gamma + z))/\sigma_a$ and $\sigma_a = \phi^{a-1}, a = 1, 2, \ldots, L-1$ and $\{\delta, \theta, \gamma, \phi, \sigma\}$ are the important structural parameters of the model.

- Pakes explains his choice behind the stochastic evolution of r_a

  1. Firm learns about the patent over time (continuing to spend money on development)

  2. May learn it is worthless – get 0

  3. May not learn anything – so expectation is δr_a where $\delta < 1$. Revenue is less because others are innovating

  4. May learn it is more valuable – z.

- Agent's maximization problem: is the value of the patent $Q_a = r_a+$ "option value" greater than the cost of renewing a patent c_a.

- Get threshold values of r_a, denoted $\bar{r}_a$, above which an agent renews (see figure 1).

- Cutoff points are due to assumptions ensuring that Q_a is increasing in r_a – so that Q_a and c_a only cross once.

- Specification also ensures that $\bar{r}_a < \bar{r}_{a+1} < \bar{r}_{a+2} < \ldots < \bar{r}_{L-1}$

2 Implementation

The paper uses aggregate data

- For cohort j (year in which patents are granted), observe the sequence $n(a,j)$, $a = f_j, f_j+1, \ldots, l_j-1, l_j$: # of cohort j patents which are *not* renewed at age a.

- f_j is the first date at which a renewal fee is observed for cohort j (e.g. UK – first renewal fee is required 6 years after patent is filed).

- l_j is the last date at which a renewal fee is observed for cohort j.

- Have left and right censoring. Don't know what happened in years before f_j – only see $\sum_{a=1}^{f_j} n(a,j)$. And don't know what happened in years after l_f.

- Note – model is fully parametric (although incorporates a flexible specification)

- Likelihood of the aggregate data is derived using $\text{Prob}(t_{ij} = a)$, the prob. that an individual patent i from cohort j is renewed up to age a.

$$\text{Prob}(t_{ij} = a) = \text{Prob}(r_a < \bar{r}_a, r_{a-1} > \bar{r}_{a-1}, \ldots, r_1 > \bar{r}_1) \tag{5}$$

$$= \int_{-\infty}^{\bar{r}_a} \int_{\bar{r}_{a-1}}^{\infty} \ldots \int_{\bar{r}_1}^{\infty} f(r_a, \ldots, r_1) dr_1 \ldots dr_a \tag{6}$$

$$\equiv \pi(a; \mathbf{c}_j) \tag{7}$$

where f is the joint density of revenues and $\mathbf{c}_j$ is the fee schedule in place for patents in cohort j.

4

- Similarly,

$$\text{Prob}(t_{ij} \leq f_j) = \sum_{a=1}^{f_j} \pi(a; \mathbf{c}_j) \equiv \mathbf{A} \tag{8}$$

$$\text{Prob}(t_{ij} \geq l_j) = 1 - \sum_{a=1}^{l_j} \pi(a; \mathbf{c}_j) \equiv \mathbf{B} \tag{9}$$

- Log-likelihood function for the aggregate date, letting ω denote the vector of parameters is

$$l(\{n(a,j)\}, \omega) = \sum_{j=1}^{J} \sum_{a=f_j}^{l_j} \log \tilde{\pi}(a; \mathbf{c}_j) n(a,j) \tag{10}$$

where

$$\tilde{pi}(a; \mathbf{c}_j) = \begin{cases} \pi(a; \mathbf{c}_j) & \text{if } f_j < a < l_j \\ \mathbf{A} & \text{if } a \leq f_j \\ \mathbf{B} & \text{if } a \geq l_j \end{cases}$$

3 Estimation Summary

Use a nested algorithm

1. Inner loop: at current parameter values $\hat{\omega}$, solve the dynamic problem and obtain the sequence of thresholds, $\bar{r}_1, \ldots, \bar{r}_{L-1}$

2. Outer loop: for the revenue cutoff values, evaluate the log-likelihood function. This is a complicated integral, evaluated by simulation.

4 Computational Details

4.1 Inner Loop

- Solve for $\bar{r}_1, \ldots, \bar{r}L - 1$ by numerical backwards induction. There are many ways to do this. The brute force way (and most simple) is *discretization*.

- Assume at each age a, returns take values within $[0, \bar{R}]$. Consider a grid of M points over this interval. We will compute the value function $V_1(r; \omega), \ldots, V_{L-1}(r; \omega)$ only on these M points. For values between points, we will approximate the value function via interpolation.

- Specifically:

  - Start with final period L,

  $$V_L(r_L; \omega) = r_L \qquad (11)$$

  for all r_L because there are profits and no chance for renewal after age L (b/c it is off patents, and assume others jump in)

  - Go to period $L - 1$

  $$V_{L-1}(r_{L-1}; \omega) = \max\left\{0, r_{L-1} + E_{r_L|r_{L-1}} V_L(r_L; \hat{\omega}) - c_{L-1}\right\} \qquad (12)$$
  $$= \max\left\{0, r_{L-1} + E_{r_L|r_{L-1}} r_L - c_{L-1}\right\} \qquad (13)$$

  where $E_{r_L|r_{L-1}} r_L$ is evaluated using the assumed parametric structure discussed earlier.

  So for each r_{L-1} on the M grid points, you calculate V_{L-1}.

  Note – in this process you will also uncover the threshold value of $\bar{r}_{L-1}$ (or determine it lies between 2 grid points).

  - Now go to period $L - 2$

  $$V_{L-2}(r_{L-2}; \omega) = \max\left\{0, r_{L-2} + E_{r_{L-1}|r_{L-2}} V_{L-1}(r_{L-1}; \hat{\omega}) - c_{L-2}\right\} \qquad (14)$$
  $$\qquad (15)$$

– For values of $V_{L-1}(r_{L-1}; \hat{\omega})$ between the grid points, approximate the value with interpolation (e.g. a straight line).

– Repeat this procedure for all periods. Will result in all the value functions and threshold values.

4.2   Outer loop

- With the cutoff rules (i.e. optimal decision rules), we can simulation the likelihood function.

- For $s = 1, \ldots, S$ (where S is the number of simulation draws)

   1. Draw a sequence of returns $r_1^s, r_2^s, \ldots, r_L^s$ according the assumed parametric specification. Start by drawing r_1^s and then drawing r_2^s given r_1^s, etc.

   2. Given this sequence, figure out the drop out age t^s, which equals the first a at which $r_a^s < \bar{r}_a(\hat{\omega})$.

   3. Then, for all $a = 1, \ldots, L-1$, you can approximate

$$\pi(a; \mathbf{c}) \approx \frac{1}{S} \sum_{s=1}^{S} \mathbf{1}(t^s = a)$$

   4. Finally, perform simulated maximum likelihood using the above approximation for $\pi(a; \mathbf{c})$.