RAND Journal of Economics
Vol. 38, No. 2, Summer 2007
pp. 373–399

Simple estimators for the parameters of discrete dynamic games (with entry/exit examples)

Ariel Pakes*

Michael Ostrovsky**

and

Steven Berry***

We estimate parameters from data on discrete dynamic games, using entry/exit games to illustrate. Semiparametric first-stage estimates of entry and continuation values are computed from sample averages of the realized continuation values of entrants and incumbents. Under certain assumptions, these values are easy-to-compute analytic functions of the parameters of interest. The entry and continuation values are used to determine the model's predictions for entry and exit conditional on the parameter vector, and the estimates compare these predictions with the data on entry and exit rates. Small-sample properties are discussed and lead to the simplest of estimators.

1. Introduction

■ This paper uses the structure of dynamic games to develop estimators for their parameters. We concentrate on games with discrete controls and, for ease of exposition, provide our results in the context of 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.

Though the costs of entry and the sell-off values (or costs) associated with exit are key determinants of the dynamics of market adjustments, data on these "sunk costs" are much harder to find than data on the determinants of current profits. As a result, we often 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 to be able 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

* Harvard University; ariel@ariel.fas.harvard.edu.

** Stanford University; ostrovsky@gsb.stanford.edu.

*** Yale University; steven.berry@yale.edu.

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.

value of continuing. Though algorithms for computing these values have been available for some time (e.g., Pakes and McGuire, 1994), they cannot be used in estimation without encountering substantial computational problems.

Consequently, the entry/exit models that have been taken to data have all been two-period models that assume away sunk costs; see Bresnahan and Reiss (1987, 1991), Berry (1992) and, more recently, Mazzeo (2002) and Seim (2005). The lack of estimates of sunk costs induced these papers to focus on characterizing differences in the number of active firms across a cross-section of markets rather than on the likely impacts of policy or environmental changes on the structure of an industry (e.g., the impact of mergers on entry or of pension and/or health-care provisions on exit).

The early entry/exit papers *did* 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 (1991) and Berry (1992) considered its impact on estimation in models where fixed costs could vary among agents. When models do not have a unique solution, it is generally 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 that differed across agents, for then the number of possible equilibria increased markedly. The original analysis here, due to Mazzeo (2002) and Seim (2005), allowed continuation values to differ with "location" and began investigating extensions that 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 (similar to those in Muth, 1961), can be used to deliver estimators that are both *easy* to compute and grounded in what *actually happened*.

□ **The underlying idea.** To determine whether a potential entrant (an incumbent) should enter (continue), we need the expected discounted value of future net cash flows should the firm enter (continue). The potential entrant (incumbent) will enter (continue) if this entry value (continuation value) is greater than the entry fee (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, which, together with the estimator's computational ease, simplifies robustness analysis greatly.

We construct the probability of entry conditional on the parameters of the entry-cost distribution as the probability of an entrant drawing an entry fee less than the estimated entry value. Similarly, the probability of an incumbent exiting conditional on the parameters of the sell-off distribution is the probability of drawing a sell-off value greater than the estimated continuation value. The parameters of these distributions are estimated to make the model's 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.

All estimators are semiparametric. There is a first stage that provides a nonparametric estimate of the entry and continuation values and a second stage that treats these estimates as true values in a parametric estimation routine. We provide assumptions under which the first stage need only be done once. That is, the estimation algorithm does not need to compute a complicated fixed point or matrix inverse each time it evaluates its objective function. As a result, the computational burden of our estimator is comparable with 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 alternative estimators suggest themselves. The alternatives have different computational and distributional properties, and so we provide fairly detailed Monte Carlo results. Finally, we investigate the robustness of our estimators to the presence of serially correlated unobservables.

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 used, and the best performing alternatives are also the *least* computationally burdensome. The computational burden of these estimators is small enough to think that the effective limitation to estimation will be the richness of the data rather than computational feasibility. Also, provided independent measures of profits are available, there is a sense in which our estimators are robust to the presence of serially correlated unobservables.

