Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry/Exit Examples)[1].

by

Ariel Pakes, Michael Ostrovsky, and Steve Berry,
(Harvard University, Harvard University and Yale University).

(First draft, June 2002, this March 2004.)

[1]We thank John Askers, Gary Chamberlain, Jeff Dominitz, Boyan Jovanovic, Jack Porter and the participants at the SITE workshop or structural econometrics for helpful remarks and the NSF for financial support. We hope to add a few more Monte Carlo examples to this version before submitting this paper.

Simple Estimators for the Parameters of Discrete Dynamic Games.
Ariel Pakes, Michael Ostrovsky, and Steve Berry.

This paper develops estimation strategies that follow from the structure of discrete dynamic games. For clarity of exposition we present all our results in the context of one example; a dynamic game of entry and exit. In addition to its importance to industrial organization, the entry/exit example illustrates rather well just why we need these estimation strategies and the major problems that arise in developing them.

In particular though the costs of entry and the selloff values (or costs) associated with exit are key determinants of the dynamics of market adjustments to policy and environmental changes, data on these sunk costs are virtually nonexistent. The decision of a firm on whether to exit is determinned by whether its continuation value is greater than its selloff value, and the latter is often associated with factors as hard to measure as "goodwill", the value of the firm's building and equipment in its "second best" alternative employment, and/or clean-up costs. The potential entrants sunk costs can be largely determinned by the time and effort required to formulate the idea to be marketed, or by an individual entreprenneur's cost in accessing startup capital and/or the requisite permissions from a local administration. As a result we have to infer the extent of sunk costs from other variables whose behavior depends on them.

The variable that seems most directly related to the costs of entry is entry itself. However to make use of the connection between actual entry and the costs of entry we need a framework which allows us to compute the value of enterring (similarly to make use of the relationship between selloff values and exit we need to be able to calculate the cost of continuing). Though such frameworks have been available for some time (see for e.g., Ericson and Pakes, 1995), their implications can not be used directly in estimation without encounterring substantial (in many cases insurmountable) computational problems.

As a consequence the models that have been used to analyze entry and exit decisions in the past have all been two-period, essentially static, models; see Bresnahan and Reiss (1987 and 1991), Berry (1992), and more recently, Mazzeo(2002) and Seim (2002). Partly because of this, the empirical work in these papers focused on providing a framework for characterizing differences in the number of active firms across a cross section of markets. More detailed analysis, say of policy issues or of the causes and/or effects of changes in the number of participants in a given industry over time, required more precise information on sunk costs.

The early frameworks for the analysis of entry and exit were also the first papers to consider the estimation issues that arise when the model used to structure the data does not generate a unique equilibrium. The uniqueness issue had been emphasized in the theoretical literature on entry, and both Bresnahan and Reiss (1991b) and Berry (1992 explicitly considered its impact on estimation in models where sunk costs could vary among agents. If anything, this issue became even more relevant once we allowed for the realism of continuation values which differed across agents. The original analysis here was in the form of differences in "locations" (see Michael Mazzeo (2002) and Katja Seim (2002)), and produced extensions which are crucial to the study of many retail and service sectors.

Our goal here is to make the transition from the two period setting to truly dynamic models of entry and exit. To do so we will provide a set of assumptions under which there is only one set of equilibrium policies that are consistent with the data generating process. We will then show how some simple ideas, ideas that can be viewed as extensions of Muth's (1961) original work, can be used to deliver estimators that are both easy to compute and grounded in what actually happened.

To determine whether a potential entrant (an incumbent) should enter (continue) we need the expected discounted value of future net cash flows should the firm enter (continue). The potential entrant will enter if this entry value is greater then the entry fee (similarly an incumbent will continue if the continuation value is greater then the sell-off value). Our measure of the entry values from a particular state is an average of the discounted value of net cash flows *actually earned* by entrants who did enter at that state. Similarly our measure of the continuation values from that state is the *actual* discounted value of net cash flows earned by incumbents who did continue from that state. These measures of entry and continuation values bear a transparent relationship to both the model, and to the data, and thus make it easy to engage in various forms of robustness analysis.

Once we have consistent estimates of entry and continuation values, the rest of the estimation problem is exceedingly simple. We obtain a consistent estimate of the probability of entry conditional on the parameters of the model as the probability of an entrant drawing an entry fee greater than the estimated entry value. Similarly the probability of an incumbent exiting is the probability of drawing a sell-off value greater then the continuation value. Given these probabilities there are a number of alternative estimation algorithms that produce (root-n) consistent and asymptotically normal estimators, and we consider differences in their computational and statistical properties in a later section of the paper.

Our use of an average of realized future values as an estimator for the expected discounted future values that decisions are based upon, turns our problem into a semi-parametric estimation problem. The first stage provides a nonparametric estimate of the entry and continuation values, and the second stage treats these nonparametric estimates as true values in a parametric estimation problem. We provide assumptions under which the first stage need only be done once. That is we do not need to compute a complicated fixed point or matrix inverse each time we evaluate the objective function at different values of the parameter vector. As a result the computational burden of our estimator is, if anything, *less than* that of the estimators for the simple static entry models.

The paper begins by making our major conceptual points in the context of the simplest entry/exit model, a model with one entry location and a fixed number of potential entrants in every period. We then show how to generalize to allow for multiple of entry locations and a random number of potential entrants. Once it becomes clear how to obtain our estimators, a number of modifications that lead to alternative estimators suggest themselves. The alternatives have different computational and distributional properties, and so are worth considering.

We provide Monte Carlo results on the performance of these alternatives in two examples (one with a single location, and the other with two locations). The Monte Carlo examples, when combined with an analytic discussion of why they occur, end up being extremely informative. Among the many alternatives we consider, there is only one, perhaps two, that should be considerred. In particular the *least* computationally burdensome of the alternative estimators performs remarkably well. This bodes well for applied work as it should be relatively easy to use this estimator even in complex environments.

Before proceeding, a word on related literature is in order. It is our ability to form nonparametric estimates of the entry and continuation values that enables us to circumvent the computational problems that have hindered the empirical analysis of dynamic models of markets. Hotz and Miller (1993) were the first to use semi-parametrics to simplify computational burdens in a dynamic problem. Their paper deals with a single agent problem, while a subsequent paper by Olley and Pakes (1996) uses semi-parametrics to a related end in a multiple agent dynamic game.

There are also a number of papers currently "in process" that present related results. Closest to our paper is a paper by Aguirreberia and Mira (2003) which introduces an estimator which under our assumptions would be one of the alternative estimators considered in our extensions[2]. Aguirreberia and Mira (2003) assume i.i.d. extreme value errors, an assumption which makes it difficult to generalize the model to multiple locations in a realistic way.

Papers by Berry and Pakes (2002) and Pakes, Porter and Wolfram (2003) use similar "Muthian" ideas to develop an estimator based on first order conditions for *continuous* controls and for (multi-unit) auctions, respectively. Bajari, Benkard, and Levin (2003) use a combination of semi-parametrics estimates of policy functions and simulation to provide estimators for models with both continuous and discrete controls, while Einav (2003) provides

[2]We discovered this paper after completing a draft of our paper, but it is clear that its initial draft was posted on the internet before ours. A less closely related paper on entry by and Pesendorfer and Schmidt-Dengler,2003, which appeared after our paper, provides an alternative semi-parametric estimator not considerred here.

assumptions which enable one to obtain estimates from the second stage of a timing game the first stage of which determines the order of entry. Hopefully these new tools will enable more detailed empirical analysis of the dynamics of markets.

1 A Simple Entry/Exit Model.

We begin with the model with only one entry location and the same number of potential entrants in each period. The generalization to multiple entry locations and a random number of potential entrants is a straightforward extension considered later.

Let n_t be the number of agents active at the beginning of each period, z_t be a vector of exogenous profit shifters which evolve as a finite state Markov process, and assume that there is a one-period profit function that is determinned by these variables, say $\pi(n, z; \theta)$, where θ is a parameter vector to be estimated.

An incumbent chooses to exit if current profits plus the discounted sellof value is greater than profits plus the discounted continuation value. So if ϕ is the selloff (or exit) value and $0 < \delta < 1$ is the discount rate, the "Bellman" equation for the value of an incumbent is

$$V(n, z; \phi, \theta) = max\left\{\pi(n, z; \theta) + \delta\phi, \pi(n, z; \theta) + \delta VC(n, z; \theta)\right\}, \tag{1}$$

where $VC(\cdot)$ is the continuation value. If the max is the first term inside the curly brackets, the incumbent exits.

If e is the number of entrants, x is the number of exitors (both of which are unknown at the time the incumbents decisions are made), and $P(\cdot)$ is notation for a probability distribution, then $VC(\cdot)$ is is just the expectation (over the possible numbers of exitors entrants and values of the profit shifters) of the next period's realization of the value function (of future $V(\cdot)$),or

$$VC(n, z; \theta) \equiv \tag{2}$$
$$\sum_{\phi', z', e, x} V(n + e - x, z', \phi'; \theta) P(d\phi'|\theta) P(de, dx|n, z, \chi = 1) P(dz'|z).$$

Note that to form this expecation we need to form the incumbent's perceptions of the likely number of entrants and exitors *conditional on the incumbent itself continuing*, perceptions that we write as the probability distribution

$$P(e, x|z, n, \chi = 1)$$

where $\chi = 1$ is notation for the incumbent continuing. We need these perceptions because the incumbent can not estimate his returns to continuing without an idea of how many other firm's will be active. It is the requirement that these perceptions be consistent with behavior that will generate our equilibrium conditions.

Analgously we assume that the entrant must commit to enterring one period before it earns any profit, so the value of entry is

$$VE(n,z;\theta) \equiv \qquad (3)$$
$$\sum_{e,x,z',\phi'} V(n+e-x,z',\phi')P(d\phi'|\theta)P(dx,de|n,z,\theta,\chi^e=1)P(dz'|z),$$

where

$$P(x,e|n,z,\chi^e=1)$$

provides the potential entrant's perceptions of the likely number of entrants and exitors *conditional on it entering*, or conditional on $\chi^e = 1$.

The potential entrant enters if

$$\delta VE(n,z;\theta) \geq \kappa$$

where κ is its sunk cost of entry[3].

We now list of our assumptions and then turn to a detailed explanation of their implications.

Assumption 1 *We will assume that entry and exit decisions are made simultaneously at the beginning of the period, and that*

1. *There are a fixed number of potential entrants in each period (denoted by $\mathcal{E}$), and the distribution over*
   - *the sunk costs of entry, say $F^{\kappa}(r|\beta)$, which has a lower bound of $\underline{\kappa} > 0$, and*
   - *the returns to exiting, say $F^{\phi}(\cdot|\beta)$, which are assumed nonegative,*

   are i.i.d. over time and across markets. Incumbents and entrants know these distribution and their own realizations, but do not know the realizations of their competitors (so there is assymetric information, as in Siem,2001).

[3]Note that we are not giving the potential entrant the possibility of waiting and then enterring in the next period with the same setup costs. To do so we would have to keep track of the value of a potential entrant, say $VPE(n,z,\kappa)$, which would be the maximum of $VE(n,z)-\kappa$ and the expected value of being a potential entrant with sunk cost κ in the next period. The analysis would then proceed much the same way as we do below, except that now we would have to keep track of the likely distribution of the potential entrant's κ's (it will depend on the value of entry in prior periods). How this is done depends on the assumption governing the evolution of a potential entrant's κ over time. Ours is the extreme case in which if a potential entrant does not enter when the entry possiblity appears, that possiblity disappears.

2. *Entrant's and incumbents perceptions of the probabilities of exit and entry by their competitors in period t depend only on (n_t, z_t) (the publically available information at that time).*

3. *The evolution of the profit shifters, z, is governed by the Markov chain $\mathcal{P}_z \equiv \{P(\cdot|z), \forall z \in Z = [0, 1, \ldots, \overline{z}]\}$, $lim_{n \to \infty} \pi(n, z) \leq 0$ for every $z \in Z$, and $\pi(\cdot)$ is bounded.*

These assumptions are *truly restrictive*, and we come back to a more detailed discussion of those restrictions directly after explaining their implications of the assumptions. However these assumptions (or simple generalizations of them that allow for multiple locations) are less restrictive than the assumptions used in any of the two-period models that have been taken to data to date. Moreover, as we will show below, these assumptions lead to an estimator of continuation and entry values that has a transparent relationship to objects in the data, and therefore has both a great deal of intutive appeal and is easy to work with.

