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, June 2005.)

Abstract

This paper considers the problem of estimating the parameters determining behavior in discrete dynamic games, focusing on models where the goal is to learn about the distribution of firms' entry and exit costs. The idea is to begin with semiparametric 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 analytic functions of the parameters of the problem, and hence are not difficult to compute. The entry and continuation values are used to determine the model's predictions for entry and exit conditional on the parameter vector, and the parameter vector is estimated by comparing these predictions to the data on entry and exit rates. Attention is given to the small sample problem of estimation error in the semiparametric estimates, and this leads to a preference for the simplest of estimators.

[1]We thank John Asker, Pat Bayer, Lanier Benkard, Gary Chamberlain, Jeff Dominitz, Boyan Jovanovic, Jack Porter, Thanasis Stengos, two referees, and the editor 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 profits (more on this below). 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 that 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 cannot 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 firms' reactions to them.

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 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*.

It is important to point out that our assumptions are not, in general, rich enough to pick out the equilibrium that would follow the introduction of a *new* policy. As a result, our estimates of sunk costs do not generally allow us to determine the entry and exit distributions that would follow a policy change. On the other hand, the sunk costs estimates should give the researcher the ability to examine what could happen after the policy change (say, by examining all possible post policy change equilibria), and would seem to be a necessary ingredient of any more detailed analysis of post policy change behavior (for further discussion see Pakes, 2005).

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 then 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 then 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 semiparametric 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 a discussion of why they occur, end up being quite 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. Our Monte Carlo discussion concludes with a discussion of the circumstances under which the estimated continuation and entry values will be robust to one of our assumptions; the absence of serially correlated unobserved state variables.

Related Literature

Hotz and Miller (1993) were the first to use semiparametrics to ease the computational burden of a dynamic estimation problem. They show that a single-agent dynamic discrete choice problem mimics a static discrete choice problem with the value functions replacing the mean utilities. Under their assumptions this implies that there is a one-to-one map between those choice probabilities and the continuation values. This enables them to obtain the continuation values nonparametrically by first estimating the agent's choice probabilities at each state and then inverting those choice probabilities to obtain the relevant continuation values. In a subsequent single-agent paper, Hotz, Miller, Sanders, and Smith (1994) suggest using the estimated probabilities to simulate sample paths. They then calculate the discounted value of utility along these paths, average those values for the paths emanating from a given state, and use these averages as the continuation values at that state.

We deal with a multiple agent problem. Also, instead of inverting the choice probabilities, or using them to simulate sample paths, we estimate the continuation values directly by computing the average of the discounted values of future net cash flows that agents starting at a particular state actually earned (at least up to the parameter vector of interest). Still, it is clearly the common idea of using the data to obtain semiparametric estimates of continuation values that simplifies the estimation problem in all three papers.[2]

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) make more restrictive assumptions on the disturbance distributions and emphasize, contrary to our small sample results, the advantages of the pseduo maximum likelihood estimator introduced below. Pesendorfer and Schmidt-Dengler (2003) make an i.i.d. probit assumption and implement the empirical analogue of the identification argument in Magnac and Thesmar (2002). Their approach to sampling error is different from 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) provide assumptions and techniques that allow them to generalize the ideas in Hotz, Miller, Sanders, and Smith (1994) to dynamic games and show that under their assumptions one can incorporate the information from the choice of continuous, as well as discrete, controls in the estimation algorithm.[3]

[2]Semiparametrics were also used to circumvent computational problems in a dynamic mutliple agent setting by Olley and Pakes (1996). They use nonparametric estimates of the choice of discrete and continuous controls to correct for selection and simultanaeity biases in estimates of production functions. They do not, however, directly use the implied nonparametric estimates of continuation values.

[3]It is possible to integrate continuous controls into the estimation algorithm outlined here, though we do not consider this extension in the current paper. We also note that there is a small related literature on estimating timing games which is just now developing; see Einav (2003) and Schmidt-Dengler (2004).

Some Limitations of the Analysis.

We have focused on certain topics and ignored others. One obvious omission is that we shall not formally consider identification issues. Many of the parameters determining behavior in dynamic games can often be estimated without ever computing an equilibrium, and those parameters that remain depend on the nature of the problem and data availability. Thus, substantive identification issues are somewhat problem-specific. There are, however, at least two ways in which one could pursue them.

First, since our mode of analysis can be used to estimate the parameters of the profit function as well as the distribution of entry and exit fees, one could investigate identification when both the profit function and the distribution of entry and exit fees are unknown. It is clear that we cannot nonparametrically identify all of these functions from observed entry and exit distributions without additional restrictions. However, measures of variable profits can often either be obtained from reported sales and costs information, or derived from estimates of demand and cost functions and a Nash in prices (or quantities) equilibrium assumption (and the latter requires only market-level data on shares, product characteristics, and prices). Both of these sources of information are more directly related to profits than the number of entrants and exitors. As a result, they can usually be used to estimate the profit function under weaker assumptions then would be required to estimate it from its implications on entry and/or exit behavior. In addition, as we show in section 4.3, when information on profits are available our estimates of continuation and entry values are robust to some familiar sources of specification error.

In contrast, it is typically quite difficult to obtain direct measures of sell-off values and/or entry costs. Sell-off values are often associated with such poorly measured objects as "good-will," the value of the firm's buildings and equipment in their "second best" alternative employment, and/or clean-up costs. Sunk entry costs are often associated with the time and effort required to formulate ideas, test markets, and access both startup capital and the requisite permissions. As a result our focus will be on providing estimates of the entry cost and sell-off value distributions.

This focus induced us to avoid making some of the assumptions on the distribution of the unobserved component of sunk costs traditionally made in dynamic estimation algorithms. Of particular importance is that we do not assume that the entry costs a given potential entrant faces in different locations in the same market are independent of one another. On the other hand, we do maintain the assumption that the sunk cost distributions have known parametric forms. Indeed, the second way of pursuing identification issues further would be to assume that profits are a known function of the state variables of the problem and ask whether, given the information on profits, observed entry and exit behavior would be enough to identify the sunk cost distributions nonparmetrically. This is a topic beyond the scope of the current paper, but it is being pursued elsewhere (see Berry and Tamer, 2005).