□ **Limitations of the analysis.** First, a conceptual point. We want to stress that, though our assumptions are sufficient to use the data to pick out the equilibrium that was played in the past, they do not allow us to pick out the equilibrium that would follow the introduction of a *new* policy. On the other hand, the sunk-costs estimates should give the researcher the ability to examine what could happen after a 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 (see Pakes, 2005).

Second, we ignore identification issues. Many of the parameters determining behavior in dynamic games can often be estimated without ever computing continuation or entry values. So though we could try and use the entry/exit analysis to estimate all of the underlying primitives, the substantive identification issues are likely to vary with the problem of interest. In particular, measures of variable profits are often available from reported sales and costs information or can be derived from a Nash equilibrium assumption and estimates of demand and cost functions. These estimates are likely to be more reliable than profit estimates obtained using only entry/exit data, and their availability enables more robust estimation strategies for dynamic parameters (see below).

In contrast, it is typically quite difficult to obtain direct measures of sell-off values, as these are determined by such poorly measured objects as "goodwill," the value of the firm's buildings and equipment in their second-best employment and clean-up costs. Similarly, measures of entry costs require estimates of the costs of formulating ideas, testing markets, and accessing both start-up capital and the requisite permissions. As a result, we have focused our discussion on estimation of the parameters of the sunk-cost distributions.

We avoid making *some* of the assumptions on the distribution of sunk costs traditionally made in dynamic estimation algorithms. In particular, 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. One could instead 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 nonparametrically. This is a topic beyond the scope of the current paper, but it is being pursued elsewhere (see Berry and Tamer, 2006).

□ **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. We deal with a multiple-agent problem and estimate continuation values directly from the average of the discounted values of realized net cash flows (rather than inverting choice probabilities). However, the use of semiparametrics to overcome computational problems underlies both papers.

There are also a number of papers currently "in process," all written independently, that present related results. The closest to our paper is the paper by Aguirregabiria and Mira (2007), who make more restrictive assumptions on the sunk-cost distributions and introduce one of the alternative estimators considered in our extensions. 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 (forthcoming) provide assumptions and techniques that allow them to generalize the ideas in Hotz et al. (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.[1]

2. 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, the true value of which is θ_0.

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

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

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

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

Note that, to form this expectation, we need to form the incumbent's perceptions of the likely number of entrants and exitors *conditional on the incumbent itself continuing*. These perceptions generate the probability distribution, $p^c(e, x \mid n, z, \chi = 1)$, where $\chi = 1$ is notation for the incumbent continuing. It is the requirement that $p^c(e, x \mid n, z, \chi = 1)$ be consistent with competitors' behavior that generates our equilibrium conditions.

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

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

[1] 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 (2006).

where $p^e(e, x \mid n, z, \chi^e = 1)$ 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.[2]

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
   - the sunk costs of entry, say $F^\kappa(\cdot \mid \theta)$, which has a lower bound of $\underline{\kappa} > 0$, and
   - the returns to exiting, say $F^\phi(\cdot \mid \theta)$, 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, 2005).
2. Agents' 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 $\mathcal{P}_z \equiv \{p(\cdot \mid z)$, for all $z \in Z = [0, 1, \ldots, \overline{z}]\}$, $\lim_{n\to\infty} \pi(n, z, \theta_0) \leq 0$ for every $z \in Z$, and $\pi(\cdot)$ is bounded.

We now discuss the relationship between players' behavior and perceptions implied by the assumptions and equilibrium restrictions and then turn to the properties of equilibrium policy functions.

□ **Equilibrium behavior.** We begin with incumbent behavior and then move to that of potential entrants. Because entry and exit decisions are simultaneous and incumbents are identical up to the draw on exit 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

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

where, for $r \geq x$,

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

and $p^c(e \mid n, z, \chi = 1)$ is consistent with the behavior of entrants.

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

$$p^e(e, x \mid n, z, \chi^e = 1) = b^x(x, n \mid n, z, \theta) p^e(e \mid n, z, \chi^e = 1), \tag{5}$$

[2] 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 who wait and an additional set of state variables that determine the number of remaining potential entrants from prior years and the distribution of their entry costs. It would also require assumptions on the formation of incumbents' and potential entrants' beliefs about these state variables; these beliefs would depend on the history of the market. Finally, calculations of potential entrants' and incumbents' value functions would have to take into account the future evolution of the number of potential entrants and their entry costs.