This model is a special case of the model in Ericson and Pakes (1995) and so has a Markov perfect equilibrium, but there may be more than one of them (see Doraszelski and Satterwaite, 2003). Each equilibrium generates a finite state Markov chain in (n, z) couples: i.e. the distribution of possible (n, z)'s in the next period depends only on the current (n, z) (and not on either prior history, or time itself). Also the last asssumption insures that there will be an $\overline{n}$ such that, provided the current n is lower than it, we will never observe an $n > \overline{n}$. The market is simply not profitable enough to induce entry if there are $\overline{n}$ incumbents.

Indeed one can go a bit further. No matter the equilibrium every possible sequence of $\{(n_t, z_t)\}$ will eventually wander into a recurrent subset of the possible (n, z) couples, say $\mathcal{R}$, and once (n_t, z_t) is in the set $\mathcal{R}$ it will stay in it forever (Freedman, 1983). All states in $\mathcal{R}$ "communicate" with each other, and will eventually be visited many times. If the distribution of exit and entry fees have unbounded supports, and all z's communicate with each other (so $\mathcal{P}_z$ is ergodic), then there is exactly one recurrent class for each equilibrium, and it is of the form $\mathcal{R} = \{(n, z) : 0 \leq n \leq \overline{n}, z \in Z\}$[4].

It is important to note that though our assumptions do not guarantee a unique equilibrium, they do insure that *there is only one equilibrium that is consistent with a given data generating process.* As a result we will be able to use the data itself to "pick out" the equilibrium that is played, and at least for large enough samples, we will pick out the correct one. This is all we require to develop consistent estimators for the parameters of the model.

To see that the data can be used to pick out the equilibrium, note that: (i) the agents only condition their perceptions on the behavior of their competitors on the publically available

[4] "Communicate" here simply means that the probability of transiting from one state to another (in any number of periods) is positive. To prove this assertion it suffices to note that any (n, z) communicates with $(0, z)$ for some $z \in Z$ and z itself is ergodic. This implies that all points in any recurrent class communicate with each other and therefore are in the same recurrent class.

information (on (n, z)), and (ii) precisely the same information is available to the econometrician. Moreover in equilibrium the realized distribution of entrants and exitors from each state must be consistent with these perceived distributions (Star and Ho, 1969).

Now recall that the data will eventually wander into the recurrent subset of points, and once in that subset will visit each point in it repeatedly. As the sample gets large we obtain an empirical distribution of entrants and exitors from each (n, z), and by the law of large numbers that distribution will converge to the distribution which generated it (almost surely). As noted this must be the distribution the agents use to form their perceptions, so we have just identified the perceived distributions needed for agents to make their decisions.

Given those perceived distributions equations (1) and (2) generate a unique best response for each incumbent and potential entrant. This is just the familiar statement that reaction functions are generically unique, and can be proven using Blackwell's theorem for single agent dynamic programs. Since there is only one policy that is consistent with both the data and our equilibrium assumptions at each $(n, z) \in \mathcal{R}$, and once we are in the set $\mathcal{R}$ we stay their forever, there is a unique equilibrium for any subgame starting from any (n, z) couple in $\mathcal{R}$ (a set which can be identified from the data) [5].

We now come back to the limitations of our assumptions. Part 2 of Assumption 1 implies that there are no state variables that the agents condition their perceptions on but the econometrician does not observe. In this context we should note that $\pi(n, z)$ can represent *expected* profits conditional on the information available at the beginning of the period. Actual profits could have additional idiosyncratic and/or common components that are *not observed* by the econometrician; indeed nothing in this paper changes if expected profits have an unobserved component that was independent over time. However if this component were serially correlated, or if there were an i.i.d unobserved component that changed equilibrium perceptions (a sunspot, for example), then we would have to condition continuation and entry values on unobserved shocks, and the simple techniques introduced below to estimate those values can not do this conditioning. Related assumptions have been extensively used and discussed in the context of estimating single agent dynamic models (see Pakes, 1994, and the literature cited their).

One more point on Assumption 1.2. We show below that this assumption enables us to identify the equilibrium chosen in the past, thus "solving" the estimation problem generated

[5] There is a detail missing here. Though points in $\mathcal{R}$ can only communicate with other points in $\mathcal{R}$ if optimal policies are followed, there are some points, "boundary points" in the terminology of Pakes and McGuire (2001), that could communicate with points outside of $\mathcal{R}$ if feasible but inoptimal policies were followed. To fully analyze equilibria for subgames in $\mathcal{R}$, boundary points need to be treated separately (as we cannot compute see Pakes and McGuire, 2001). In our case the only decisions that involve boundary points are the decisions of entrants at the maximum n observed for any given z; thus we can easily isolate them, and not use them in the estimation algorithm.

by the possibility of multiple equilibria. It does not, however, enable us to use the results to evaluate policy changes or other changes in the environment. Once we change the environment a new equilibrium will be chosen, so predictions cannot be made without a more detailed analysis of the relationship between the equilibria chosen in the past and those that are likely to prevail in the future, a topic which, though important, is beond the scope of this paper.

1.1 Equilibrium Behavior.

We now characterize equilibrium behavior, beginning with that of incumbents, and then moving to that of potential entrants.

Since entry and exit decisions are simultaneous and incumbents (potential entrants) are identical up to the draw on exit (entry) fees, for an incumbent's behavior to be based on *equilibrium* perceptions it *must* perceive all competing incumbents to have the same probability of exit, that probability being the probability that the random draw on the exit fee is greater than the value of continuing. I.e. the perceptions in 1 that define continuation values must be formed as

$$P(de, dx|n, z, \chi = 1) = b^x(x, n-1|n, z, \theta)p(e|n, z, \chi = 1, \theta) \tag{4}$$

where for $r \geq x$

$$b^x(x, r|n, z, \theta) = \binom{r}{x} F^{\phi}\{VC(n, z, \theta)|\theta\}^{r-x} \left[1 - F^{\phi}\{VC(n, z, \theta)|\theta\}\right]^x$$

and

$$p(de|n, z, \chi = 1, \theta)$$

is consistent with the behavior of entrants.

Equilibrium requires that all potential entrants have the same probability of entering, that probability being the probability that the random draw on the entry fee is less than the value of entry. Consequently in *equilibrium* the perceptions required to calculate entry values satisfy

$$p(e, x|n, z, \chi^e = 1) = b^x(x, n|n, z)p(e|n, z, \chi^e = 1)$$

where $b^x(x, n|n, z)$ is defined as in equation 4, and

$$p(e|n, z, \chi^e = 1) = b^e(e-1, \mathcal{E} - 1|z, n), \tag{5}$$

where $\forall R \geq e$

$$b^e(e, R|n, z) \equiv \binom{R}{e} F^{\kappa}\{\delta VE(n, z, \theta)|\theta\}^e \left[1 - F^{\kappa}\{\delta VE(n, z, \theta)|\theta\}\right]^{R-e}.$$

Note that this implies that for incumbents

$$p(e|n, z, \chi = 1) \equiv b^e(e, \mathcal{E}|z, n).$$

2 Equilibrium Perceptions and Estimation.

In equilibrium the perceptions of potential entrants and incumbents must be consistent with what is actually observed. This fact leads directly to a number of alternative two step semiparametric estimators for the parameters of the model.

We begin with the simplest of the alternatives. Its first step computes simple averages of the the realized continuation (entry) values of all firms who did continue (enter) at alternative values of (n, z). Since agents expectations must be consistent with average realizations, the sample average of discounted future continuation (entry) values will converge to expected continuation (entry) value we are after. The second step of the estimation procedure treats these estimates of continuation (entry) values as the actual continuation (entry) values, and estimates the parameters of the distribution of selloff and entry values by fitting the model's predictions for the solutions to the binary choice problems facing incumbents (should they continue?) and potential entrants (should they enter?) to the actual data.

Conditional on our estimates of entry and continuation values both these binary choice problems have closed form expressions for the choice threshold ($VC(\cdot)$ and $VE(\cdot)$). Moreover, at least under convenient specification for the distribution of exit values (and regardless of the assumption on the distribution of entry values), our estimates of $VC(\cdot)$ and $VE(\cdot)$ are linear functions of variables that can be constructed directly from the data and held fixed for the entire estimation run. Thus though our estimator is a two step estimator, it *is not* a nested fixed point estimator (the data transformation which is required to obtain the estimates of $VC(\cdot)$ and $VE(\cdot)$ need not be redone every time we evaluate the objective function at a different value of the parameter vector). This makes the computational burden of the second step the same as that of standard binary choice models whose "cutoffs" are analytic functions of the parameters of interest. We illustrate with a pseudo maximum likelihood estimator.

2.1 Estimates of $VC(\cdot)$ and $VE(\cdot)$.

When we index function by t we will mean we are evaluating the function at $(n, z) = (n_t, z_t)$. An incumbent at t who continues to $t+1$ draws $\tilde{\phi}$ from $F^{\phi}(\cdot)$ and has a value of

$$\pi_{t+1} \ + \ \delta \, max \left[VC_{t+1}(\theta), \ \tilde{\phi} \right].$$

Taking the expectation of this expression conditional on (n_{t+1}, z_{t+1}) provides the continuation value of *firms who actually continue* from period t *conditional on publically available information available in period* $t+1$, say $V\tilde{C}R_t(\theta)$

$$V\tilde{C}R_t(\theta) = \pi_{t+1}(\theta) + \delta \int_{\phi} max\{\phi, VC_{t+1}(\theta)\} F(d\phi) \tag{6}$$

Let

$$\mu^x_{t+1} \equiv Pr\{\phi > VC_{t+1}(\theta)\}$$

be the exit probability, and initially assume that ϕ distributes exponentially so that

$$E[\phi|\phi > V_{t+1}(\theta)] = \sigma + V_{t+1}(\theta),$$

is the expected selloff value conditional on exiting (we show how to generalize on this assumption below). Substituting these expressions into (6)

$$\begin{aligned} V\tilde{C}R_t(\theta) &= \pi_{t+1}(\theta) + \delta[1-\mu^x_{t+1}]VC_{t+1}(\theta) + \delta\mu^x_{t+1}[\sigma + VC_{t+1}(\theta)] \\ &= \pi_{t+1}(\theta) + \delta VC_{t+1}(\theta) + \delta\sigma\mu^x_{t+1}. \end{aligned} \tag{7}$$

If we average $V\tilde{C}R_t(\theta)$ over observed sample paths emanating from a particular value of (n_t, z_t), say (n, z), we average using the observed distribution of entrants and exitors, which, from our equilibrium assumptions, is given by

$$b(x; n, z) \times b(e; n, z).$$

The incumbent computes continuation values conditions on itself continuing, so it averages with

$$b(x; n_t - 1, z_t) \times b(e; n_t, z_t).$$

So to obtain an unbiased estimate of the continuation values used by incumbents when they make their decisions we need to form

$$V\hat{C}R_t(\theta) = \frac{b(x; n_t - 1, z_t)}{b(x; n_t, z_t)} V\tilde{C}R_t(\theta) = [1-\mu^x_t]/[1-(x_t/n_t)]V\tilde{C}R_t(\theta) \equiv w^c_t \times V\tilde{C}R_t(\theta).$$

It follows that

$$E[V\hat{C}R_t(\theta)|n_t, z_t] = VC_t(\theta),$$

the unknown continuation value. So if we let

$$T(n, z) = \{t : (n_t, z_t) = (n, z)\}$$

be the set of periods with the same (n, z), and then form

$$\hat{\mathcal{VC}}(n,z) \equiv \frac{1}{\#T(n,z)} \sum_{t \in T(n,z)} V\hat{C}R_t(\theta),$$

we have

$$E[\hat{\mathcal{VC}}(n,z)] = VC(n,z).$$

Moreover the Markov property insures that the realizations of $V\hat{C}R_t(\theta)$ in $T(n, z)$ are mutually independent, so

$$Var[\hat{\mathcal{VC}}(n,z)] \to 0 \quad \text{as} \quad \#T(n,z) \to \infty.$$

That is for all points in $\mathcal{R}$ our estimate of the continuation value, $\hat{\mathcal{VC}}(\cdot)$, converges in mean sqaure to the true continuation value.