Finally, we do not provide the details of the asymptotic distributions of our parameter estimates. This is because the asymptotic results are obtainable using standard semiparametric econometric procedures (see, e.g., Pakes and Olley, 1995, and the literature cited there) and, as will be explained, consistent estimates of the variance-covariance of the parameter estimates can be easily obtained from parametric bootstrap procedures. On the other hand, we do present extensive Monte Carlo results on the small sample properties of our estimators. This is because the first stage of our estimation algorithm requires nonparametric estimates at a number of states that is often quite large relative to the moderately sized samples that Industrial Organization researchers use. The Monte Carlo results, when combined with an analysis of how they are generated, are quite informative in this respect, as they indicate that some nonparametric estimators work quite well in our context while others do not.

1 A Simple Entry/Exit Model.

We begin with a Markov Perfect (Maskin and Tirole, 1988) 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 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 which might need 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

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

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

If e is the number of entrants, x is the number of exitors (both of which are unknown at the time the incumbents' decisions are made), and $p(\cdot)$ is notation for a probability distribution, then $VC(\cdot)$ is 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

$$VC(n, z; \theta) \equiv \tag{2}$$

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

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.* These perceptions generate the probability distribution

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

where $\chi = 1$ is notation for the incumbent continuing. We need this distribution 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

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

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 entry.[4]

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 is a fixed number of potential entrants in each period (denoted by $\mathcal{E}$), and the distributions over*

[4]Note that we are not giving the potential entrant the possibility of waiting to enter in some future period. It is possible to do so, but it would require assumptions on the evolution of sunk costs for those whose wait, and an additional set of state variables which determine the number of remaining potential entrants from prior years and the distribution of their entry costs.

- *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 distributions and their own realizations, but do not know the realizations of their competitors (so there is asymmetric information, as in Seim, 2002).

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* $\mathcal{P}_z \equiv \{p(\cdot|z), \forall z \in Z = [0, 1, \ldots, \overline{z}]\}$*,* $lim_{n\rightarrow\infty}\pi(n, z) \leq 0$ *for every* $z \in Z$*, and* $\pi(\cdot)$ *is bounded.*[5]

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 $\overline{n}$ such that, provided the current n is lower than it, we will never observe an $n > \overline{n}$. The market is simply not profitable enough to induce entry if there are $\overline{n}$ 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 "infinitely often."[6]

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.

[5]For 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 $\mathcal{R} = \{(n, z) : 0 \leq n \leq \overline{n}, z \in Z\}$. 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.

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).

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]

Before concluding this section we want to point out two 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 the potential entrants and incumbents in a given period would condition their behavior on the unobserved states, and the simple econometric techniques introduced below do not – a difference which can bias the parameter estimates. We come back to a more detailed discussion of the consequences of serially correlated unobserved state variables when they

[7] There is a detail missing here. Though points in $\mathcal{R}$ can only communicate with other points in $\mathcal{R}$ if optimal policies are followed, there are some points, "boundary points" in the terminology of Pakes and McGuire (2001), that could communicate with points outside of $\mathcal{R}$ if feasible but 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.

do exist in our Monte Carlo analysis below. There we show that the impact of serially correlated unobserved state variables is likely to depend on the nature of the data available to the researcher.

Finally, we remind the reader of a consequence of Assumption 1.2 pointed out in our introduction. Though 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 in itself enable us to predict the equilibria that will be played after a policy or environmental change. On the other hand, the estimated sunk cost distribution would seem to be a necessary ingredient of any analysis of the likely implications of such changes.

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

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

where for $r \geq x$

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

and

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

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

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

where $\forall R \geq e$

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

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 specifications 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, even 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, its (expected) continuation value is equal to

$$VC(n,z;\theta) = E^c_{n',z'}\left[\pi(n',z') + \delta E_{\phi'}\left[\max\{VC(n',z';\theta),\phi'\}|n',z'\right]\right], \tag{6}$$

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
- $E^c_{n',z'}(\cdot)$ 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

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

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

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

$$\begin{aligned} VC(n,z;\theta) &= E^c_{n',z'}\left[\pi(n',z') + \delta\left(1 - p^x(n',z')\right)VC(n',z';\theta) + \delta p^x(n',z')\left(VC(n',z';\theta) + \sigma\right)\right] \\ &= E^c_{n',z'}\left[\pi(n',z') + \delta VC(n',z';\theta) + \delta p^x(n',z')\sigma\right]. \end{aligned} \tag{8}$$

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

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

Equation (9) computes $VC(\cdot)$ as the sum of expected current returns and the future continuation value (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

$$VC(\theta) = M_c\,[\pi + \delta\sigma p^x] + \delta {M_c}^2\,[\pi + \delta\sigma p^x] + \delta^2 {M_c}^2 VC(\theta) = \tag{10}$$

$$\ldots \quad = \sum_{\tau=1}^{\infty} \delta^\tau M_c^\tau[\pi + \delta\sigma p^x] = [I - \delta M_c]^{-1} M_c[\pi + \delta\sigma p^x],$$

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

That is, continuation values can be computed by finding the expected discounted future returns that the firm would earn on alternative possible future sample paths. This implies that we can obtain a consistent estimate of the continuation value by averaging over the discounted 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, as a result, will converge to $p^x(n, z)$ provided $\#T(n, z) \to \infty$.

