Econometrica, Vol. 55, No. 5 (September, 1987), 999-1033

OPTIMAL REPLACEMENT OF GMC BUS ENGINES: AN EMPIRICAL MODEL OF HAROLD ZURCHER

BY JOHN RUST[1]

This paper formulates a simple *regenerative optimal stopping* model of bus engine replacement to describe the behavior of Harold Zurcher, superintendent of maintenance at the Madison (Wisconsin) Metropolitan Bus Company. The null hypothesis is that Zurcher's decisions on bus engine replacement coincide with an optimal stopping rule: a strategy which specifies whether or not to replace the current bus engine each period as a function of observed and unobserved state variables. The optimal stopping rule is the solution to a stochastic dynamic programming problem that formalizes the trade-off between the conflicting objectives of minimizing maintenance costs versus minimizing unexpected engine failures. The model depends on unknown "primitive parameters" which specify Zurcher's expectations of the future values of the state variables, the expected costs of regular bus maintenance, and his perceptions of the customer goodwill costs of unexpected failures. Using ten years of monthly data on bus mileage and engine replacements for a subsample of 104 buses in the company fleet, I estimate these primitive parameters and test whether Zurcher's behavior is consistent with the model. Admittedly, few people are likely to take particular interest in Harold Zurcher and bus engine replacement *per se*. I focus on a specific individual and capital good because it provides a simple, concrete framework to illustrate two ideas: (i) a "bottom-up" approach for modelling replacement investment, and (ii) a "nested fixed point" algorithm for estimating dynamic programming models of discrete choice.

KEYWORDS: Optimal replacement, regenerative optimal stopping models, dynamic programming, controlled stochastic processes, nested fixed point algorithm.

1. INTRODUCTION

THIS PAPER FORMULATES a simple *regenerative optimal stopping* model of bus engine replacement to describe the behavior of Harold Zurcher, superintendent of maintenance at the Madison (Wisconsin) Metropolitan Bus Company. The null hypothesis is that Zurcher's decisions on bus engine replacement coincide with an optimal stopping rule: a strategy which specifies whether or not to replace the current bus engine each period as a function of observed and unobserved state variables. The optimal stopping rule is the solution to a stochastic dynamic programming problem that formalizes the trade-off between the conflicting objectives of minimizing maintenance costs versus minimizing unexpected engine failures. The model depends on unknown "primitive parameters" which specify Zurcher's expectations of the future values of the state variables, the expected costs of regular bus maintenance, and his perceptions of the customer goodwill costs of unexpected failures. Using ten years of monthly data on bus mileage and engine replacements for a subsample of 104 buses in the company fleet, I estimate these primitive parameters and test whether Zurcher's behavior is consistent with the model.

[1] This research was made possible by financial support from the Graduate School of the University of Wisconsin and National Science Foundation Grant SES-8419570. I thank Alice Wilcox for an excellent job typing the manuscript and Tom Rust for his careful work in coding the data. I am especially grateful to Harold Zurcher for providing the data used in this study, and for his assistance in interpreting the estimation results.

Admittedly, few people are likely to take particular interest in Harold Zurcher and bus engine replacement, *per se.* I focus on a particular individual and a specific capital good because it provides a simple, concrete framework to illustrate two ideas: (i) a "bottom-up" approach for modelling replacement investment and (ii) a "nested fixed point" algorithm for estimating dynamic programming models of discrete choice.

The "bottom-up" approach uses a micro-theoretic model to derive aggregate replacement investment from individual optimizing behavior. Most existing econometric models use a "top down" approach to derive replacement investiment. This approach, often identified with the work of Jorgenson (1973), requires a measure of a hypothetical continuous aggregate capital stock K, and computes replacement investment using variations of Wicksell's (1934) original proportional decay specification δK. The limitations of the approach are well known; see, for example, the critique by Feldstein and Rothschild (1974). The bottom-up approach, on the other hand, generates replacement investment by explicitly aggregating individual replacement demands for specific capital goods, including bus engines.[2] The seemingly continuous demand for replacement investment at the aggregate level is actually the sum of a large number of binary-valued stochastic processes $\{i_t\}$, where $i_t = 1$ if a replacement occurs at time t and $i_t = 0$ otherwise. Taken to its logical extreme, the bottom-up approach requires us to begin our investigation at the level of individual capital goods, and even individual decision-makers, including Harold Zurcher.[3] The idea is to use economic theory to "explain" the joint stochastic process $\{i_t, x_t\}$, where x_t denotes observed state variables associated with the replacement investment decision. I model $\{i_t, x_t\}$ as a *regenerative stochastic process,* where a regeneration corresponds to replacing an existing used asset with a new one. Under the hypothesis of expected discounted profit maximization, $\{i_t, x_t\}$ is also a *controlled stochastic process* generated from the solution to a dynamic programming problem. The unknown parameters of this stochastic process will generally be a complicated, nonlinear function of the "primitive parameters" of the model, namely, the parameters of the individual's (or firm's) objective (profit) function, and the parameters of the stochastic processes governing observed (and unobserved) state variables. Unfortunately, since the controlled process $\{i_t, x_t\}$ is the solution to a discrete

[2] The bottom-up approach has also been applied to study new investment; a good example is Peck's (1974) model of investment in new electric generators.

[3] Of course, given current limitations on data and computational capacity, I do not pretend that the bottom-up approach can offer a practical approach for forecasting aggregate replacement investment for the forseeable future. However, to the extent that the approach offers an alternative *theory* of aggregate replacement investment, I thought it would be best to test its validity by constructing a narrower, but more precise model at the level of a single individual and single capital good. This way I avoid the econometric problems of aggregation bias (such as the use of aggregate capital stock measures), and heterogeneity bias that plague studies that use aggregate time-series and disaggregate cross-sectional data. By using nonexperimental data I also avoid problems encountered in laboratory tests of choice under uncertainty: lack of incentives and insufficient time to learn. Harold Zurcher, a professional with over 20 years experience in bus maintenance at Madison Metro, has had plenty of time to learn. Furthermore, to the extent Zurcher values his job at Madison Metro, the incentives for behaving rationally ought to be quite high.

stochastic control problem, one will rarely find a closed-form solution for its probability density or any sort of "first order condition" convenient for estimation. In general the solution can only be described recursively using Bellman's principle of optimality. The second objective of this paper is to illustrate a new estimation method that allows me to compute maximum likelihood estimates of the primitive parameters of a class of controlled stochastic processes, even though there is no analytic formula for the associated likelihood function.

The analysis begins in Section 3 where I derive a regenerative optimal stopping model of bus engine replacement which *does* have a simple analytic solution. The model shows how one *derives* the sample likelihood function $\ell(i_1, \ldots, i_T, x_1, \ldots, x_T; \theta)$ for the regenerative process $\{i_t, x_t\}$ as the solution to a regenerative optimal stopping problem.[4] I argue, however, that models with closed-form solutions have certain inherent limitations which make them poor candidates for empirical work and discuss the deficiencies of the analytic model of replacement investment. In Section 4 I describe a *nested fixed point* maximum likelihood algorithm which does not require a closed-form solution to the stochastic control problem, avoiding many of the limitations of current methods which depend critically on the existence of an analytic solution. In Section 5 I generalize the model of Section 3, removing restrictive assumptions about functional forms and incorporating *unobserved state variables.* The regenerative stochastic process $\{i_t, x_t\}$ derived from the solution to this more general model has no closed-form solution, but can be estimated using the nested fixed point algorithm described in Section 4. Using this algorithm, I compute maximum likelihood estimates of the primitive parameters of the model. I conclude in Section 6 by deriving the implied demand curve for replacement investment by aggregating over the individual regenerative processes $\{i_t, x_t\}$.

2. THE DATA

Before I prejudice you with a theoretical model, it's useful to give a simple description of the data. Harold Zurcher was kind enough to provide me with maintenance records on 162 buses in the fleet of Madison Metro over the period December, 1974 (or date of purchase for buses purchased after 12/74) until May, 1985. The data consist of monthly observations on the mileage (odometer reading) on each bus, plus a maintenance diary which records the date, mileage, and list of components repaired or replaced each time a bus visits the company shop.

[4] There is a vast literature in operations research on optimal maintenance and replacement of stochastically deteriorating assets (see, for example, the surveys by Pierskalla and Voelker (1976) and Sherif and Smith (1981)). By and large the focus of this literature is normative: starting from specific assumptions about the objective function and the stochastic process governing deterioration, one derives an optimal replacement strategy from the solution to a stochastic control problem. This paper can be viewed as solving the "inverse" problem: given observations on a sequence of asset states and replacement decisions, I go backwards and infer the objective function and the stochastic process governing deterioration whose associated optimal replacement strategy coincides with the observed data.

TABLE I

BUS TYPES INCLUDED IN SAMPLE

Bus Group	Number of Buses	Manufacturer	Engine	Model	Year	Seats	Empty Weight	Purchase Price	Estimated Value as of 10/1/84
1	15	Grumman	V6-92 series	870	1983	48	25,800	$145,097	$145,097
2	4	Chance	3208 CAT	RT-50	1981	10*	N.A.	100,775	124,772
3	48	GMC	8V71	T8H203	1979	45	25,027	92,668	125,000
4	37	GMC	8V71	5308A	1975	53	20,955	62,506	55,000
5	12	GMC	8V71	5308A	1974	53	20,955	49,975	48,000
6	10	GMC	6V71	4523A	1974	45	19,274	45,704	48,000
7	18	GMC	8V71	5308A	1972	51	20,955	43,856	45,000
8	18	GMC	6V71	4523A	1972	45	19,274	40,542	40,000

Note All buses are diesel powered and have air conditioning

* Handicap bus, outfitted with 4 long benches and accommodations for 6 wheelchairs.

Maintenance operations fall into three categories: (i) routine, periodic maintenance (examples are brake adjustments and tire rotation), (ii) replacement or repair of individual components at time of failure, and (iii) major engine overhaul and/or replacement. This study focuses on the third component of maintenance investment, which can be regarded as part of a general "preventive maintenance" strategy in the following sense. The bus engine can be viewed as a portfolio of individual components each of which has its own individual stochastic failure or "hazard" rate as a function of accumulated use (as measured by the bus odometer). If a particular component fails when a bus has relatively low mileage, then it seems reasonable to simply replace or repair the failed component and put the bus back on the road. However when a particular component fails on a bus with relatively high mileage, then to the extent that one wants to minimize unexpected failures it seems reasonable to expect that other components will fail in the near future, so it might make sense to replace the entire engine with a "new" engine freshly rebuilt in the company machine shop (Zurcher claims that rebuilt engines are every bit as good, if not better, than engines purchased brand new). Under the maintained hypothesis that this preventive maintenance strategy is optimal, I focus on constructing a model which predicts the time and mileage at which engine replacement occurs.

Table IIa summarizes the replacement data for the subsample of buses which had at least one engine replacement. On average, bus engines were replaced after 5 years with over 200,000 elapsed miles. Data for the full sample are also summarized visually in Figure 1, which shows considerable variation in the time and mileage at which replacement occurs. Looking across the different bus groups, we notice large differences in the mean age and mileage at replacement, although it is difficult to tell whether these differences are significant given the large standard deviations and small numbers of observations. A statistical problem with the simple tabulation in Table IIa is that although the use of complete spells avoids bias due to censoring, it fails to account for possible selection bias. Table IIb looks at the subsample of buses for which no replacements occurred. These data

TABLE IIa

SUMMARY OF REPLACEMENT DATA

(Subsample of buses for which at least 1 replacement occurred)

	Mileage at Replacement				Elapsed Time (Months)				
Bus Group	Max	Min	Mean	Standard Deviation	Max	Min	Mean	Standard Deviation	Number of Observations
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0
3	273,400	124,800	199,733	37,459	74	38	59.1	10.9	27
4	387,300	121,300	257,336	65,477	116	28	73.7	23.3	33
5	322,500	118,000	245,291	60,258	127	31	85.4	29.7	11
6	237,200	82,400	150,786	61,007	127	49	74.7	35.2	7
7	331,800	121,000	208,963	48,981	104	41	68.3	16.9	27
8	297,500	132,000	186,700	43,956	104	36	58.4	22.2	19
Full Sample	387,400	83,400	216,354	60,475	127	28	68.1	22.4	124

are *right censored* since we do not observe the final age and mileage at which replacement occurs. We can see from Table IIb that despite the right censoring, both the mean elapsed age and mileage are significantly higher for this subsample. The data for bus groups 7 and 8 are also *left censored* since these buses were acquired in 1972 and my data begin in December, 1974. The presence of these biases makes it difficult to summarize the unconditional population distribution of the age and mileage at replacement. The empirical analysis in Section 5 implicitly accounts for censored spells through the use of a conditional likelihood function given the observed sample of data. I account for selection bias by allowing for heterogeneity in parameter estimates across bus groups.

