



HIER

Harvard Institute of Economic Research

Discussion Paper Number 2036

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

by

Ariel Pakes, Michael Ostrovsky
And Steve Berry

May 2004

Harvard University
Cambridge, Massachusetts

This paper can be downloaded without charge from:
<http://post.economics.harvard.edu/hier/2004papers/2004list.html>

The Social Science Research Network Electronic Paper Collection:
<http://ssrn.com/abstract=566864>

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 draft May 2004.)

Abstract

This paper considers the problem of estimating the distribution of payoffs in a discrete dynamic game, focusing on models where the goal is to learn about the distribution of firms' entry and exit costs. The idea is to begin with non parametric first stage estimates of entry and continuation values obtained by computing sample averages of the realized continuation values of entrants who do enter and incumbents who do continue. Under certain assumptions these values are linear functions of the parameters of the problem, and hence are not computationally burdensome to use. Attention is given to the small sample problem of estimation error in the non parametric estimates and this leads to a preference for use of particularly simple estimates of continuation values and moments.

¹We thank John Asker, Pat Bayer, Lanier Benkard, Gary Chamberlain, Jeff Dominitz, Boyan Jovanovic, Jack Porter, and Thanasis Stengos for helpful remarks and the NSF for financial support.

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

This paper develops estimation strategies based on the structure of discrete dynamic games. For ease 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 sunk costs of entry and the sell off 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 much harder to find than data on the determinants of current profits². As a result we have to infer the extent of sunk costs from other variables whose behavior depends on them. The variable that is most directly related to the costs of entry is entry itself. To use the connection between actual entry and the costs of entry in estimation we need a framework which allows us to compute the value of entering (similarly, to make use of the relationship between sell off values and exit, we need to be able to calculate the value of continuing). Though such frameworks have been available for some time (e.g., Ericson and Pakes, 1995), their implications can not be used directly in estimation without encountering substantial computational problems (with current computational abilities these problems are often insurmountable) .

As a consequence the models that have been used to analyze entry and exit decisions in the past have all been two-period models; see Bresnahan and Reiss (1987 and 1991), Berry (1992), and more recently, Mazzeo (2002) and Seim (2002). Two-period models are essentially static and assume away sunk costs. Partly as a result the empirical work in these papers stayed away from examining the impacts of policy or environmental changes on the structure of a given industry over time, focusing instead on providing a framework for characterizing differences in the number of active firms across a cross section of markets. More detailed analysis of environmental or policy issues, say of the likely effects of a merger on subsequent entry or of a change in pension or health care rules on exit, required more detail on sunk costs and how firms react to them.

²The decision of a firm on whether to exit is determined by whether its continuation value is greater than its sell off 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 determined by the time and effort required to formulate ideas, test markets, and access both startup capital and the requisite permissions.

The early frameworks for the analysis of entry and exit were also the first papers to explicitly 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) considered its impact on estimation in models where sunk costs could vary among agents. When models do not have a unique solution it is, in general, not possible to determine the probability of a given outcome conditional on observables and the value of a parameter vector. This rules out many standard estimators. The uniqueness issue became even more important once we allowed for the realism of continuation values which differed across agents, for then the number of possible equilibria increased markedly. The original analysis here, due to Mazzeo (2002) and Seim (2002), allowed continuation values to differ with “location” and began investigating 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*.

The Underlying Idea.

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 than the entry fee (similarly, an incumbent will continue if the continuation value is greater than 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 make the relationship between the model and the data transparent, a fact, which together with the estimator’s computational ease, simplifies robustness analysis greatly.

Once we have consistent estimates of entry and continuation values, the rest of the estimation problem is 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 less than the estimated entry value. Similarly, the probability of an incumbent exiting is the probability of drawing a sell-off value greater than the continuation value.

We then find the value of the model's parameters that make its predictions for entry and exit rates "as close as possible" to the rates observed in the data. Alternative metrics for closeness produce alternative (root-n) consistent and asymptotically normal estimators, and we provide an extensive discussion of the differences in their computational and statistical properties.

Our use of the 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 non (or semi) parametric estimate of the entry and continuation values. The second stage treats these 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 this estimator is, if anything, *less than* that of the estimators for the simple static entry models.

The paper begins with 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 entry locations and a random number of potential entrants. Once conceptual issues are clarified, a number of modifications that lead to alternative estimators suggest themselves. The alternatives have different computational and distributional properties, and so are worth considering.

As a result, we provide fairly detailed Monte Carlo results on two examples: one with a single location, and the other with two locations. The Monte Carlo results, when combined with an analytic discussion of why they occur, end up being extremely informative. Among the alternatives we consider only one, perhaps two, should be considered by researchers, and the best performing alternatives are also the *least* computationally burdensome. Moreover, the computational burden of these estimators is small enough to think that the effective limitation to the empirical analysis of entry and exit costs will become the richness of the data, rather than computational feasibility.

As we will point out our procedures carry costs both in terms of assumptions, and in terms of the ability to use the results in subsequent analysis. We conclude with a discussion of the robustness of our estimator with respect to assumptions that are likely to be problematic.

Related Literature

Before proceeding to the details, a word on the 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; they use semi-parametric estimates of the choice of discrete controls to simplify their estimation algorithm. The Hotz and Miller paper deals with a single agent problem, while a subsequent paper by Olley and Pakes (1996) uses semi-parametric estimates of the choice of discrete and continuous to simplify estimates of a multiple agent dynamic game.

There are also a number of papers currently “in process”, all written independently, that present related results. The closest to our paper is a paper by Aguirregabiria and Mira (2003) which introduces one of the alternative estimators considered in our extensions. Aguirregabiria and Mira (2003) focus on a model with “logit” (extreme value) errors and emphasize, contrary to our conclusions, the efficiency advantages of a maximum likelihood estimator. Pesendorfer and Schmidt-Denglar(2003) make an i.i.d. probit assumption and implement the empirical analog of the identification argument in Magnac and Thesmar (2002). Their approach to sampling error is different than that in the other papers as they place a tight restriction on the distribution of *endogenous* outcomes, rather than taking the traditional approach of restricting primitives and working out the implications of those restrictions on the endogenous quantities of interest. Bajari, Benkard and Levin (2004) consider mixed continuous discrete controls, but also have an entry exit example with logit errors. Their approach is different in that they first non-parametrically estimate how the controls are determined by the state variables of the problem (as in Hotz and Miller,1993 and Olley and Pakes,1996), and then use those estimates to simulate continuation values³.

All these estimators could be used to estimate the parameters of the variable profit function, as well as the distribution of entry and exit fees. Throughout, however, we focus on estimating the distribution of entry and exit fees as it is these parameters that are difficult to estimate without using the structure of a dynamic game. This leads us to avoid i.i.d. extreme value or i.i.d. normal disturbances. Those assumptions imply that the entry costs a given potential entrant faces in different locations in the same market are independent of one another and have full support, and these implications can be quite unrealistic. Also, at least in special cases, our simple analytic solutions do not require the Hotz-Miller inversion or value function iterations. Finally, we present extensive Monte Carlo results on the moderately sized samples that Industrial Organization researchers face. Then the first stage requires non-parametric estimates at a number of states that is potentially larger than the number of underlying observations on markets. We show that our simpler methods work quite well in this case and explain why, while the other methods we consider, including estimators based

³The continuous control case is also considered by Berry and Pakes (2002) and Pakes, Porter and Wolfram (2003) who use “Muthian” ideas similar to those used in this paper to develop an estimator based on first order conditions for a dynamic game and for (multi-unit) auctions, respectively. Einav (2003) provides 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.

on likelihood functions, do not.

1 A Simple Entry/Exit Model.

We begin with a Markov Perfect Equilibrium (Maskin and Tirole, 1988) for 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 determined 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 sell off value is greater than profits plus the discounted continuation value. So if ϕ is the sell off (or exit) value and $0 < \delta < 1$ is the discount rate, the ‘‘Bellman’’ equation for the value of an incumbent is

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

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 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, $V(\cdot)$, or

$$VC(n, z; \theta) \equiv \sum_{e, x, z'} \int_{\phi'} V(n + e - x, z', \phi'; \theta) p(d\phi' | \theta) p^c(e, x | n, z, \chi = 1) p(z' | z). \quad (2)$$

Note that to form this expectation 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

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

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

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

$$VE(n, z; \theta) \equiv \sum_{e, x, z'} \int_{\phi'} V(n + e - x, z', \phi') p(d\phi' | \theta) p^e(e, x | n, z, \chi^e = 1) p(z' | z), \quad (3)$$

where

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

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 entry⁴.