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 $\#T(n, z) \to \infty$, 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 periods.[8]

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

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

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 $\hat{VC}(\theta)$. If δ is known (and we usually think that the prior information we have on δ is likely to swamp the information on δ available from estimating an entry model), then

$$\hat{VC}(\theta) = \tilde{A}\pi + \tilde{a}\sigma \tag{12}$$

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

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

[8] There 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 by $b(x; n_t - 1, z_t)/b(x; n_t, z_t) = [1 - (x_t/n_t)]/[1 - p^x_t]$ which is the weight above once we substitute $\tilde{p}^x_t = \sum_{t\in T(n,z)} x_t / \sum_{t\in T(n,z)} n_t$ for p^x_t.

or, in matrix notation,

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

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 by[9]

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

Accordingly, we obtain our consistent estimate of VE as

$$\hat{VE}(\theta) = \tilde{B}\pi + \tilde{b}\sigma, \tag{14}$$

where

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

The simplicity of the form of the solution for $(\hat{VC}(\theta), \hat{VE}(\theta))$ 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 for 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

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

where $(1 - \tilde{p}^x) \times \hat{\mathcal{VC}}(\theta)$ is the vector formed by multiplying each $1 - \tilde{p}^x(n, z)$ with the corresponding $\hat{\mathcal{VC}}(\theta)(n, z)$, while $\tilde{p}^x \times E[\phi|\phi > \hat{\mathcal{VC}}(\theta)]$ is the vector formed by multiplying each element of $\tilde{p}^x(n, z)$ by the corresponding $E[\phi|\phi > \hat{\mathcal{VC}}(\theta)(n, z)]$. This equation system

[9]These weights can also be derived as 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 $\frac{1}{\#T(n,z)} \sum_{t \in T(n,z)} \frac{[e_t/\mathcal{E}]}{[\tilde{p}^e(n,z)]} 1_{[(n_{t+1}, z_{t+1})=(n',z')]}$, where $\tilde{p}^e(n, z) = \frac{1}{\#T(n,z)} \sum_{t \in T(n,z)} \frac{e_t}{\mathcal{E}}$.

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 less than $1/\delta$ 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 $(\hat{VC}(\theta), \hat{VE}(\theta))$, we can form consistent estimates of the probability of exit and of entry conditional on θ as $(1 - F^{\phi}(\hat{VC}(\theta)|\theta))$ and $F^{\kappa}(\hat{VE}(\theta)|\theta))$, 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 of these estimators is comparable to that of estimators for the simplest static entry/exit models.

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, ..., or do 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* $\mathcal{E}$*,*

[10] To see this let $T(x)$ be the operator that produces the right hand side of this equation when $\hat{VC} = x$ 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.

- *potential entrants can enter in only one of the two locations and have entry cost* (κ_1, κ_2) *in the first and second locations respectively, where the vector* (κ_1, κ_2) *is a draw from the distribution*
$$Pr\{\kappa_1 \leq r_1 \ \ and \ \ \kappa_2 \leq r_2\} \equiv F^{\kappa}(r_1, r_2|\theta)$$
which is independent over time and across agents, and

- *once in a particular location, the entrant cannot switch locations, but can exit to receive an exit fee of* ϕ *which is an i.i.d. drawn from* $F_1^{\phi}(\cdot|\theta)$ *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

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

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

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

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

$$m_0 \equiv Pr\{\kappa_1 > \delta VE_1(n_1, n_2, \cdot) \ \ and \ \ \kappa_2 > \delta VE_2(n_2, n_1, \cdot)\}, \qquad (16)$$

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

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

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

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

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

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

[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.

and

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

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 $[1-(x^l_t/n_l)]/[1-\tilde{p}_l{}^x(n_l,n_{-l},z)]$ and $e^l_t/[\hat{\tilde{e}}^l(n_l,n_{-l},z)]$, where $\hat{\tilde{e}}^l(n_l,n_{-l},z) = \#T(n_l,n_{-l},z)^{-1} \sum_{t\in T(n_l,n_{-l},z)} e^l_t$.[12] 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 inverse needed 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. This grows polynomially in the number of distinct states (which typically increases in the number of locations). As we will show in our Monte Carlo examples, given the simplicity of the rest of the estimation procedure, this "setup" time can easily become the dominant computational burden.

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, 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). We show below that there are good reasons to expect the

[12] Using this expression for the weights, the two formulas are, respectively,

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

and

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

small sample biases caused by this variance to be more noticeable in some of the alternative possible consistent estimators than in others.

Panel Data.

Note that the argument given here for consistency of the estimators of the continuation and entry values required that; (i) the realizations of the transitions from two observations that start at the same state are independent of one another, (ii) that there are a subset of states that will be visited repeatedly (formally they need to be visited infinitely often), and (iii) that equilibrium perceptions depend only on the state variables observed by the econometrician. Thus if we are using panel data and do not make any allowance for the panel nature of the data, we are assuming; independence over the markets whose data we use[13], that the sampling process generates repeated observations from the same states, and that the same equilibrium is played in each of the markets.

3 Alternative Two-Step Estimators.

Under our assumptions, the ideas discussed above can be combined and/or modified in different ways to obtain alternative $\sqrt{n}$-consistent and asymptotically normal (or CAN) estimators. This section will discuss several of these alternatives.

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 $\hat{VC}$ and $\hat{VE}$) 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.

[13] If for example 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 expectations 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.

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, the variance in the first stage nonparametric estimates can be large. The objective function used in estimation is a nonlinear functional of the first stage estimation errors, and, as a result, the variance in the first stage estimates will cause a small smaple bias in the parameter estimates. The extent of that bias will differ (sometimes dramatically) with how the objective function transforms the estimation errors, and since there are a number of alternative objective functions to choose from, one might want to take this into account in choosing among them.[14]

It is possible to use standard semiparametric techniques to obtain an analytic formula for the asymptotic variance-covariance of our parameter estimates. The fact that this variance-covariance matrix depends on the variance of the first stage nonparametric component (i.e., that the estimator is not "adaptive") follows from the fact that the derivative of the objective function with respect to the estimates of $\hat{VC}$ and $\hat{VE}$ do not have a conditional expectation of zero. Since we have a complete model and it is relatively easy to both simulate data from that model and to use this data to estimate new values for the parameters, it is perhaps easiest to obtain consistent estimates of the needed variance covariance matrix from a parametric bootstrap.[15]

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

[14] An 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.