The empirical analysis in Section 5 focuses on a subsample of the full data set, bus groups 1-4. The buses in these groups were the most recent acquisitions

TABLE IIb

CENSORED DATA

(Subsample of buses for which no replacements occurred)

	Mileage at May 1, 1985				Elapsed Time (months)				
Bus Group	Max	Min	Mean	Standard Deviation	Max	Min	Mean	Standard Deviation	Number of Observations
1	120,151	65,643	100,117	12,929	25	25	25	0	15
2	161,748	142,009	151,183	8,530	49	49	49	0	4
3	280,802	199,626	250,766	21,325	75	75	75	0	21
4	352,450	310,910	337,222	17,802	118	117	117.8	0.45	5
5	326,843	326,843	326,843	0	130	130	130	0	1
6	299,040	232,395	265,264	33,332	130	128	129.3	1.15	3
7	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0
Full Sample	352,450	65,643	207,782	85,208	130	25	66.4	34.6	49

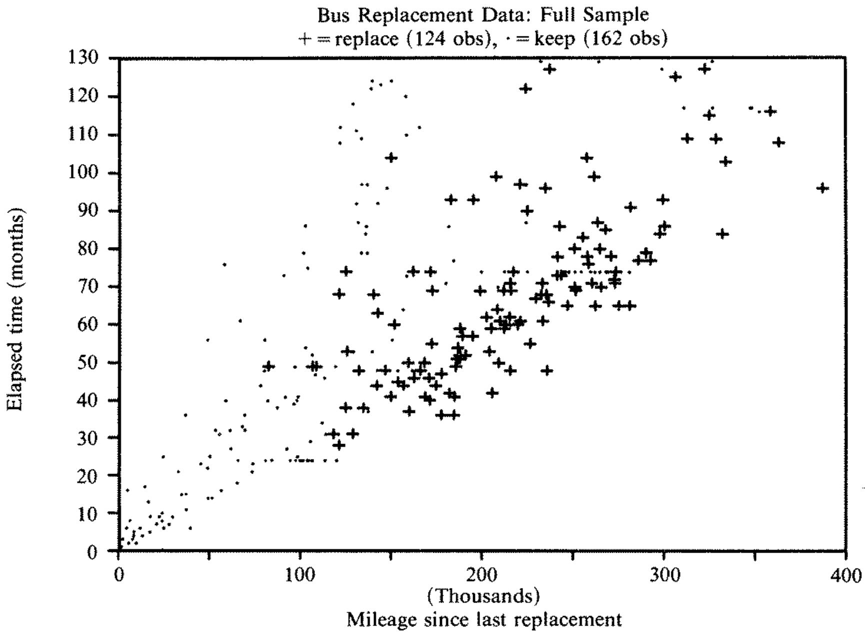

FIGURE 1

at Madison Metro, the main "workhorses" on the company's most active bus routes. I focus on this subsample for two reasons: (a) data on actual engine replacement costs were available for these groups, (b) utilization, summarized by the monthly mileage distributions for each bus, is fairly homogeneous within each of the four groups. Since the estimation procedure allows for heterogeneity *between* groups, but does not account for differences in buses *within* each group, I wanted to minimize the possible heterogeneity bias by selecting bus groups which appeared to be most homogeneous. Estimates of discretized monthly mileage given in Table VI in Section 5 show that we cannot reject the hypothesis that the monthly mileage distributions for the individual buses within each of these groups are identical. On the other hand the older 1972 and 1974 GMC buses in groups 5–8 have been utilized less intensively since the acquisition of the new GMC model 203 buses in 1979. The fixed effects regression results in Table VII of Section 5 (see equation (5.6)) show that monthly mileage for the newer groups 1–3 is significantly higher, by 308 miles. The policy of putting older buses "out to pasture" on charter assignments and low mileage routes suggests that a simple replace/no replace model which treats utilization as exogenous is not strictly correct. Less intense utilization is an obvious substitute for more frequent maintenance. Older buses can also be kept in inventory as back-ups or "spares", providing another substitution possibility. Although utilization and replacement are best viewed as jointly endogenous decisions in a comprehensive maintenance policy, I decided that since a joint model is substantially more

TABLE III

AVERAGE ENGINE REPLACEMENT COSTS[a]

	Bus Group		
Operation	1, 2, 3	4	1, 2, 3, 4
Labor time[b] to drop engine	\$ 150	\$ 150	\$ 150
Labor time[b] to overhaul engine	3373	2870	3032
Parts required to overhaul engine	5826	4343	4730
Labor time[b] to re-install engine	150	150	150
Total cost of replacement	\$9499	\$7513	\$8062

[a] Based on 1985 replacement cost data supplied by Harold Zurcher.
[b] Includes fringe benefits.

complex, and since these rather subtle interrelationships would be difficult to identify given my limited sample, it would be best to focus on the simplest model capable of explaining the major features of the data.

Table III shows the average engine replacement costs for bus groups 1 to 4. These data will be used in Section 5 to identify additional parameters of an expected cost function which specifies Zurcher's perceptions of the combined costs of monthly maintenance and lost customer goodwill due to unexpected breakdowns. Notice that total replacement costs for the newer buses (groups 1, 2, 3) is about 25 per cent higher than the older 1975 GMC buses. Despite these higher replacement costs we can see from Table IIa that engine replacements for the newer buses occur on average 57,600 miles and 14.6 months *earlier* than for the older 1975 GMC buses. Presumably the operating and maintenance costs for the newer buses must increase faster than for the older buses in group 4 in order to warrant this behavior.

3. OPTIMAL REPLACEMENT OF BUS ENGINES

My objective is to explain the bus data by deriving a regenerative stochastic process $\{i_t, x_t\}$ with an associated likelihood function $\ell(i_1, \ldots, i_T, x_1, \ldots, x_T; \theta)$ formed from the solution to a particular *regenerative optimal stopping* problem.

Let the *state variable* x_t denote the accumulated mileage (since last replacement) on the bus engine at time t and suppose that expected per period operating costs are given by an increasing, differentiable function of x_t, $c(x_t, \theta_1)$. Operating costs are the sum of maintenance, fuel, and insurance costs (which are potentially observable), plus Zurcher's estimate of the costs of lost ridership and goodwill due to unexpected breakdowns. The latter costs are generally not directly observable, so I attempt to infer them by postulating a total cost function $c(\cdot, \theta_1)$ and estimating θ_1. The function c can be decomposed as follows:

$$c(x, \theta_1) = m(x, \theta_{11}) + \mu(x, \theta_{12}) b(x, \theta_{13}) \tag{3.1}$$

where $m(x, \theta_{11})$ is the conditional expectation of normal maintenance and operating expenses, $\mu(x, \theta_{12})$ is the conditional probability of an unexpected engine failure, and $b(x, \theta_{13})$ is the conditional expectation of towing costs, repair costs, and the perceived dollar cost of lost customer goodwill in the event of an unexpected engine failure. Given actual maintenance and operating cost data, one could directly estimate m by nonlinear regression. Unfortunately I do not have these data, nor do I have data on the occurrence of unexpected breakdowns, so I am unable to separately identify the functions m, μ, and b. Therefore the best I can do is to specify and estimate their sum, c.

Suppose that the mileage travelled each month by a given bus is exponentially distributed with parameter θ_2, independently of mileage driven in previous periods. Each month, Zurcher faces the discrete decision: (i) perform "normal maintenance" on the current bus engine and incur operating costs $c(x_t, \theta_1)$, or (ii) "cannibalize" the old bus engine for scrap value $\underline{P}$, install a new (or rebuilt) bus engine at cost $\bar{P}$, and incur operating costs $c(0, \theta_1)$. I assume that Zurcher chooses an optimal replacement policy to minimize the expected discounted costs of maintaining his fleet of buses. Let i_t denote Zurcher's replacement decision at time t, $i_t = 0$ (keep), $i_t = 1$ (replace). It follows that the stochastic process governing $\{i_t, x_t\}$ is the solution to the following *regenerative optimal stopping* problem:

$$(3.2) \qquad V_\theta(x_t) = \sup_{\Pi} E\left\{ \sum_{j=t}^{\infty} \beta^{j-t} u(x_j, f_j, \theta_1) \,\middle|\, x_t \right\}$$

where the *utility function* u is given by

$$(3.3) \qquad u(x_t, i_t, \theta_1) = \begin{cases} -c(x_t, \theta_1) & \text{if} \quad i_t = 0, \\ -[\bar{P} - \underline{P} + c(0, \theta_1)] & \text{if} \quad i_t = 1, \end{cases}$$

and where Π is an infinite sequence of *decision rules* $\Pi = \{f_t, f_{t+1}, \ldots\}$ where each f_t specifies Zurcher's replacement decision at time t as a function of the entire history of the process, $i_t = f_t(x_t, i_{t-1}, x_{t-1}, i_{t-2}, x_{t-2}, \ldots)$ and the expectation in (3.2) is taken with respect to the controlled stochastic process $\{x_t\}$ whose probability distribution is defined from Π and the transition probability $p(x_{t+1} | x_t, i_t, \theta_2)$. The utility function (3.3) shows why I call the model a *regenerative* optimal stopping model of engine replacement: once the bus engine is replaced the system "regenerates" to state $x_t = 0$. This regeneration property is formally defined by the stochastic process governing the evolution of $\{x_t\}$ given by the transition probability $p(x_{t+1} | x_t, i_t, \theta_3)$ below:

$$(3.4) \qquad p(x_{t+1} | x_t, i_t, \theta_2) = \begin{cases} \theta_2 \exp\{\theta_2(x_{t+1} - x_t)\} & \text{if} \quad i_t = 0 \quad \text{and} \quad x_{t+1} \geqslant x_t, \\ \theta_2 \exp\{\theta_2(x_{t+1})\} & \text{if} \quad i_t = 1 \quad \text{and} \quad x_{t+1} \geqslant 0, \\ 0 & \text{otherwise.} \end{cases}$$

According to (3.4) if the decision is made to keep the current bus engine ($i_t = 0$), then next period accumulated mileage x_{t+1} is given by a draw from the exponential distribution $1 - \exp\{\theta_2(x_{t+1} - x_t)\}$. However if the decision is made to replace the bus engine ($i_t = 1$), then x_t regenerates to state 0 and next period accumulated mileage x_{t+1} is a draw from the exponential distribution $1 - \exp\{\theta_2(x_{t+1} - 0)\}$.

The function $V_\theta(x_t)$ defined in (3.2) is the *value function* and is the unique solution to *Bellman's equation* given by[5]

$$V_\theta(x_t) = \max_{i_t \in C(x_t)} [u(x_t, i_t, \theta_1) + \beta EV_\theta(x_t, i_t)] \tag{3.5}$$

where $C(x_t) = \{0, 1\}$ and where the function $EV_\theta(x_t, i_t)$ is defined by

$$EV_\theta(x_t, i_t) \equiv \int_0^\infty V_\theta(y) p(dy \mid x_t, i_t, \theta_2). \tag{3.6}$$

Using Bellman's equation, I have shown elsewhere (Rust, (1986a)) that there is an optimal stationary, Markovian replacement policy $\Pi = (f, f \ldots)$ where f is given by

$$i_t = f(x_t, \theta) = \begin{cases} 1 & \text{if} \quad x_t > \gamma(\theta_1, \theta_2), \\ 0 & \text{if} \quad x_t \leqslant \gamma(\theta_1, \theta_2), \end{cases} \tag{3.7}$$

where $\gamma(\theta_1, \theta_2)$ is the unique solution to

$$(\bar{P} - \underline{P})(1 - \beta) = \int_0^{\gamma(\theta_1, \theta_2)} [1 - \beta \exp\{-\theta_2(1-\beta)y\}] \, \partial c(y, \theta_1)/\partial y \, dy. \tag{3.8}$$

The constant γ represents a threshold value of mileage (optimal stopping barrier) such that whenever current mileage on the bus x_t exceeds γ it is optimal to incur the replacement costs $RC = (\bar{P} - \underline{P})$ and replace the old bus engine with a new one.

The likelihood function $\ell(i_1, \ldots, i_T, x_1, \ldots, x_T, \theta)$ specifies the conditional probability density of observing the sequence of states and replacement decisions for a single bus in periods 1 to T. Under the assumption that monthly mileage and replacement decisions are independently distributed across buses, the likelihood function $L(\theta)$ for the full sample of data is simply the product of the individual bus likelihoods ℓ. The precise functional form of this likelihood function can be easily derived from the optimal stopping rule (3.7) and (3.8) using the regeneration property, the fact that the distribution of monthly mileage is exponential, and the easily proven result that the distribution of the optimal stopping time (i.e. the first passage time from $x = 0$ to the optimal stopping barrier γ) is Poisson with parameter $\theta_2 \gamma$. This structural model has two key features which distinguish it from traditional *reduced-form* models of replacement investment: (i) the parametric specification occurs at the level of the primitive objects of the model, namely, the utility function $u(x_t, t_i, \theta_1)$ and the transition probability $p(x_{t+1} \mid x_t, i_t, \theta_2)$, (ii) the sample likelihood function is not specified directly, but rather is *derived* from the solution to the underlying optimization problem. Thus, ℓ is simply the probability density of the controlled stochastic process $\{i_t, x_t\}$.