We now list 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 $F^\kappa(r | \theta)$, which has a lower bound of $\underline{\kappa} > 0$, and*
- *the returns to exiting, say $F^\phi(\cdot | \theta)$, which are assumed nonnegative,*

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 asymmetric information, as in Seim, 2001).

⁴Note that we are not giving the potential entrant the possibility of waiting and then entering 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 entrants' 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 possibility appears, that possibility disappears.

2. Entrants' and incumbents' perceptions of the probabilities of exit and entry by their competitors in period t depend only on (n_t, z_t) (the publicly available information at that time).
3. The evolution of the profit shifters, z , is governed by the Markov chain $\mathcal{P}_z \equiv \{p(\cdot|z), \forall z \in Z = [0, 1, \dots, \bar{z}]\}$, $\lim_{n \rightarrow \infty} \pi(n, z) \leq 0$ for every $z \in Z$, and $\pi(\cdot)$ is bounded⁵.

We come back to a more detailed discussion of the restrictions implied by these assumptions directly after explaining the assumptions' implications.

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 Satterthwaite, 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), and there is an \bar{n} such that, provided the current n is lower than it, we will never observe an $n > \bar{n}$. The market is simply not profitable enough to induce entry if there are \bar{n} or more incumbents.

Indeed, one can go a bit further. For any 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 times⁶.

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 publicly available 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 (Starr and Ho, 1969).

⁵For simplicity, throughout we will assume that there are no unknown parameters in the $\pi(\cdot)$ function, and focus on estimating the parameters determining the entry and exit distributions; see the introduction.

⁶"Communicate" here simply means that the probability of transiting from one state to another (in any number of periods) is positive. If, in addition, 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 $\mathcal{R} = \{(n, z) : 0 \leq n \leq \bar{n}, z \in Z\}$. 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.

Now recall that the data will eventually wander into the recurrent subset of points (\mathcal{R}), and once in \mathcal{R} will visit each point in it repeatedly. As the sample gets large, we obtain an empirical distribution of entrants and exitors from each $(n, z) \in \mathcal{R}$, 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 there 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) ⁷.

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 there).

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

⁷There 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 suboptimal policies were followed. To fully analyze equilibria for subgames in \mathcal{R} , boundary points need to be treated separately (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.

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 beyond 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 needed to form continuation values are formed as

$$p^c(e, x|n, z, \chi = 1) = b^x(x, n - 1|n, z, \theta)p^c(e|n, z, \chi = 1) \quad (4)$$

where for $r \geq x$

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

and

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

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

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

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

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

where $\forall R \geq e$

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

Note that this implies that for incumbents

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

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 semi-parametric estimators for the parameters of the model, and we begin with the simplest of them. Its first step computes averages of 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, these averages will converge to the true expected continuation (entry) values 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 model's parameters by fitting the model's predictions for entry and exit conditional on alternative parameter values to the data on entry and exit rates.

Conditional on our estimates of entry and continuation, there are closed form expressions for the entry and exit rates predicted by the model. 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 is the reason the estimator does not have a significant computational burden.

We start with the only detailed calculation of the paper—that required to compute our estimates of $VC(\cdot)$ and $VE(\cdot)$. We then come back to the intuition underlying these estimators and a discussion of their implications for the estimation algorithm.

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

Consider an incumbent in a market with $n - 1$ other incumbents and particular value of z ; i.e. the market is in state (n, z) . If the incumbent decides to continue, his (expected) continuation value is equal to

$$VC(n, z; \theta) = E_{n', z'}^c [\pi(n', z') + \delta E_{\phi'} [\max\{VC(n', z'; \theta), \phi'\} | n', z']], \quad (6)$$

where

- n' and z' are the next period's values of (n, z) and ϕ' is the incumbent's draw on the exit value in that period, and

- $E_{n',z'}^c(\cdot)$ takes the expectation of the future state conditional on the incumbent itself continuing.

Given a realization of (n', z') the incumbent will exit if $\phi' > VC(n', z'; \theta)$, so the expectation of the continuation value from a realization of (n', z') is given by

$$E_{\phi'}[\max\{VC(n', z'; \theta), \phi'\} | n', z'] = \quad (7)$$

$$Pr\{\phi < VC(n', z'; \theta)\}VC(n', z'; \theta) + Pr\{\phi > VC(n', z'; \theta)\}E[\phi' | \phi' > VC(n', z'; \theta)].$$

To simplify this expression let

$$p^x(n', z') \equiv Pr\{\phi > VC(n', z'; \theta)\}$$

be the exit probability (this is an object we can estimate), and initially assume that ϕ distributes exponentially ($F(\phi) = 1 - e^{-(1/\sigma)\phi}$) so that

$$E[\phi | \phi > VC(n', z'; \theta)] = VC(n', z'; \theta) + \sigma,$$

(we generalize on this assumption below). Substituting these values into (7), and the result into (6) we get

$$\begin{aligned} VC(n, z; \theta) &= E_{n',z'}^c[\pi(n', z') + \delta(1 - p^x(n', z'))VC(n', z'; \theta) + \delta p^x(n', z')(VC(n', z'; \theta) + \sigma)] \\ &= E_{n',z'}^c[\pi(n', z') + \delta VC(n', z'; \theta) + \delta p^x(n', z')\sigma]. \end{aligned} \quad (8)$$

We now need some matrix notation. Arrange $VC(n, z; \theta)$ into the vector $VC(\theta)$, exit probabilities into the vector p^x , and incumbents' perceived transition probabilities into the matrix M_c . Then

$$VC(\theta) \equiv M_c[\pi + \delta VC(\theta) + \delta \sigma p^x] = M_c[\pi + \delta \sigma p^x] + \delta M_c VC(\theta). \quad (9)$$

Equation (9) computes $VC(\cdot)$ as the sum of expected current returns and itself (where now current returns include the expected excess returns from the possibility of exit). To solve for $VC(\theta)$ substitute the expression for $VC(\theta)$ in (9) into the right hand side of that same equation and iterate to get

$$\begin{aligned} VC(\theta) &= M_c[\pi + \delta \sigma p^x] + \delta M_c^2[\pi + \delta \sigma p^x] + \delta^2 M_c^3 VC(\theta) = \\ \dots &= \sum_{\tau=1}^{\infty} \delta^\tau M_c^\tau[\pi + \delta \sigma p^x] = [I - \delta M_c]^{-1} M_c[\pi + \delta \sigma p^x], \end{aligned} \quad (10)$$

the continuation value in terms of the expected discounted value of current returns.

That is continuation values can be computed by taking the expectation over the discounted future returns that the firm could earn over alternative possible future sample paths. This implies that we can obtain a consistent estimate of the continuation value by averaging over the discounted “future” returns actually earned by the firms who continued from state (n, z) . More precisely we will compute consistent estimates of the transition and exit probabilities, i.e. of M_c and p^x , and substitute them into (10). By the continuous mapping theorem this will generate a consistent estimate of VC .

Let

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

be the set of periods with the same (n, z) . Then, by the Markov property,

$$\tilde{p}^x(n, z) = \frac{1}{\#T(n, z)} \sum_{t \in T(n, z)} \frac{x_t}{n}$$

is a sum of (conditionally) independent draws on the exit probability and converges to $p^x(n, z)$ provided $\#T(n, z) \rightarrow \infty$.

Let $M_{c,(n,z),(n',z')}$ be an incumbent’s probability of transiting (in the next period) to state (n', z') , conditional on not exiting in state (n, z) , i.e. the element of matrix M_c in the row corresponding to state (n, z) and column corresponding to state (n', z') . Then, provided $\#T(n, z) \rightarrow \infty$, we can obtain consistent estimate of this probability as the fraction of incumbents not exiting in state (n, z) who transit to state (n', z') in the next period, that is

$$\tilde{M}_{c,(n,z),(n',z')} = \frac{\sum_{t \in T(n,z)} (n - x_t) 1_{[(n_{t+1}, z_{t+1}) = (n', z')]}]}{\sum_{t \in T(n,z)} (n - x_t)},$$

where $1_{[(n_{t+1}, z_{t+1}) = (n', z')]}$ is the indicator function which takes on the value of one when $(n_{t+1}, z_{t+1}) = (n', z')$ and zero elsewhere. Note that, to account for the fact that the incumbents condition their calculations on themselves continuing, we weight the transitions from (n, z) in the different periods by the number of incumbents who actually continue in those periods⁸.