[15] This explains why the analytic formula for the standard errors, which appeared in an early version of this paper, has been omitted here. The parametric bootstrap is constructed as follows. Use the estimates of the continuation and entry values conditional on (n, z), and the estimates of θ and $p(z'|z)$ to generate independent samples of size equal to our sample size. Then estimate θ from each of these pseudo random samples and take the variance of those estimates.

(b) a method of moments estimator 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 moments estimator 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.[16]

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

   (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 multistage 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 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

[16] In 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.

inversion in (3a) or the nested fixed point in (3b).[17] We note, however, that there is no guarantee that the iterations improve the estimates or that they will converge to anything (see below). Moreover, if they do converge there is a question of whether 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 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. This 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 the estimated and not the true entry and exit values), events which have nonzero probability in the true likelihood and hence can occur, can be assigned a zero probability in the pseudo likelihood (for any $\theta \in \Theta$). If only one such event does occur the pseudo likelihood will be undefined (the log-likelihood

[17] 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 consider a multi-step version of their estimator which iterates in this way. Aguirregabiria 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 positive probability).

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.

The "pseudo minimum χ^2" objective function in (1b) is, like the log-likelihood, likely to have problems due to the nonlinearities it induces. To see this, let s index states, $\hat{F}^\kappa_s(\theta) = F^\kappa(\hat{VE}_s(\theta))$ index estimated entry probabilities, $\hat{q}_e(s)$ denote the entry rates in the data, and note that the first order condition for the entry parameters in the the pseudo minimum χ^2 estimator is a weighted average of

$$[\hat{q}_e(s) - \hat{F}^\kappa_s(\theta)]\frac{\partial \hat{F}^\kappa_s(\theta)}{\partial\theta} \equiv$$

$$[\hat{q}_e(s) - F^\kappa_s(\theta)]\frac{\partial \hat{F}^\kappa_s(\theta)}{\partial\theta} + [F^\kappa_s(\theta) - E\hat{F}^\kappa_s(\theta)]\frac{\partial \hat{F}^\kappa_s(\theta)}{\partial\theta} + [E\hat{F}^\kappa_s(\theta) - \hat{F}^\kappa_s(\theta)]\frac{\partial \hat{F}^\kappa_s(\theta)}{\partial\theta},$$

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 $\hat{VE}$ (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., $\frac{\partial \hat{F}^\kappa_s(\theta)}{\partial\theta}$), which is correlated with the error in the estimate of the probability. Thus, if we replaced $\frac{\partial \hat{F}^\kappa_s(\theta)}{\partial\theta}$ 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 estimators.[18] 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

[18] The 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.

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

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

Then use the the binomial formula

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

and

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

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 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 estimators.[19] 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 will be larger the larger the number of state variables in the model.

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

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

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

where

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

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

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 algorithm.[20]

The extra computational burden of the fixed point grows exponentially in the number of state variables (or locations),[21] 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 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.

This section is divided into three. The first subsection provides Monte Carlo results on the one location model, and the second does the same on a two location model. These two subsections are designed to give the reader a sense of both the computational burden and the distributions associated with the various estimators. The third subsection examines the likely impact of serially correlated unobservables. For the reasons discussed above, throughout we focus on estimating the distribution of entry fees and sell-off values.[22]

Sample Design.

Much of the earlier work on entry was on cross-sections (or panels) of relatively isolated markets, and focussed on the relationship between market size and the number of active firms, or, in the multiple location case, between market size and the relationship between

[20]If 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.

[21]Actually as the product of two exponentials; one for the number of points which need to be evaluated in the fixed point calculation, and one for computing the future value at each point.

[22]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 evaluated in the results presented below (while the matrix inversion only occurs once).

the number of firms in the different "locations" (see Bresnahan and Reiss, 1987, Mazzeo, 2002, and Seim, 2002). Since we want designs which generate entry into and exit from an existing industry, we construct panel data sets. However, the primary source of variation in our data will be across markets, and our exogenous variables (i.e., z in our notation) will be a variable that determines current market size and another that determines the growth rate of that market size variable.

The first six sample designs varied 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. There are about 250 isolated markets in South Dakota, and we chose the support of our population variable to give us the approximate range of population sizes in the South Dakota markets. The growth rate of the market size variable was allowed to be serially correlated, as it is in the actual data. The seventh market design was C=50 and T=5. This design was suggested by a referee and gives us a chance to examine the problems that arise in data sets that are quite small by recent standards.[23]

Throughout we compute our equilibrium using a variant of the Pakes and McGuire (1994) algorithm for computing equilbria (the variant simply shuts down the investment decision in that model). We then use the equilibrium policies to simulate a long time series of data. The data sets used for estimation by initiating C separate time series of length T from the approximate ergodic distribution of states generated by the simulation. The estimators use the data on entry and exit at each tuple of state variables visited, regardless of how many times that state was visited. When we report more than one estimator for a given sample design, we use the same data for all estimators. The reported standard errors are computed from the distribution of estimators over independent Monte Carlo data sets, each sampled as described above.

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

[23]Mazzeo (2002) has 492 interstate exits as his markets, while Seim (2002) has data from 151 city groups with an average of 21 related markets per city group.

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

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

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

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

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.

We computed an equilibrium for $a = .3$ and $\sigma = .75$. State variable z was allowed to take on 45 values at .05 increments and we allowed three growth rates $(.05, 0, -.05)$.[24] The maximum number of firms ever active in our parameterization was nineteen. 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.

4.1.1 Results from the Single Location Model.

Tables 1 and 2 provide a selection of the results that convey what we learned. The first panel of Table 1 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 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

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

$(VC(\cdot), VE(\cdot))$ 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 of Table 1 provides the results when C=1000 and T=15.[25] 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. However, even in this sized panel there is some indication that OF=0 estimators are preferred to those from the other objective functions (though the poor performance of the estimate of σ in the first two columns of the OF = 2 estimator are driven by outliers—without the worst 2% of the runs, the average estimates of σ for both estimators are between .75 and .76, and the standard deviations of the estimates are approximately 0.01).