Although this simple model leads directly to a convenient, analytic formula for the likelihood function, I have serious reservations about using it for empirical

[5] A good reference on dynamic programming and stochastic control which derives Bellman's equation is Bertsekas (1976).

work. The solution for the likelihood function depends critically on specific choice of functional form: namely, that monthly mileage $(x_{t+1}-x_t)$ has an i.i.d. exponential distribution. Unfortunately, my sample of data flatly refutes this assumption: the exponential distribution constrains the mean and standard deviation of monthly mileage to be equal, whereas the data show that the standard deviation is less than one third of mean monthly mileage. If I try to use a more realistic mileage distribution (such as the log-normal distribution which has separate parameters for mean and variance), I can no longer obtain an explicit solution for the stochastic control problem (3.2) and the associated likelihood function.[6] Perhaps even more restrictive is the basic model formulation which assumes that the physical state of a bus is completely described by a single variable, accumulated mileage x_t. This formulation implies a degenerate hazard function for bus engine replacement: the probability of replacing a bus engine is 0 in the interval $(0, \gamma)$ and 1 thereafter. Looking back at the replacement data summarized in Figure 1 we can see that there is clear evidence against the hypothesis of a single fixed optimal stopping barrier γ: mileage at replacement varies from a minimum of 82,400 to a maximum of 387,300. This variation is too large to be consistent with a threshold replacement rule. More realistically, we might assume that the odometer value x_t might be only one indicator of the physical state of the bus, and Harold Zurcher might base his replacement decisions on other information ε_t which we have not observed. Unfortunately, my attempts to formulate a more realistic model which included such *unobserved state variables* lead to models which had no analytical solution.

The problem of statistical degeneracy caused by a failure to account for unobserved state variables is not unique to this model; it is a problem common to the majority of models in decision theory. A basic result in Markovian decision theory (cf. Blackwell (1968)) shows that under quite general conditions the solution to the class of infinite horizon Markovian decision problems takes the general form

$$i_t = f(x_t, \theta) \tag{3.9}$$

where f is some deterministic function relating the agent's state variables x_t to his optimal action i_t. Suppose we assume that there are no unobserved state variables, i.e. that the econometrician observes all of x_t. The theory then implies that the data obey the *deterministic* relation (3.9) for some unknown parameter value θ^*. However in general, real data will never exactly obey (3.9) for any value of the parameter θ: the data *contradict* the underlying optimization model. The typical solution to this problem is to "add an error term" ε_t in order to reconcile the difference between $f(x_t, \theta)$ and the observed choice i_t

$$i_t = f(x_t, \theta) + \varepsilon_t. \tag{3.10}$$

[6] The solution requires computation of the fixed point V_θ to the functional equation (3.4) and computing the optimal stopping boundary γ_θ by solving the nonlinear equation $V_\theta(\gamma_\theta) = \bar{P} - \underline{P} + V_\theta(0)$.

By making a convenient distributional assumption for ε_t, one might use the model (3.10) to estimate θ. The difficulty with this procedure is that it is *internally inconsistent*: the structural model was formulated on the hypothesis that the agent's behavior is described by the solution of a dynamic optimization problem, yet the statistical implementation of that model implies that the agent randomly departs from this optimal solution. If error terms ε_t are to be introduced to a structural model in an internally consistent fashion, they must be explicitly incorporated into the solution of the dynamic optimization problem. When this is done, a correct interpretation of the "error term" ε_t is that it is an *unobservable*, a state variable which is observed by the agent but not by the statistician.[7]

4. STRUCTURAL ESTIMATION WITHOUT CLOSED-FORM SOLUTIONS

Rust (1987) has developed a maximum likelihood estimation algorithm for a class of dynamic discrete choice models which (i) does not require closed-form solutions for the agent's stochastic control problem and associated likelihood function, and (ii) treats unobservables ε_t in an internally consistent fashion by explicitly incorporating them into the formulation and solution of the model. Although the algorithm can estimate a considerably wider class of models than the regenerative processes considered here, the basic notation for the general case is no more complicated, so I present the general notation below:

$C(x_t)$:	Choice set; a finite set of allowable values of the control variable i_t when state variable is x_t.
$\varepsilon_t = \{\varepsilon_t(i) \mid i \in C(x_t)\}$:	A $\# C(x_t)$-dimensional vector of state variables observed by agent but not by the econometrician. $\varepsilon_t(i)$ is interpreted as a component of utility of an alternative i in time period t which is known by the agent but not by the econometrician.
$x_t = \{x_t(1), \ldots, x_t(K)\}$:	K-dimensional vector of state variables observed by both the agent *and* the econometrician.
$u(x_t, i, \theta_1) + \varepsilon_t(i)$:	Realized single-period utility of decision i when state variable is (x_t, ε_t). θ_1 is a vector of unknown parameters to be estimated.
$p(x_{t+1}, \varepsilon_{t+1} \mid x_t, \varepsilon_t, i_t, \theta_2, \theta_3)$:	Markov transition density for state variable (x_t, ε_t) when alternative i_t is selected. θ_2 and θ_3 are vectors of unknown parameters to be estimated.
$\theta = (\beta, \theta_1, \theta_2, \theta_3)$:	The complete $(1 + K_1 + K_2 + K_3)$ vector of parameters to be estimated.

Given the stochastic evolution of the state variables (x_t, ε_t) embodied by the transition probability p, the agent must choose a sequence of decision rules or

[7] Besides increased realism, the addition of unobservable state variables offers another benefit: additional parameters can be estimated. The rigid specification of the replacement model without unobservables condenses all information about replacement behavior into the single constant γ. As a result, at most 1 cost function parameter θ_1 is identifiable in this model (see equation (3.8)). Addition of unobservables produces a model which can be consistent with a wide variety of shapes for the implied replacement hazard function, enabling us to identify more cost function parameters, as well as the replacement cost parameter RC and possibly the discount factor β.

controls $f_t(x_t, \varepsilon_t, \theta)$ to maximize expected discounted utility over an infinite horizon. Define the value function V_θ by

$$(4.1)\qquad V_\theta(x_t, \varepsilon_t) = \sup_{\Pi} E\left\{\sum_{j=t}^{\infty} \beta^{(j-t)}[u(x_j, f_j, \theta_1) + \varepsilon_j(f_j)] \| x_t, \varepsilon_t, \theta_2, \theta_3\right\}$$

where $\Pi = \{f_t, f_{t+1}, f_{t+2}, \ldots\}$, $f_t \in C(x_t)$ for all t, and where the expectation is taken with respect to the controlled stochastic process $\{x_t, \varepsilon_t\}$ whose probability density is defined from Π and the transition probability p by

$$(4.2)\qquad dp\{x_{t+1}, \varepsilon_{t+1}, \ldots, x_{t+N}, \varepsilon_{t+N} | x_t, \varepsilon_t\}$$
$$= \prod_{i=t}^{N-1} p(x_{i+1}, \varepsilon_{i+1} | x_i, \varepsilon_i, f_i(x_i, \varepsilon_i), \theta_2, \theta_3).$$

Problem (4.1) is known as an infinite-horizon, discounted Markovian decision problem. Under certain regularity assumptions described in Rust (1987) the solution to this problem is given by a stationary decision rule

$$(4.3)\qquad i_t = f(x_t, \varepsilon_t, \theta)$$

which specifies the agent's optimal decision when the state variables are (x_t, ε_t). The optimal value function V_θ is the unique solution to Bellman's equation given by

$$(4.4)\qquad V_\theta(x_t, \varepsilon_t) = \max_{i \in C(x_t)} [u(x_t, i, \theta_1) + \varepsilon_t(i) + \beta EV_\theta(x_t, \varepsilon_t, i)]$$

where

$$(4.5)\qquad EV_\theta(x_t, \varepsilon_t, i) \equiv \int_y \int_\eta V_\theta(y, \eta) p(dy, d\eta | x_t, \varepsilon_t, i, \theta_2, \theta_3)$$

and the optimal control f is defined by

$$(4.6)\qquad f(x_t, \varepsilon_t, \theta) \equiv \operatorname*{argmax}_{i \in C(x_t)} [u(x_t, i, \theta_1) + \varepsilon_t(i) + \beta EV_\theta(x_t, \varepsilon_t, i)].$$

As it stands, there are two difficulties which hamper direct econometric implementation of the model $i_t = f(x_t, \varepsilon_t, \theta)$ given by the solution to (4.4) and (4.6). First, many commonly chosen distributions for the unobservable ε_t will be continuously distributed with unbounded support. However, this raises serious dimensionality problems since the optimal stationary policy f will ordinarily be computed by solving for the fixed point V_θ from Bellman's equation. Even taking a rough grid approximation to the true continuous distribution of ε_t, the dimensionality of the resulting finite approximation will still be too large to be computationally tractable. Secondly, since ε_t appears nonlinearly in the unknown function EV_θ, we face the additional problem of integrating out over the ε_t distribution to obtain choice probabilities. Since EV_θ is an unknown function, this will require the dual task of integrating V_θ with respect to a finite grid approximation of the density $p(x_{t+1}, \varepsilon_{t+1} | x_t, \varepsilon_t, i, \theta_2, \theta_3)$ to obtain EV_θ, and then numerically integrating Bellman's equation (4.4) to obtain the conditional choice probability $P(i_t|x_t, \theta)$. The following assumption (number (A6) in Rust (1987)) enables us to circumvent these problems.

CONDITIONAL INDEPENDENCE ASSUMPTION (CI): *The transition density of the controlled process* $\{x_t, \varepsilon_t\}$ *factors as*

$$(4.7) \qquad p(x_{t+1}, \varepsilon_{t+1}|x_t, \varepsilon_t, i, \theta_2, \theta_3) = q(\varepsilon_{t+1}|x_{t+1}, \theta_2)p(x_{t+1}|x_t, i, \theta_3).$$

Assumption (CI) involves two restrictions. First, x_{t+1} is a sufficient statistic for ε_{t+1}, which implies that any statistical dependence between ε_t and ε_{t+1} is transmitted entirely through the vector x_{t+1}. Second, the probability density of x_{t+1} depends only on x_t and not ε_t. Intuitively, the $\{\varepsilon_t\}$ process can be regarded as noise superimposed on the underlying $\{x_t\}$ process, since in each period t, ε_t is drawn according to the density $q(\varepsilon_t|x_t, \theta_2)$ given the realized value of x_t. Admittedly, (CI) is a strong assumption.[8] The payoff is twofold. First, (CI) implies that EV_θ is not a function of ε_t, so that required choice probabilities will not require integration over the unknown function EV_θ. Second, (CI) implies that EV_θ is a fixed point of a separate contraction mapping on the reduced state space $\Gamma = \{(x, i)|x \in R^K, i \in C(x)\}$, eliminating the need to compute the fixed point V_θ on the much larger full state space $S = \{(x, \varepsilon)|x \in R^K, \varepsilon \in R^{\# C(x)}\}$ and avoiding the numerical integration required to obtain EV_θ from V_θ. These results are summarized in the following theorem proven in Rust (1987).

THEOREM 1: *Assume that (CI) holds. Let* $P(i|x, \theta)$ *denote the conditional probability of choosing action* $i \in C(x)$ *given state variable* x. *Let* $G([u(x, \theta_1) + \beta EV_\theta(x)]|x, \theta_2)$ *denote the social surplus function corresponding to the density* $q(\varepsilon|x, \theta_2)$, *defined by*

$$(4.8) \qquad G([u(x, \theta_1) + \beta EV_\theta(x)]|x, \theta_2)$$

$$\equiv \int_\varepsilon \max_{j \in C(x)} [u(x, j, \theta_1) + \beta EV_\theta(x, j)]q(d\varepsilon|x, \theta_2).$$

Then $P(i|x, \theta)$ *is given by*

$$(4.9) \qquad P(i|x, \theta) = G_i([u(x, \theta_1) + \beta EV_\theta(x)]|x, \theta_2)$$

where G_i *denotes the partial derivative of* G *with respect to* $u(x, i, \theta_1)$ *and the function* EV_θ *is the unique fixed point to a contraction mapping* T_θ, $T_\theta(EV_\theta) = EV_\theta$, *defined for each* $(x, i) \in \Gamma$ *by*

$$(4.10) \qquad EV_\theta(x, i) = \int_y G([u(y, \theta_1) + \beta EV_\theta(y)]|y, \theta_2)p(dy|x, i, \theta_3).$$

[8] I present a specification test for (CI) in Section 5.