Substituting these estimates into equation (10), we get our consistent estimate of VC as

$$\hat{VC}(\theta) = \sum_{\tau=1}^{\infty} \delta^\tau \tilde{M}_c^\tau [\pi + \delta \sigma \tilde{p}^x] = [I - \delta \tilde{M}_c]^{-1} \tilde{M}_c [\pi + \delta \sigma \tilde{p}^x]. \quad (11)$$

⁸There are alternative ways to get to this formula. From our equilibrium assumptions the unweighted transitions from (n, z) are generated by $b(x; n, z) \times b(e; n, z)$. The incumbent computes continuation values conditional on itself continuing, so it averages with $b(x; n_t - 1, z_t) \times b(e; n_t, z_t)$. As a result to obtain an unbiased estimate of the continuation values used by incumbents when they make their decisions we need to multiply each realization with $b(x; n_t - 1, z_t)/b(x; n_t, z_t) = [1 - (x_t/n_t)]/[1 - p_t^x]$ which is the weight above once we substitute $\tilde{p}_t^x = \sum_{t \in T(n,z)} x_t / \sum_{t \in T(n,z)} n_t$ for p_t^x .

2.2 Implications and Generalizations.

Note first that our estimates of continuation values are just the averages of the discounted values 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 makes the relationship between the data and the model transparent, thus simplifying robustness analysis. It is also the reason we expect our estimates to make empirical sense. Unless incumbents have perceptions that are systematically biased, the actual average of realized continuation values should be close to the expected continuation values used by the agents in making their decisions.

Second, note how easy it is to compute our estimates of continuation values, or $\hat{V}C(\theta)$. 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

$$\hat{V}C(\theta) = \tilde{A}\pi + \tilde{a}\sigma \quad (12)$$

for $\tilde{A} = [I - \delta\tilde{M}_c]^{-1}\tilde{M}_c$ and $\tilde{a} = \delta[I - \delta\tilde{M}_c]^{-1}\tilde{p}_x$. 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. For a potential entrant, the expected value of entry in state (n, z) is

$$VE(n, z; \theta) = E_{n', z'}^e [\pi(n', z') + \delta VC(n', z'; \theta) + \delta \sigma p^x(n', z')]$$

or, in matrix notation,

$$VE(\theta) = M_e(\pi + \delta VC(\theta) + \delta p^x \sigma), \quad (13)$$

where the elements of the matrix M_e , say $M_{e, (n, z), (n', z')}$, provide a potential entrant's probability of starting operations at state (n', z') conditional on it entering in state (n, z) .

Consistent estimates of these probabilities are obtained as the fraction of those who enter in state (n, z) who then begin operations state (n', z') , that is by⁹

$$\tilde{M}_{e, (n, z), (n', z')} = \frac{\sum_{t \in T(n, z)} (e_t) \mathbf{1}_{[(n_{t+1}, z_{t+1}) = (n', z')]}}{\sum_{t \in T(n, z)} (e_t)}.$$

⁹The weights are derived from the ratio of probabilities used by the potential entrant to form its expected entry value (these condition on the entrant entering) to the observed entry probabilities or $b^e(e-1, \mathcal{E}-1)/b^e(e, \mathcal{E})$, which can be written as $\frac{1}{\#T(n, z)} \sum_{t \in T(n, z)} \frac{[e_t/\mathcal{E}]}{[\tilde{p}^e(n, z)]} \mathbf{1}_{[(n_{t+1}, z_{t+1}) = (n', z')]}$, where $\tilde{p}^e(n, z) = \frac{1}{\#T(n, z)} \sum_{t \in T(n, z)} \frac{e_t}{\mathcal{E}}$.

Accordingly we obtain our consistent estimate of VE as

$$\hat{V}E(\theta) = \tilde{B}\pi + \tilde{b}\sigma, \quad (14)$$

where

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

The simplicity of the form of the solution for $(\hat{V}\mathcal{C}(\theta), \hat{V}E(\theta))$ did not depend *at all* on the distribution of entry costs. This fact generalizes to the model with multiple locations, and enables us to use realistic joint distributions of entry costs into the multiple locations of that model without increasing the computational burden of the estimator significantly.

On the other hand, those solutions can become somewhat more complex when the distribution of exit fees is not exponential. The property of the distribution of exit fees that enables the use of the matrix inversion is that $E[\phi|\phi > \phi_0]$ is linear in ϕ_0 (the exponential is a special case of this). Though this linearity assumption may be a good first approximation for the distribution of sell off values we would like to be able to generalize (at least for robustness analysis). If we use any other form for $F^\phi(\cdot)$ and repeat the logic in the last subsection, then after substituting our consistent estimates for their theoretical counterparts, the fixed point analogous to equation (10) becomes

$$\hat{\mathcal{V}}\mathcal{C}(\theta) = \tilde{M}_c \left\{ \pi + \delta \left[(1 - \tilde{p}^x) \times \hat{\mathcal{V}}\mathcal{C}(\theta) \right] + \delta \left[\tilde{p}^x \times E[\phi|\phi > \hat{\mathcal{V}}\mathcal{C}(\theta)] \right] \right\},$$

where $(1 - \tilde{p}^x) \times \hat{\mathcal{V}}\mathcal{C}(\theta)$ is the vector formed by multiplying each $1 - \tilde{p}^x(n, z)$ with the corresponding $\hat{\mathcal{V}}\mathcal{C}(\theta)(n, z)$, while $\tilde{p}^x \times E[\phi|\phi > \hat{\mathcal{V}}\mathcal{C}(\theta)]$ is the vector formed by multiplying each element of $\tilde{p}^x(n, z)$ by the corresponding $E[\phi|\phi > \hat{\mathcal{V}}\mathcal{C}(\theta)(n, z)]$. This equation system is contraction mapping, and therefore is easy to solve, if the derivative of $E[\phi|\phi > x]$ with respect to x is less than or equal to one everywhere (actually, $< 1/\delta$ will do). This will be true if the distribution $F^\phi(\cdot)$ is log-concave (see Heckman and Honore (1990), Proposition 1)¹⁰, an assumption satisfied by most of the distributions used in empirical work (normal, logistic, extreme value, gamma, beta, Weibull, and so on).

Once we have $(\hat{V}\mathcal{C}(\theta), \hat{V}E(\theta))$, we can form consistent estimates of the probability of exit and of entry conditional on θ as $(1 - F^\phi(\hat{V}\mathcal{C}(\theta)|\theta))$ and $F^\kappa(\hat{V}E(\theta)|\theta)$, respectively. The second stage of the algorithm fits these probabilities to the exit and entry rates observed in the data. We come back to a more detailed discussion of the properties of estimators which do this in section 3 below. For now all we want to note is that the computational complexity

¹⁰To see this let $T(x)$ be the operator that produces the right hand side of this equation when $\hat{V}\mathcal{C} = x$ and let $\|x\|$ denote the maximum element of the vector x . Then since M_c is a Markov matrix, $\|T(x_1) - T(x_2)\| \leq \delta \|[(1 - \tilde{p}^x) \times (x_1 - x_2)] + \delta [\tilde{p}^x \times (E[\phi|\phi > x_1] - E[\phi|\phi > x_2])]\|$. Under the log concave conditions $|E[\phi|\phi > x_1] - E[\phi|\phi > x_2]| \leq |x_1 - x_2|$ which, given that $0 \leq \tilde{p}^x \leq 1$, proves the result.

of these estimators 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).

We want to conclude this section by reminding the reader of an implication of our assumptions that may be problematic in particular applications. Our model assumes that the sample paths emanating from a given (n, z) have: (i) a distribution which does not depend on other features of the firms' environment and (ii) realizations which are independent of one another. It is for this reason that the average of these realizations converges to the continuation and entry values of interest. 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 determined sell off decisions. Of course if we observed repeated changes in regulations, and the process generating those changes were ergodic, then our averages would be consistent.

2.3 Multiple Locations and Random \mathcal{E} .

We generalize to allow for multiple entry locations and a random number of potential entrants. Allowing for multiple locations changes the entry model from a binomial to a multinomial model with the mutually exclusive and exhaustive outcomes being: enter in location 1, enter in location 2, \dots , 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 the number of potential entrants (or the size of the sample) for the multinomial draws. However, all the other aspects of the two-step estimation strategy remain intact, so the reader who is not interested in the details can skip this subsection and move directly to the section discussing alternative estimators.

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} ,*
- *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*

$$Pr\{\kappa_1 \leq r_1 \text{ and } \kappa_2 \leq r_2\} \equiv F^\kappa(r_1, r_2|\theta)$$

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 $F_1^\phi(\cdot|\theta)$ if the incumbent was in the first location and an i.i.d. draw from $F_2^\phi(\cdot|\theta)$ if the incumbent was in the second location.