Panel B provides estimates from the sample with C=250 and T=5. At this sample size we learn more about the estimators' performance. All the estimates of a in panel B seem to have an upward bias. 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 often be problematic. Note also that though there may be a bias problem in some of the estimators, 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 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 have negative estimation error and hence be lower than the minimum true entry value. This 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 observations this derivative is positive, and that accounts for the positive deviations in those estimators.

No matter the sample design the $OF = 0$ (or simple method of moments) estimators had both smaller biases *and* smaller variances then those of either the pseudo mle ($OF = 1$), or the pseudo minimum χ^2 (or $OF = 2$) estimators. For intuition on why this occurs compare

[25] As noted if we use structural probabilities we will need estimates of entry and exit probabilities at points not observed in the data. Here is how we obtain them. Assume that for a given z we only observe behavior from $[n_1(z), n_2(z)]$. Then: (i) if $n \in [1, n_1(z)]$ the probability of entry is equal to the probability of entry in state $n_1(z)$ and the probability of exit is zero; if $n \in [n_2(z), max]$ 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 row labeled $\#(n, g, z)$, which provides the number of states ever visited in the run (424 when T=5 and 630 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. 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 even 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 transited to, so *each* observed transition is implicitly *averaging* over all sample paths from the point transited 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 probabilities.[26] The nested fixed point times are always larger than the matrix inverse times, but the difference is much larger when we use structural probabilities.

So far the results 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 $OF = PF = 0$ 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 fixed point for calculating values ($VF = 1$).[27]

[26]These ratios are much worse for the OF=1 and the OF=2 estimators (up to a factor of 15). Note however that though the estimates for the PR=1 case have to impute entry and exit rates for over eighty percent of the states, the imputation does not seem to cause a noticeable bias in the estimates.

[27]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

Note that neither of these two estimators is particularly computationally burdensome. The compute time with T=5 and C=50 was about ten seconds, and it was only about a half a minute on our largest sample. Compute times in this range should allow researchers to engage in quite a bit of robustness analysis.

We now move to Table 2 which begins with estimates from our smallest sample (T=5, C=50). Given the results in Table 1, we only use the OF=PR=VF=0 estimators in this table. However, because the sample is so small, we examine the possibility of reducing the variance of the estimates of VC and VE by using kernel estimates of their values. That is, we estimated continuation and entry values at each point visited as in the text, and then used normal kernels with a diagonal covariance matrix and different bandwidths to obtain estimates of VC and VE at each point that use the information from nearby points. These "smoothed" continuation and entry values were then used in the estimation algorithm. For comparison, we also present the OF=PR=VF=0 estimators that use kernel estimates of the entry and continuation values for the T=5, C=250 sample.

When we do not use kernels the problems in the Table 1 estimators for the C=250, T=5 samples are accentuated when C=50 and T=5. As expected the use of the kernels causes a marked reduction in the variance of our estimates of VE and VC, but it causes an even more marked increase in their bias (so most kernel estimates actually have a larger mean square error then do the non-smoothed estimators). Still, once we move to the kernels, the small sample problems with the parameter estimates largely *disappear*. So the use of kernels seems to significantly improve the performance of the estimator in small samples. What is small? Once we get to the (C=250, T=5) samples it is no longer obvious that the kernels do better than the non-smoothed estimators.[28]

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

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 smaller samples.

[28]The interplay between the number of states at which we estimate continuation and entry values, sample size, and the appropriate estimation method is clearly a topic which deserves more research. In particular there are a number of alternative estimation procedures which might improve performance when samples are quite small including alternative forms of kernels (we tried local linear kernels with results similar to those above), and small sample bias corrections.

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 chooses a quantity to market in its 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 $F^\kappa(\cdot, \cdot \mid \theta)$ and can enter in at most one of the markets. Since κ_1 and κ_2 reflect differences in a given individual's cost of building the high and the low quality motel in a particular market, we allow them to be correlated. Indeed, we make the reasonable assumption that the cost to a given individual of building the high quality motel in a given market is larger than that individual's cost of building the low quality motel in the same market, i.e., we assume

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

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.[29] Thus, there are two entry parameters to estimate: (a_1, a_2). Exit fees are i.i.d., exponential with parameters (σ_1, σ_2) in the two locations.

[29] Actually, we use a discretized version of the density in equation (17) for the r distribution.

4.2.1 Results from the Two Location Example.

The Monte Carlo results for large samples are split between Table 3 (for the OF=1 or pseudo mle estimators) and Table 4 (for the OF=0 and OF=2 estimators). There is a "start" row in the pivot table for some of the two location runs; when start=1, we start the search from the $OF = PR = VF = 0$ estimator.[30]

Table 3 summarizes the 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 all.[31] 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 a given data set, we tried another initial condition, and continued until we started the search at ten different randomly selected points, none of which resulted in a positive likelihood. At that point we called the search on the data set unsuccessful, and moved to the next data set. The sub-panel labeled "success rate" provides the fraction of the time when this subroutine recorded a "success". The pseudo mle does not work *most* of the time.

The reason for the zero pseudo likelihood is that the first stage is producing a $\hat{VE}_1(\cdot) > \hat{VE}_2(\cdot)$ for candidate parameter values when there is in fact entry in location 2. Since the cost of entry in location 2 is always higher than in location 1, if our estimated entry values were true, entry in location 1 would never happen (hence the zero likelihood). The reason it does happen is because $VE_1(\cdot) < VE_2(\cdot)$; i.e., the disturbances in the first stage estimates have reversed the order of the two entry values. 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

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

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

that one should not be using pseudo mle, at least not without some auxiliary procedure that ameliorates the problems caused by imprecise first stage estimates of entry values. Consequently, we do not present them in what follows.

Table 4 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 that results from the functional form of the objective function in the OF=2 estimators (see the discussion in Section 3.2),[32] 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.

In contrast, with this sample size, all the $OF = 0$ estimators do well. The estimates of σ that use structural probabilities (i.e., OF=0, but PR=1) do have larger mean square errors, 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), but even those results are quite good. However there are 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 with the structural probabilities (VF=PR=1) are twenty times more computationally burdensome. On the other hand, the estimates that 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 5 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, 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

[32] We 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.

rapidly.