The significance of Theorem 1 is that the conditional choice probabilities $P(i|x,\theta)$ can be computed using the same formulas used in the static case with the addition of the term $\beta EV_\theta(x,i)$ to the usual static utility term $u(x,i,\theta_1)$. Notice that McFadden's (1973), (1981), static model of discrete choice appears as a special case of Theorem 1 when $p(\cdot|x,i,\theta_3)$ is independent of i. In that case the expected utilities $EV_\theta(x,i)$ are also independent of i which implies that G is a function of $\{u(x,j,\theta)|j\in C(x)\}$ alone. This implies that $P(i|x,\theta)=G_i(u(x,\theta)|x,\theta_2)$ can be interpreted at the usual static choice probability. The intuition behind this result is clear; when $p(\cdot|x,i,\theta_3)$ is independent of i, current choices do not affect the evolution of the state variables $\{x_t,\varepsilon_t\}$ and so have no future consequences. Therefore, it is optimal to behave myopically each period and choose the alternative i which maximizes single period utility $u(x_t,i,\theta_1)+\varepsilon_t(i)$. When current choices do have future consequences, the term $\beta EV_\theta(x,i)$ provides the appropriate "shadow price" for the future consequences of each action and must be added to the current utility in order to correctly describe the optimal behavior of the agent. Specific functional forms for $q(\varepsilon|y,\theta_2)$ yield more concrete formulas for the choice probability $P(i|x,\theta)$ and the contraction mapping T_θ. For example, if $q(\varepsilon|y,\theta_2)$ is given by a multivariate extreme value distribution

$$(4.11)\qquad q(\varepsilon|x,\theta_2)=\prod_{j\in C(x)}\exp\{-\varepsilon(j)+\theta_2\}\exp\{-\exp\{-\varepsilon(j)+\theta_2\}\}$$

$$\theta_2=\gamma=0.577216,$$

then the social surplus function G is given by

$$(4.12)\qquad G([u(x,\theta_1)+\beta EV_\theta(x)]|x,\theta_2)$$
$$=\ln\left\{\sum_{j\in C(x)}\exp[u(x,j,\theta_1)+\beta EV_\theta(x,j)]\right\},$$

$P(i|x,\theta)$ is given by the well-known multinomial logit formula

$$(4.13)\qquad P(i|x,\theta)=\frac{\exp\{u(x,i,\theta_1)+\beta EV_\theta(x,i)\}}{\sum_{j\in C(x)}\exp\{u(x,j,\theta_1)+\beta EV_\theta(x,j)\}},$$

and EV_θ is given by the unique solution to the functional equation

$$(4.14)\qquad EV_\theta(x,i)=\int_y\log\left\{\sum_{j\in C(y)}\exp[u(y,j,\theta_1)+\beta EV_\theta(y,j)]\right\}p(dy|x,i,\theta_3).$$

We now are in position to state exactly how the structural parameters of the controlled process $\{i_t,x_t\}$ can be estimated. Given time series observations $\{(i_0,x_0),(i_1,x_1),\ldots,(i_T,x_T)\}$ for a single individual we form the likelihood function $\ell^f(i_1,x_1,\ldots,i_T,x_T|i_0,x_0,\theta)$ and estimate the unknown parameters θ by the method of maximum likelihood. The following theorem of Rust (1987) shows that under Assumption (CI) this likelihood function has an especially simple form.

THEOREM 2: *Under Assumption* (*CI*) *the likelihood function* ℓ^f is given by

$$(4.15)\qquad \ell^f(x_1,\dots,x_T,i_1,\dots i_T\,|\,x_0,i_0,\theta)=\prod_{t=1}^{T}P(i_t\,|\,x_t,\theta)p(x_t\,|\,x_{t-1},i_{t-1},\theta_3),$$

where $P(i_t\,|\,x_t,\theta)$ *is given by* (4.9).

Formula (4.15) shows how previous period choices i_{t-1} can affect current period choices i_t by altering the probability distribution of the state variable x_t. Thus the model reflects what Heckman (1981) terms *structural state dependence.* Given a cross-section of individuals each of whom has $T\geqslant 2$ periods data, we can compute the likelihood for the full panel by simply multiplying the likelihoods ℓ^f for each individual. Theorems 9 and 11 of Rust (1987) prove that as the number of individuals in the cross-section tends to infinity, the corresponding sequence of maximum likelihood estimators are consistent and asymptotically normally distributed. Alternatively, one can invoke the martingale limit theorems of Billingsley (1961) to prove consistency and asymptotic normality of the estimator for a single individual as the number of time periods T tends to infinity.

Although Theorems 1 and 2 suggest that *in theory* one can estimate a wide class of discrete choice processes, in practice the range of estimable models will be much more limited. The virtue of the approach is that it frees us from using restrictive and contrived functional forms just because they yield closed-form solutions. However the drawback is the computational burden of numerical solution of the contraction fixed point EV_θ needed to solve the stochastic control problem. If the approach is to be of any practical use, we must find an efficient algorithm to compute the maximum likelihood estimates. Theorem 1 and 2 suggest the following *nested fixed point* algorithm: an "inner" fixed point algorithm computes the unknown function EV_θ for each value of θ, and an "outer" hill climbing algorithm searches for the value of θ which maximizes the likelihood function. Rust (1987) showed that the contraction mapping T_θ is Fréchet differentiable. This enables us to use the highly efficient Newton-Kantorovich algorithm to compute EV_θ, and as a by-product, yields analytic solutions for the θ derivatives of EV_θ needed to compute the derivatives of the likelihood function. If the vector x contains components that are continuously distributed, it will be necessary to *discretize* these components in order to compute EV_θ on a digital computer. The discretization procedure approximates the function EV_θ, an element of an infinite-dimensional Banach space B, by a suitable vector in a high-dimensional Euclidean space.[9] I have programmed the nested fixed point algorithm on the IBM-PC, and used it to compute fixed points of several hundred dimensions. The contraction property guarantees that the Newton-Kantorovich iteration is numerically well-conditioned so that the resulting fixed point is

[9] Theorems 1 and 2 implicity assume that the fixed point EV_θ is computed exactly. A referee has pointed out that if EV_θ can only be computed approximately, the choice of discretization may affect the asymptotic distribution of the parameter estimates. The referee suggests expanding the number of grid points in the discretization as a function of the sample size in order to deduce the correct asymptotic distribution. The point is well-taken: I think this suggestion is an important area for further research.

TABLE IV

APPROXIMATE SPEED OF THE NESTED FIXED POINT ALGORITHM WRITTEN IN GAUSS FOR THE IBM-PC

Fixed point time[b] (90 dimensions to tolerance 10^{-16})	60 seconds
Function evaluation and moment matrix time (16,000 bus/month obs.)	180 seconds
Total time required per likelihood function evaluation	240 seconds[a]

[a] During line search, we do not require computation of the moment matrix of first derivatives, resulting in a savings of 30 seconds Thus approximately 210 seconds are required per likelihood function evaluation during line search.

[b] The fixed point algorithm can run up to 4 times faster by using special linear equation algorithms which account for the banded structure of the transition probability matrix used to compute the Newton-Kantorovich iterations. The results here used a general Crout decomposition algorithm to solve the linear system taking no account of the special structure of the problem.

insensitive to round-off-errors as long as β is less than 1. The performance of the algorithm for the 90-dimensional fixed point problem solved in Section 5 is presented in Table IV. Using standard linear algebra routines written in the *Gauss* programming language, the nested fixed point algorithm can compute a 90-dimensional fixed point to within 10^{-16} in two Newton-Kantorovich iterations in an average of 60 seconds. Using the full data set with approximately 16,000 bus/month observations, the time required to evaluate the likelihood and compute the moment matrix of first derivatives averaged about four minutes. Thus, internal computation of the fixed point amounted to about 1/4 of the total time required for each likelihood function evaluation. Notice that these timings are based on an algorithm which ignores the special banded structure of the Markov transition matrix of the regenerative optimal stopping problem. If I use special band-matrix linear algebra routines which exploit this special structure, I can compute the fixed point in less than 30 seconds. For details about computation using the nested fixed point algorithm, see Rust (1985b).[10]

5. APPLYING THE NESTED FIXED POINT ALGORITHM TO BUS ENGINE REPLACEMENT

In this section I generalize the regenerative optimal stopping model presented in Section 3, eliminating restrictive assumptions about functional form and incorporating unobserved state variables. The regenerative stochastic process $\{i_t, x_t\}$ derived from the solution to this more general model has no closed-form solution, but can be estimated using the nested fixed point algorithm described in Section 4. In terms of the general approach of Section 4 the choice set is binary, $C(x_t)=\{0,1\}$. I incorporate unobserved state variables by assuming that unobserved costs $\{\varepsilon_t(0), \varepsilon_t(1)\}$ follow a specific stochastic process, to be described below. Let RC denote the expected cost of a replacement bus engine. In terms of my earlier notation, $RC=\bar{P}-\underline{P}$ and I can write Harold Zurcher's implied

[10] At least four other studies have constructed computable maximum likelihood algorithms which, similar to the nested fixed point algorithm, require internal computation of the likelihood *function*, as well as its value. The studies by Gotz and McCall (1986), Miller (1984), Pakes (1986), and Wolpin (1984) are, to my knowledge, the first examples of estimable econometric models which are derived from discrete stochastic control problems which do not have closed-form solutions.

utility function as follows:

$$(5.1)\qquad u(x_t, i, \theta_1)+\varepsilon_t(i)=\begin{cases}-RC-c(0,\theta_1)+\varepsilon_t(1) & \text{if } i=1,\\ -c(x_t,\theta_1)+\varepsilon_t(0) & \text{if } i=0.\end{cases}$$

I relax the restrictive functional form assumptions of Section 3 by allowing monthly mileage $(x_{t+1}-x_t)$ to have an arbitrary parametric density function g, which implies a transition density of the form

$$(5.2)\qquad p(x_{t+1}|x_t, i_t, \theta_3)=\begin{cases}g(x_{t+1}-x_t,\theta_3) & \text{if } i_t=0,\\ g(x_{t+1}-0,\theta_3) & \text{if } i_t=1.\end{cases}$$

If i_t is fixed at 0, formula (5.2) implies that total bus mileage follows a random walk (with drift), where monthly incremental mileage is given by the density g with support on $(0,\infty)$. When the behavior of the optimal control $i_t=f(x_t,\epsilon_t,\theta)$ is taken into account, (5.2) defines a *regenerative random walk* for the controlled process $\{x_t\}$.

In summary, the data consist of $\{i_t^m, x_t^m\}$ $(t=1,\ldots,T_m;\ m=1,\ldots,M)$ where i_t^m is the engine replacement decision in month t for bus m and x_t^m is the mileage since last replacement of bus m in month t. I assume that the data are a realization of a controlled Markov process generated from the solution to the infinite horizon stochastic control problem (4.1). My procedure is to estimate the unknown parameters $\theta=(\beta,\theta_1,RC,\theta_3)$ by maximum likelihood using the nested fixed point algorithm. To do this I had to (i) discretize the state variable x_t to enable me to compute the fixed point EV_θ on the IBM-PC, (ii) specify functional forms for c, q, and g. I discretized mileage into 90 intervals of length 5,000, which implies that the fixed point EV_θ is an element of the Banach space $B=R^{90}$. Using the discretized mileage data, the distribution g reduces to a multinomial distribution on the set $\{0,1,2\}$, corresponding to monthly mileage in the intervals $[0,5000)$, $[5000, 10{,}000)$ and $[10{,}000,+\infty)$, respectively. Thus, the distribution is completely specified by two parameters $(\theta_{30},\theta_{31})$. The functional forms for c which I estimated include (i) polynomial: $c(x,\theta_1)=\theta_{11}x+\theta_{12}x^2+\theta_{13}x^3$, (ii) exponential: $c(x,\theta_1)=\theta_{11}\exp(\theta_{12}x)$, (iii) hyperbolic: $c(x,\theta_1)=\theta_{11}/(91-x)$, and (iv) square root $c(x,\theta_1)=\theta_{11}\sqrt{x}$. The exponential and hyperbolic forms were estimated under the hypothesis that costs are a convex function of mileage, as opposed to the square root form which implies a concave cost function. I included the polynomial form, which can be concave, convex, or both, in order to check that my results were not artifacts of restrictive a priori assumptions about functional form. The disadvantage of the polynomial form is that collinearity among the terms can lead to imprecise estimates of the coefficients $(\theta_{11},\theta_{12},\theta_{13})$. Notice that none of the specifications for c include a constant term. This is due to the fact that the absolute level of c is not identified since subtracting a constant from the utility function (5.1) will not affect the choice probabilities. Clearly, the most we can hope to identify is the value of the *change* in operating costs as a function of mileage, so I normalize by setting $c(0,\theta_1)=0$.

I assume that the unobservable state variables $\{\varepsilon_t(0),\varepsilon_t(1)\}$ obey an i.i.d. bivariate extreme value process, with mean normalized to $(0,0)$ and variance normalized to $(\pi^2/6, \pi^2/6)$. $\varepsilon_t(0)$ should be interpreted as an unobserved component of maintenance and operating costs for the bus in period t. A large negative value for $\varepsilon_t(0)$ could be interpreted as an unobserved component failure which sends the bus into the shop for repair, whereas a large positive value could be interpreted as a bus driver's report that the bus is operating smoothly. $\varepsilon_t(1)$ should be interpreted as an unobserved component of cost associated with replacing an old bus engine with a newly rebuilt engine. A large negative value for $\varepsilon_t(1)$ could indicate that all available service bays in the company shop are occupied, or alternatively, that there are no available rebuilt engines at time t. A large positive value for $\varepsilon_t(1)$ could indicate empty service bays and surplus inventories of rebuilt engines.[11] Neither the location nor the scale of these observed costs are identifiable without additional information, the reason for my arbitrary normalizations of the mean and variance. Later I will use data on the cost of replacement engines given in Table III to identify the scale of unobserved costs $\{\varepsilon_t(0), \varepsilon_t(1)\}$.