Since κ_1 and κ_2 are draws on the entry costs of the same agent in alternative locations, the fact that our assumptions allow the two entry costs to be freely correlated adds realism to the model (see below).

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

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

where

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

and

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

provides the type l incumbent's 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 binomial distribution of exitors from each location, 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

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

where

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

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

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

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

where $n = (n_1, n_2)$ and

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

provides the equilibrium distribution of the number of entrants 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 than *both* not entering at all, and entering into location $-l$, i.e. *iff*

$$\delta VE_l(n_l, n_{-l}, z, \theta) \geq \kappa_l \quad \text{and} \quad \delta VE_l(n_l, n_{-l}, z, \theta) - \kappa_l \geq \delta VE_{-l}(n_{-l}, n_l, z, \theta) - \kappa_{-l}. \quad (15)$$

Using this fact, we find the equilibrium entry distribution in two steps: we first find the equilibrium entry distribution conditional on a particular number of potential entrants (on E), and then integrate 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

$$\begin{aligned} m_0 &\equiv Pr\{\kappa_1 > \delta VE_1(n_1, n_2, \cdot) \quad \text{and} \quad \kappa_2 > \delta VE_2(n_2, n_1, \cdot)\}, \\ m_1 &\equiv Pr\{\kappa_1 \leq \delta VE_1(n_1, n_2, \cdot) \quad \text{and} \quad \kappa_2 > \delta VE_2(n_2, n_1, \cdot) - \delta VE_1(n_1, n_2, \cdot) + \kappa_1\}, \quad \text{and} \\ m_2 &\equiv Pr\{\kappa_2 \leq \delta VE_2(n_2, n_1, \cdot) \quad \text{and} \quad \kappa_1 > \delta VE_1(n_1, n_2, \cdot) - \delta VE_2(n_2, n_1, \cdot) + \kappa_2\}, \end{aligned} \quad (16)$$

i.e. (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, then a potential entrant who conditions on $E - 1$ other potential entrants and enters in location l will set

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

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

Integrating out over the distribution of E , we obtain equilibrium perceptions as

$$p^{e,l}(e_l, e_{-l} | n, z, \chi_l^e = 1) = \sum_{E \geq (e_l + e_{-l})} m(e_l - 1, e_{-l}, E - 1; m_0, m_1, m_2) \frac{EP(E | \theta)}{\sum_E EP(E | \theta)}.^{11}$$

The incumbent's perceived entry probabilities are given by

$$p^{e,l}(e_l, e_{-l} | n_l, n_{-l}, z, \chi_l = 1) = p(e_l, e_{-l} | n, z, \theta) = \sum_{E \geq (e_l + e_{-l})} m(e_l, e_{-l}, E; m_0, m_1, m_2) P(E | \theta).$$

The adjustments to the estimation procedure required for the generalized model are also straightforward. Our first stage estimates of continuation and entry values are obtained by conditioning on the set of periods that have a particular value of (n_1, n_2, z) , computing weighted sample averages of the realized entry and continuation values from the two locations at each such state, and then applying the matrix inversion formula to simplify the calculation of entry and exit values in terms of the data and θ . Estimators of θ are derived by fitting the entry and exit rates from the different locations predicted by these continuation and entry values and different values of θ to the entry and exit rates in the data.

Consistent estimates of an incumbent's and an entrant's perceived transition probabilities from state (n_l, n_{-l}, z) to state (n'_l, n'_{-l}, z') are given by

$$\tilde{M}_{(n_l, n_{-l}, z), (n'_l, n'_{-l}, z')}^{e,l} = \frac{\sum_{t \in T(n_l, n_{-l}, z)} (n^l - x_t^l) 1_{[(n_{l,t+1}, n_{-l,t+1}, z_{t+1}) = (n'_l, n'_{-l}, z')]}]}{\sum_{t \in T(n_l, n_{-l}, z)} (n^l - x_t^l)}$$

and

$$\tilde{M}_{(n_l, n_{-l}, z), (n'_l, n'_{-l}, z')}^{e,l} = \frac{1}{\#T(n_l, n_{-l}, z)} \frac{\sum_{t \in T(n_l, n_{-l}, z)} e_t^l 1_{[(n_{l,t+1}, n_{-l,t+1}, z_{t+1}) = (n'_l, n'_{-l}, z')]}]}{\sum_{t \in T(n_l, n_{-l}, z)} e_t^l}.$$

¹¹ $\frac{EP(E | \theta)}{\sum_E EP(E | \theta)}$ is a potential entrant's perceived probability that there are $E - 1$ other potential entrants.

As before these numbers are not equal to the empirical frequency of transition from state (n_l, n_{-l}, z) to state (n'_l, n'_{-l}, z') or to $\#T(n_l, n_{-l}, z)^{-1} \sum_{t \in T(n_l, n_{-l}, z)} 1_{[(n_{l,t+1}, n_{-l,t+1}, z_{t+1}) = (n'_l, n'_{-l}, z')]$, but rather to a weighted average of these transitions. The weights, which account for the fact that the incumbent (potential entrant) computes continuation values conditional on continuing (entering), can be shown to equal $[1 - (x_t^l/n_l)]/[1 - \tilde{p}_l^x(n_l, n_{-l}, z)]$ and $e_t^l/[\hat{e}^l(n_l, n_{-l}, z)]$, where $\hat{e}^l(n_l, n_{-l}, z) = \#T(n_l, n_{-l}, z)^{-1} \sum_{t \in T(n_l, n_{-l}, z)} e_t^{l12}$. Appendix 1 derives the formula for the entry weights.

Two final points about the model with multiple locations. First, as in the single location model, given the matrix inversion formula for continuation values, the computational burden of obtaining estimates for the parameters of this model is minimal. There is, however, the burden of obtaining the Markov transition matrix and computing its inverse, which grows polynomially in the number of distinct states (and this increases in the number of locations). 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 it will be excessive if we increase the number of locations enough.

Second, and probably more important, though our estimators remain consistent when we increase the number of entry locations (or, for that matter, the number of states per location), their small sample properties will change. In particular, both the small sample bias and variance of our estimator will depend on the variances of the nonparametric component (a point we come back to in greater detail presently). 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 is likely to be. As a result, as we increase the number of states for a given sized sample we may have to worry more about 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. As we will see, the small sample biases are predictably more noticeable in some of the alternative estimators analyzed than in others.

¹²Using this expression for the weights the two formula are, respectively

$$\frac{1}{\#T(n_l, n_{-l}, z)} \sum_{t \in T(n_l, n_{-l}, z)} \frac{[1 - (x_t^l/n_l)]}{[1 - \tilde{p}_l^x(n_l, n_{-l}, z)]} 1_{[(n_{l,t+1}, n_{-l,t+1}, z_{t+1}) = (n'_l, n'_{-l}, z')]$$

and

$$\frac{1}{\#T(n_l, n_{-l}, z)} \sum_{t \in T(n_l, n_{-l}, z)} \frac{e_t^l}{[\hat{e}^l(n_l, n_{-l}, z)]} 1_{[(n_{l,t+1}, n_{-l,t+1}, z_{t+1}) = (n'_l, n'_{-l}, z')].$$

3 Alternative Two-Step Estimators.

The ideas discussed above can be combined and/or modified in different ways to obtain alternative estimators. This section will discuss several of these alternatives. Throughout we will assume that the number of observations at each of a finite set of (n, z) states grows without bound with sample size. As noted our modeling assumptions insure this would occur if the data came from a single long time series. It would also occur if our data consisted of a panel of different markets (or industries) with initial conditions restricted in a way that insures sufficient visits to a small number of states. Either set of conditions insure that all estimators to be introduced here are \sqrt{n} -consistent and asymptotically normal (or CAN).

The alternative estimators will have both different computational burdens and different distributions (large as well as small sample, though as we shall see this is a case where the properties of their small sample distributions are likely to be particularly important for applied work). After categorizing the alternative estimators and providing an informal theoretical discussion of their properties, we provide rather extensive Monte Carlo results on how they perform in two examples: a single and a two location example.

Two general points should be kept in mind while considering the alternatives. All our estimators are semi-parametric estimators with the properties that

- the nonparametric components (the Markov transition matrices and exit and entry probabilities needed to compute $\hat{V}C$ and $\hat{V}E$) enter the objective function in a non-linear way, and
- the variance of the parametric component (of θ) depends on the variance of the first stage nonparametric estimates.

The nonlinearity of the objective function in the first stage estimation errors implies that our parameter estimates are likely to be biased in small samples. Since, as we will see, in some samples the average number of observations per (n, z) “cell” can be as small as two or three, this small sample bias can be noticeable. That bias will differ (sometimes dramatically) with the extent of the nonlinearity induced by the form of the objective function, and since there are a number of alternative objective functions to choose from, one might want to take this into account in choosing among them¹³.