To complete the argument we need some matrix notation. Arrange; $\hat{\mathcal{VC}}(n, z; \theta)$ into the vector $\hat{\mathcal{VC}}(\theta)$, the average exit probabilities into the vector $\tilde{\mu}^x$, and the weighted (by the $\hat{w}_t^c$) transition probabilities into the matrix $\tilde{M}_c$. Then from above

$$\hat{\mathcal{VC}}(\theta) \equiv \tilde{M}_c\left[\pi(\theta) + \delta VC(\theta)\right] + \delta\sigma\tilde{M}_c\tilde{\mu}^x = \tilde{M}_c\left[\pi(\theta) + \delta\sigma\tilde{\mu}^x\right] + \delta\tilde{M}_c VC(\theta). \tag{8}$$

(8) gives us our estimate of $VC(\cdot)$ in terms of itself. Subsituting the estimate of $VC(\cdot)$ (the left hand side of 8) for the unknown $VC(\cdot)$ on its right hand side, and iterating we get

$$\hat{\mathcal{VC}}(\theta) = \tilde{M}_c\left[\pi(\theta) + \delta\sigma\tilde{p^x}\right] + \delta\tilde{M}_c^{\,2}\left[\pi(\theta) + \delta\sigma\mu^x\right] + \delta^2\tilde{M}_c^{\,2} VC(\theta) = \tag{9}$$

$$\ldots \quad = \sum_{\tau=0}^{\infty} \delta^\tau \tilde{M}_c^\tau[\pi(\theta) + \delta\sigma\tilde{\nu}^x] = [I - \delta\tilde{M}_c]^{-1}\tilde{M}_c[\pi(\theta) + \delta\sigma\tilde{\mu}^x] \to_P VC(\theta),$$

where $\to_P$ reads converges in probability[6].

Characteristics of the First Stage Estimates.

There are a couple of points to stress here. First our estimate of continuation values is just the discounted value of the returns of the incumbents who did continue (adjusted to account for the fact that the incumbent conditions on itself continuing). This is the sense in which our estimator has a transparent relationship to the objects of interest to us and as a result

[6] Convergence in probability follows from the continuous mapping theorem and the consistency of $\tilde{M}_c$ and $\tilde{\mu}_x$.

is likely to produce values which make empirical sense. That is we often expect the actual average of realized continuation values to be close to the continuation values used by the agents in making their decisions.

Second but equally important, note how easy it is to compute our estimates of continuation values, or $\hat{VC}(\theta)$. If δ is known (and we usually think that the prior information we have on δ is likely to swamp the information on δ available from estimating an entry model), then

$$\hat{VC}(\theta) = \tilde{A}\pi(\theta) + \sigma\tilde{a}$$

for $\tilde{A} = [I - \delta\tilde{M}_c]^{-1}\tilde{M}_c$ and $\tilde{a} = \delta[I - \delta\tilde{M}_c]^{-1}\tilde{\mu}_x$. Both $\tilde{A}$ and $\tilde{a}$ are independent of the parameter vector and can therefore be computed once at the beginning of the estimation routine and held in memory thereafter. So if profits were linear functions of θ, the first stage estimates of continuation values are also.

An analogous calculation produces consistent first stage estimates of entry values. The value of entry at t *conditional on information available at* $t+1$ is

$$V\breve{E}R_t(\theta) = \pi_{t+1}(\theta) + \delta VC_{t+1}(\theta) + \delta\sigma\mu_{t+1}^x. \tag{10}$$

The average of the $V\breve{E}R_t(\theta)$ observed in the data converges to the average of the entry values of the potential entrants who did enter. As noted, when deciding whether or not to enter a potential entrant computes entry values *conditional on itself actually entering*, and hence forms expectations using the distribution

$$b(x, n_t; n_t, z_t) \times b(e-1, \mathcal{E}-1; n_t, z_t)$$

Thus to obtain an unbiased estimate of the entry values actually used by potential entrants in deciding whether to enter we average estimates of the form

$$V\tilde{E}R_t(\theta) = \frac{b(e-1, \mathcal{E}-1; n_t, z_t)}{b(e, \mathcal{E}; n_t, z_t)} V\breve{E}R_t(\theta) = [e/\mathcal{E}]/\mu_e V\breve{E}R_t(\theta) \equiv w_t^e \times V\breve{E}R_t(\theta),$$

Making the appropriate substitutions we have, as our consistent estimate of VE

$$\hat{VE}(\theta) = \tilde{B}\pi(\theta) + \sigma\tilde{b}, \tag{11}$$

where

$$\tilde{B} \equiv \tilde{M}_e + \delta\tilde{M}_e\tilde{A}, \qquad \tilde{b} \equiv \delta\tilde{M}_e\tilde{a} + \delta\tilde{M}_e p^x,$$

and $\tilde{M}^e$ is the Markov transition matrix formed after weighting the observed transitions with w^e.

Note that the simplicity of the solution for $(\hat{VC}(\theta), \hat{VE}(\theta))$ did not depend *at all* on the distribution of entry values; a fact which will enable us to use realistic generalization to models with multiple entry locations (see below). On the other hand, those solutions do become somewhat more complex when the distribution of exit fees implies that $E[\phi|\phi > a]$ is not linear in a (as it is when $F^{\phi}(\cdot)$ is exponential or uniform). Though exponential distributions may be a good first approximation for the distribution of selloff values we would like to know what would be required were we to use a richer distribution. It is easy to show if we substitute any other form of $F^{\phi}(\cdot)$ into (**??**), follow the argument above and set the left hand side of (8) equal to $VC(\cdot)$, that equation becomes a contraction mapping. As a result it can be solved relatively quickly for a consistent estimate of $\hat{\mathcal{VC}}(\theta)$. That is though lifting the exponential assumption does move us to a nested fixed point algorithm, it produces a fixed point which is relatively easy to solve.

2.2 The Second Step.

Once we have $(\hat{VC}(\theta), \hat{VE}(\theta))$, a number of alternate estimation algorithms are available. To illustrate consider the esimator which maximizes the likelihood formed by treating $(\hat{VC}, \hat{VE})$ "as if" they were the true continuation and entry values conditional on θ. These "pseudo" maximum likelihood estimators are defined by

$$max_{\theta\in\Theta}\mathcal{L}(\theta) \equiv max_{\theta\in\Theta} \sum_t l(x_t, e_t|\theta),$$

where

$$l(x_t, e_t|\theta) = (n_t - x_t) log F^{\phi}\{\hat{VC}_t(\theta)|\theta\} + x_t log\left[1 - F^{\phi}\{\hat{VC}_t(\theta)|\theta\}\right]$$
$$+ e_t log F^{\kappa}\{\hat{VE}_t(\theta)|\theta\} + (\mathcal{E} - e_t)\left[1 - F^{\kappa}\{\hat{VE}_t(\theta)|\theta\}\right]^{e_t}.$$

We will provide a more detailed discussion of the properties of this estimator along with those of a more general class of method of moments estimators introduced below. For now all we want to note is that the computational complexity of the pseudo mle is comparable to that of estimators for the simplest static entry/exit models (and these models are neither dynamic nor do they allow for any heterogeneity). I.e. we need only analyze a binomial likelihood in an analytic function of the parameters of interest.

2.3 Generalizations of the Model.

We consider generalizations to the model that allow for multiple entry locations and a random number of potential entrants. Allowing for multiple locations changes the entry decision from a binomial to a multinomial decision; the mutually exclusive and exhaustive outcomes are enter in location 1, enter in location 2, ..., or not enter at all. Allowing for a random number of potential entrants changes the model for observations on entry from a standard multinomial model into a mixture of multinomials where we mix over the "size of the sample" for the multinomial draws. However all the other aspects of the two step estimation strategy remain in tact, so the reader who is not interested in the details can skip this subsection.

We detail a model with two locations and a random number of potential entrants (the extension to a finite number of entry locations is straightforward). To keep matters as simple as possible we maintain all of Assumption 1 (with obvious differences in notation to allow for two locations) except 1.2 (which deals with the sunk costs of entry and exit).

Assumption 2 *Instead of assumption 1.2 we assume*

- *the number of potential entrants in each period is an independent random draw from the distribution* $\{p(E|\theta)\}_{E=0}^{\mathcal{E}}$ *for a finite* $\mathcal{E}$*; and*
- *potential entrants can enter in only one of the two locations and have entry cost* (κ_1, κ_2) *in the first and second locations respectively, where the vector* (κ_1, κ_2) *is a draw from the distribution*
$$Pr\{\kappa_1 \leq r_1 \ \ and \ \ \kappa_2 \leq r_2\} \equiv F^{\kappa}(r_1, r_2|\theta)$$
which is independent over time and across agents, and
- *once in a particular location the entrant cannot switch locations, but can exit to receive an exit fee of* ϕ *which is an i.i.d. drawn from* $F_1^{\phi}(\cdot|\beta)$ *if the incumbent was in the first location and an i.i.d. draw from* $F_2^{\phi}(\cdot|\beta)$ *if the incumbent was in the second location.*

Note that κ_1 and κ_2 are allowed to be freely correlated. These are draws on the entry costs of the same agent in alternative locations. If, for example, the agent has an easier time in accessing finance capital, the agent will have a lower entry cost in all locations.

We begin with the incumbent's problem. Letting l index the different locations and making obvious notational changes, the Bellman equation for an incumbent in the two location model is

$$V_l(n_l, n_{-l}, z; \phi, \theta) = max\left\{\pi_l(n_l, n_{-l}, z; \theta) + \delta\phi, \pi_l(n_l, n_{-l}, z; \theta) + \delta VC_l(n_l, n_{-l}, z; \theta)\right\},$$

where

$$VC_l(n_l, n_{-l}, z; \theta) =$$
$$\sum_{z', e_l, e_{-l}, x_{-l}, x_l} \int_{\phi} V(n_l + e_l - x_l, n_{-l} + e_{-l} - x_{-l}, z', \phi) P(d\phi) p(e_l, e_{-l}, x_l, x_{-l}|n_l, n_{-l}, z, \chi_l = 1) p(z'|z),$$

and

$$p(e_l, e_{-l}, x_l, x_{-l}|n_l, n_{-l}, z, \chi_l = 1)$$

provides the type l incumbents perceived probability of $(e_l, e_{-l}, x_l, x_{-l})$ conditional on that incumbent continuing.

Just as in the model with a single location the incumbent views all its competitors in a particular location as identical. Consequently in equilibrium it perceives a distribution of exitors from each location which is binomial with the binomial probability determined by the fraction of draws on the exit fee that are larger than the location's continuation value. More formally in equilibrium

$$p(e_l, e_{-l}, x_l, x_{-l}|n_l, n_{-l}, z, \chi_l = 1) =$$
$$p(e_l, e_{-l}|n_l, n_{-l}, z, \chi_l = 1) b_l(x_l, n_l - 1|n_l, n_{-l}, z) b_{-l}(x_{-l}, n_{-l}|n_l, n_{-l}, z),$$

where, if we let a variable which is not subscripted equal the vector of its values in the two locations

$$b_l^x(x, r|n, z, \theta) \equiv \binom{r}{x} F^\phi \left\{VC_l(n_l, n_{-l}, z, \theta)|\ \theta\right\}^{r-x} \left[1 - F^\phi \left\{VC_l(n_l, n_{-l}, z, \theta)|\ \theta\right\}\right]^x,$$

an analogous definition holds for $b_{-l}(x_{-l}, n_{-l}|n, z)$, and the perceived entry probabilities, i.e. $p(e_l, e_{-l}|n, z, \chi_l = 1)$, must equal the equilibrium entry probabilities defined below.

Since incumbents become entrants at the beginning of the period after entry and have exit perceptions that are consistent with equilibrium behavior

$$VE_l(n_l, n_{-l}, z; \theta) =$$
$$\sum_{e,x,z'} \int_\phi V_l(n_l + e_l - x_l, n_{-l} + e_{-l} - x_{-l}, z', \phi) P(d\phi) b_l^x(x_l, n_l|n, z, \theta) b_{-l}^x(x_{-l}, n_{-l}|n, z, \theta) p(e|n, z, \chi_l^e = 1) p(z'|z),$$

where

$$p(e|n, z, \chi_l^e = 1)$$

provides the equilibrium probabilities of entry conditional on the potential entrant *entering in location l.*

The only behavioral difference in the more general model is that now a potential entrant will enter into location l if and only if it is a better alternative then *either* not entering at all, or entering into location $-l$, i.e. *iff*

$$\delta VE_l(n_l, n_{-l}, z, \theta) \geq \kappa_l \quad \textit{and} \quad \delta VE_l(n_l, n_{-l}, z, \theta) - \kappa_l \geq \delta VE_{-l}(n_{-l}, n_l, z, \theta) - \kappa_{-l}. \tag{12}$$

Recall that in equilibrium the distributions perceived by the agents must be equal to the objective distributions generated by their decision rules. We find this distribution by first conditioning on a particular number of potential entrants (on E) and then integrating out over the distribution of potential entrants given in Assumption 2.