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

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

where, for all $R \geq e$,

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

Note that this implies that, for incumbents,

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

□ **Properties of equilibrium policy functions.** This model is a special case of the model in Ericson and Pakes (1995). It has a Markov perfect equilibrium, but there may be more than one of them. Moreover, there is an $\overline{n}$ such that, if $n_t \leq \overline{n}$, we will never observe an $n > \overline{n}$ (the market is not profitable enough to induce entry if $n \geq \overline{n}$). Thus, each equilibrium generates a *finite*-state Markov chain in (n, z) couples; 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). The finiteness ensures that 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."[3]

It is important to note that, though our assumptions do not guarantee a unique equilibrium, they do ensure that, essentially, *there is only one profile of equilibrium policies that is consistent with a given data-generating process in the recurrent class*, i.e., as Proposition 1 below shows, for a given data-generating process, for any state (n, z) inside the recurrent class $\mathcal{R}$, the policies of players are the same in all equilibria consistent with the DGP. 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.

Proposition 1. Suppose the firms play an equilibrium, S, that has a recurrent class $\mathcal{R}$ and gives rise to some data-generating process. Let P be any set of equilibrium policy functions on states in $\mathcal{R}$ that gives rise to the behavior consistent with that data-generating process. Then, for any state (n, z) inside $\mathcal{R}$, the policy functions of any player have to be the same in S and P.

Proof. As we noted earlier, the agents only condition their perceptions of their competitors' behavior on the publicly available information (i.e., on (n, z)) and, in any 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, with probability 1, the data will eventually wander into the finite recurrent subset of points ($\mathcal{R}$) and, once in $\mathcal{R}$, will visit each point in it infinitely often. As the sample gets large, by the law of large numbers, the empirical distribution of entrants and exitors from each $(n, z) \in \mathcal{R}$ will converge to the distribution that generated it (almost surely). As noted, this must be the distribution the agents use to form their perceptions in any equilibrium and so, for the states in $\mathcal{R}$, the perceived behavior of competitors in the equilibrium with policy functions P must be the same as in equilibrium S.[4]

[3] "Communicate" here simply means that the probability of transiting from one state to another (in any number of periods) is positive. If, in addition, we place restrictions on primitives that ensure that any (n, z) communicates with $(0, z)$ for some $z \in Z$ and z itself is ergodic, then there is exactly one recurrent class, and it is of the form $\mathcal{R} = \{(n, z) : 0 \leq n \leq \overline{n}, z \in Z\}$. The relevant propositions here are Ericson and Pakes (1995), Theorems 1 and 2, and Doraszelski and Satterthwaite (2007), Proposition 2. For a review, see Doraszelski and Pakes (forthcoming), Section 3.1.

[4] In other words, for the states in the recurrent class $\mathcal{R}$, the DGP uniquely pins down each player's perception of the behavior of each of his competitors. Following equations (4) and (5), this, in turn, uniquely determines each player's perceived transition probabilities conditional on his own behavior.

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. Hence, inside the recurrent class $\mathcal{R}$, the policy functions of any player are the same in S and P.[5] *Q.E.D.*

Before concluding this section, we want to point out that 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. At the cost of a small increase in complexity, we could have weakened this assumption slightly and allowed expected profits to have an unobserved component that was independent over time. However, if this component were serially correlated, then the potential entrants and incumbents in a given period would condition their behavior on the unobserved states and the distribution of that unobserved state would depend on the current number of active firms, a complication that would require a more complex estimation algorithm. We come back to a discussion of the consequences of serially correlated unobserved state variables in our Monte Carlo analysis below, where we show that their impact is likely to depend on the nature of the data available to the researcher.

3. Equilibrium perceptions and estimation

■ In equilibrium, the perceptions of potential entrants and incumbents must be consistent with what is actually observed. This fact leads directly to a number of alternative two-step semiparametric estimators for the parameters of the model, and we begin with the simplest of them. Its first step computes averages of the realized continuation (entry) values of all firms that did continue (enter) at alternative values of (n, z). Because 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 values, 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 that 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.