We note that the large means and standard deviations of the estimates of a_2 under empirical transition probabilities (PR=0) in the $(T = 5, C = 250)$ sample are driven by outliers: for those runs, the value of a_2 was estimated to lie on the boundary of our search space, which we set at 100. There were 9 such runs for the (PR = 0, VF = 0) estimator, and an additional run for the (PR = 0, VF = 1) estimator, out of the total of 500. Excluding these outliers, the average estimates of a_2 are 0.51 for the first estimator and 0.52 for the second, and the standard deviations are 0.41 and 0.42, respectively; the statistics for the estimates of other parameters remain virtually unchanged. Hence, the estimators work most of the time even for our small samples, but an econometrician who uses them has to use some judgement. Below we suggest a remedy for the outlier problem in small samples: kernel smoothing.

Comparing the various $OF = 0$ estimators, we see that in the $(T = 5, C = 250)$ samples, even excluding the outliers, the estimators that use structural probabilities (PR=1) do better on the a's but much worse on the σ's than the simplest estimator (PR=VF=0). As we increase sample size, the problem with the latter's estimates of the a's disappears rather rapidly, much more rapidly than the problems with the (PR=1) 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 estimators that use structural probabilities is twenty or more times that of the corresponding estimators that 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 then 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 in two ways; (i) by adopting kernels to smooth the estimates of VE and VC, and (ii) by iterating on the initial estimates and seeing whether the structure imposed by the iterations helps. Table 6 presents the results. The kernels do away with the problem of extreme outliers that appears in the unsmoothed estimators. Thus, with two locations the use of kernels with samples of size C=250 T=5 definitely helps, and we would expect to need to use kernels at larger sample sizes when there are more locations. Apparently, when there are a small number of observations per location, say under two or three, the improvement in the variance in the estimates of VC and VE generated by the kernels can have very desirable effects on the estimators.

The iterated estimators used the $OF = 0$ but $PR = VF = 1$ estimators of θ to construct structural probabilities, and then used the fixed point to get the continuation values generated by the structural transition probabilities and different values of θ to minimize the OF=0 objective function. Comparing the $(1,1)$ column in Table 5 to the first iteration estimator in Table 6 we see that 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 initial estimators for the $a's$ but now the estimators of the $\sigma's$ get much worse. When we move to the third and fourth iterations we find that the estimators seem to oscillate, with oscillations which, if anything, are of increasing magnitude (thus moving us farther away from the true values of the parameters).[33]

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. Moreover, the computational burden of the iterated estimators is huge: even the single iteration makes them twenty times more burdensome than the $PR = VF = 0$ 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.

4.3 Robustness Analysis; Serially Correlated Unobservables.

The Monte Carlo results presented thus far were designed to investigate the distributional and computational properties of our estimators. In empirical work we typically worry as much about the specification of the model estimated as about the distributional or computational properties conditional on the chosen model. Indeed, one of the advantages of our estimator is that its light computational burden 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 cannot accommodate.

In this context it is important to distinguish between the two types of data sets we considered in our introduction; data in which we have measures of profits conditional on the state of the system, and data where no such information is available (so that the parameters

[33] We only did 50 runs for the iterated estimators, because the negative results for these runs were already sufficiently conclusive.

determining profits must be estimated along with the sunk cost parameters from the entry/exit model). In the case where separate information on profits is not available, we would expect the presence of a serially correlated state variable to generate particular biases.

Typically one of the observed state variables is the number of incumbents, and the researcher is interested in the extent profits per firm falls as the number of incumbents increases. If there is a state variable which increases profits that is unobserved to the econometrican but observed to the agents, then entry will be higher and exit will be lower when that state variable is higher. Since the model attributes increases in entry rates to higher profits, if the current and lagged value of the unobservable are positively correlated, the estimation algorithm will infer that the positive correlation between entry and profits is a result of profits being higher because the number of incumbents are higher (i.e. when there was more entry and less exit in prior periods). Consequently, there will be a positive bias in the estimated relationship between the number of incumbents and profits.[34]

Not only does this bias not occur when there is an independent measure of profits, but there is a sense in which serially correlated unobserved state variables do not cause a bias in our estimates of continuation and entry values. It is easiest to see this in the one location model where the exogenous variable (our z) is both serially correlated and unobserved by the econometrician. Recall that the Bellman equation for that model is

$$VC(n,z) = \mathcal{E}[\pi(n',z') + \beta\sigma p^x(n',z')|n,z] + \beta\mathcal{E}[VC(n',z')|n,z].$$

To obtain the average of the continuation values associated with a given n, say $\tilde{VC}(n)$ we simply integrate both sides of this equation with respect to the distribution of z conditional on n that is relevant for the observed equilibrium and recurrent class, say $p(z|n)$. That is,

$$\tilde{VC}(n) = \sum_z \{\mathcal{E}[\pi(n',z') + \beta\sigma p^x(n',z')|n,z] + \beta\mathcal{E}[VC(n',z')|n,z]\}\, p(z|n).$$

Moreover, since the empirical distribution of z conditional on n will converge to $p(z|n)$, our estimates of continuation values conditional only on n are consistent estimates of an average of the *true* continuation values at that n, where the averaging is done using $p(z|n)$. This is useful in itself, as robust estimates of continuation and entry values are desirable in many policy applications.

Of course, potential entrants and incumbents condition on the realziation of z at the time they make their decisions, and the probabilities of entry and exit are convex function of the entry and continuation values. So even if we were to use just the average entry and exit rates at each n to estimate the distributions of entry and exit costs (as we do in our OF=0

[34]For a discussion of the procedures available for ameliorating biases due to serially correlated unobserved state variables see Pakes (1994) and more recently Ackerberg, Benkard, Berry, and Pakes (2005).

estimator), we would be incurring an error. I.e., we would use the probability of entry/exit conditional on the average continuation value at a particular n, instead of the average of these probabilities over z conditional on n, to generate the entry and exit rates that we fit to data at the different possible values of the parameter vector. However, this is the type of error that one might think is "second order".