It is possible to use standard semi-parametric formula to obtain the asymptotic variance of our parameter estimates, but there is little reason to focus on those formula here. This is because we have a complete model and the computational burden of obtaining estimates

¹³An alternative would be to develop small sample bias corrections, a route we do not consider here. Also throughout we ignore the problem of developing tests of our model, even though it clearly is possible to develop such tests.

from that model is minimal. Thus it is relatively easy to obtain estimates of standard errors from a parametric bootstrap¹⁴. The fact that the asymptotic variances of the parameter estimates do depend on the variance of the nonparametric component (that the estimator is not “adaptive”) follows from the fact that the derivative of the objective function with respect to the estimates of $\hat{V}C$ and $\hat{V}E$ do not have a conditional expectation of zero.

Alternative CAN estimators can be obtained by varying each of the three different major components of the algorithm. We consider “natural” suggestions for each of these components. Estimator can be obtained by combining any suggestion for each of the three components. We consider all possibilities in the Monte Carlo examples. Here is a listing of the components and our suggestions for each.

1. The objective function used in the second stage. In this context we consider
 - (a) the pseudo log-likelihood function, which in the single location model would be formed by summing

$$l(x_t, e_t|\theta) = (n_t - x_t) \log F^\phi\{\hat{V}C_t(\theta)|\theta\} + x_t \log [1 - F^\phi\{\hat{V}C_t(\theta)|\theta\}] \\ + e_t \log F^\kappa\{\hat{V}E_t(\theta)|\theta\} + (\mathcal{E} - e_t) \log [1 - F^\kappa\{\hat{V}E_t(\theta)|\theta\}]$$

over t , and

- (b) a method of moment estimators that minimizes a norm in the difference between the data on the state specific entry and exit rates and the entry and exit rates predicted by the model for different values of θ (and so might be called a pseudo minimum χ^2 estimator), and
- (c) a method of moment estimators that minimizes a norm in the (average over all states) of the difference between the actual entry and exit rates and the entrants/exits predicted by the model for different values of θ ¹⁵.

2. The estimation of the transition probabilities between states. In this context we consider

¹⁴Since we have estimates of continuation and entry values conditional on (n, z) , we can use our estimates of θ and $p(z'|z)$ to generate independent samples of size equal to our sample size. To obtain the parametric bootstrap we simply estimate θ from each of these pseudo random samples and take the variance of those estimates.

¹⁵In our Monte Carlo examples the number of parameters equals the number of location specific entry and exit rates so that these moments will just identify the parameters. If there were more parameters we would have to add covariances between the prediction errors and the value of the state variables to identify the model.

- (a) using the empirical Markov matrix (as above), and
 - (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 function 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 (see below for details).

Finally note that given any one of these estimators we could always iterate on to a multi-stage estimator. The second stage uses the first stage parameter estimates to compute entry and exit values which are used in conjunction with the first stage estimates are to compute entry and exit probabilities, which, in turn, are used to compute structural transition probabilities analogous to those 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 Aguirregabiria and Mira (2003) noted in our introduction is, in our terminology, a pseudo maximum likelihood estimator as in (1a) that uses the structural transition matrices in (2b) and the nested fixed point to compute entry and continuation values as in (3b), and they favor a multi-step version of their estimator which iterates in this way¹⁶. We should note that there is no guarantee that the iterations improve the estimates or that they will converge to anything (we come back to both these points below), or, if they do converge, they converge to something that is consistent with our assumptions on the choice of equilibrium.

3.1 Comments on the Alternative Estimators.

Alternative Objective Functions.

We already noted that the variance of all of our estimators will be determined, in part, by the variance of the first step estimators of $VC(\cdot)$ and $VE(\cdot)$. As a result the pseudo maximum

¹⁶Aguirregabiria and Mira require i.i.d. extreme value distributions for both the entry fees and the sell off values (as in Rust,1987). This assumption implies that the entry costs for the same agent in different locations are independent and have full support (i.e. they will be large negative numbers with some probability).

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 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).

So 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 the sensitivity of the pseudo maximum likelihood estimator to the estimation error in the continuation and exit values is determined by the derivative of the log of the probabilities of exit and entry with respect to these values; i.e. by one over the true probabilities. When those probabilities are small, and they often are in entry/exit data sets, that derivative will be very large and maximum likelihood will tend to produce estimators with poor finite sample performance.

Second, and conceptually quite distinct, since the pseudo likelihood's probabilities are not the true probabilities (they are conditioned on estimated and not the true entry and exit values), 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. Our examples below illustrate that this is likely to happen even in relatively simple models.

Moving to the “pseudo minimum χ^2 ” objective function in (1b) it, like the log-likelihood, is also likely to have problems with nonlinearities in its objective function. To see this, let s index states, $\hat{F}_s^\kappa(\theta) = F^\kappa(\hat{V}E_s(\theta))$ index estimated entry probabilities, and note that the first order condition for the entry parameters in the the pseudo minimum χ^2 estimator is a weighted average of

$$[\hat{g}_e(s) - \hat{F}_s^\kappa(\theta)] \frac{\partial \hat{F}_s^\kappa(\theta)}{\partial \theta} \equiv$$

$$[\hat{g}_e(s) - F_s^\kappa(\theta)] \frac{\partial \hat{F}_s^\kappa(\theta)}{\partial \theta} + [F_s^\kappa(\theta) - E\hat{F}_s^\kappa(\theta)] \frac{\partial \hat{F}_s^\kappa(\theta)}{\partial \theta} + [E\hat{F}_s^\kappa(\theta) - \hat{F}_s^\kappa(\theta)] \frac{\partial \hat{F}_s^\kappa(\theta)}{\partial \theta},$$

where E integrates over sampling variance of the estimates of the entry and exit values. At $\theta = \theta_0$, the true value of θ , the expectation of the first term in the last equation is zero. However since $\hat{F}^\kappa(\cdot)$ is a nonlinear function of the first stage estimation error, and $\hat{F}^\kappa(\cdot)$ and $\frac{\partial \hat{F}^\kappa(\cdot)}{\partial \theta}$ are constructed from the same estimates of $\hat{V}E$ (and hence are correlated), both of

the last two terms have a nonzero expectation. As a result a value of $\theta = \theta_0$ should not be expected to produce a minimum to the first order conditions (at least in finite samples).

The fact that our first stage estimates of entry values contain errors implies that an analogue to the bias caused by the first of the last two terms will be present in all method of moments estimators based on the difference between the observed and our estimates of the entry rates. However the bias caused by the last term is a result of the fact that pseudo minimum χ^2 is using an “instrument” (i.e. $\frac{\partial \hat{F}_s^k(\theta)}{\partial \theta}$) which is correlated with the error in the estimate of the probability. Thus if we replaced $\frac{\partial \hat{F}_s^k(\theta)}{\partial \theta}$ in the top equation with any known function of the observed state variables, the last term would have expectation zero for all θ and s . Consequently it would average out across states and tend not to produce a problem in our estimators¹⁷. The simpler method of moment’s estimator in (1c) is a special case in which the “instrument” becomes “one” for all observations, and hence it has one less “bias” term than the pseudo minimum χ^2 estimator in (1b).

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 entering and exiting at that state or by

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

Then use the the binomial formula

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

and

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

together with the Markov process generating z , to compute the probabilities of $(n' = n - x + e, z')$ given (n, z) from the point of view of an incumbent. Analogous expressions are used to compute transition probabilities from the point of view of a potential entrant.

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

¹⁷The argument here is very similar to the argument against using non-linear least squares to estimate regression functions when the regression function itself is simulated with a finite number of simulation draws; see Laffont, Ossard and Vuong, 1995.

in the data. To compute the first stage estimates of $VC(\cdot)$ and $VE(\cdot)$ we will then have to impute the entry and exit rates from those states (since we do not observe transitions from these states), and the imputation errors will affect our estimators¹⁸. Second, to go from the binomial probabilities to the probabilities needed for the transitions from n to n' requires the computation 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 $\hat{V}C(\cdot)$ and $\hat{V}E(\cdot)$.

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

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

where

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

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

At least in cases where the matrix inversion formula in (9) is available, the nested fixed point calculation will 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 sell off values or profits in the search algorithm¹⁹.

The extra computational burden of the fixed point grows exponentially in the number of state variables (or locations)²⁰ making it harder to use in more complex problems, and will increase in the number of parameters to be estimated (as this will typically require the estimation algorithm to do more function evaluations in its search procedure). The extra

¹⁸There 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.