□ **Estimates of $VC(\cdot)$ and $VE(\cdot)$.** This is the only detailed calculation of the paper. 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

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

where

[5] It is important to note that, 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.

- 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

$$\begin{aligned} &E_{\phi'}[\max\{VC(n', z'; \theta), \phi'\} \mid n', z'] \\ &\quad = \Pr\{\phi < VC(n', z'; \theta)\} VC(n', z'; \theta) + \Pr\{\phi > VC(n', z'; \theta)\} E[\phi' \mid \phi' > VC(n', z'; \theta)]. \end{aligned} \tag{7}$$

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

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

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

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

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

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

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 that continued from state (n, z). More precisely, we compute consistent estimates of the transition and exit probabilities (of M_c and p^x) and substitute them into (10) to produce 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 $\#T(n, z) \to \infty$. Similarly, 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 $\#T(n, z) \to \infty$,

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

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, and $\to_P$ denotes convergence in probability. 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.[6]

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

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

These estimates of continuation values are the averages of the discounted values of the returns *actually earned* by agents. This makes the relationship between the data and our estimates transparent, thus simplifying robustness analysis. It also is the reason we expect our estimates to make empirical sense; unless agents have systematically biased perceptions, the actual average of realized continuation values should be close to the expected continuation values used in decision making.

Also, 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

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

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 hence need only be computed once at the beginning of the estimation routine. So, given profits, the first-stage estimates of continuation values are linear in θ.

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

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

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).

[6] 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.

Consistent estimates of these probabilities are obtained as the fraction of those that enter in state (n, z) that then begin operations in state (n', z'); that is, by[7]

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

Accordingly, we obtain our consistent estimate of VE as

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

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

The simplicity of the form of the solution for $(\widehat{VC}(\theta), \widehat{VE}(\theta))$ 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 $E[\phi \mid \phi > \phi_0]$ 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

$$\widehat{VC}(\theta) = \tilde{M}_c\{\pi + \delta[(1 - \tilde{p}^x) \times \widehat{VC}(\theta)] + \delta[\tilde{p}^x \times E[\phi \mid \phi > \widehat{VC}(\theta)]]\},$$

where $(1-\tilde{p}^x) \times \widehat{VC}(\theta)$ is the vector formed by multiplying each $1-\tilde{p}^x(n, z)$ with the corresponding $\widehat{VC}(\theta)(n, z)$, while $\tilde{p}^x \times E[\phi \mid \phi > \widehat{VC}(\theta)]$ is the vector formed by multiplying each element of $\tilde{p}^x(n, z)$ by the corresponding $E[\phi \mid \phi > \widehat{VC}(\theta)(n, z)]$. This equation system is contraction mapping and therefore is easy to solve, if the derivative of $E[\phi \mid \phi > x]$ with respect to x is less than or equal to $1/\delta$ everywhere. This will be true if the distribution $F^{\phi}(\cdot)$ is log concave, an assumption satisfied by most distributions used in empirical work.[8]

Once we have $(\widehat{VC}(\theta), \widehat{VE}(\theta))$, we can form consistent estimates of the probability of exit and of entry conditional on θ as $(1 - F^{\phi}(\widehat{VC}(\theta) \mid \theta))$ and $F^{\kappa}(\widehat{VE}(\theta) \mid \theta))$, respectively. The second stage of the algorithm fits these probabilities to the exit and entry rates observed in the data (see Section 4 below). Note that the computational complexity of these estimators is comparable with that of estimators for the simplest static entry/exit models.

□ **Multiple locations and random $\mathcal{E}$.** We now 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 for the multinomial draws. However, there are no other new conceptual points, so the reader who is not interested in these details can skip to the next section.

---

[7] 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 $(1/\#T(n, z)) \sum_{t\in T(n,z)}([e_t/\mathcal{E}]/[\tilde{p}^e(n, z)])1_{[(n_{t+1},z_{t+1})=(n',z')]}$, where $\tilde{p}^e(n, z) = (1/\#T(n, z)) \sum_{t\in T(n,z)} e_t/\mathcal{E}$.