To any potential entrant the remaining potential entrants draw from the same distribution of entry fees. Consequently the probability of (e_l, e_{-l}) entrants conditional on E is determined by the multinomial probabilities induced by the decision rule above. That is if we let

$$m_0 \equiv Pr\{\kappa_1 > VE_1(n_1, n_2, \cdot) \ \ and \ \ \kappa_2 > VE_2(n_2, n_1, \cdot)\}, \tag{13}$$

$$m_1 \equiv Pr\{\kappa_1 \leq VE_l(n_1, n_2, \cdot) \ \ and \ \ \kappa_2 > VE_2(n_2, n_1, \cdot) - VE_1(n_1, n_2, \cdot) + \kappa_1\}, \ \ and$$

$$m_2 \equiv Pr\{\kappa_2 \leq VE_2(n_2, n_1, \cdot) \ \ and \ \ \kappa_1 > VE_1(n_1, n_2, \cdot) - VE_2(n_2, n_1, \cdot) + \kappa_2\},$$

then (m_0, m_1, m_2) are the probabilities of a potential entrant not entering, entering in location 1, and entering in location 2, respectively. Consequently a potential entrant who conditions on $E-1$ other potential entrant and enters in location l will set

$$p(e_l, e_{-l}|n, z, \chi_l = 1, E) = m(e_l - 1, e_{-l}, E - 1; m_0, m_1, m_2)$$

where $m(r_1, r_2, r; m_0, m_1, m_2)$ is the multinomial probability of cell sizes $(r - r_1 - r_2, r_1, r_2)$ given cell probabilities of m_0, m_1, m_2 and a sample size (number of potential entrants) of r, i.e.

$$m(r_1, r_2, r; m_0, m_1, m_2) \equiv \frac{r!}{(r - r_1 - r_2)! r_1! r_2!} m_0^{r - r_1 - r_2} m_1^{r_1} m_2^{r_2},$$

provided $e_l + e_{-l} \geq E$ (otherwise $m(\cdot) = 0$).

The unconditional probability are obtained by integrating out with respect to the distribution of E or

$$p(e_l, e_{-l}|n, z, \chi_l = 1) = \sum_{E \geq (e_l + e_{-l})} m(e_l - 1, e_{-l}, E - 1; m_0, m_1, m_2) \frac{EP(E|\ \theta)}{\sum_{E \geq (e_l + e_{-l})} EP(E|\ \theta)}$$

Note that this implies that the incumbents perceived entry probabilities are given by

$$p(e_l, e_{-l}|n, z, \theta) = \sum_{E \geq (e_l + e_{-l})} m(e_l, e_{-l}, E; m_0, m_1, m_2) P(E|\ \theta).$$

The adjustments to the estimation procedure required for the generalized model are also straightforward. To obtain our first stage estimates of continuation and entry values, we consider the continuation value of those who do continue conditional on the information realized by $t+1$ from each t. We then on the set of periods that have a particular value of $(n_1, n_2, z) = (n_1, n_2, z)$, and take a weighted sum of period $t+1$ realizations with weights given by

$$w_t^{c,l} = \frac{[1 - x_t^l]}{[1 - \hat{p}_l^x(n_t, z_t)]} \quad and \quad w_t^{e,l} = \frac{e_t^l}{\hat{\hat{E}}^l(n_t, z_t)} \quad for \;\; l = 1, 2,$$

where $\hat{\hat{E}}^l \equiv \#T(n,z)^{-1} \sum_{t \in T(n,z)} e_t^l$. These sums converge to the entry and exit values used by the incumbents and potential entrants from the two locations. The matrix inversion formula is then used to simplify the calculation of the entry and continuation values implied by the average of the sample paths realized in our data.

As before our weights correct for the fact that the sample averages of the continuation and entry values from a particular (n_1, n_2, z) converges to the expectation of the continuation and entry values of those who continue (enter) from the given state. The continuation and entry values that determine continuation and entry behavior condition on a particular incumbent (potential entrant) continuing (entering). The appropriate weight for $w_t^{c,l}$ is derived just as it was in the model with one location, while the formula for the entry weights in developed in appendix 1.

Given the estimates of the continuation and entry values the pseudo likelihood is

$$\mathcal{L}(\theta) \equiv \sum_t l(x_t, e_t | \theta),$$

where $l(x_t, e_t|\theta)$ is proportional to

$$(n_{1,t} - x_{1,t}) log F_1^{\phi}\{\hat{VC}_{1,t}(\theta)|\theta\} + x_{1,t} log \left[1 - F_1^{\phi}\{\hat{VC}_{1,t}(\theta)|\theta\}\right]$$

$$(n_{2,t} - x_{2,t}) log F_2^{\phi}\{\hat{VC}_{2,t}(\theta)|\theta\} + x_{2,t} log \left[1 - F_2^{\phi}\{\hat{VC}_{2,t}(\theta)|\theta\}\right]$$

$$+ log \sum_{E \geq (e_1 + e_2)} m(e_l, e_2, E; \hat{m}_{0,t}(\theta), \hat{m}_{1,t}(\theta), \hat{m}_{2,t}(\theta)) P(E|\, \theta).$$

and $\hat{m}_{l,t}$ is the estimate of the multinomial probabilities obtained by placing the first stage estimates of the entry and continuation values into the formula in 13.

2.4 Limitations of the Framework.

This subsection considers a number of issues that users should be aware of.

Providing our model's assumptions are satisfied, our data could come from either a single long time series, or a panel of different markets (or industries) with initial conditions in the recurrent class (or restricted some other way to insure sufficient visits to a small number of states). In addition to requiring that there be repeated observations at a given (n, z), this assumes that conditional on (n, z) the process generating the different continuation values are both; (i) similar, and (ii) have independent realizations. That is any differences across time or markets must be adequately capturred by our $z's$, and conditional on a current value of z the realizations of future $z's$ must be independent. If, for example, we were working with panel data, and there was a national regulation which influenced the realizations of z's in all markets in a particular period, the average of realized continuation values across these markets in that period would not converge to the continuation values which determinned selloff decisions. Of course if we observed repeated changes in regulations, and the process generating those changes were ergodic, we could average over those changes....

Second we noted that there was no formal problem in generalizing the framework to allow for either multiple entry locations, or a larger number of states per location. Though in principal this is true, all of the estimators considerred in this paper are semiparametric estimators with variances for the parametric component of interest (of θ in our case) that depends on the variance of the first stage estimates of continuation and entry values[7]. For a given sized data set the larger the number of distinct states (locations times states per location) the fewer the number of observations per state and the larger the variance in the first stage estimates. Our estimator for the parametric component "averages" over these first stage estimators, and as a result converges at the rate given by the square root of sample size. However consistency requires the number of observations per distinct state to grow large, and as we raise the number of states for a given sized sample we may well incur increased small sample biases as well as larger variances. It is for this reason that we report on a reasonably extensive Monte Carlo study of just how these estimators perform below.

Finaly we stressed that given the matrix inversion formula in (8) the computational burden of obtaining estimates for our model is minimal. However there is the burden of obtaining the Markov transition matrix and computing its inverse; a burden which grows polynomially in the number of distinct states. As we will show in our Monte Carlo examples below, given the simplicity of the rest of the estimation procedure, this "setup" time can easily become the dominant computational burden, and for large enough problems may

[7] That is none of the estimators are "adaptive". This also implies that the variances that come out of standard minimization routines that do not take account of the variance in the first stage estimates will be wrong. To see that the estimators are not adaptive take the derivative of the objective function which is being maximized in the second step with respect to the first step estimates of continuation and entry values and note that the conditional expectation of that derivative, conditional on (n, z), is nonzero. We discuss estimation of the variance of the estimators below.

become a limiting factor.

3 Alternative Two-Step Estimators.

There are different components of the estimation algorithm described above that can be changed without changing the $\sqrt{n}$-consistency or asymptotic normality of the estimator[8]. The alternative estimators these changes generate will have both different distributions and different computational burdens. We begin by simply categorizing the changes that can be made. We then provide an informal theoretical discussion of the properties of the alternative estimators, and conclude (in the next section) by examining how the different estimators perform in two Monte Carlo examples[9].

Changes can be made in each of three components of the algorithm. We compare our original suggestion to at least one natural aternative for each of these components. An estimator can be obtained by combining either suggestion for each of the three components. All of these estimators will be examined in the Monte Carlo examples.

We could change each of the following components of the algorithm.

1. The objective function used in the second stage. In this context we consider
   (a) the pseudo likelihood function (as above), and
   (b) alternative method of moment estimators that minimizes a norm in the difference between the actual number of entrants/exits and the entrants/exits predicted by the model for different values of θ. We look at two such estimators, one which compares average entry and exit probabilities, and the other which compares state specific entry and exit probabilities.
2. The estimation of the transition probabilities between states. In this context we consider
   (a) using the empirical Markov matrix (as above), and

[8]Estimates of the standard errors from all techniques considered can be obtained from a parametric bootstrap. The consistent parameter estimates enable us to build consistent estimates of entry and continuation values, and these in turn make it easy to obtain our parametric bootstrap. That is we simply use the estimates, and their implications for the entry and continuation values, to build random samples of entry and exit sequences that have the same dimensions as the original sample. We then calculate the estimator for each of these samples, and obtain a consistent estimate of the variance-covariance matrix by computing the variance-covariance matrix of the estimates across the Monte Carlo samples.

[9]Throughout we ignore the problem of developing tests of our model, even though it clearly is possible to develop such tests.

(b) computing estimates of the entry and exit probabilities at each location at each (n, z), and then using the binomial (or multinomial) formula to generate the Markov matrix these probabilities imply. We call the estimates obtained in this way the "structural" transition matrix.

3. The computation of first stage continuation and entry values conditional on the estimated transition probabilities. In this context we consider

   (a) using the discounted sum of future profits given by the formula above, and

   (b) using a single agent dynamic programming nested fixed point algorithm; i.e. we substitute the profit functiona and estimates of the transition probabilities into the contraction mapping defining a single agent's value function (equation 1) and compute it for each different θ.

Finally note that given any one of these estimators we could always iterate on to a multi-stage estimator based on the structural transition probabilities in (2b). This would involve using the first iteration's estimates to recompute the "structural" transition probabilities in (2b). The new estimates of transition probabilities are then used to produce new first stage estimates of continuation and entry values, either using the matrix inversion in (3a) or the nested fixed point in (3b). Indeed the estimator in Aguirreberia and Mira (2003) noted in our introduction is a modifed version of the pseudo maximum likelihood estimator in (1a) that uses the structural transition matrices in (2b) and the nested fixed point to compute entry and continuation values (3b), and they favor a multi-step version of their estimator which iterates in this way. The modification used by Aguirreberia and Mira is that they require i.i.d. extreme value distributions for both the entry fees and the selloff values (as in Rust,1987); an assumption which implies that the entry cost for the same agent in different locations are independent and that exit fees are independent across locations[10].

3.1 Comments on the Alternative Estimators.

Alternative Objective Functions.

We already noted that the variance of all of our estimators will be determinned, in part, by the variance of the first step estimators of $VC(\cdot)$ and $VE(\cdot)$. As a result the pseudo maximum likelihood estimator will not be asymptotically efficient even under standard regularity conditions. In addition if there is a lower bound on the distribution of entry fees that

[10] It also implies that entry fees have full support, i.e. they can be any number between plus and minus infinity.

must be estimated (and depending on the profit function such a bound might be required to insure that the Markov chain generated by the model is finite, see assumption 1.2), one of the parameters to be estimated defines a point of support of the likelihood (which invalidates the regularity conditions required for the efficiency of the maximum likelihood estimators for multinomials even if the continuation and entry values were computed without error).