The estimation procedure consists of three stages corresponding to each of the likelihood functions ℓ^1, ℓ^2, and ℓ^f, where ℓ^f is the full likelihood function given in (4.15), and ℓ^1 and ℓ^2 are "partial likelihood" functions given by

$$\ell^1(x_1, \ldots, x_T, i_1, \ldots, i_T \mid x_0, i_0, \theta) = \prod_{t=1}^{T} p(x_t \mid x_{t-1}, i_{t-1}, \theta_3), \tag{5.3}$$

$$\ell^2(x_1, \ldots, x_T, i_1, \ldots, i_T \mid \theta) = \prod_{t=1}^{T} P(i_t \mid x_t, \theta). \tag{5.4}$$

The first stage is to estimate the parameters θ_3 of the transition probability $p(x_{t+1} \mid x_t, i_t, \theta_3)$ using the likelihood function ℓ^1. This stage does not require computation of the fixed point EV_θ, and reduces to a standard parametric maximum likelihood problem. One can easily show (using the "principle of conditionality", Cox and Hinkley (1974)) that the resulting partial likelihood estimator is consistent and asymptotically normally distributed. Given our discretization of the state space, this transition probability is fully specified by two parameters $(\theta_{30}, \theta_{31})$ where $\theta_{3j} = \Pr\{x_{t+1} = x_t + j \mid x_t, i_t = 0\}$, $j = 0, 1$.

The results of the stage 1 estimation of θ_3 are presented in Tables V and VI . Table V includes a likelihood ratio test of the hypothesis that the mileage process $p(x_{t+1} \mid x_t, i_t, \theta_3)$ is the same for each bus within a given bus group. As can be seen from the bottom row of Table V, there is no evidence against this hypothesis.

[11] Note that I have implicitly assumed that the stochastic processes $\{x_t^j, \varepsilon_t^j\}$ are independently distributed across different buses, j. In a perceptive comment, a referee noted that this assumption may not be valid if Zurcher is attempting to optimize the use of his service bays. A lack of available service bays may cause Zurcher to simultaneously delay servicing several buses in need of new engines. This induces correlation across j in the stochastic processes for $\{x_t^j, \varepsilon_t^j\}$. While I think this is a useful insight, I think its impact is minor relative to other sources of specification error, particularly relative to assumption (CI). To properly handle the referee's problem, I would need to formulate a more complicated model of joint maintenance operations, including optimal scheduling of buses to service bays. Given my limited data set, this more ambitious model is beyond the scope of this paper.

TABLE V

WITHIN GROUP ESTIMATES OF MILEAGE PROCESS
WITHIN GROUP HETEROGENEITY TESTS

(Standard errors in parentheses)

	Group 1 1983 Grumman	Group 2 1981 Chance	Group 3 1979 GMC	Group 4 1975 GMC	Group 5 1974 GMC (8V)	Group 6 1974 GMC (6V)	Group 7 1972 GMC (8V)	Group 8 1972 GMC (6V)
θ_{31}	.197	.391	.307	.392	.489	.618	.600	.722
	(.021)	(.035)	(.008)	(.007)	(.013)	(.014)	(.010)	(.009)
θ_{32}	.789	.599	.683	.595	.507	.382	.397	.278
	(.021)	(.035)	(.008)	(.007)	(.013)	(.014)	(.010)	(.009)
θ_{33}	.014	.010	.010	.013	.005	.000	.003	.000
	(.006)	(.007)	(.002)	(.002)	(.002)	(0)	(.001)	(0)
Restricted Log, Likelihood	−203.99	−138.57	−2219.58	−3140.57	−1079.18	−831.05	−1550.32	−1330.35
Unrestricted Log Likelihood	−187.71	−136.77	−2167.04	−3094.38	−1068.45	−826.32	−1523.49	−1317.69
Likelihood ratio test statistic	32.56	3.62	105.08	92.39	21.46	9.46	53.67	25.31
Degrees of Freedom	42	9	141	108	33	18	51	34
Marginal Significance Level	.852	.935	.990	.858	.939	.948	.372	.859

TABLE VI

BETWEEN GROUP ESTIMATES OF MILEAGE PROCESS
BETWEEN GROUP HETEROGENEITY TESTS

(Standard errors in parentheses)

	1, 2, 3	1, 2, 3, 4	4, 5	6, 7	6, 7, 8	5, 6, 7, 8	Full Sample
θ_{31}	.301	.348	.417	.607	.652	.618	.475
	(.007)	(.005)	(.006)	(.008)	(.006)	(.006)	(.004)
θ_{32}	.688	.639	.572	.392	.347	.380	.517
	(.007)	(.005)	(.007)	(.008)	(.006)	(.006)	(.004)
θ_{33}	.011	.012	.011	.002	.001	.002	.007
	(.002)	(.001)	(.001)	(.001)	(.004)	(.001)	(.000)
Restricted Log Likelihood	−2575.98	−5755.00	−4243.73	−2384.50	−3757.76	−4904.41	−11,237.68
Unrestricted Log Likelihood	−2491.51	−5585.89	−4162.83	−2349.81	−3668.50	−4735.95	−10,321.84
Likelihood ratio test statistic	168.93	338.21	161.80	69.39	180.52	336.93	1,831.67
Degrees of Freedom	198	309	144	81	135	171	483
Marginal Significance Level	.934	.121	.147	.818	.005	1.5E−17	7·7E−10

Table VI shows the extent to which buses in different groups can be pooled. One can see that although bus groups 1, 2, and 3, and possibly 6 and 7, appear homogeneous, further aggregation of bus groups appears to be contra-indicated by the data. On the basis of these results I decided to pool groups 1, 2, and 3 and estimate group 4 separately.

The maintained hypothesis that bus mileage follows a regenerative random walk is examined in Table VII. Under the random walk hypothesis, the coefficients β_1 and β_2 in the regression

$$m_{it} = \beta_0 + \beta_1 m_{it-1} + y_{it}\beta_2 + e_{it} \tag{5.5}$$

should converge to zero, where $m_{it} = x_t^i - x_{t-1}^i$ is the mileage travelled in month t by bus i and y_{it} are other explanatory variables. However, if there are unobserved bus-specific differences in monthly mileage, then $e_{it} = \alpha_i + u_{it}$ and it is well known that OLS estimates of β_1 will be upward biased. Consistent estimates of β_1 and β_2 can be obtained from the *fixed-effect regression*

$$m_{it} - \bar{m}_i = \beta_1(m_{it-1} - \bar{m}_i) + (y_{it} - \bar{y}_i)\beta_2 + e_{it} - \bar{e}_i, \tag{5.6}$$

where $\bar{m}_i$, $\bar{y}_i$, and $\bar{e}_t$ are the time averages of m_{it}, y_{it}, and e_{it}. The significance of the coefficient β_1 for lagged mileage in the fixed effect regressions in Table VII is inconsistent with the random walk hypothesis, suggesting a higher order Markov process for bus mileage. However, when I performed similar fixed effect logit estimations using the discretized data, lagged mileage was insignificantly different from zero at the 1 per cent level. This discrepancy is likely due to the loss of information inherent in discretizing the underlying continuous mileage data. Given that I am estimating the structural model using the discretized data, I

TABLE VII

FIXED EFFECTS REGRESSION RESULTS DEPENDENT VARIABLE: MONTHLY MILEAGE (LESS BUS-SPECIFIC MEAN MILEAGE)

(Sample: Bus Groups 1-4)

Variable		Estimate	Standard Error	t-Statistic	Marginal Significance Level
December		−135.26	119.13	−1.13	0.256
January		203.79	119.22	1.71	0.087
February		−216.08	119.27	−1.81	0.070
March		−167.23	119.56	−1.40	0.162
April		−12.00	119.18	−0.10	0.920
May		−111.28	123.09	−0.90	0.364
June		−185.79	127.39	−1.46	0.145
July		12.77	127.07	0.10	0.920
August		103.18	121.50	0.85	0.393
September		−104.67	120.55	−0.87	0.383
October		−8.42	120.55	−0.07	0.944
Time		0.6755	2.52	0.26	0.792
Post 1979 Dummy		308.04	102.67	3.00	0.003
Odometer		−0.00168	0.00093	−1.83	0.067
Mileage $(t-2)$	0	.17949	0.02664	6.74	0.000
Mileage $(t-1)$	0	.41807	0.02724	15.37	0.000

decided to proceed despite the negative regression results for the underlying "raw data."

Using the estimates of θ_3 from the likelihood function ℓ^1 as initial consistent starting values, in stage 2 I estimated the remaining structural parameters (β, θ_1, RC) using the partial likelihood function ℓ^2 given in equation (5.4). Maximization of this likelihood function requires internal calculation of the fixed point EV_θ at each evaluation of the likelihood according to the nested fixed point algorithm outlined in Section 4. The final stage 3 estimation used the initial consistent estimates of θ computed in stages 1 and 2 in order to produce efficient maximum likelihood estimates using the full likelihood function ℓ^f. This estimator also yields a consistent estimator of the asymptotic covariance matrix for $\hat{\theta}$. Note that the estimated covariance matrix for the stage 2 estimates is not guaranteed to be consistent due to the use of the estimated values of θ_3 instead the true value θ_3^*. However, after computing the fully efficient estimates using l^f, I found that the estimated covariance matrix was almost perfectly block diagonal, and the parameter estimates and standard errors produced using ℓ^2 were nearly identical to the fully efficient estimates. Nevertheless, I present the efficient stage 3 estimates below.

The stage two estimation results from the nested fixed point algorithm for the partial likelihood function ℓ^2 are presented in Table VIII. I estimated the structural model $P(i|x, \theta)$ on various subsamples of the data set and for various parametric specifications of the cost function $c(x, \theta_1)$, yielding a number of alternative models which are summarized in Table VIII. In order to test for possible heterogeneity biases, I estimated separate models for the new buses (groups 1, 2, 3) and the older 1975 buses (group 4) and compared the results to the pooled model (groups 1, 2, 3, 4). By comparing the log-likelihood value of the pooled model (restricted log-likelihood) to the sum of the log-likelihoods of groups 1, 2, 3, and 4 separately (unrestricted log-likelihood), I could calculate likelihood ratio tests of the hypothesis of parameter homogeneity between groups. I also estimated a variety of alternative functional forms for the cost function $c(x, \theta_1)$ in order to insure that my conclusions were not artifacts of restrictive a priori choices of functional form. A completely nonparametric estimation was performed which allowed $c(x, \theta_1)$ to be any function. This is essentially equivalent to estimating a 90-dimensional coefficient vector $\theta_1 = (\theta_{1,1}, \ldots, \theta_{1,90})$ where $\theta_{1,x} = c(x, \theta_1)$, $x = 1, \ldots, 90$. These estimates yield a nonparametric estimate of the hazard function $P(1|x, \theta)$ (equal to the sample average of replacements in each mileage category x), which in turn produces the maximum attainable value for the log-likelihood function. Each of the parametric models can be regarded as restricted versions of the nonparametric model. For example, the cubic model (with 3 free parameters) can be regarded as the nonparametric model conjoined with 87 linear restrictions on the coefficients $(\theta_{1,1}, \ldots. \theta_{1,90})$. By comparing the log-likelihood values of a particular parametric model to the log-likelihood value of the nonparametric model, I could perform a likelihood ratio specification test of my a priori choice of functional form. As you can see from Table VIII, even with a sample size of 8,156 observations this likelihood ratio or "Kullback-Leibler"

TABLE VIII

SUMMARY OF SPECIFICATION SEARCH[a]

	Bus Group		
Cost Function	1, 2, 3	4	1, 2, 3, 4
Cubic	Model 1	Model 9	Model 17
$c(x, \theta_1) = \theta_{11}x + \theta_{12}x^2 + \theta_{13}x^3$	−131.063	−162.885	−296.515
	−131.177	−162.988	−296.411
quadratic	Model 2	Model 10	Model 18
$c(x, \theta_1) = \theta_{11}x + \theta_{12}x^2$	−131.326	−163.402	−297.939
	−131.534	−163.771	−299.328
linear	Model 3	Model 11	Model 19
$c(x, \theta_1) = \theta_{11}x$	−132.389	−163.584	−300.250
	−134.747	−165.458	−306.641
square root	Model 4	Model 12	Model 20
$c(x, \theta_1) = \theta_{11}\sqrt{x}$	−132.104	−163.395	−299.314
	−133.472	−164.143	−302.703
power	Model 5[b]	Model 13[b]	Model 21[b]
$c(x, \theta_1) = \theta_{11}x^{\theta_{12}}$	N.C.	N.C.	N.C.
	N.C.	N.C.	N.C.
hyperbolic	Model 6	Model 14	Model 22
$c(x, \theta_1) = \theta_{11}/(91 - x)$	−133.408	−165.423	−305.605
	−138.894	−174.023	−325.700
mixed	Model 7	Model 15	Model 23
$c(x, \theta_1) = \theta_{11}/(91 - x) + \theta_{12}\sqrt{x}$	−131.418	−163.375	−298.866
	−131.612	−164.048	−301.064
nonparametric	Model 8	Model 16	Model 24
$c(x, \theta_1)$ any function	−110.832	−138.556	−261.641
	−110.832	−138.556	−261.641

[a] First entry in each box is (partial) log likelihood value ℓ^2 in equation (5.2)) at $\beta = .9999$. Second entry is partial log likelihood value at $\beta = 0$.