¹⁹If we assume the sell off value distributes i.i.d. type II extreme value as in Aguirregabiria and Mira, 2003, the integral over ϕ has an analytic form which makes each calculation of the fixed point easier.

²⁰Actually 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

burden will also be compounded when we use it in combination with the structural transition matrix, as this will require us to calculate the fixed points at more states.

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.

Throughout 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 examples 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 (our “C”) between C=250, C=500, and C=1000. Much of the earlier work on entry was on relatively isolated markets (see Bresnahan and Reiss, 1987, and Mazzeo, 2002). The number of relatively isolated markets in South Dakota is 250, which explains our smallest sample size, and the support of our z variable was chosen to give us the approximate range of population sizes in the South Dakota markets.

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.

4.1 The 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

$$\pi(n, z) = 2 \frac{Z^2}{(1+n)^2},$$

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

$$f(\kappa = r) = a^2(r - 1/a)\exp[-a(r - 1/a)] \quad (17)$$

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 . 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 .

An 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). State variable z was allowed to take on 45 values at .05 increments and we allowed three growth rates $(.05, 0, -.05)$ ²¹. 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, g, z) vectors possible, is about two thousand seven hundred. The “data” was obtained by using the computed equilibrium policies to simulate sample paths. The same data was used for all estimators.

4.1.1 Results from the Single Location Model.

Table 1 provides the results. The first panel of the table is a “pivot” table which defines the estimators in the different columns. The first row of that panel specifies the objective function (OF) used. If $OF = 0$ we fit the mean (over all observations) of the entry and exit probabilities predicted by the model to the data, if $OF = 1$ we use the pseudo likelihood, and if $OF = 2$ we minimize the sum of squares of the difference between the empirical and the estimates of the state specific entry probabilities weighted by the inverse of the number of times that the data visited that state. The PR row of the pivot table indicates which first stage probabilities are used in estimation. If $PR = 1$ we used “structural” transition probabilities (we estimate entry and exit probabilities and use them and the binomial formula to

²¹When $g_t = .05$, $g_{t+1} = .05$ 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.

compute transition probabilities), whereas if $PR = 0$ we used the empirical transition probabilities (or \tilde{M} in the notation above). The VF row 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 when $C=1000$ and $T=15$. These results are reassuring since they indicate that, at least in the one location model, with a big enough sample all estimators “work” reasonably well (though there is some indication in this panel that pseudo maximum likelihood does worse than the other estimators)²². 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 in panel B seem to have an upward bias and the bias appears larger in the $OF=1$ (or pseudo mle) estimators. 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 quite precise (the exception here is the $OF=2$ estimate of σ which sometimes is not well estimated). Moreover as Panel C shows, once we increase the size of the data set the bias problem disappears rather rapidly.

An upward bias in a implies a downward bias in the estimate of $1/a$. For the $OF=1$ estimator or pseudo mle estimator the estimate of $1/a$ has to be lower than the lowest estimated entry value at which some potential entrant entered. In small samples the minimum estimated entry value will tend to be lower than the minimum true entry value, and will tend to make $\hat{a} > a$. One can use a second order expansion to look at the small sample bias in the $OF = 0$ estimator of a and show 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.

No matter the sample size the $OF = 0$ (or simple method of moments) estimators had both smaller biases *and* smaller variances than those of either the pseudo mle ($OF = 1$), or the pseudo minimum χ^2 (or $OF = 2$) estimators. Moreover this was true for all sample designs. For intuition on why this occurs compare the row labeled $\#(n, g, z)$, which provides

²²As 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 $n \in [n_2(z), max]$ 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.

the number of states ever visited in the run (423 when $T=5$ and 629 when $T=15$), to the number of observations (or $C \times T$, 1250 when $T=5$ and 3750 when $T=15$). That is when $T=5$ we are constructing the rows of the empirical transition probabilities which determine the Markov transition matrix, and thus through the inverse formula the continuation and entry values, with three observed transitions (on average). As a result one might expect the estimates of continuation and entry values to contain nontrivial sampling error, and as explained in the last section that error is likely to have much larger impacts on the pseudo maximum likelihood and the pseudo minimum χ^2 estimators than on the $OF=0$ estimator.

What is striking is just how well the $OF=0$ estimator does with this number of observations. At $T=15$ there is an average of six observed transitions per state visited and that seems to be enough to “nail” the parameters, and with three transitions per state (at $T=5$) we do reasonably well. Of course, each observed transition averages over all sample points reached from the state it transitioned to, so *each* observed transition is implicitly *averaging* over all sample paths from the point transitioned to.

The differences in the distributions generated by the alternative $OF = 0$ estimators are small. 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 $\#(n, g, z)$ 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 probabilities²³. 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. This is true even though the imputed states are well over eighty percent of the total number of states. 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 (the $OF = 0$ estimators) have both better distributions and impose less of a computational burden. There is not much difference in the distributions of the $OF = 0$ estimators, and since the computational burden of the $PR = 1$ estimator is so much larger, there is a clear preference for estimators that set $OF = PR = 0$ estimator. There is very little difference in either performance or in computational burden between the simplest estimator, the estimator with $OF = PR = VF = 0$, and the estimator which uses $OF = PR = 0$ but the nested

²³These ratios are much worse for the $OF=1$ and the $OF=2$ estimators (up to a factor of 15).

fixed point for calculating values ($VF = 1$).

As noted, were we doing a more complex problem or estimating more parameters, we would see more of a computational preference for the $OF = PR = VF = 0$ over the $OF = PR = 0$ but $VF = 1$ estimator. On the other hand, there does seem to be a slight improvement in the statistical performance of the estimator with $VF = 1$ over $VF = 0$ in the smallest sample (panel B), and this held up in the limited number of runs we did with smaller sample sizes than those presented here. So it is probably the case that the choice between the $OF = PR = VF = 0$ and the $OF = PR = 0$ but $VF = 1$ estimators depends on the characteristics of the problem and the sample being analyzed. The other important point to note is that neither of these two estimators is particularly computationally burdensome; the compute time with our smallest sample size averaged about ten seconds, and it was only about a half a minute on our largest sample. I.e. the framework should enable researchers to analyze much more complex problems and/or engage in a quite bit of robustness analysis.

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 consumes one of the goods marketed, 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

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

where c_k is the marginal cost of product k , p_k is its equilibrium price, and c_k^f is its fixed cost. 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 $F^\kappa(\cdot, \cdot | \theta)$ 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

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

In particular the cost of entry into the low quality product distributes as does the entry cost in the one location model (see equation 17) with parameter a_1 while the cost of entry into the high quality product is given by

$$\kappa_2 = \kappa_1 + r$$

where the distribution of r is given by equation (17) with parameter a_2 ²⁴. 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. 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.

4.2.1 Results from the Two Location Example.

The Monte Carlo results for large samples are split between table 2 (for the OF=1 or pseudo mle estimators) and table 3 (for the OF=0 and OF=2 estimators). There is a "start" row in the pivot table for the some of the two location runs; when start=1 we start the search from the $OF = PR = VF = 0$ estimator²⁵.

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" at all²⁶. The answer is pseudo mle *does not* work. We labeled a search starting from a certain point "unsuccessful" when the Nelder-Mead search algorithm in Matlab could not find a positive value for the likelihood. When there was an unsuccessful initial condition for

²⁴Actually we use a discretized version of the density in equation (17) for the r distribution.

²⁵It 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.

²⁶Since 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 $R = 14$.

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 $\hat{V}E_1(\cdot) > \hat{V}E_2(\cdot)$ 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. 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. Note that all this has to do is happen once among the estimates of the entry values for all the states visited for the pseudo mle to fail.

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 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). 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 follows²⁷.

Table 3 provides the estimators obtained from OF=0 and OF=2 when $T = 15$ and $C = 1000$. The striking point from this table is that when we use $OF = 2$ (we use the pseudo minimum χ^2 estimator) the estimates of σ 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 less incumbents in each state (since they are now split between two locations). The actual outcomes for each transition behave like a multinomial with n_j draws, and with n_j smaller they have more variance. As a result, even with relatively large samples, the preliminary estimators of continuation and entry values can be quite noisy. That noise, when combined with the accentuation of the error

²⁷Though 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.

that results from the functional form of the objective function in the OF=2 estimators(see the discussion in section 3.2)²⁸, makes those estimators problematic. 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.

On the other hand, with this sample size all the $OF = 0$ estimators do fairly well. The estimates of σ that use structural probabilities (i.e. OF=0, but PR=1) do seem to have larger 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 (compare $\#(n, g, z)$ to $\#\hat{p}$ for these columns). More striking, however, is the 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. On the other hand the estimates which use the empirical probabilities and value function iterations are only about 10% more burdensome than those that use the matrix inversion and the empirical probabilities (though, as noted, this would increase were we also estimating profit function parameters).