To pursue this further we did a Monte Carlo investigation of the sensitivity of the parameter estimates to the presence of serially correlated unobserved state variables in our two location model. We proceeded as follows. First, we computed an 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. Finally, we "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 the 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 researcher can 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. We note that these serially correlated unobserved 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.

Table 7 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. Recall that the model is in fact misspecified and the parameter estimates should be inconsistent; it is just that we had reasons to believe that the bias caused by the misspecification would be small. Apparently, at least with our model specification, the extent of the asymptotic bias in the misspecified estimators is negligible (and we have tried several more simulations than those reported in Table 7, all with similar results).

4.4 Monte Carlo Results: A Summary.

The results from our Monte Carlo experiments are unusually clear cut. We prefer estimators based on (i) an objective function whose moments are constructed by multiplying state specific differences between observed and empirical entry and exit rates and a deterministic function of the state (our OF=0 case), (ii) the empirical transition matrix, and, if available, (iii) the matrix inversion (though use of the contraction mapping instead of the matrix inversion only causes minor increments in compute time and produces estimators with roughly the same properties). When the number of states at which we have to estimate entry and continuation values is large relative to the size of the sample, our results show a preference for using kernel averages over nearby entry and continuation values, rather then pointwise estimates of the continuation values, to construct the predictions for the entry and exit behavior that go into the objective function.

What is striking is that estimators with these properties not only seem to have better small sample distributions, they are also the least computationally burdensomen of the estimators we tried. This fact, together with the property that when profit data is available, continuation values conditional on observed states should be approximately correct, simplifies the analysis of robustness of the estimator to model specification.

In Monte Carlo work the computational burden associated with estimating on repeated samples induces researchers (including us) to stick with models that are not too detailed. 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. In this context we note that there are two reasons to expect the Monte Carlo results reported here to extend to more complex situations.

First, the way in which both the computational burden and the data requirements associated with a given level of precision increase 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. So the argument for using empirical transitions should just get stronger in more complex environments.

Second, the increased computational burden associated with evaluating the fixed point relative to using the matrix inverse formula (that is in using VF=1 rather than VF=0 in the notation of the tables) impacts on the compute time as a multiple of the number of times different values of the parameter vector are evaluated in an estimation run. As we increase the number of parameters estimated, the number of objective function evaluations is likely to grow. So the argument for use of the matrix inversion also gets stronger in more complex environments.

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 that ensure 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 that 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 semiparametric 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 semiparametric estimates of those functions, affects the distributions of the alternative estimators in different ways. Also, the computational burdens of the alternatives estimators 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, though at small enough samples the use of kernels to decrease the variance of the pointwise estimates of entry and continuation values seems warranted. Moreover, the preferred estimators performed well in robustness checks and had minimal computational burdens.

Indeed, their computational burdens were small enough for us to expect the effective barrier to the empirical analysis of discrete dynamic games to shift from being the computational burden involved in obtaining parameter estimates to the richness of the data available to support the analysis.

References.

Ackerberg, D., L. Benkard, S. Berry, and A. Pakes (2005); "Econometric Tools for Analyzing Market Outcomes", *mimeo* Harvard University.

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.

Berry, S. and E. Tamer (2005); "Identification in Models of Entry", *mimeo* Yale University.

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

Bresnahan, T. and P. 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, R. and A. Pakes (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. (2005); "Econometrics and Econmetric Theory in Applied I.O.", Fisher-Schultz Lecture, The World Congress of The Econometric Society, London.

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. and S. Olley, (1995); " A Limit Theorem for a Smooth Class of Semiparametric Estimaors", the *Journal of Econometrics*, Vol.65 (6), pp 1263-98.

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.

Schmidt-Dengler, P. (2004); "The Timing of New Technology Adoption: The Case of MRI", *mimeo* Yale University.

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=500												
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	0.01	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	0.83	0.86	0.75	0.75
SD(σ)	0.01	0.01	0.01	0.01	0.02	0.07	0.01	0.07	0.56	1.00	0.01	0.11
t(total)	30.0	33.1	70.6	104.7	37.4	65.3	77.2	341.0	30.8	60.1	79.1	407.9
Panel B: T=5, C=250, R=500												
a=.3	0.37	0.37	0.38	0.37	0.40	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	0.02	0.02	0.03	0.03	0.03	0.03
$\sigma = .75$	0.77	0.75	0.76	0.74	0.75	0.71	0.74	0.71	9.30	1.85	5.02	0.78
SD(σ)	0.04	0.04	0.04	0.04	0.04	0.07	0.04	0.03	27.15	7.20	19.92	0.23
t(setup)	8.9	8.9	55.5	55.5	8.9	8.9	55.5	55.5	8.9	8.9	55.5	55.5
t(search)	0.0	1.3	0.1	32.9	0.5	8.8	0.5	207.4	0.3	10.0	6.5	251.8
t(total)	8.9	10.3	55.6	88.4	9.4	17.7	56.0	263.0	9.2	18.9	62.0	307.3
$\#(n,g,z)$	424	424	424	424	424	424	424	424	424	424	424	424
$\#\hat{p}$	465	465	2580	2580	465	465	2580	2580	465	465	2580	2580
Panel C: T=15, C=250, R=500												
a=.3	0.32	0.32	0.32	0.32	0.35	0.33	0.33	0.32	0.31	0.32	0.32	0.32
SD(a)	0.01	0.01	0.01	0.01	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.01
$\sigma = .75$	0.75	0.75	0.75	0.74	0.74	0.72	0.74	0.73	1.28	0.97	0.97	0.75
SD(σ)	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	6.29	1.16	4.44	0.02
t(setup)	18.4	18.4	63.7	63.7	18.4	18.4	63.7	63.7	18.4	18.4	63.7	63.7
t(search)	0.0	2.2	0.1	34.1	1.6	17.5	1.5	240.8	0.5	18.2	7.2	282.9
t(total)	18.4	20.6	63.8	97.9	20.0	35.9	65.3	304.6	18.9	36.6	70.9	346.6
$\#(n,g,z)$	630	630	630	630	630	630	630	630	630	630	630	630
$\#\hat{p}$	640	640	2689	2689	640	640	2689	2689	640	640	2689	2689