[b] No convergence. Optimization algorithm attempted to drive $\theta_{12} \to 0$ and $\theta_{11} \to +\infty$.

specification test cannot reject any of the particular parametric functional forms which I tried. As a result, I adopted more intuitive criteria in order to select a "best fit" model from the array of alternative functional forms. My decision was a compromise between the objectives of (i) choosing the functional form with the highest likelihood value, (ii) choosing a functional form which is parsimonious, yet consistent with my priors and other nonquantitative information about the bus replacement problem. These criteria lead me to choose the linear and square root functional forms as the "best fit" specifications.

Tables IX and X present the structural parameter estimates computed by maximizing the full likelihood function ℓ^f using the nested fixed point algorithm. In Table IX I present structural estimates for the unknown parameters (RC, θ_{11}) of the linear specification for two alternative discount factors, $\beta = 0$ and $\beta = .9999$. The estimation results for $\beta = 0$ can be interpreted as a "myopic model" of bus engine replacement, under which a replacement occurs only when current operating costs $c(x_t, \theta_1)$ exceed the current cost of replacement $RC + c(0, \theta_1)$. The

TABLE IX

STRUCTURAL ESTIMATES FOR COST FUNCTION $c(x, \theta_1) = .001\theta_{11}x$

FIXED POINT DIMENSION = 90

(Standard errors in parentheses)

Parameter		Data Sample			Heterogeneity Test	
Discount Factor	Estimates/ Log-Likelihood	Groups 1, 2, 3 3864 Observations	Group 4 4292 Observations	Groups 1, 2, 3, 4 8156 Observations	LR Statistic ($df = 4$)	Marginal Significance Level
$\beta = .9999$	RC	11.7270 (2.602)	10.0750 (1.582)	9.7558 (1.227)	85.46	1.2E − 17
	θ_{11}	4.8259 (1.792)	2.2930 (0.639)	2.6275 (0.618)		
	θ_{30}	.3010 (.0074)	.3919 (.0075)	.3489 (.0052)		
	θ_{31}	.6884 (.0075)	.5953 (.0075)	.6394 (.0053)		
	LL	−2708.366	−3304.155	−6055.250		
$\beta = 0$	RC	8.2985 (1.0417)	7.6358 (0.7197)	7.3055 (0.5067)	89.73	1.5E − 18
	θ_{11}	109.9031 (26.163)	71.5133 (13.778)	70.2769 (10.750)		
	θ_{30}	.3010 (.0074)	.3919 (.0075)	.3488 (.0052)		
	θ_{31}	.6884 (.0075)	.5953 (.0075)	.6394 (.0053)		
	LL	−2710.746	−3306.028	−6061.641		
Myopia test:	LR Statistic ($df = 1$)	4.760	3.746	12.782		
$\beta = 0$ vs. $\beta = .9999$	Marginal Significance Level	0.0292	0.0529	0.0035		

TABLE X

STRUCTURAL ESTIMATES FOR COST FUNCTION $c(x, \theta_1) = .001\theta_{11}x$

FIXED POINT DIMENSION = 175

(Standard errors in parentheses)

Parameter		Data Sample			Heterogeneity Test	
Discount Factor	Estimates Log-Likelihood	Groups 1, 2, 3 3864 Observations	Group 4 4292 Observations	Groups 1, 2, 3, 4 8156 Observations	LR Statistic (df = 6)	Marginal Significance Level
$\beta = .9999$	RC	11.7257 (2.597)	10.896 (1.581)	9.7687 (1.226)	237.53	1.89E − 48
	θ_{11}	2.4569 (.9122)	1.1732 (0.327)	1.3428 (0.315)		
	θ_{30}	.0937 (.0047)	.1191 (.0050)	.1071 (.0034)		
	θ_{31}	.4475 (.0080)	.5762 (.0075)	.5152 (.0055)		
	θ_{32}	.4459 (.0080)	.2868 (.0069)	.3621 (.0053)		
	θ_{33}	.0127 (.0018)	.0158 (.0019)	.0143 (.0013)		
	LL	−3993.991	−4495.135	−8607.889		
$\beta = 0$	RC	8.2969 (1.0477)	7.6423 (.7204)	7.3113 (0.5073)	241.78	2.34E − 49
	θ_{11}	56.1656 (13.4205)	36.6692 (7.0675)	36.0175 (5.5145)		
	θ_{30}	.0937 (.0047)	.1191 (.0050)	.1070 (.0034)		
	θ_{31}	.4475 (.0080)	.5762 (.0075)	.5152 (.0055)		
	θ_{32}	.4459 (.0080)	.2868 (.0069)	.3622 (.0053)		
	θ_{33}	.0127 (.0018)	.0158 (.0019)	.0143 (.0143)		
	LL	−3996.353	−4496.997	−8614.238		
Myopia tests:	LR Statistic ($df = 1$)	4.724	3.724	12.698		
$\beta = 0$ vs. $\beta = .9999$	Marginal Significance Level	0.0297	0.0536	.00037		

estimation results for $\beta = .9999$ (a discount factor which corresponds to a very low annual real interest rate of .1 per cent) can be interpreted as a "dynamic model" of bus engine replacement which recognizes that replacing a bus engine is an investment which not only reduces current costs, but future costs as well. The "myopia test" on the bottom two rows of Table IX shows that the data reject the hypothesis that Harold Zurcher behaves as a myopic decisionmaker: the dynamic model with $\beta = .9999$ produces a statistically significant improvement in the ability of the model to fit the data. Although the data clearly reject the myopic model, I was not able to precisely estimate the discount factor β. Changing β to .98 or .999999 produced negligible changes in the likelihood function and parameter estimates of (RC, θ_{11}). The reason for this insensitivity is that β is highly collinear with the replacement cost parameter RC: both parameters induce similar effects on replacement behavior. For example, raising RC tends to postpone engine replacement, an effect which can also be achieved by lowering the discount factor β. Thus, if I treated β as a free parameter, the estimated information matrix was nearly singular, causing difficulties for the maximization algorithm. I did note a systematic tendency for the estimated value of β to be driven to 1. This curious behavior may be an artifact of computer round-off errors, or it could indicate a deeper result. By Abel's Theorem (also known as the final value theorem for Z-transforms (Howard (1971)), we have $\lim_{\beta\to 1} (1-\beta)\sum_{t=0}^{\infty} \beta u_t = \lim_{T\to\infty} (1/T)\sum_{t=0}^{T} u_t$ (for a formal proof of this result in the context of stochastic dynamic programming models, see Bhattacharya and Majumdar (1986)). This suggests that if Harold Zurcher is actually minimizing *long run average costs*, an estimation algorithm based on discounted costs would use Abel's theorem and attempt to drive β to 1. This might be what's happening here.[12]

The "heterogeneity test" in the last two columns of Table IX shows that the data reject the hypothesis that the structural coefficients (RC, θ_{11}) are the same for bus groups 1, 2, 3, and 4. The data show that Zurcher perceives the new GMC model 203 buses to have both higher engine replacement costs and a faster rate of increase in maintenance costs as a function of accumulated mileage. Using the replacement cost data from Table III, I can actually identify the scale of the coefficients (RC, θ_{11}). For groups 1, 2, 3 the average observed replacement cost was \$9499. Computing the ratio of the actual to estimated replacement cost we obtain a scale estimate of $\sigma = \$9499/11.7257$. Multiplying this scaling constant times θ_{11} I obtain a dollar estimate for θ_{11} for groups 1, 2, 3 of \$3.75. Thus, the estimates imply that Zurcher perceives average *monthly* maintenance costs to increase \$3.75 for every 5,000 accumulated miles on the bus. Thus, the expected maintenance costs for a bus with 300,000 miles are \$225.00 per month higher than for a bus with a newly replaced engine. In comparison, monthly maintenance

[12] The identification of β depends on a priori specification of the utility function u. Actually β is *nonparametrically unidentified*: in the absence of a priori knowledge of the form of u it is impossible to infer β. This can be seen in Table VIII where the difference in the log-likelihoods for $\beta = 0$ vs. $\beta = .9999$ disappears as I generalize the specification of the cost function, c. While this theoretical result might appear disturbing at first, on reflection it is clear we often do have substantial a priori information about β itself. In the case of Zurcher, we know that β must be "large" because $\beta = 0$ implies an implausibly large rate of increase in monthly operating costs. See Figure 2.

costs for buses in group 4 are estimated to increase only \$1.70 for every 5000 accumulated miles on the bus. These results appear to resolve the puzzle raised in Section 2. The reason that bus engines are replaced earlier on the newer 1979 GMC buses despite their 25 per cent higher replacement cost seems to be due to Zurcher's perception that monthly maintenance costs for the new buses increase more than twice as fast as a function of mileage.

At this point it is reasonable to ask: how sensitive are the inferences of this model with respect to (a) choice of cost function, and (b) choice of grid size for the discretization of bus mileage? Table X, which presents estimation results for model 11 with a fixed point dimension of 175, gives us some insight into the latter question. By dividing mileage into nearly twice the number of cells (of length 2,571 as opposed to 5,000) we obtain a multinomial distribution for monthly mileage which now depends on 4 parameters: $\theta_{3j}=\Pr\{x_{t+1}=x_t+j\,|\,x_t, i_t=0\}$, $j=0,1,2,3$. At first sight, Table X seems to show significant changes in the parameter estimates with a significant *deterioration* in the value of the log-likelihood function. However on closer inspection we see that both choices of grid size fit the data nearly identically. The decrease in the log-likelihood function is due to the fact that the finer grid size produces more observations in low probability cells (corresponding to parameters θ_{30} and θ_{34}) which have low log-likelihood values. Notice also that while the estimates of the cost function parameter θ_{11} change significantly, the estimates of RC are nearly identical using either grid size. Furthermore, the cost function parameter θ_{11} behaves exactly as we would expect due to a halving of the grid size: it is cut *almost exactly in half.* This produces estimated value and hazard functions which are nearly identical for either choice of grid size. I ran plots of these functions for the 175-dimensional case and the plots were visually identical to the plots for the 90-dimensional case presented in Figures 2 and 3. I also ran a model with a 45-dimensional fixed point and as expected the coefficient estimates of RC were nearly unchanged but the estimates of θ_{11} were nearly double the estimates in the 90 dimensional case. Notice also that the "myopia test" statistics are nearly identical to the values in the 90-dimensional case. Only the heterogeneity test statistics change significantly. This is simply an indication of the increased information content obtained by finer discretization of the mileage distribution. Nearly the entire increase in the heterogeneity test statistics can be ascribed to the increased ability to discriminate among mileage distributions using a finer discretization of the mileage variable. Thus, I conclude that my inferences are basically unaffected by the choice of discretization. The parameter estimates may change significantly, but only in such a way as to maintain a constant estimate of the value and hazard functions which are basically invariant to the choice of grid size.