Table 4 provides the estimates from smaller samples. Since the number of incumbents per transition is about a third of what it was in a similar sized sample in the one-location model (on average, there are around 6 incumbents in the one-location example and around 2 incumbents in each location in the two-location case), we should not be surprised when we see larger small sample biases and larger standard errors than 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.

Comparing the various $OF = 0$ estimators we see that 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 σ '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. Since the problem when using the empirical probabilities is in the precision of the first stage estimates, and the problem in the structural probabilities is compounded by the fact that it has to impute entry and exit rates for about 85% of the states it uses, this result should have been expected. On the other hand it implies that even at ($T = 5, C = 500$) it is pretty clear that we prefer estimators based on empirical probabilities, and this conclusion is reinforced at larger sample sizes. Moreover the computational burden of the

²⁸We note that the results on the OF=2 estimator were not due to a few outliers. That is the actual Monte Carlo distribution of the OF=2 estimators did not contain a single or a few estimators that drove the variance results.

estimators which use structural probabilities is ten or more times that of the corresponding estimators which use the empirical probabilities.

Comparing the two estimators which use $OF = PR = 0$ we find that in the smallest two samples it is clear that we prefer the estimator which does the matrix inversion ($VF = 0$) to the estimator which uses the nested fixed point ($VF = 1$). However by the time we get to a sample size of $C = 500$ and $T = 15$, when both estimators are doing reasonably well, the nested fixed point estimator seems to have marginally better performance. Once again the $VF = 1$ estimator is only minimally more burdensome than the $VF = 0$ estimator, a finding which is likely to change were we estimating more parameters or a more complex model.

We tried to improve the estimates from the smallest sample by iterating on them. The iterated estimators used the $OF = 0$ but $PR = VF = 1$ estimators to construct the exit and entry probabilities those estimates implied. Given these probabilities we use the $OF = 0$ but $PR = VF = 1$ estimation algorithm to obtain the iterated values of θ ²⁹. Table 5 provides results from iterating in this way one to four times.

Comparing the (0, 1, 1) column in Table 4 to the first iteration estimator in Table 5 we see that for our smallest sample size the estimates of the a 's worsen but the estimates of the σ 's improve. Iterating the second time moves us approximately back to the (0, 1, 1) estimators for the a 's but now the estimators of the σ 's get much worse. When we move to the third and fourth iterations we find that the estimators seem to oscillate, and the oscillations seem to be of increasing magnitude, moving us farther away from the truth. When we look at the slightly larger sample with $C = 500$ and $T = 5$ there are no apparent advantages from iterating at all.

Since the iterations need not converge, perhaps the oscillations should not be a surprise. More fundamentally, since we are no longer holding the transition probabilities fixed, we can no longer rule out multiple equilibria for a given value of the parameter vector. As a result it is possible for the algorithm to find a value for the parameter vector that optimizes the objective function for a different equilibrium than the one that generated the data. Moreover, the computational burden of the iterated estimators is huge: even the single iteration makes them twenty times more burdensome than the $OR = PR = VF = 0$ estimators, and when we iterate five times they are over one hundred times more burdensome. There seems not to be any argument for iterating at all.

Indeed if there is any argument for using anything but the simplest estimator it is only for our smallest sample sizes and then only if the researcher is interested in particular parameters more than others. On the other hand, even with moderate sample sizes the simple estimators

²⁹That is we constructed the implied structural probabilities, and then used the fixed point to get the continuation values generated by these structural transition probabilities and different values of θ to minimize the $OF=0$ objective function.

both perform reasonably well and have a minimal computational burden: under a half a minute for the smallest sample size and just over a minute for the largest sample.

4.3 Monte Carlo: Summary and Extensions.

The results from our Monte Carlo experiments are unusually clear cut. We prefer estimators based on (i) the simplest objective functions (moments that multiply state specific differences between observed and empirical entry and exit rates and a deterministic function of the state), (ii) the empirical transition matrix, and, if available, (iii) the matrix inversion. What is striking is that estimators with these properties not only seem to have better small sample distributions, they are also the least computationally burdensome.

There is one further advantage of these estimators. In Monte Carlo work the computational burden associated with estimating on repeated samples induces researchers (including us) to stick with relatively simple estimation problems. The hope, however, is that the estimators can be used in empirical work that analyzes more complex environments, in our case environments in which there are more state variables (more locations or more states per location), and perhaps more unknown parameters.

The computational burden associated with an increase in the number of state variables depends on how the number of first stage estimates of the discounted continuation and entry values increases in the number of state variables. In models that use empirical transition probabilities this grows as does the size of the recurrent class, which in I.O. models tends to be linearly in the number of state variables (see Pakes and McGuire, 2001). If we were forced to use structural transition probabilities the number of states would grow exponentially in the number of state variables.

The increased computational burden associated with estimating more parameters equals the increase in the number of times we have to evaluate the objective function at different θ in larger dimensional searches times the computational burden of doing each evaluation. When we use the nested fixed point algorithm, each time we evaluate the objective function we need to compute a fixed point, while all we need to do with the matrix inversion is multiply a known inverse times a known function of the parameters.

All this bodes well for future empirical research as it implies that the effective limitation on analyzing entry and exit fees should now be much more closely associated with what the data can support than what can be computed. However, there is at least one more issue of importance to empirical work, and it may well be more important.

In empirical work we typically worry about the specification of the model estimated as much as about distributional or computational properties conditional on the chosen model. The light computational burden of the estimators presented here should facilitate robustness analysis with respect to many aspects of our model's specification (the importance of different

observed state variables, the number of locations needed, etc). However, as noted earlier, there is at least one aspect of the specification, the possible presence of serially correlated unobserved state variables, that the estimators presented here can not accommodate. We noted above that since our estimates of entry and continuation values are approximations to the average entry and continuation values actually earned, there was some reason to believe that provided incumbents and potential entrants were not systematically biased in their evaluations, our estimators of entry and exit costs should not be too far off the mark even if there was a misspecification of this sort.

To illustrate we conclude with a Monte Carlo investigation of the sensitivity of the parameter estimates to the presence of serially correlated unobserved state variables. We proceed as follows. First we compute the equilibrium of a model that has four serially correlated state variables. We then use the equilibrium policies to simulate data from industries in which these four state variables determined behavior. We then “pretend” that this data was generated from a model with only three state variables, and use our estimators assuming that the *misspecified* three state variable model generated this data.

The details are as follows. The equilibrium model is our two location model except that now we assume that the whole distribution of entry and exit fees is shifted up and down over time with local market conditions that the estimator does not control for. In particular, we assume that the distributions of entry and exit fees are subjected to a simultaneous serially correlated shock which takes on three values: a positive (negative) shock which increases (decreases) both entry and exit fees by 25%, and zero. The positive (negative) shock is followed by the same value with probability .75 and returns to zero with probability .25, while a value of zero is followed by a zero with probability .5 and moves to each of the other values with probability .25.

In this specification the possible values of the shocks have large impacts on entry and exit behavior. A positive shock yields exit rates which increase by about 40% in one market and 50% in the other, and going from a positive to a negative shock doubles the exit rates and cuts the entry rate to 1/4 to 1/5 of its original value. The average value of the shock is zero, however, so we can interpret what the misspecified model estimates as the average value of the entry and exit fees.

Table 6 presents the results from fifty repetitions of the estimator on a very large sample. They are striking: the serially correlated unobserved state variable results in asymptotic biases in the parameter estimates of at most one or two per cent. There is some intuition for this result. It is clear that if the continuation values were obtained as the average over realized sample paths of the discounted value of returns for continuing conditional on the three states (i.e. on our “observed” states) even the misspecified model would produce a “true” average of the continuation values. Here we are defining the true average conditional on the observed states to be the true value conditional on the four states (on the unobserved

as well as the observed state) averaged over the limiting distribution of the unobserved state conditional on the observed states.

We still, however, should not expect to get consistent parameter estimates. What the (misspecified) objective function fits is the average fraction of exitors from the observed states to the probability of obtaining an exit fee greater than our estimate of the average continuation value from those states. In fact the average probability of exit from the observed states is an average over the average probability of exit from the unobserved states conditional on the observed states. Since the probability of exit is itself a nonlinear function of the continuation value conditional on the unobserved states, when we average over it we do not get the probability of the exit fee being greater than the average continuation value. As a result even if we do have the “true” average continuation values, the model should deliver inconsistent parameter estimates. Apparently, however, with our specification the extent of the asymptotic bias in the misspecified estimators is small (and we have tried several more runs than those reported in table 6 always getting this same result). We leave an investigation of why this is the case in our example, and when it is likely to be true in other examples, for future research.