[8] For the derivative property, see Heckman and Honore's (1990) Proposition 1. To prove the contraction property, let $T(x)$ be the operator that produces the right-hand side of this equation when $\widehat{VC} = x$ and let $\|x\|$ denote the maximum element of the vector x. Then, because M_c is a Markov matrix, $\|T(x_1)-T(x_2)\| \leq \delta\|[(1-\tilde{p}^x)\times(x_1-x_2)]+\delta[\tilde{p}^x \times(E[\phi \mid \phi > x_1] - E[\phi \mid \phi > x_2])]\|$. Under the log-concave conditions, $\mid E[\phi \mid \phi > x_1] - E[\phi \mid \phi > x_2] \mid\leq\mid x_1 - x_2 \mid$, which, given that $0 \leq \tilde{p}^x \leq 1$, proves the result.

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 obtain a simple multiple-location model, replace Assumption 1.2 (which deals with the sunk costs of entry and exit) with the following.

Assumption 2. Instead of Assumption 1.2, assume

- the number of potential entrants in each period is an independent random draw from the distribution $\{p(E \mid \theta)\}_{E=0}^{\mathcal{E}}$ for a finite $\mathcal{E}$,
- potential entrants can enter in only one of the two locations and have entry cost (κ_1, κ_2) in the first and second locations respectively, where the vector (κ_1, κ_2) is a draw from the distribution

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

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

Because 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. Let ℓ index the locations. Then the Bellman equation for an incumbent in location ℓ is

$$V_{\ell}(n_{\ell}, n_{-\ell}, z, \phi; \theta) = \max\{\pi_{\ell}(n_{\ell}, n_{-\ell}, z) + \delta\phi, \pi_{\ell}(n_{\ell}, n_{-\ell}, z) + \delta VC_{\ell}(n_{\ell}, n_{-\ell}, z; \theta)\},$$

where

$$\begin{aligned} &VC_{\ell}(n_{\ell}, n_{-\ell}, z; \theta) \\ &\quad = \sum_{z', e_{\ell}, e_{-\ell}, x_{-\ell}, x_{\ell}} \int_{\phi'} V_{\ell}(n_{\ell} + e_{\ell} - x_{\ell}, n_{-\ell} + e_{-\ell} - x_{-\ell}, z', \phi'; \theta) p(d\phi') \\ &\qquad\qquad \times p^{c,\ell}(e_{\ell}, e_{-\ell}, x_{\ell}, x_{-\ell} \mid n_{\ell}, n_{-\ell}, z, \chi_{\ell} = 1) p(z' \mid z), \end{aligned}$$

and $p^{c,\ell}(e_{\ell}, e_{-\ell}, x_{\ell}, x_{-\ell} \mid n_{\ell}, n_{-\ell}, z, \chi_{\ell} = 1)$ provides the type-ℓ incumbent's perceived probability of $(e_{\ell}, e_{-\ell}, x_{\ell}, x_{-\ell})$ conditional on that incumbent continuing.

The incumbent views all its competitors in a particular location as identical. Consequently, 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, for $\ell = 1, 2$,

$$\begin{aligned} &p^{c,\ell}(e_{\ell}, e_{-\ell}, x_{\ell}, x_{-\ell} \mid n_{\ell}, n_{-\ell}, z, \chi_{\ell} = 1) \\ &\quad = p^{c,\ell}(e_{\ell}, e_{-\ell} \mid n_{\ell}, n_{-\ell}, z, \chi_{\ell} = 1) b_{\ell}(x_{\ell}, n_{\ell} - 1 \mid n_{\ell}, n_{-\ell}, z) b_{-\ell}(x_{-\ell}, n_{-\ell} \mid n_{\ell}, n_{-\ell}, z), \end{aligned}$$

where

$$b_{\ell}^{x}(x, r \mid n_{\ell}, n_{-\ell}, z; \theta) \equiv \tbinom{r}{x} F_{\ell}^{\phi}\{VC_{\ell}(n_{\ell}, n_{-\ell}, z; \theta) \mid \theta\}^{r-x} \left[1 - F_{\ell}^{\phi}\{VC_{\ell}(n_{\ell}, n_{-\ell}, z; \theta) \mid \theta\}\right]^{x},$$

and the perceived entry probabilities, i.e., $p^{c,\ell}(e_{\ell}, e_{-\ell} \mid n_{\ell}, n_{-\ell}, z, \chi_{\ell} = 1)$, must equal the equilibrium entry probabilities defined below.