Thus the usual limiting arguments in favor of maximum likelihood do not apply here. Moreover there are two arguments which should lead us to worry about the small sample properties of the pseudo maximum likelihood estimator. First since the pseudo likelihood's probabilities are not conditioned on the true entry and exit values, but rather just on estimates of them, events can occur which the pseudo likelihood assigns zero probability to (for any $\theta \in \Theta$), even though the same events must have nonzero probability in the true likelihood (else they could not occur). If only one such event occurs the pseudo likelihood will be undefined (the log-likelihood is negative infinity for all θ), and the estimation procedure will break down.

Second the pseudo maximum likelihood estimator is obtained by setting the derivative of the log-likelihood to zero. That derivative requires an estimate of one over the probability of entry (exit). When those probabilities are close to zero (and they often are in entry/exit data sets), the derivative of the likelihood will be quite sensitive to errors in the first stage estimates of entry and exit values, accentuating the impact of the first stage esimation errors on the parameters of interest.

Neither of these two problems will occur with alternative method of moments estimators provided we do not choose a norm which weights with the inverse of the probability of entry and/or exit.

Empirical vs Structural Transition Matrices.

The empirical transition matrices are computed as described above. We obtain the structural transition matrix as follows. First compute maximum likelihood estimates of the the entry and exit probabilities for each state. In the simple case of a fixed number of potential entrants these are given by the average fraction enterring and exiting at that state of by

$$\hat{g}_x(z,n) = \frac{\sum_{t \in T(z,n)} x_t}{\sum_{t \in T(z,n)} n_t} \text{ and } \hat{g}_e(z,n) = \frac{\sum_{t \in T(z,n)} e_t}{\sum_{t \in T(z,n)} \mathcal{E}}.$$

Then use the the binomial formulae

$$\hat{p}(x|n,z,\chi=1) = \hat{b}_x(x, n-1|n,z) \equiv \binom{n-1}{x} \hat{g}_x(n,z)^x \left[1 - \hat{g}_x(z,n)\right]^{n-1-x},$$

and

$$\hat{p}(e|n,z) = \hat{b}_e(e, \mathcal{E}|n,z) \equiv \binom{\mathcal{E}}{e} \hat{g}_e(z,n)^e \left[1 - \hat{g}_e(z,n)\right]^{\mathcal{E}-e},$$

together with the Markov process generating z, to compute the probabilities of ($n' = n - x + e, z'$) given (n, z).

In finite samples use of the structural transition matrix is likely to generate two problems. First the transitions estimated from the binomial formula will take us to states not observed in the data. To compute the first stage estimates of $VC(\cdot)$ and $VE(\cdot)$ we will then have to impute (since there is no data to estimate) the entry and exit rates from those states, and the imputation errors will affect our estimators[11]. Second, to go from the binomial probabilities to the probabilities needed for the transitions from n to n' requires the computaion of a convolution of probability distributions, and then the inversion of a larger Markov matrix (or integration over a larger number of future states if we use the fixed point in 3b). This increases the computational burden of the estimator, and the increase is likely to be larger the larger the number of state variables in the model.

Discounted Sample Paths vs. Nested Fixed Points for $\hat{VC}(\cdot)$ and $\hat{VE}(\cdot)$.

The sample path calculation is explained above. The nested fixed point algorithm finds its estimates of continuation values by computing the contraction mapping

$$\hat{V}(n, z; \phi, \theta) = max \left\{ \pi(n, z; \theta) + \delta\phi, \pi(n, z; \theta) + \delta \hat{VC}(n, z; \theta) \right\}$$

where

$$\hat{VC}(n, z; \theta) \equiv \sum_{z', e, x} \int \hat{V}(n + e - x, z', \phi) P(d\phi|\theta) \hat{p}(x, e|n, z, \chi = 1) \hat{p}(z'|z),$$

and $\hat{p}(\cdot)$ refers to estimated probabilities.

At least in cases where the matrix inversion formula in (8) is available, the nested fixed point calculation is likely to substantially increase the computational burden of the estimator. When we use the matrix inversion the inversion itself is only done once. When we use the nested fixed point the fixed point calculation needs to be done every time we evaluate a different vector of the parameters determining selloff values or profits in the search algorithm. A single fixed point calculation is typically about as computationally burdensome as a matrix inversion[12].

[11] There may be a similar problem for the empirical transition matrix but it is much less severe. If we follow a single time series there is the issue of whether the last observation is an observation which has been visited before. If it has we have estimates of all required transitions. If not we need to impute estimates of the transitions from this last state. If we have a panel of firms then we might have to impute transitions for the last states of each panel member.

[12] If we assume the sell off value distributes i.i.d. type II exterme value as in Aguirreberia and Mira, 2003, or exponential as above, the integral over ϕ has an analytic form and this simplifies the fixed point calculation somewhat.

The extra burden in the nested fixed point is compounded when we use it in combination with the structural transition matrix. The computational burden of the fixed point depends on the number of states and the number of states will, for the reasons noted above, be larger when we use structural probabilities. For similar reasons the computational burden of the fixed point grows exponentially in the number of state variables (or locations); actually as the product of two exponentials (one for the number of points which need to be evaluated in the fixed point calculation, and one for computing the future value at each point).

These computational problems are exascerbated when we iterate on the nested fixed point algorithm, using the first stage estimates to compute a new set of structural transition probabilities. Moreover there is no guaranteee that the iterations improve the estimates, or indeed that they will converge to anything.

If there is an advantage of the nested fixed point it is that the structure it provides might lead to more precise first stage estimate of the continuation and entry values. We examine this possibility below.

4 Monte Carlo.

The Monte Carlo results are designed to give the reader a sense of both the computational burden and the distributions associated with the various estimators. We begin with a single location example and then move to two locations. Thoughout we focus on estimating the distribution of entry fees and sell-off values, as these are the parameters that cannot be estimated from static models of markets. Note that this choice minimizes the increase in computational burden in moving from the matrix inversion to the fixed point (from 3a to 3b above), as were we also to estimate parameters of the profit function, the fixed point would have to be evaluated many more times than it will be evaluate in the results presented below (while the matrix inversion only occurs once).

In both cases we have done all calculations for six sample designs, all of which are panels. We consider varying the time dimension (T) between T=5 and T=15, and the number of markets or the cross sectional dimension (or C) between C=250, C=500, and C=1000[13]. Much of the earlier work on entry was on relatively isolated markets (see Bresnahan and Reiss,1987, and Mazzeo,2002) and this sample design mimics that. We report a selection of the results that seem sufficient to convey the problems that can arise in using these techniques. The reported standard errors are computed from the distribution of estimators over independent Monte Carlo data sets.

[13] The smallest sample size was obtained by counting up the number of relatively isolated markets in South Dakota.

4.1 A Single Location Example.

Our single location example uses a Cournot model with linear demand to determine output and profits conditional on (n, z). Changes in z shift the demand curve over time; they play the role of population size in Bresnahan and Reiss (1987). Appropriate choice of parameter values gives us the following single period profit function

$$\pi(z, n) = Z^2/(1 + n)^2,$$

and we assume that $z = log[Z]$ is the second order Markov process

$$z_{t+1} = z_t + g_{t+1}$$

where g, the growth rate, is a first order Markov process (this generates persistence in growth rates, a phenomena typically observed for the populations of small towns). Thus the state variables for the dynamic problem are (n, g, z).

We assume that the density of the distribution of entry fees is given by

$$f(\kappa = r) = a^2(r - 1/a)exp[-a(r - 1/a)] \tag{14}$$

for $r \in (1/a, \infty)$. This is a unimodal distribution with positive density only at points $r > 1/a$ and a mode at $2/a$. Note that a defines a boundary of the support for κ, and that the existence of this boundary insures that there will be no entry when there are a sufficient number of incumbents. The sell-off value is distributed exponentially with parameter σ.

The actual equilibrium of the model for values of $a = .3$ and $\sigma = .75$ was computed using a variant of the algorithm presented in Pakes and McGuire (1994) (the variant simply shuts down the investment decision in that model). z was allowed to take on 45 values at .05 increments, producing a range of values which roughly corresponds to a nine-fold difference in population size, and we allowed three growth rates $(.05, 0, -.05)$[14]. The maximum number of firms for our parameterization was nineteen, and this implies that the "cardinality" of the state space, or the number of distinct (n, z) couples possible, is about two thousand eight hundred. The "data" was obtained by using the computed equilibrium policies to simulate sample paths. The same data was used in each estimation routine (i.e. the alternative estimators have exactly comparable data, but are not mutually independent).

[14]When $g_t = .05$, $g_{t+1} = .05$ with probability .75 and $g_{t+1} = 0$ with probability .25. The transition probabilities when $g_t = -.05$ are analogous, and when $g_t = 0$, $g_{t+1} = 0$ with probability .5 and moves to each of the alternatives with probability .25. At corners of the permissible z values, the probability of moving out of Z is set to zero and its probability is added to the next closest number.

4.1.1 Results from the Single Location Model.

The first three rows of table 1 constitute a "pivot" table which defines the alternative estimators. OF signifies the objective function used in estimation. If $OF = 1$ we use the pseudo likelihood, if $OF = 0$ we fit the mean (over all observations) of the entry and exit probabilities predicted by the model to the data, and if $OF = 2$ we minimize the sum of squares of the difference between the empirical and the estimate of the entry and exit probabilities at each state weighted by the inverse of the number of times that the data visited that state[15]. PR indicates which first stage probabilities are used in estimation. If $PR = 1$ we calculate "structural" transition probabilities (we estimate entry and exit probabilities and use them and the binomial formula to compute transition probabilities), whereas if $PR = 0$ the transition probabilities are those observed in the data (or $\tilde{M}$ in the notation above). VF indicates how the first stage estimates of $(VC(\cdot), VE(\cdot))$ are computed. If $VF = 1$ then $(VC(\cdot), VE(\cdot))$ are computed via a nested fixed point, whereas if $VF = 0$ these values are found by a single matrix inversion at the beginning of each run.

Panel A provides the results from our largest sample. The esimates are reassuring since they indicate that, at least in the one location model, all estimators "work" reasonably well[16]. The only noticeable difference among the estimators is that pseudo maximun likelihood does a bit worse than the other estimators.

Panel B provides estimates from our smallest sample and at that sample size we learn more about the estimators' performance. All the estimates of a seem to have an upward bias and the bias appears larger in the OF=1 (or pseudo mle) estimator. On the other hand the OF=0 estimates of σ are "right on", and the OF=1 estimates are close, but the OF=2 estimates of σ can be problematic. Note also that though there may be a bias problem in some of the estimators, even at our smallest sample size, the parameter estimates are

[15] Note that the exit probability just depends on $\hat{VC}(\cdot)$ and σ, and the continuation value does not depend on the parameters of the entry distribution. As a result the OF=0 runs can be estimated by finding the value of σ that "zeroes" the exit moment (provided such a value exists), and then a can be chosen to zero the entry moment conditional on this value of $\hat{sigma}$ (provided such a value exists). This makes the OF=0 runs particularly simple computationally, and a similar procedure can be used in the two locations model.

[16] As noted if we use structural probabilities we will need estimates of entry and exit probabilities at points not observed in the data. Here is how we obtain them. Assume that for a given z we only observe behavior from $[n_1(z), n_2(z)]$. Then: (i) if $n \in [1, n_1(z)]$ the probability of entry is equal to the probability of entry in state $n_1(z)$ and the probability of exit is zero; if $n \in [n_2(z), nmax]$ the entry probability is set to zero, and the exit probability is set to that at $n_2(z)$; and (iii) if there is a hole inside the set $[n_1(z), n_2(z)]$ the exit probability is set equal to the closest observed exit probability below it, and the entry probability is set to the closest observed entry probability above it. If we use empirical probabilities and we sometimes get to terminal conditions which are not visited prior to the terminal period, and hence do not have empirical estimates of transition probabilities. For these transitions we take the average transition probabilities for the cells nearest to the terminal cell weighted by the number of times these cells were observed.

quite precise (the exception here is the OF=2 estimate of σ which sometimes is not well estimated). Moreover as Panel B shows, once we move to longer panels the bias problem disappears rather rapidly (and in results not shown here, the same thing happens when we increase C, or the cross sectional dimension of the panel).