I now turn to an analysis of the sensitivity of my results with respect to the choice of cost function, *c*. The estimation results for the square root form of the cost function turned out to be nearly identical to the linear case. The change in functional form yields slightly higher likelihood values, but does not otherwise alter any of the basic qualitative results found in the linear case. Figure 2 displays

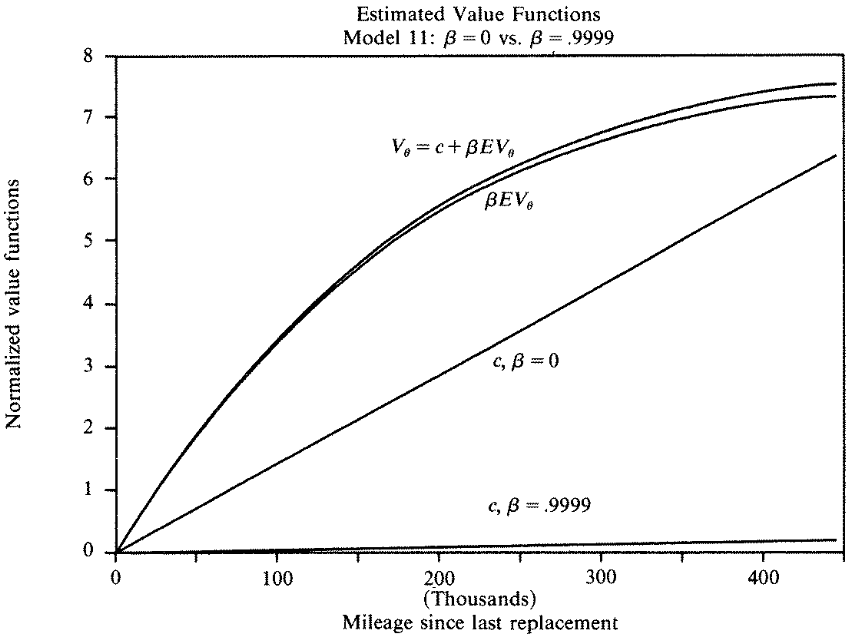

FIGURE 2

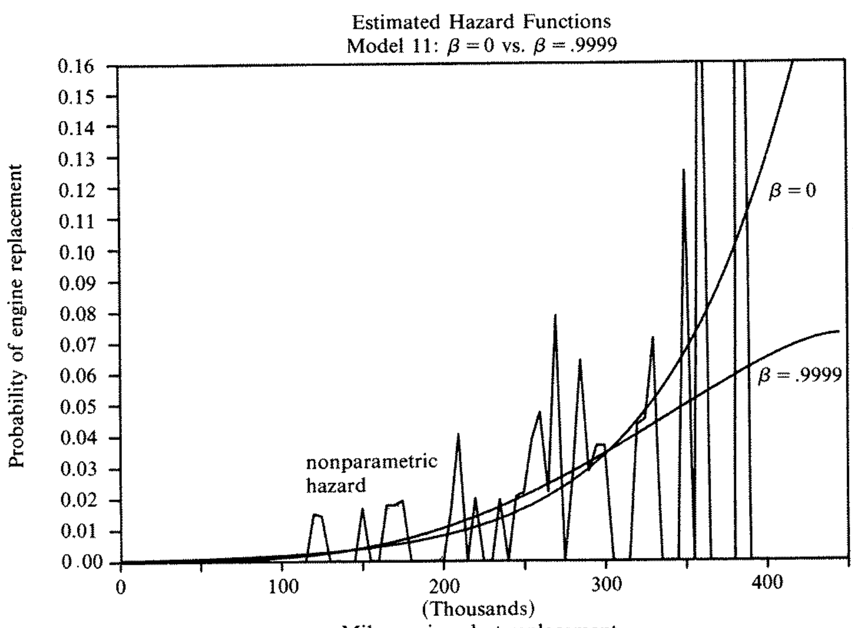

FIGURE 3

the estimated value function for the linear case, model 11 (note that the value function is shown in terms of the original unscaled coefficient estimates). Figure 3 displays the estimated hazard function for model 11, including the nonparametric and myopic ($\beta = 0$) hazard functions for comparison. We can see that the linear specification leads to a gently rising hazard function that appears to flatten out at a hazard rate of about 7 per cent at 450,000 miles. These estimates stand in marked contrast to the myopic model which implies a rapidly rising hazard function, with a hazard rate of over 20 per cent at 450,000 miles. It is unwise to use the nonparametric hazard estimate to try to decide whether or not the tail behavior of the dynamic model is more realistic than the tail behavior of the myopic model. Almost all of the observations are concentrated in bus mileages less than 100,000 and in fact we have very few observations for mileages beyond 300,000. As a result, the upper tail of the nonparametric hazard is estimated very erratically, leading ultimately to hazard rate estimates of 0 or 1 depending upon whether a single bus in a high mileage cell did or did not experience a replacement. This erratic "Dirac" behavior of the nonparametric hazard makes it unwise to try to infer anything about the precise nature of the tail behavior of the true underlying hazard function. Although the problem can be alleviated somewhat by choosing wider "windows" over which the nonparametric hazard is calculated, the basic problem is due to lack of observations in the upper tail and can only be addressed by increasing the size of the sample.

The lack of observations is reflected in the estimated value and hazard functions for the cubic and quadratic specifications. A positive estimated coefficient θ_{13} on the x^3 term in the cubic model leads to a sharply rising hazard function beyond 300,000 miles. A negative estimated coefficient θ_{12} for the quadratic model leads to the opposite behavior, leading to a hazard rate which actually *decreases* after 350,000 miles. The wide divergence in the tail behavior of these two specifications was not accompanied by a significant change in the value of the log-likelihood function. Although the hazard function is precisely estimated until about 300,000 miles, the tail is essentially an artifact of the particular functional form chosen for $c(x_t, \theta_1)$. My prior belief that the hazard function should never decrease leads me to reject the quadratic specification, and conversations with Harold Zurcher lead me to reject the cubic model with its sharply rising hazard function. When asked to choose the hazard function which best represents his engine replacement behavior, Zurcher chose the hazards derived from the linear and square root specifications which flatten out at about 7 or 8 per cent after 350,000 miles. According to Zurcher, monthly maintenance costs increase very slowly as a function of accumulated mileage. If the mechanical reliability of a bus deteriorates only very gradually with accumulated mileage, then it makes sense that the hazard would flatten out instead of abruptly increasing after 400,000 miles as it does in the myopic and cubic models. Remember that the alternative to not replacing a bus engine is to replace individual components at time of failure. Eventually such a "replace on failure" strategy yields bus engines with a significant fraction of new components, even though some components may have significant accumulated mileage. Thus, even though a given bus may have gone 400,000 miles since last engine replacement, the cumulative maintenance on the bus significantly reduces the chance that it would suddenly "fall apart." These considerations ultimately lead me to reject the cubic and quadratic specifications and to choose the linear and square root forms as my "best fit" specifications.

Although I have examined the sensitivity of my results with respect to choice of cost function and grid size, it is very difficult to assess the impact of the crucial Assumption (CI) used to produce a computationally tractable model. Recall that (CI) implies that ε_{t+1} is independent of ε_t given x_t. Thus, lagged $\{x_{t-j}, i_{t-j}\}$ $j \geqslant 1$ do not "cause" i_t conditional on the current observed state variable x_t. This suggests the following specification test of Assumption (CI): include the lagged control variable i_{t-1} as an explanatory variable in the choice model (4.13). If we let α be the coefficient of i_{t-1}, then under the null hypothesis (CI), the maximum likelihood estimate of α should converge to zero with probability 1. Under the alternative that (CI) does not hold, ε_t and ε_{t-1} will not be independent given x_t. Thus, in this case we would expect that the lagged control variable $i_{t-1} = f(x_{t-1}, \varepsilon_{t-1}, \theta)$ will be correlated with the current unobserved state variable ε_t and hence, the estimated value of α will converge to a nonzero value. Table XI presents a Lagrange multiplier test of the hypothesis that $\alpha = 0$.

We can see from Table XI that for group 4 there is no strong evidence that (CI) is violated, while for groups 1, 2, and 3 and the combined groups 1-4 there is strong evidence that (CI) does not hold. The reason for rejection in the latter cases may be due to the presence of "fixed-effects" heterogeneity which induces serial correlation in the error terms. This suggests that by separating the buses into more homogeneous subgroups (such as group 4), we can minimize violations of (CI).

I conclude with Figures 4 and 5 which display the confidence bands for the estimated value and hazard functions for model 11. Figure 4 shows a uniform 95 per cent confidence band and a "one standard deviation band" about the estimated value function $V_{\hat{\theta}}$, the latter which was derived by computing the standard deviation of $V_{\hat{\theta}}(x)$ at each point x. Interestingly, this "one standard deviation" band about the Banach-valued random element $V_{\hat{\theta}}$ contains the true

TABLE XI

LAGRANGE MULTIPLIER SPECIFICATION TESTS OF INDEPENDENCE ASSUMPTION (CI)[a]

COST FUNCTION $c(x, \theta_1) = .001\theta_{11}x$

FIXED POINT DIMENSION = 90

Statistic	Discount Factor	Groups 1, 2, 3	Group 4	Groups 1, 2, 3, 4
LM Statistic	$\beta = .9999$	8.154	2.047[b]	21.425[c]
	$\beta = 0$	27.086	33.174	60.250
Marginal Significance Level	$\beta = .9999$	0.0043	0.1526	3.68E−6
	$\beta = 0$	1.96E−7	8.44E−9	1.54E−9

[a] Hypothesis test of $\alpha = 0$, where α is coefficient of lagged control variable $i_{t-1} = f(x_{t-1}, \varepsilon_{t-1}, \theta)$ in choice probability formula (4.13).
[b] Corresponding Wald and Likelihood Ratio test statistics are 2.073 and 1.267, respectively.
[c] Corresponding Likelihood Ratio statistic is 17.416.

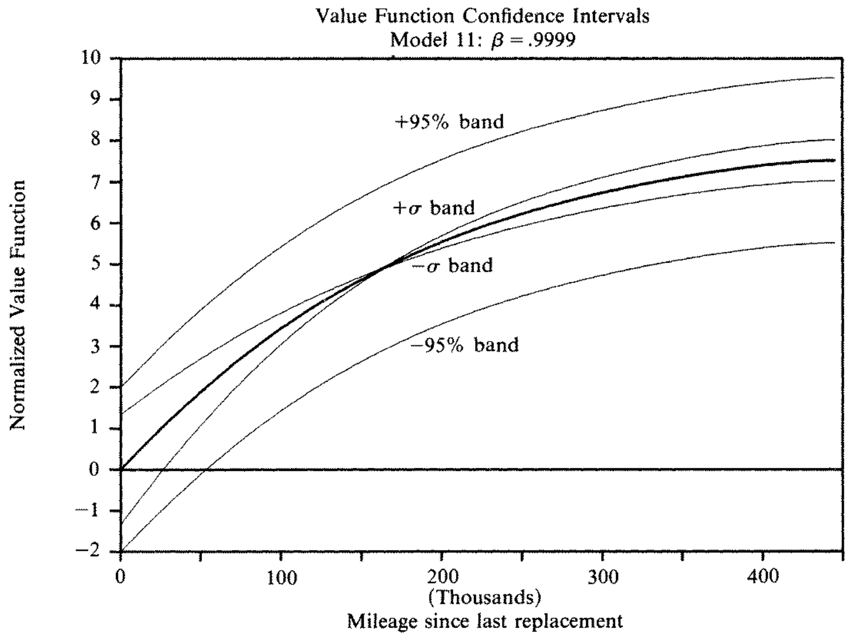

FIGURE 4

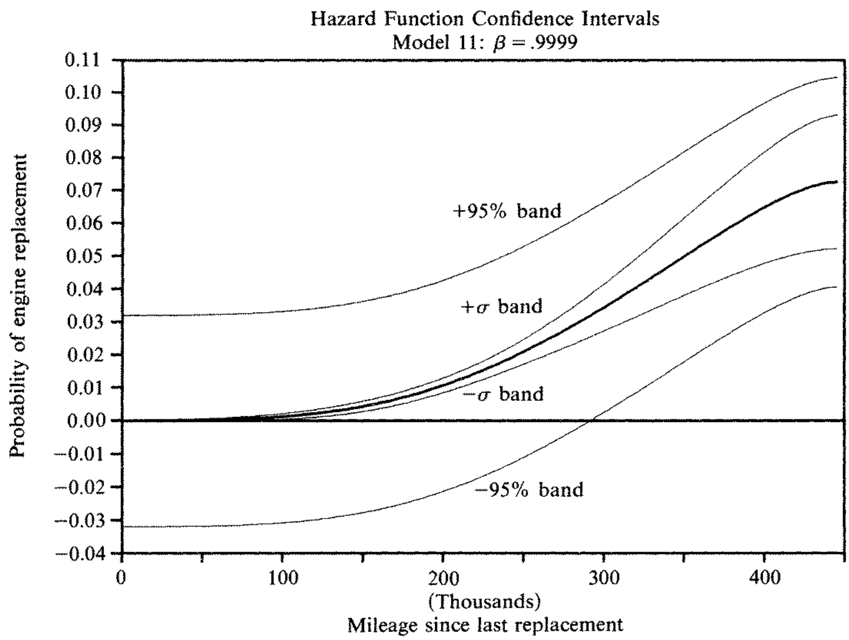

FIGURE 5

value function with probability 25.5 per cent as opposed to the 68.36 per cent probability for a standard univariate one standard deviation band. Similarly, the one standard deviation band about the estimated hazard function in Figure 5 contains the true hazard function with probability 25.5 per cent. The discrepancy is explained by the fact that the standard deviation band is based on the univariate distribution of $P(1|\bar{x}, \hat{\theta})$, and $V_{\hat{\theta}}(\bar{x})$ at a particular point $\bar{x}$, which has no necessary connection to the distribution of P and V as elements of the Banach space B (for details on the derivations of the infinite dimensional asymptotic distributions of p and v, see Rust (1988a)). Notice how the one-standard deviation band diverges in the tail of the hazard function. This is yet one more indication of the lack of high mileage observations which prevents accurate inference of the tail behavior of the hazard function.

The foregoing empirical results lead to two main conclusions: (i) the nested fixed point algorithm can be a practical, efficient, and numerically stable method for estimating certain structural models lacking closed-form solutions, (ii) the data are by and large consistent with my simple regenerative optimal stopping model of bus engine replacement. Despite the simplicity of the model, it leads to a wealth of interesting behavioral implications. In particular, the model can be used to perform a wide wide variety of "policy experiments" which forecast how changes in various structural parameters such as β, RC, and θ_3 affect the timing and frequency of bus engine investment. In Section 6 I show how this is done by deriving a demand curve for bus engine replacement.