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

$$VE_\ell(n_\ell, n_{-\ell}, z;\theta) = \sum_{e,x,z'} \int_{\phi'} V_\ell(n_\ell + e_\ell - x_\ell, n_{-\ell} + e_{-\ell} - x_{-\ell}, z', \phi';\theta) p(d\phi') b^x_\ell(x_\ell, n_\ell \mid n, z, \theta) \times b^x_{-\ell}(x_{-\ell}, n_{-\ell} \mid n, z, \theta) p^{e,\ell}(e \mid n, z, \chi^e_\ell = 1) p(z' \mid z),$$

where $n = (n_1, n_2)$ and $p^{e,\ell}(e \mid n, z, \chi^e_\ell = 1)$, 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 $-\ell$, i.e., *iff*

$$\delta VE_\ell(n_\ell, n_{-\ell}, z;\theta) \geq \kappa_\ell \quad \text{and} \quad \delta VE_\ell(n_\ell, n_{-\ell}, z;\theta) - \kappa_\ell \geq \delta VE_{-\ell}(n_{-\ell}, n_\ell, z;\theta) - \kappa_{-\ell}. \tag{15}$$

To find the equilibrium entry distribution, use this equation to compute the equilibrium entry distribution conditional on a particular number of potential entrants (on E), and then integrate out over the distribution of the number of potential entrants.

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

$$m_0 \equiv \Pr\{\kappa_1 > \delta VE_1(\cdot) \text{ and } \kappa_2 > \delta VE_2(\cdot)\},$$
$$m_1 \equiv \Pr\{\kappa_1 \leq \delta VE_1(\cdot) \text{ and } \kappa_2 > \delta VE_2(\cdot) - \delta VE_1(\cdot) + \kappa_1\},$$

and

$$m_2 \equiv \Pr\{\kappa_2 \leq \delta VE_2(\cdot) \text{ and } \kappa_1 > \delta VE_1(\cdot) - \delta VE_2(\cdot) + \kappa_2\}, \tag{16}$$

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

$$p^{e,\ell}(e_\ell, e_{-\ell} \mid n_\ell, n_{-\ell}, z, \chi^e_\ell = 1, E) = m(e_\ell - 1, e_{-\ell}, E - 1; m_0, m_1, m_2),$$

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 possible entrants) of r or

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

provided $e_\ell + e_{-\ell} \leq E$ (otherwise, $m(\cdot) = 0$).

Integrating out over the distribution of E, equilibrium perceptions are

$$p^{e,\ell}(e_\ell, e_{-\ell} \mid n_\ell, n_{-\ell}, z, \chi^e_\ell = 1) = \sum_{E \geq (e_\ell + e_{-\ell})} m(e_\ell - 1, e_{-\ell}, E - 1; m_0, m_1, m_2) \frac{EP(E \mid \theta)}{\sum_E EP(E \mid \theta)}$$

and

$$p^{c,\ell}(e_\ell, e_{-\ell} \mid n_\ell, n_{-\ell}, z, \chi_\ell = 1) = p(e_\ell, e_{-\ell} \mid n_\ell, n_{-\ell}, z, \theta) = \sum_{E \geq (e_\ell + e_{-\ell})} m(e_\ell, e_{-\ell}, E; m_0, m_1, m_2) P(E \mid \theta)$$

for the potential entrant and incumbent, respectively.[9]

[9] $EP(E \mid \theta)/\sum_E EP(E \mid \theta)$ is a potential entrant's perceived probability that there are $E - 1$ other potential entrants.

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.[10] 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.

As in the single-location model, given the matrix inverse needed for continuation values, the computational burden of obtaining estimates for the parameters 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 "set-up" time can easily become the dominant computational burden.