An upward bias in a implies a downward bias in the estimate of $1/a$. For the OF=1 estimator or pseudo mle, this illustrates the rather striking differences betweeen semiparametric and parametric estimators of support points. The estimate of a is one over the minimum entry value at which we observed an entrant. We will never observe an entrant with a draw on the entry value which was less than $1/.3$ (the true lower bound on the support of the entry value). Thus in the parametric case our estimator would have to approach .3 from below. Moreover in that case standard arguments prove that the estimator of a would be "superefficient"; it would converge at a faster rate than $\sqrt{n}$. In our model the entry values are estimated, and the error in the first stage estimates insure that our final parameter estimators converge at the standard rate. Moreover once we allow for errors in our estimates of the entry values the argument for the direction of small sample bias reverses itself. If we underestimate the true entry value we could see an entrant entering when the estimated entry value is less than $1/.3$, which would lead the mle to infer that the true value of $1/a$ must be less than $1/.3$ (the true value of a is higher than .3). Indeed with enough (n, z) values this should invariably occur, and as our results show, it does[17].

Throughout the simple $OF = 0$ estimator outperforms both the pseudo mle ($OF = 1$), and the $OF = 2$ estimator. All the estimators have first order conditions which are nonlinear functions of the first step estimates of $(VC(\cdot), VE(\cdot))$. As noted as the sample size grows the estimates converge to their true values, however for samples the size of ours the ratio of the number of estimates (which is reported as $\#\hat{V}$ in the tables) to the number of observations on entry and exit ($= C \times T$) is still quite small. Thus the form of the nonlinearlity will effect the small sample properties of the estimator.

As noted the by taking logarithms the likelihood accentuates the error in the entry and exit probability when these probabilities are small (which is the case in our Monte Carlo, and in most real, data sets). To examine the $OF = 2$ estimator let s index the (n, z) states and E be the expectation opertaor. Then the expectation of its first order condition for, say, the entry probabilities is

$$\sum_s \frac{1}{n_s} \left(\left[\hat{g}_e(s) - E\hat{F}^\kappa(s) \right] E\frac{\partial \hat{F}^\kappa(s)}{\partial a} + \left[\hat{F}^\kappa(s) - E\hat{F}^\kappa(s) \right] \left[\frac{\partial \hat{F}^\kappa(s)}{\partial a} - E\frac{\partial \hat{F}^\kappa(s)}{\partial a} \right] \right).$$

[17] Appendix2 uses a second order expansion to look at the small sample bias in the $OF = 0$ estimator of a and shows that it depends on the derivative of the density of a at the points which generate entry. In almost all of these observations this derivative is positive, and that accounts for the positive deviations in those estimators.

Since $\hat{F}^{\kappa}(s)$, and $\frac{\partial \hat{F}^{\kappa}(s)}{\partial a}$ are constructed from the same estimates of $\hat{VE}$ they are correlated, and the second term is nonzero. Since this term is a function of θ it accentuates the small sample bias in the $OF = 2$ estimators. In contrast the first order condition in the simple $OF = 0$ estimator is linear in the error in the estimates of $\hat{F}$[18].

There is no obvious difference in the distributions generated by the various $OF = 0$ estimators. However there are *large* differences in their computational burdens. When we use structural probabilities (PR=1) we have to compute either a matrix inverse or a fixed point with four to six times the number of states (compare the $\#(n, z)$ row, which is the number of states when PR=0, and the $\#\hat{p}$ row, which is the number when PR=1). In the case of the matrix inverse (i.e. $VF = 0$) this causes an increase in compute time of factors between 4 and 7, and once we go to the nested fixed point calculation the computational burden of the structural probabilities increases further, to between 6 and 10 times the compute time for the empirical probabilities. These ratios are much worse for the OF=1 and the OF=2 estimators (up to a factor of 15). The one positive surprise in the estimates from the structural probabilities is the fact that the imputation of entry and exit rates at the points not actually visited does not seem to cause a noticeable bias in the estimates (and there are alot of them; $\#(n, z)$ provides the number of states actually visited while $\#\hat{p}$ provides the number of probabilities estimated). The nested fixed point times are always larger than the matrix inverse times, but the difference is much more noticeable when we use structural probabilities.

The Monte Carlo results from the model with one entry location are pretty clear. The method of moments estimators that fit the average entry and exit rates from the various states always does best. There is not much difference in the distributions of the $OF = 0$ estimators, but their computational burdens differ dramatically. Indeed the simplest estimator, the one that uses the empirical probabilities and the matrix inverse to calculate entry and continuation values, is clearly the preferred estimator for the single location examples and our sample sizes. We did do a limited number of runs with smaller sample sizes. When $T = 5$ and we decreased C below C=200 sometimes it was useful to do one iteration on the simplest estimators. The second iteration used the initial iteration's estimates as the starting values, OF=0, and PR=0, but did the value function iteration (i.e. VF=1). The gain in precision from the second step was small, but so was the computational cost of using VF=1 provided PR=0 (of course this cost would go up if we were estimating parameters from the profit funciton in addition to those determining the entry and exit costs).

[18] The argument here is very similar to the argument against using non-linear least sqares to estimate regression functions when the regression function itself is simulated with a finite number of simulation draws; see Laffont, Ossard and Vuong, 1995.

4.2 The Two Location Example.

Our two location example is in the spirit of Mazzeo(2002) who estimates a model of competition among vertically differentiated (i.e. high and low quality) motels. The demand curve is derived from a discrete choice model. If the consumer chooses to buy a good in this market, it can choose either the low or high quality good. Consumers' are differentiated by their price coefficient (meant to mimic their marginal utility of income), and the inverse of that coefficient (which should be increasing in income) distributes exponentially. The model generates demand for the low and high quality options, respectively, as

$$Q_1 = M\left(e^{-\lambda\frac{p_1}{\delta_1}} - e^{-\lambda\frac{p_2-p_1}{\delta_2-\delta_1}}\right)$$

and

$$Q_2 = M\left(e^{-\lambda\frac{p_2-p_1}{\delta_2-\delta_1}}\right),$$

provided $\frac{p_2-p_1}{\delta_2-\delta_1} > \frac{p_1}{\delta_1}$ (otherwise, $Q_1 = 0$).

Each of the (n_1, n_2) firms choose a quantity to market in their location, and prices adjust to the (unique) Cournot equilibrium price vector. The profit of firm i manufacturing product k are computed "offline" as

$$\pi_{k,i} = (p_k - c_k)q_{k,i},$$

where c_k is the marginal cost of product k and p_k is its equilibrium price. We set $\frac{c_2}{c_1} > \frac{\delta_2}{\delta_1}$, as this guarantees positive equilibrium quantities.

We now list the assumptions on entry and exit. There is a uniform distribution of the number of potential entrants with $P(E) = 1/4$ for $E \in [0, 1, 2, 3]$ in each period. When a potential entrant appears it receives an independent draw on $\kappa = (\kappa_1, \kappa_2)$ from $F^{\kappa}(\cdot,\cdot \mid \theta)$ and can enter in at most one of the markets. Since κ_1 and κ_2 reflect differences in a given individual's cost of building the high and the low quality motel in a particular market, we allow them to be correlated. Indeed we make the reasonable assumption that the cost to a given individual of building the high quality motel in a given market is larger than that individual's cost of building the low quality motel in the same market, i.e. we assume

$$\kappa_2 > \kappa_1 \text{ with probability one.}$$

More precisely we assume that the cost of entry into the low quality product distributes as does the entry cost in the one location model (see equation 14) while the cost of entry into the high quality product is given by

$$\kappa_2 = \kappa_1 + r$$

where r distributes as does κ_1[19]. Thus there are two entry parameters to estimate (a_1, a_2). Exit fees are distributed i.i.d. exponential with parameters (σ_1, σ_2) in the two locations.

4.2.1 Results from the Two Location Example.

Computation of the equilibrium and generation of the data used for estimation is done in a manner analogous to how it was done for the single location example. There is a "start" row in the pivot table for the OF=1 and OF=2 two location runs; when start=1 we start the search from the $OF = PR = VF = 0$ estimator[20].

Table 2 summarizes results on pseudo mle estimates from panels with $T = 15$, and $C = 5000$. Given this amount of data, the table is designed to tell us whether pseudo mle "works"[21]. The answer is pseudo mle *does not* work. We labeled a search starting from a certain point "unsuccessful" when our Nelder-Mead simplex (direct) search could not find a positive value for the likelihood. When there was an unsuccessful initial condition for a given data set, we tried another initial condition, and continued until we started the search at ten different randomly selected points none of which resulted in a positive likelihood. At that point we called the search on the data set unsuccessful, and moved to the next data set. The sub-panel labeled "success rate" provides the fraction of the time when this subroutine recorded a "success". The pseudo mle does not work *most* of the time.

The reason for the zero pseudo likelihood is that the first stage is producing a $\hat{VE}_1(\cdot) > \hat{VE}_2(\cdot)$ for candidate parameter values when there is in fact entry in location 2. Since the cost of entry in location 2 is always higher than in location 1, if our estimated entry values were true entry in location 1 would never happen (hence the zero likelihood). The reason it does happen is because $VE_1(\cdot) < VE_2(\cdot)$; i.e. the disturbances in the first stage estimates have reversed the order of the two entry values. I.e. since the pseudo likelihood does not recognize the possibilities generated by first stage estimation error, it can record a probability of zero for events which do happen.

Since the precision of the first stage estimates at a point are a function of the number of times that point was visited, we thought we might improve the performance of the pseudo likelihood estimators if we trimmed points which were visited infrequently. The "success sub-panel" presents success ratios when we trimmed the one-half of the states visited which were visited the least number of times ("trim=.5 states"), and when we trimmed all states that were visited less than ten times ("trim=10 visits"). Trimming does improve our success

[19] Actually we use a discretized version of the density in equation (14) for the r distribution.

[20] It is straightforward to find a zero to the first order condition in the $OF = 0$ runs, so we did not try alternative starting values for them.

[21] Since with this much data Monte Carlo repetitions take quite a bit of computer time, and we knew the answer with a relatively small number of repetitions, we stopped this run at $R = 14$.

rate, but it is still noticeably below one. The next panel provides the estimates from the trimmed run with the highest success rate. It is clear that the trimming is both biasing the estimates and causing their variance to go up (in some cases dramatically). The performance of the pseudo mle was no better at smaller sample sizes. We conclude that one should not be using pseudo mle, at least not without some auxiliary procedure that ameliorates the problems caused by imprecise first stage estimates of entry values. Consequently we do not present them in what follows[22].

Table 3 provides the estimators obtained from OF=0 and OF=2 with a quite large sample ($T = 15, C = 1000$). Consider first the OF=2 columns. It is clear that when using this objective function the estimators can be very different from their true values, and even when they are not so different their variances are markedly higher than those of the OF=0 estimators. In the two location model there are fewer draws on entry and continuation values at each state. Thus even with relatively large samples the preliminary estimators of these values can be quite noisy. That noise, when combined with the accentuation of the error which results from multiplication of the error in the probability and the error in the derivative of the probability in the OF=2 estimators makes them problematic (see the discussion of the Monte Carlo estimates in the one location model)[23]. As one might expect these problems only get worse with smaller samples, so we focus the rest of the discussion on the OF=0 estimates.

There $OF = 0$ estimators all do fairly well. The estimates of σ that use structural probabilites seem to have a slightly large mean square error, which likely corresponds to the fact that they have to impute entry and exit rates for about 85% of the states they compute entry and continuation values for.

There are, however rather *stark* differences in the computational burdens of the alternative OF=0 estimators. Estimates which use structural probabilities and VF=0 are fifteen times as computationally burdensome as estimates which use empirical probabilities and VF=0, and estimates that use the fixed point combined (VF=1) with the structural probabilities are twenty times more computationally burdensome. The estimates which use the empirical probabilities and value function iterations are only about 10% more burdensome then those that use the matrix inversion (though again this would increase were we also estimating profit function parameters). Note also that the simplest estimator (which is the estimator with the preferred distribution) computes in just over a minute, even with this, relatively large, data set.

Table 4 provides the estimates from smaller samples. Since the ratio of estimated entry

[22] Though we did compute them (with results similar to those presented above). We note that when you iterate on the pseudo mle estimators they only get worse, as the iterated success rate can only go down.