6. CALCULATING THE IMPLIED DEMAND FOR REPLACEMENT INVESTMENT

I conclude by demonstrating the bottom-up approach to demand for replacement investment. Conceptually, the approach is quite simple. The replacement demand for a specific capital good is simply the sum of the replacement demands generated by individual decision makers. Multiplying the total replacement demand by the replacement cost RC of each capital good, I obtain a common unit of measurement, dollars, which allows me to sum over heterogeneous capital goods to obtain aggregate replacement investment.

Thus, my problem reduces to computing replacement demand for specific capital goods and specific decision makers. In the case of Harold Zurcher, annual demand for bus engines is a random function $\tilde{d}(RC)$ given by the sum

$$\tilde{d}(RC) = \sum_{t=1}^{12} \sum_{i=1}^{M} \tilde{i}_t^m \tag{6.1}$$

where each $\tilde{i}_t^m$ is a realization of the regenerative process $\{i_t^m, x_t^m\}$. Given an initial distribution $\pi_m(x_0^m, i_0^m)$ for the initial states of each bus m, I can compute the probability distribution of the random function $\tilde{d}(RC)$ using the controlled transition density $P(i_t|x_t, \theta)p(x_t|x_{t-1}, i_{t-1}, \theta_3)$ by integrating out the unnecessary state variables x_t. Then, by varying bus engine replacement costs RC, I can trace out how the entire probability distribution for replacement investment varies as a function of replacement costs.

To simplify my presentation, I will focus on calculating the expected replacement demand function $d(RC) = E\{\tilde{d}(RC)\}$, which I expect to be a nicely behaved, downward sloping function of RC. Suppose that the initial distribution π is the long run stationary (or *equilibrium*) distribution of the controlled process $\{i_t, x_t\}$. π is given by the unique solution to the functional equation

$$\pi(x, i) = \int_y \int_j P(i \mid x, \theta) p(x \mid y, j, \theta_3) \pi(dy, dj). \tag{6.2}$$

From (6.2) you can see that the equilibrium distribution π is an implicit function of the structural parameters θ, which I emphasize by the notation π_θ. Under the additional hypothesis that the regenerative processes $\{i_t^m, x_t^m\}$ and $\{i_t^k, x_t^k\}$ are independent if $m \neq k$, I obtain the following simple formula for $d(RC)$:

$$d(RC) = 12M \int_0^\infty \pi_\theta(dx, 1). \tag{6.3}$$

Thus, the problem further reduces to computing the equilibrium distribution π_θ. Figure 6 presents the equilibrium distribution for model 11 in the form of the conditional densities of π_θ, $\pi_\theta(x \mid 1)$ and $\pi_\theta(x \mid 0)$. Using these densities, the predicted mean mileage at replacement is estimated to be 287,892 which is within half a standard deviation of the actual value of 257,336 in Table IIa. The predicted

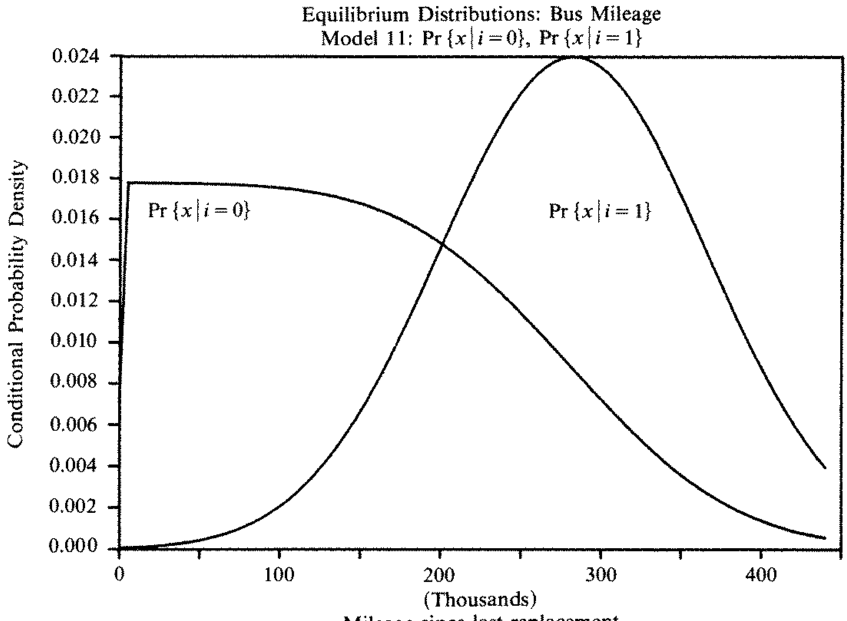

FIGURE 6

value of mean mileage given that replacement hasn't yet occurred is 159,305 which is also within half a standard deviation of the actual value of 134,862. Thus, use of a stationary distribution to compute replacement demand does not appear to be greatly at odds with the data.

By parametrically varying replacement costs, I can trace out the equilibrium distribution π_θ as a function of RC. In particular, using formula (6.3) I can compute the expected demand curve for replacement investment. Figure 7 presents the expected demand function $d(RC)$ for model 11 for a fleet containing a single bus, $M = 1$. For comparison, I also present the implied demand curve for the static model with $\beta = 0$. We can see significant differences in the predictions of the two models. As one might expect, the demand curve for the myopic model is much more sensitive to the cost of replacement bus engines, overpredicting demand at low prices, underpredicting demand at high prices. Notice, however, that the maximum likelihood procedure insures that both models generate the same predictions at the actual replacement cost of \$4343.

Figure 7 summarizes the value of the "bottom-up" approach to replacement investment. Since engine replacement costs have not varied much in the past, estimating replacement demand by a "reduced-form" approach which, for example, regresses engine replacements on replacement costs, is incapable of producing reliable estimates of the replacement demand function. In terms of Figure 7, all the data would be clustered in a small ball about the intersection of the two demand curves: obviously many different demand functions would

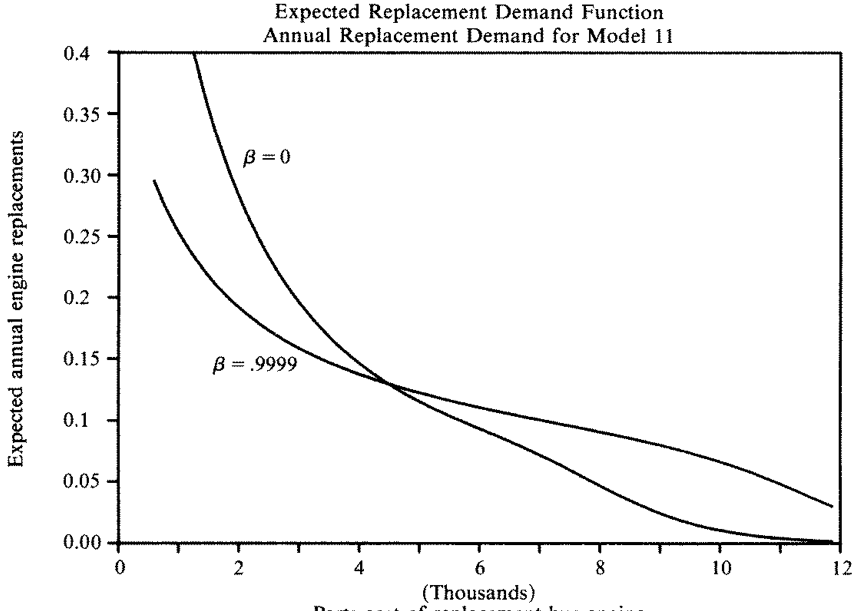

FIGURE 7

appear to fit the data equally well. The structural approach, on the other hand, efficiently concentrates additional information contained in the sequences $\{i_t^m, x_t^m\}$ into estimates of a small number of primitive parameters. Despite the relatively small number of such parameters, we obtain a rich behavioral model that can be used to answer a wide range of "what if?" policy questions.[13]

Department of Economics, University of Wisconsin—Madison, Madison, Wisconsin 53706, U.S.A.

Manuscript received December, 1985; final revision received October, 1986.

REFERENCES

BASAWA, I. V., AND B. L. S. PRAKASA RAO (1980): *Statistical Inference for Stochastic Processes.* New York: Academic Press.

BERNDT, E., J. HAUSMAN, B. HALL, AND R. HALL (1976): "Estimation and Inference in Nonlinear Structural Models," *Annals of Economic and Social Measurement*, 3, 653-665.

BERTSEKAS, D. (1976): *Dynamic Programming and Stochastic Control.* New York: Academic Press.

BHATTACHARYA, S., AND M. MAJUMDAR (1986): "Dynamic Programming for Discounted and Long-Run Average Rewards", manuscript, Cornell University.

BILLINGSLEY, P. (1961): *Statistical Inference for Markov Processes.* Chicago: University of Chicago Press.

BLACKWELL, D. (1968): "Discounted Dynamic Programming," *Annals of Mathematical Statistics*, 36, 226-235.

CHAMBERLAIN, G. (1980): "Analysis of Covariance with Qualitative Data," *Review of Economic Studies*, 67, 225-238.

COX, D. R. (1975): "Partial Likelihood," *Biometrika*, 62, 269-276.

COX, D. R., AND D. V. HINKLEY (1974): *Theoretical Statistics.* London: Chapman and Hall.

COX, D. R., AND H. D. MILLER (1965): *The Theory of Stochastic Processes.* London: Chapman and Hall.

FELDSTEIN, M. S., AND M. ROTHSCHILD (1974): "Towards an Economic Theory of Replacement Investment," *Econometrica*, 42, 393-423.

GIHMAN, I. I., AND A. V. SKOROHOD (1980): *Controlled Stochastic Processes.* New York: Springer-Verlag.

GOTZ, G. A., AND J. J. MCCALL (1986): "Estimating the Stay/Leave Decision in a Sequential Decisionmaking Model," manuscript, The Rand Corporation.

GRENANDER, U. (1980): *Abstract Inference.* New York: Wiley.

HECKMAN, J. J. (1981): "Statistical Models for Discrete Panel Data," in *Structural Analysis of Discrete Data with Econometric Applications*, edited by C. Manski and D. McFadden. Cambridge: M.I.T. Press.

HOWARD, R. A. (1971): *Dynamic Probabilistic Systems Volume I: Markov Models.* New York: Wiley.

JORGENSON, D. (1973): "The Economic Theory of Replacement and Depreciation," in *Econometrics and Economic Theory*, ed. by W. Sellykaerts. New York: Macmillan, 189-221.

JORGENSON, D. W., J. J. MCCALL, AND R. RADNER (1967): *Optimal Replacement Policy.* Chicago: Rand McNally and Company.

MCFADDEN, D. (1973): "Conditional Logit Analysis of Qualitative Choice Behavior" in *Frontiers of Econometrics*, ed. by P. Zarembka. New York: Academic Press.

——— (1981): "Econometric Models of Probabilistic Choice," in *Structural Analysis of Discrete Data*, ed. by C. Manski and D. McFadden. Cambridge: M.I.T. Press.

MILLER, R. (1984): "Job Matching and Occupational Choice," *Journal of Political Economy*, 92, 1086-1120.

[13] A diskette containing the complete data set used in this paper and a documented version of the nested fixed point algorithm (written in the *Gauss* matrix programming language) are available from the author for a fee which covers production and mailing costs.

PAKES, A. (1986): "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, 54, 755–784.

PECK, S. C. (1974): "Alternative Investment Models for Firms in the Electric Utilities Industry," *Bell Journal of Economics*, Autumn, 420–458.

PIERSKALLA, W. P., AND J. A. VOELKER (1976): "A Survey of Maintenance Models: The Control and Surveillance of Deteriorating Systems," *Naval Research Logistics Quarterly*, 23, 353–388.

RAO, C. R. (1973): *Linear Statistical Inference*. New York: Wiley.

RUST, J. (1985a): "Stationary Equilibrium in a Market for Durable Assets," *Econometrica*, 53, 783–805.

——— (1985b): "Nested Fixed Point Algorithm Documentation Manual," Software for IBM-PC, University of Wisconsin—Madison.

——— (1985c): "Equilibrium Holding Distributions in Durable Asset Markets," *Transportation Research*, 19, 331–346.

——— (1986): "When Is It Optimal to Kill Off the Market for Used Durable Goods?," *Econometrica*, 54, 65–86.

——— (1987): "Maximum Likelihood Estimation of Discrete Control Processes," forthcoming, *SIAM Journal on Control and Optimization.*

——— (1988a): "The Asymptotic Distribution of Randomly Parameterized Elements of Separable Banach Spaces," manuscript.

——— (1988b): "An Inverse Estimation Algorithm for Discrete Control Processes," manuscript.

SARGENT, T. J. (1981): "Interpreting Economic Time Series," *Journal of Political Economy*, 89, 213–248.

SHERIF, Y. S., AND M. L. SMITH (1981): "Optimal Maintenance Models for Systems Subject to Failure—A Review," *Naval Research Logistics Quarterly*, 32, 47–74.

WICKSELL, K. (1934): *Lectures on Political Economy*, Volume I. London: George Routledge and Sons.

WOLPIN, K. I. (1984): "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality," *Journal of Political Economy*, 92, 852–874.