* 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. One Location Using Kernel Estimates of VE and VC.*

Sample Size	T=5, C=50, R=500			T=5, C=250, R=500		
Bandwidth	0.00	1.00	0.50	0.00	1.00	0.50
a=.3	0.47	0.28	0.28	0.37	0.28	0.28
SD(a)	0.08	0.03	0.03	0.03	0.01	0.01
σ=.75	0.82	0.71	0.72	0.77	0.71	0.72
SD(σ)	0.11	0.08	0.08	0.04	0.04	0.04
Average % Bias, % Standard Error, and % MSE of VE and VC Estimates.						
VE bias	0.03	0.20	0.18	0.01	0.16	0.14
VE "se"	0.21	0.15	0.18	0.13	0.06	0.07
VE "$\sqrt{MSE}$"	0.21	0.29	0.29	0.13	0.24	0.22
VC bias	0.02	0.05	0.04	0.01	0.03	0.03
VC "se"	0.24	0.11	0.13	0.15	0.05	0.06
VC "$\sqrt{MSE}$"	0.25	0.18	0.18	0.15	0.16	0.15

* All estimates use OF=PR=VF=0 (see the legend to Table 1). If the bandwidth is zero, then there is no smoothing of our estimates of VE and VC. If the bandwidth is x, we smooth our estimates with a normal kernel with a diagonal covariance matrix with bandwidth parameters equal to x times the variance of the corresponding variable. The average (i) percent bias, (ii) standard error and (iii) square root of mean square error (mse), are calculated from the average (over the distribution of realized states) of: (i) the difference between the estimate of VE or VC and their true value, (ii) the sampling variance of the estimate about its estimated value, and (iii) the mean square error of the estimate.

Table 3: 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
σ1=1	1.17	0.96	0.85	1.11	1.10	1.07
SD(σ1)	0.26	0.22	0.24	0.29	0.29	0.26
σ2=.5	0.51	0.54	0.40	0.51	0.57	0.53
SD(σ2)	0.08	0.13	0.11	0.09	0.15	0.14

* Legend. See the footnote to Table 1.

Table 4: Two Locations, OF=0 and OF=1.

Pivot Table*

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

Pivot Table*

PR	0	0	1	1	0	0	1	1
VF	0	1	0	1	0	1	0	1
Data	T=5, C=250, R=500				T=5, C=500, R=500			
a1=.3	0.39	0.39	0.28	0.28	0.34	0.34	0.28	0.28
SD(a1)	0.11	0.10	0.04	0.04	0.04	0.04	0.03	0.03
a2=.3	2.30	2.49	0.34	0.33	0.39	0.39	0.31	0.31
SD(a2)	13.25	13.95	0.66	0.52	0.08	0.08	0.06	0.05
σ1=1	1.35	3.52	2.34	2.37	1.10	1.36	1.74	1.57
SD(σ1)	1.84	9.04	1.87	2.25	0.73	2.33	0.97	1.15
σ2=.5	0.68	1.90	1.10	1.08	0.64	0.81	0.82	0.71
SD(σ2)	0.85	4.45	0.90	1.62	0.88	2.57	0.55	0.51
t(setup)	10	10	396	396	15	15	514	514
t(search)	8	9	71	210	9	11	85	210
t(total)	18	19	467	605	24	26	599	724
$\#(n,g,z)$	321	321	321	321	404	404	404	404
$\#\hat{p}$	347	347	2682	2682	427	427	2918	2918
Data	T=15, C=250, R=500				T=15, C=500, R=500			
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	0.34	0.34	0.31	0.31	0.32	0.32	0.31	0.31
SD(a2)	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03
σ1=1	1.10	1.10	1.34	1.17	1.06	1.02	1.17	1.08
SD(σ1)	0.28	0.84	0.42	0.33	0.16	0.11	0.21	0.13
σ2=.5	0.55	0.52	0.62	0.55	0.53	0.51	0.56	0.52
SD(σ2)	0.11	0.13	0.16	0.11	0.07	0.04	0.09	0.05
t(setup)	19	19	534	534	26	26	730	730
t(search)	10	12	86	196	13	15	104	251
t(total)	30	32	620	730	39	41	834	981
$\#(n,g,z)$	458	458	458	458	531	531	531	531
$\#\hat{p}$	465	465	2854	2854	537	537	3160	3160

* Legend. See the footnote to Table 1.

Table 6. Kernel and Iterated Estimators for T=5, C=250.

Two Locations, Kernel Estimates of VE and VC				
Bandwidth	0	0-no bad runs	1	0.5
a1=.3	0.39	0.39	0.18	0.18
SD(a1)	0.11	0.11	0.03	0.03
a2=.3	2.30	0.51	0.30	0.29
SD(a2)	13.25	0.41	0.15	0.14
$\sigma 1 = 1$	1.35	1.35	0.66	0.73
SD(σ1)	1.84	1.86	0.08	0.13
$\sigma 2 = .5$	0.68	0.67	0.33	0.35
SD(σ2)	0.85	0.83	0.03	0.04
# runs	500	491	500	500
Iterated Estimates: From OF=0, PR = 1, VF = 1; R=50				
Iteration	1	2	3	4
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
σ1=1	1.13	7.21	1.43	7.53
SD(σ1)	0.31	8.26	0.97	5.97
σ2=.5	0.54	2.53	0.62	5.46
SD(σ2)	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

* Legend. See the footnotes to Tables 1 and 2.

Table 7: "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
σ1=1	1.02	1.01	1.03	1.02
SD(σ1)	0.03	0.03	0.03	0.03
σ2=.5	0.50	0.50	0.51	0.50
SD(σ2)	0.01	0.01	0.02	0.01

* Legend. See the footnote to Table 1.

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

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

and is consistently estimated by substituting

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

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,

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

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

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

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

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

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

as desired. ♠