[23] We note that the results in this table are not a result of a few outliers. Indeed the Monte Carlo distribution of the OF=2 estimators looked close to normal.

and continuation values to observations $((2 \times \#(n,z))/(C \times T)))$ is about twice that ratio in the one location example with the same sample sizes, we should not be surprised when we see larger small sample biases and larger standard errors then in Table 1. However, just as in that table, as we increase either the length of the panel or the size of its cross sectional dimension these biases and standard errors go down rather rapidly.

In the $(T = 5, C = 250)$ samples the estimators that use structural probabilities (PR=1) do better on the a's but much worse on the $\sigma's$. As we increase sample size the problem with the empirical probability's estimates of the a's disappears rather rapidly, much more rapidly than the problems with the structural probability's estimates of the σ's. Even at $(T = 5, C = 500)$ it seems pretty clear that we prefer the estimates which use the empirical probabilities, and this conclusion is reinforced at larger sample sizes.

There is a sense in which this should have been expected. All estimators have the problems generated by the imprecision of the first stage estimates, but that problem gets smaller as sample size increases. Only the stuctural probabilities have the problem of imputing entry and exit rates at out of sample points, and this problem does not tend to get smaller as sample size increases. Indeed a comparison of $\#(n,z)$ to $\#\hat{p}$ for the two location samples in Table 4 indicates that regardless of sample size the structural probability estimates have to impute entry and exit rates for eighty five to ninety percent of the states needed for the first stage estimates.

We also tried estimates which iterate on the PR=1 VF=1 estimates. The iterations use the initial parameter estimates and the value function to construct entry and exit rates and the transition matrices they imply for *all* points. The hope was that this might improve on the "imputed" estimates of transition probabilities at the states not actually visited in the $PR = 1$ estimators. However the iterated estimates did not perform any better than those in the table, and did not seem to converge to anything.

Note also that the computational burden of the estimators that use structural probabilities is quite large; when VF=1 they require fifteen to twenty times the cpu time required by the estimates that use the empirical probabilities. Moreover when we iterate on the $PR = 1$ estimator it becomes computationally prohibitive (over two hundred times more burdensome than any of the estimators that use the empirical probabilities).

The two remaining estimators, the estimators with OF=0, PR=0, and VF=1 or 0, perform similarly. The VF=0 estimator seems preferred for our smallest sample, but for sample sizes between $(T = 5, C = 500)$ and $(T = 15, C = 1000)$ there seems to be a preference for the VF=1 estimator, whereupon preferences move back to the VF=0 estimator. On the other hand neither the differences in the performances of these two estimators, nor the differences in their computational burdens are very large. As noted if we also had to estimate profit function parameters, a more significant difference in computational burdens would arise, and it would favor the OF=0, or matrix inversion, estimator.

4.3 Conclusions form the Monte Carlo Examples.

The results from our Monte Carlo experiments are unusually clear cut.

- Among the three objective functions we tried, the one that fits predicted to actual average (over the entire sample) entry and exit rates is clearly preferred. The pseudo mle has conceptual problems which often make it impossible to use, and even when these problems do not arise the pseudo mle tends to under-perform the alternative objective functions (probably because the use of the logarithm of estimated probabilities tends to exaggerate the estimation error in the estimates of those probabilities). Estimators based on minimizing a metric in state specific entry and exit rates run into the problem that the estimation error in the relevant probabilities resulting from the first stage estimates is correlated with the estimation error in the derivatives of those probabilities, thus accentuating both the small sample bias and the variance of the estimators.

- The results favor the use of empirical over structural probabilities. Structural probabilities are at least a factor of ten more computationally burdensome and have the added disadvantage that they have to impute transition probabilities for the states not visited (imputations that can generate errors in the estimators).

- Among the favored estimators, i.e. those that fit entry and exit rates and use empirical instead of structural probabilities, we could use either the matrix inversion formula or the fixed point formula to obtain the first stage estiamtes of the entry and continuation values. What differences there are in the performance of these two estimators seems to depend on sample size, and, provided we are not estimating profit function as well as sunk cost parameters, the difference in their computational burdens is small.

Finally note that since our estimators use empirical, and not structural, probabilities the empirical transitions will be confined to lie in a recurrent subclass of all states. This implies that the increase in computational burden as we increase the number of state variables is more likely to be linear in the number of states, then exponential (see Pakes and McGuire,2001). This bodes well for our ability to use these models with realistic assumptions.

5 Conclusions.

This paper provided estimators for discrete dynamic games that are easy to use and examined their properties. The estimators rely on assumptions which insure that there is a unique equilibrium associated with the given data generating process. Given those assumptions it is shown that one can obtain consistent estimates of entry and continuation values by simply accumulating the discounted value of net returns *actually earned* by the entrants who entered at particular states, and the discounted value of net returns *actually earned* by incumbents who continued from those states. These discounted values can be consistently estimated up to a parameter to be estimated from a matrix inversion which need only be done once at the beginning of the estimation run. This makes the computational burden of our estimator similar to the burden of estimating a multinomial model in probabilities which are known functions of the data.

We considered alternative ways to form estimates for the parameters of our semi-parametric multinomial model. A theoretical discussion showed that the fact that our multinomials are formed from estimated (rather than known) probabilities, should effect both the distributional and computational burdens of the alternatives. The Monte Carlo examples were designed to push the investigation of the computational and distributional properties of the alternatives further.

The results from the Monte Carlo examples showed that the simplest estimators were also the most effective. Moreover, the favored estimators were so simple computationally that for most of our sample designs estimation took under thirty seconds, and even the two location model with a data set of seventy five thousand observations took under ten minutes to estimate.

References.

Aguirreberia, V. and P. Mira (2003); "Identification and Estimation of Dynamic Discrete Games", *mimeo* Boson University.

Andrews, D. and S. Berry (2003)"Using Necessary Conditions to Place Bounds on the Parameters of Entry Models Without Unique Equilibria" *mimeo*, Yale University.

Bajari, P., L. Benkard, and J. Levin (2003); "Estimating Dynamic Models of Imperfect Competition", *mimeo*, Stanford GSB.

Berry, S. (1992); "Estimation of a Model of Entry in the Airline Industry", *Econometrica*, vol. 60, no. 4, pp. 889-917.

Berry, S. and A. Pakes (2002); "Estimation From the First Order Conditions for Dynamic Controls" *mimeo*, Department of Economics, Harvard and Yale Universities.

Brenahan, T. and P. Reiss (1987); "Do Entry Conditions Vary Across Markets" *Brookings Papers on Economic Activity, Microeconomics Annual* ,1, pp.833-882.

Bresnahan, Timothy and Peter Reiss (1991): "Entry and Competition in Concentrated Markets", *Journal of Political Economy*, vol. 99, no. 5. pp. 977-1009.

Einav, L. (2003); “Not All Rivals Look Alike: Estimating an Equilibrium Model of the Release Date Timing Game”, *mimeo* Stanford, Department of Economics.

Ericson, Richard and Pakes, Ariel (1995): “Markov-Perfect Industry Dynamics: A Framework for Empirical Work”, *Review of Economic Studies*, vol. 62, no. 1, pp. 53-82.

Freedman, D.(1983), *Markov Chains*, Springer Verlag, New York.

Hotz, J. and R. Miller (1993); “Conditional Choice Probabilities and the Estimation of Dynamic Models”, *The Review of Economic Studies*,60, pp. 265-89.

Laffont, J.J., H. Ossard, Q. Vuong (1995); ” Econometrics of First-Price Auctions”, *Econometrica* (63) pp.,953-980.

Maskin, E. and J. Tirole (1988); “A Theory of Dynamic Oligopoly, I and II; *Econometrica*, vol. 56, no. 3, pp. 549-99.

Maskin, E. and J. Tirole (1994) “Markov Perfect Equilibrium”, Harvard Institute of Economic Research Discussion Paper.

Mazzeo, M. (2002); “Product Choice and Oligopoly Market Structure”, *The Rand Journal of Economics.* Summer, pp.221-42.

Muth, J. (1961)

S. Olley and A. Pakes (1996); “The Dynamics of Productivity in the Telecommunications Equipment Industry”, *Econometrica*, vol. 64, no. 6, pp. 1263-1297.

Pakes, A. (1994);”Dynamic Structural Models; Problems and Prospects”, in C. Sims (ed.) *Advances in Econometrics. Sixth World Congress of the Econometric Society*, Cambridge University Press.

Pakes, A. and P. McGuire, (1994); “Computing Markov-Perfect Nash Equilibria: Numerical Implications of a Dynamic Differentiated Product Model”, *RAND*, vol. 25, no. 4, pp. 555-589.

Pakes, A. and P. McGuire, (2001); “Stochastic Algorithms, Symmetric Markov Perfect Equilibria and The ’Curse’ of Dimensionality” *Econometrica*, Vol 69 (5), pp1261-81.

Pakes, A., J. Porter, and C. Wolfram (2003); “Cost Functions From Profit Maximization Conditions; The Electric Utility Case”, *mimeo* Harvard University.

Pesendorfer, M., and P. Schmidt-Dengler, 2003; “Identification and Estimation of Dynamic Games”, *mimeo* London School of Economics and Yale University.

Rust, J. (1987) “Optimal Replacement of GMC Bus Engines; An Empirical Model of Harold Zurcher”, *Econometrica*, 55 (no.5) pp. 999-1033.

Rust, J. (1994); “Estimation of Dynamic Structural Models; Problems and Prospects” in C. Sims (ed.) *Advances in Econometrics. Sixth World Congress of the Econometric Society*, Cambridge University Press.

Seim, K.(2002);"Geographic Differentiation and Firms' Entry Decisions: The Video Retail Industry", *mimeo*, GSB, Stanford.

Star, A.W. and Y.C. and Ho (1969); "Nonzero-Sum Differential Games", *Journal of Optimization Theory and Applications*, pp.1984-208.

Tamer, E. (2000); "Incomplete Simultaneous Discrete Response Model wih Multiple Equilibria" *mimeo*, Department of Economics, Princeton University.

Table 1; One Location.

Pivot Table*

OF	0	0	0	0	1	1	1	1	2	2	2	2
PR	0	0	1	1	0	0	1	1	0	0	1	1
VF	0	1	0	1	0	1	0	1	0	1	0	1
Panel A: T=15,C=1000, R=100.												
a=.3	0.30	0.30	0.30	0.30	0.33	0.31	0.31	0.31	0.30	0.30	0.30	0.30
SD(a)	0.00	0.00	0.00	0.00	0.03	0.00	0.01	0.00	0.00	0.00	0.00	0.00
$\sigma = .75$	0.75	0.75	0.75	0.75	0.74	0.72	0.74	0.74	0.76	0.75	0.75	0.75
SD(σ)	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01
t(total)	32.1	35.6	84.7	130.0	36.6	62.9	89.2	415.9	32.8	56.3	93.0	405.4
Panel B: T=5,C=250, R=100												
a=.3	0.36	0.36	0.37	0.37	0.41	0.37	0.38	0.37	0.36	0.36	0.37	0.37
SD(a)	0.03	0.03	0.03	0.03	0.05	0.03	0.03	0.02	0.03	0.03	0.03	0.03
$\sigma = .75$	0.77	0.74	0.75	0.74	0.74	0.70	0.73	0.70	1.04	0.81	0.81	0.77
SD(σ)	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.60	0.36	0.06	0.05
t(setup)	9.4	9.4	65.6	65.6	9.4	9.4	65.6	65.6	9.4	9.4	65.6	65.6
t(search)	0.0	1.5	0.1	41.3	0.3	8.6	0.3	223.4	0.2	9.3	6.0	241.5
t(total)	9.4	10.9	65.7	106.9	9.7	18.0	65.9	289.0	9.6	18.7	71.6	307.2
$\#(n,z)$	423	423	423	423	423	423	423	423	423	423	423	423
$\#\hat{p}$	464	464	2597	2597	464	464	2597	2597	464	464	2597	2597
Panel C: T=15,C=250, R=100.												
a=.3	0.32	0.32	0.32	0.32	0.34	0.32	0.33	0.32	0.31	0.32	0.32	0.32
SD(a)	0.01	0.01	0.01	0.01	0.03	0.01	0.01	0.01	0.01	0.02	0.01	0.01
$\sigma = .75$	0.75	0.75	0.75	0.74	0.74	0.72	0.74	0.72	0.83	0.91	0.77	0.75
SD(σ)	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.36	0.86	0.03	0.02
t(setup)	19.2	19.2	74	74	19.2	19.2	74	74	19.2	19.2	74	74
t(search)	0.0	2.4	0.1	42.3	0.9	15.8	1.0	265.2	0.4	15.2	6.8	271.0
t(total)	19.3	21.6	74.1	116.3	20.1	35.1	75.0	339.2	19.6	34.4	80.8	345.0
$\#(n,z)$	629	629	629	629	629	629	629	629	629	629	629	629
$\#\hat{p}$	638	638	2691	2691	638	638	2691	2691	638	638	2691	2691

* Legend. OF=Objective Function. $OF = 0,1,2 \Rightarrow$ MOM fitting average entry and exit rates, MLE, MOM fitting state specific entry and exit rates. PR=Estimates of Probabilities. $PR = 0,1 \Rightarrow$ empirical probabilities, structural probabilities. VF=value function. $VF = 0,1 \Rightarrow$ matrix inversion, nested fixed point. $\#(n,z)$ is the number of states visited, while $\#\hat{p}$ is the number of states for which we must compute probabilities when $PR = 1$.