5 Conclusions.

This paper provided estimators for the parameters of discrete dynamic games that are easy to use, and then 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. If the conditional expectation of the exit fee, conditional on the exit fee being greater than the continuation value, is linear in the continuation value (as it is in the exponential case), then 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 (a result which does not depend on the distribution of entry costs). This makes the computational burden of the estimator similar to the burden of estimating a multinomial model in probabilities which are known functions of the data. For richer distributions of exit fees our estimator is a nested fixed point estimator, but the fixed point is a contraction mapping and need not be computed when we vary the entry fee distribution.

Given these ideas, a number of alternative estimators for our semi-parametric model suggested themselves. A theoretical discussion showed that the fact that the multinomial

probabilities that go into the various objective functions being minimized in the estimation algorithms are not known functions of the data and parameter vector, but rather semi-parametric estimates of those functions, affects the distributions of the alternative estimators in different ways. Moreover, the computational burdens of the alternatives also varied (sometimes markedly). As a result, Monte Carlo examples were designed to push the investigation of the computational and distributional properties of the estimators further.

The results from the Monte Carlo examples showed that the simplest estimators also tended to have the preferred distributions. Moreover, those estimators performed quite well in reasonably sized samples, and had minimal computational burdens. The hope is that now the effective barrier to the empirical analysis of these problems can be shifted from being the computational burden of obtaining the estimates, to the richness of the data available to support the analysis.

References.

- Aguirregabiria, V. and P. Mira (2003); "Identification and Estimation of Dynamic Discrete Games", *mimeo* Boson 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.
- Bresnahan, 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.
- Doraszelski, U. and M. Satterthwaite (2003); "Foundations of Markov Perfect Equilibrium: Existence, Purification, and Uniqueness" *mimeo* Northwestern University and the Hoover Institution.
- 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.

- Heckman, J. and B. Honore (1990); "The Empirical Content of the Roy Model" *Econometrica* (58) pp.1121-49.
- 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.
- Magnac T., and D. Thesmar (2002); "Identifying Dynamic Discrete Decision Processes", *Econometrica*(70) pp. 801-16.
- Maskin, E. and J. Tirole (1988); "A Theory of Dynamic Oligopoly, I and II; *Econometrica*, vol. 56, no. 3, pp. 549-99.
- Mazzeo, M. (2002); "Product Choice and Oligopoly Market Structure", *The Rand Journal of Economics*. Summer, pp.221-42.
- Muth, J. (1961); "Rational Expectations and the Theory of Price Forecasts", *Econometrica*.
- 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.
- Seim, K.(2002);"Geographic Differentiation and Firms' Entry Decisions: The Video Retail Industry", *mimeo*, GSB, Stanford.
- Starr, A.W. and Y.C. Ho (1969); "Nonzero-Sum Differential Games", *Journal of Optimization Theory and Applications*, pp.1984-208.

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
$\#(n, g, z)$	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
$\#(n, g, z)$	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. $\#(n, g, z)$ is the number of states visited, while $\#\hat{p}$ is the number of states for which we must compute probabilities when $PR = 1$. C and T are the cross sectional and time dimensions of the panel and R is the number of monte carlo repetitions.

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=5000. "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(σ_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(σ_2)	0.08	0.13	0.11	0.09	0.15	0.14

* Legend. See the footnote to Table 1.

Table 3: Two Locations.

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
$\sigma_1=1$	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(σ_1)	0.09	0.07	0.10	0.07	31.13	26.22	0.83	2.44	33.02	36.22	3.20	3.61
$\sigma_2=.5$	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(σ_2)	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
$\#(n, g, z)$	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 values for this estimator are the estimates for the simplest model (OF=PR=VF=0).

Table 4: Two Locations.

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
$\sigma_1=1$	1.29	1.99	2.55	2.27	1.09	1.14	1.82	1.68
SD(σ_1)	1.82	2.72	3.10	2.11	0.41	1.18	0.76	1.18
$\sigma_2=.5$	0.57	2.96	1.11	1.45	0.61	0.77	0.75	0.65
SD(σ_2)	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
$\#(n, g, z)$	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
$\sigma_1=1$	1.09	1.08	1.33	1.15	1.05	1.01	1.17	1.07
SD(σ_1)	0.35	0.77	0.38	0.28	0.10	0.08	0.18	0.10
$\sigma_2=.5$	0.54	0.51	0.61	0.55	0.51	0.50	0.55	0.52
SD(σ_2)	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
$\#(n, g, z)$	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 From OF=0, PR = 1, VF = 1.

Pivot Table*

Iteration	1	2	3	4
Estimates	T=5, C=250, R=50			
a1=.3	0.21	0.28	0.15	0.26
SD(a1)	0.04	0.03	0.04	0.04
a2=.3	0.38	0.29	0.39	0.30
SD(a2)	0.24	0.06	0.41	0.08
$\sigma_1=1$	1.13	7.21	1.43	7.53
SD(σ_1)	0.31	8.26	0.97	5.97
$\sigma_2=.5$	0.54	2.53	0.62	5.46
SD(σ_2)	0.14	2.44	0.21	8.07
time setup	351	351	349	352
time search	183	286	204	326
time cumulative	1088	1725	2278	2956
Estimates	T=5, C=500, R=50			
a1=.3	0.26	0.28	0.19	0.25
SD(a1)	0.03	0.03	0.05	0.03
a2=.3	0.33	0.31	0.33	0.33
SD(a2)	0.13	0.05	0.23	0.07
$\sigma_1=$	1.08	3.75	1.81	7.50
SD(σ_1)	0.19	3.15	2.01	5.38
$\sigma_2=.5$	0.54	1.93	0.64	4.58
SD(σ_2)	0.12	2.22	0.20	5.27
time setup	468	467	466	468
time search	202	334	256	406
time cumulative	1360	2161	2883	3757

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

Table 6: “Misspecified” Model.

Pivot Table*

OF	0	0	0	0
PR	0	0	1	1
VF	0	1	0	1
Data	T=15, C=5000, R=50			
a1=.3	0.31	0.31	0.30	0.30
SD(a1)	0.00	0.00	0.00	0.00
a2=.3	0.31	0.31	0.30	0.30
SD(a2)	0.01	0.01	0.01	0.01
$\sigma_1=1$	1.02	1.01	1.03	1.02
SD(σ_1)	0.03	0.03	0.03	0.03
$\sigma_2=.5$	0.50	0.50	0.51	0.50
SD(σ_2)	0.01	0.01	0.02	0.01

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

Appendix 1. Entry Weights With a Random Number of 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,

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

where

$$w^{e,l} = \frac{e_l}{\bar{e}_l(n_l, n_{-l}, z)} = \frac{e_l}{\sum_E m_l(n_l, n_{-l}, z) EP(E|\theta)},$$

and is consistently estimated by substituting

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

in the above formula. Recall that

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

Proof. From the text,

$$\begin{aligned} p^{e,l}(e_l, e_{-l}|n_l, n_{-l}, z, \chi_l^e = 1, \theta) &= \sum_{E \geq (e_l + e_{-l})} m(e_l - 1, e_{-l}, E - 1; m_0, m_1, m_2) \frac{Ep(E|\theta)}{\sum Ep(E|\theta)} \\ &= \frac{1}{\sum Ep(E|\theta)} \sum_{E \geq e_l + e_{-l}} \frac{(E-1)! \times E}{(E - e_l - e_{-l})!(e_l - 1)!e_{-l}!} m_0^{E - e_l - e_{-l}} m_l^{e_l - 1} m_{-l}^{e_{-l}} p(E|\theta). \end{aligned}$$

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 $m_l^{e_l - 1} \times m_l = m_l^{e_l}$, the equation is is equal to

$$\begin{aligned} &\frac{e_l}{m_l \sum Ep(E|\theta)} \sum_{E \geq e_l + e_{-l}} \frac{E!}{(E - e_l - e_{-l})!(e_l - 1)!e_{-l}!} m_0^{E - e_l - e_{-l}} m_l^{e_l} m_{-l}^{e_{-l}} p(E|\theta) \\ &= w^{e,l} \sum_{E \geq (e_l + e_{-l})} m(e_l, e_{-l}, E; m_0, m_1, m_2) p(E|\theta) \\ &= w^{e,l} p(e_l, e_{-l}|n_l, n_{-l}, z), \end{aligned}$$

as desired. ♠