Table 2: Two Locations, Pseudo MLE.

Pivot Table*.						
PR	0	0	0	1	1	1
VF	0	1	1	0	1	1
Start	0	0	1	0	0	1
T=15, C=1000; "Success" Rate						
pseudo mle	0/14	4/14	4/14	0/14	4/14	5/14
trim=.5 states	6/14	8/14	11/14	13/14	10/14	11/14
trim=10 visits	5/14	5/14	11/14	9/14	9/14	13/14
Estimates from Best Trim.						
a1=.3	0.37	0.35	0.34	0.30	0.30	0.30
sd(a1)	0.05	0.03	0.06	0.01	0.01	0.02
a2=.3	0.35	0.34	0.37	0.32	0.31	0.33
sd(a2)	0.05	0.04	0.07	0.03	0.02	0.11
$\sigma 1=1$	1.17	0.96	0.85	1.11	1.10	1.07
SD($\sigma 1$)	0.26	0.22	0.24	0.29	0.29	0.26
$\sigma 2=.5$	0.51	0.54	0.40	0.51	0.57	0.53
SD($\sigma 2$)	0.08	0.13	0.11	0.09	0.15	0.14

* Legend. See the footnote to Table 1.

Table 3; Two Location.

Pivot Table*

OF	0	0	0	0	2	2	2	2	2	2	2	2
PR	0	0	1	1	0	0	0	0	1	1	1	1
VF	0	1	0	1	0	0	1	1	0	0	1	1
Start	0	0	0	0	0	1	0	1	0	1	0	1
T=15,C=1000, R=50												
a1=.3	0.30	0.30	0.29	0.29	0.29	0.29	0.29	0.29	0.21	0.21	0.27	0.27
SD(a1)	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.11	0.11	0.05	0.05
a2=.3	0.30	0.30	0.30	0.30	0.27	0.27	0.28	0.28	4.42	9.29	4.29	4.28
SD(a2)	0.02	0.02	0.03	0.02	0.04	0.04	0.02	0.02	19.36	27.22	19.74	19.74
σ1=1	1.02	1.00	1.06	1.02	16.09	16.33	1.36	1.47	21.72	31.52	1.75	1.93
SD(σ1)	0.09	0.07	0.10	0.07	31.13	26.22	0.83	2.44	33.02	36.22	3.20	3.61
σ2=.5	0.51	0.50	0.52	0.50	10.71	11.46	0.87	0.62	11.78	19.34	0.64	1.26
SD(σ2)	0.03	0.03	0.04	0.03	24.38	20.72	1.79	0.38	26.47	28.23	0.16	2.86
t(setup)	47	47	920	920	47	47	47	47	920	920	920	920
t(search)	22	27	177	385	146	125	199	121	950	933	3527	2653
t(total)	69	75	1096	1305	193	173	246	168	1870	1853	4446	3573
$\#(n,z)$	595	595	595	595	595	595	595	595	595	595	595	595
$\#\hat{p}$	602	602	3463	3463	602	602	602	602	3463	3463	3463	3463

* Legend. See the footnote to table 1. Start=1 indicates that the starting value for this estimator are the estimates for the simplest model (OF=PR=VF=0).

‖ ‖

Table 4; Two Location.

Pivot Table*								
OF	0	0	0	0	0	0	0	0
PR	0	0	1	1	0	0	1	1
VF	0	1	0	1	0	1	0	1
Data	T=5,C=250, R=100				T=5,C=500, R=100			
a1=.3	0.41	0.40	0.28	0.28	0.34	0.34	0.29	0.29
SD(a1)	0.16	0.14	0.04	0.04	0.05	0.05	0.03	0.03
a2=.3	0.49	0.49	0.31	0.31	0.39	0.39	0.30	0.30
SD(a2)	0.19	0.18	0.07	0.07	0.08	0.08	0.05	0.04
σ1=1	1.29	1.99	2.55	2.27	1.09	1.14	1.82	1.68
SD(σ1)	1.82	2.72	3.10	2.11	0.41	1.18	0.76	1.18
σ2=.5	0.57	2.96	1.11	1.45	0.61	0.77	0.75	0.65
SD(σ2)	0.24	10.89	0.81	4.66	0.39	2.46	0.28	0.27
t(setup)	14	14	350	350	21	21	477	477
t(search)	10	13	101	249	12	15	112	250
t(total)	24	26	451	599	33	36	589	727
$\#(n,z)$	322	322	322	322	403	403	403	403
$\#\hat{p}$	347	347	2638	2638	427	427	2837	2837
Data	T=15,C=250, R=100				T=15,C=500, R=100			
a1=.3	0.34	0.34	0.28	0.28	0.32	0.32	0.29	0.29
SD(a1)	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02
a2=.3	0.35	0.35	0.31	0.31	0.32	0.32	0.31	0.31
SD(a2)	0.04	0.04	0.05	0.05	0.03	0.03	0.03	0.03
σ1=1	1.09	1.08	1.33	1.15	1.05	1.01	1.17	1.07
SD(σ1)	0.35	0.77	0.38	0.28	0.10	0.08	0.18	0.10
σ2=.5	0.54	0.51	0.61	0.55	0.51	0.50	0.55	0.52
SD(σ2)	0.09	0.06	0.14	0.08	0.05	0.05	0.06	0.04
t(setup)	27	27	532	532	36	36	667	667
t(search)	14	17	126	268	18	22	148	303
t(total)	40	44	658	799	53	57	816	970
$\#(n,z)$	463	463	463	463	529	529	529	529
$\#\hat{p}$	470	470	2938	2938	536	536	3128	3128

* Legend.See the footnotes to Tables 1 and 3.

Table 5; Iterated MLE, Two Locations,T=5, C=500, R=36.

Start=(0, 0, 0). Iterations PR=1,VF=1.						
iterations	0	1	2	3	4	5
Estimates.						
a1=.3	above	0.28	0.26	0.29	0.23	0.26
sd(a1)	above	0.03	0.02	0.03	0.03	0.03
a2=.3	above	0.31	0.32	0.31	0.33	0.31
sd(a2)	above	0.07	0.05	0.05	0.08	0.06
σ1=1	above	1.20	0.99	1.21	0.97	1.20
sd(σ1)	above	0.22	0.12	0.17	0.10	0.19
σ2=.5	above	0.57	0.49	0.59	0.49	0.59
sd(σ2)	above	0.08	0.05	0.06	0.06	0.08
Time (seconds).						
time setup	above	692	710	707	707	697
time search	above	3677	2003.	2368	2805	2554
time cummulative	3216	7585	10298	13373	16885	20136

Appendix 1: Entry Weights With Random Potential Entrants.

We go directly to the the model with two entry locations. The result for a single entry location is a special case ($e_{-l} \equiv 0, m_{-l} \equiv 0$).

Proposition. In the model with two locations

$$p(e_l, e_{-l}|n, z, \chi_l^e = 1, \theta) = w_t^{e,l} p(e_l, e_{-l}|n, z)$$

where

$$w_t^{e,l} = \frac{e_t^l}{\bar{E}_l(z_t, n_t)} = \frac{e_t^l}{\sum_E m_l(n_t, z_t) EP(E|\ \theta)},$$

and is consistently estimated by substituting

$$\hat{\bar{E}} \equiv \#T(n,z)^{-1} \sum_{t \in T(n,z)} e_t^l \quad \text{for} \quad \sum_E m_l(n_t, z_t) EP(E|\ \theta)$$

in the above formula. Recall that

$$p(e_l, e_{-l}|n, z) = \sum_{E \geq (e_l + e_{-l})} m(e_l, e_{-l}, j; m_0, m_1, m_2) \frac{EP(E|\ \theta)}{\sum EP(E|\ \theta)}.$$

Proof. From the text

$$p_l(e_l, e_{-l}|n, z, \chi_l^e = 1, \theta) = \sum_{E \geq (e_l + e_{-l})} m(e_l - 1, e_{-l}, E - 1; m_0, m_1, m_2) \frac{Ep(E|\ \theta)}{\sum Ep(E|\ \theta)}$$

$$= \frac{1}{\sum Ep(E|\ \theta)} \sum_{E \geq e_l + e_{-l}} \frac{(E-1)! \times E}{(E - e_l - e_{-l})!(e_l - 1)! e_{-l}!} m_0^{E - e_l - e_{-l}} m_l^{e_l - 1} m_{-l}^{e_{-l}} p(E|\ \theta).$$

Multiply both the numerator and denominator of this equation by $e_l \times m_l$ and note that since $e_l \times (e_l - 1)! = e_l$ and $m_l^{e_l - 1} \times m_l = m_l^{e_l}$, that equation

$$= \frac{e}{m_l \sum Ep(E|\ \theta)} \sum_{E \geq e_l + e_{-l}} \frac{E!}{(E - e_l - e_{-l})!(e_l - 1)! e_{-l}!} m_0^{E - e_l - e_{-l}} m_l^{e_l} m_{-l}^{e_{-l}} Ep(E|\ \theta)$$

$$= w_t(e) \sum_{E \geq (e_l + e_{-l})} m(e_l, e_{-l}, E; m_0, m_1, m_2) p(E|\ \theta)$$

$$= w_t(e) p(de_l, de_{-l}|n, z),$$

as desired. ♠

Appendix 2; Small Sample Bias of $\hat{a}$ When $OF = 0$.

Just look at the entry equation, assuming we know σ. Then our equation is setting

$$\frac{1}{NT}\sum_{t,i}\left[F^{\kappa}(\tilde{VE}_{i,t}|a) - e_{i,t}\right] = 0.$$

Now expand about the true $VE_{i,t}$ and let

$$\tilde{VE}_{i,t} = VE_{i,t} + \epsilon_{i,t}.$$

Then

$$\frac{1}{NT}\sum_{t,i}\left[F^{\kappa}(VE_{i,t}|a) + c_{i,t}\epsilon_{i,t} + d_{i,t}\epsilon_{i,t}^2 - e_{i,t}\right] = 0$$

where

$$c_{i,t} = \frac{\partial F^{\kappa}(VE_{i,t}|a)}{\partial VE_{i,t}}$$

and

$$d_{i,t} = \frac{\partial^2 F^{\kappa}(VE_{i,t}|a)}{\partial VE_{i,t}^2}.$$

If we multiply by $\sqrt{NT}$ and assume

$$\frac{1}{\sqrt{NT}}\sum_{i,t} c_{it}\epsilon_{i,t} \approx \mathcal{N}(0, V_c)$$

$$\frac{1}{\sqrt{NT}}\sum_{i,t}[d_{it}\epsilon_{i,t}^2 - \mu_d] \approx \mathcal{N}(0, V_d)$$

where

$$\mu_d = \frac{1}{NT}B(a).$$

For fixed "a" then

$$\frac{1}{NT}\sum_{t,i}\left[F^{\kappa}(\tilde{VE}_{i,t}|a) - e_{i,t}\right]$$

$$\approx \frac{1}{NT}\sum_{t,i}[F^{\kappa}(VE_{i,t}|a) - e_{i,t}] + \frac{1}{\sqrt{NT}}\nu(a) + \frac{1}{NT}B(a)$$

where $\nu(a)$ is mean zero and $B(a)$ is the bias term. There is an issue as to whether there is a two-stage procedure which subtracts out an estimate of $B(a^0)$ in a second stage. It should follow from the multinomial sampling probabilities.