Finally, note that, though our estimators remain consistent when we increase the number of entry locations (or the number of states per location), their small sample properties will change. In particular, both the small-sample bias and variance of our estimator depend on the variances of the nonparametric component. For a given sized data set, the larger the number of states the fewer the number of observations per state and the larger the variance in the first-stage estimates is likely to be. So if we fix sample size and increase the number of states, we should worry more about small-sample biases and variances. We show below that certain consistent estimators are likely to have larger small-sample biases than others.

Panel data. Note that the argument given here for consistency of the estimators of the continuation and entry values required (i) that the realizations of the transitions from two observations that start at the same state are independent of one another, (ii) that the data process generates repeated observations on the same states and (iii) that equilibrium perceptions depend only on the state variables observed by the econometrician. Thus, if we are using panel data on different markets and do not make any allowance for the panel nature of the data, we are assuming independence over markets and that the same equilibrium is played in each of them.[11]

4. Alternative two-step estimators

■ The ideas discussed above can be used to obtain several different $\sqrt{n}$-consistent and asymptotically normal (or CAN) estimators. These alternatives will have different computational burdens

[10] Consistent estimates for the incumbents' and the entrants' transition probabilities in the two-location model are given by

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

and

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

As in the one-location model, these numbers are not equal to the empirical frequency of transition from state $(n_\ell, n_{-\ell}, z)$ to state $(n'_\ell, n'_{-\ell}, z')$ but rather to a weighted average of these transitions. The weights in the formula for $\tilde{M}^{c,\ell}$ are proportional to the number of incumbents continuing in location ℓ from period t to $t+1$, due to the fact that an incumbent's perceived transition probabilities are conditional on the incumbent himself staying in the market. Likewise, the weights in the formula for $\tilde{M}^{e,\ell}$ are proportional to the number of entrants actually entering in location ℓ in period $t+1$, due to the fact that a potential entrant's perceived transition probabilities are conditional on the potential entrant himself entering. The working paper version of this article (Pakes, Ostrovsky, and Berry, 2005) contains a formal derivation of these weights.

[11] If, for example, there were a national regulation that 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 that 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.

and have different large-, as well as small-sample, distributions. This section describes those estimators and discusses their properties, while the next section provides 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 semiparametric estimators with the property that the nonparametric components (the Markov transition matrices and exit and entry probabilities needed to compute $\widehat{VC}$ and $\widehat{VE}$) enter the objective function in a nonlinear way. As a result, the variance in the nonparametric component can cause small sample bias in $\hat{\theta}$, and the extent of that bias varies with the way the objective function transforms the errors in the nonparametric component. Because, as we will see, in some samples the average number of observations per (n, z) "cell" can be as small as two or three, we will want to consider this small-sample bias in choosing among estimators.[12]

Second, the limiting variance of $\hat{\theta}$ depends on the variance of the first-stage nonparametric estimates (i.e., the estimator is not "adaptive"). This follows from the fact that the derivatives of the objective function with respect to the estimates of $\widehat{VC}$ and $\widehat{VE}$ do not have a conditional expectation of zero. It is possible to use standard semiparametric techniques to obtain an analytic formula for the asymptotic variance–covariance of our parameter estimates. However, because we have a complete model and it is relatively easy to simulate data from that model and then use it to estimate new values for the parameters, we obtain consistent estimates of the needed variance–covariance matrix from a parametric bootstrap.[13]

Alternative CAN estimators can be obtained by varying each of three different components of the algorithm. We consider "natural" suggestions for each of these components, and the Monte Carlo examples consider all possible combinations thereof. We first list the components and our suggestions and then comment on them.

I. The objective function. We consider using:

1. a pseudo log-likelihood function, which, in the single-location model, would be formed by taking the sum over t of

$$\ell(x_t, e_t \mid \theta) = (n_t - x_t)\log F^{\phi}\{\widehat{VC}_t(\theta) \mid \theta\} + x_t \log[1 - F^{\phi}\{\widehat{VC}_t(\theta) \mid \theta\}] \\ + e_t \log F^{\kappa}\{\widehat{VE}_t(\theta) \mid \theta\} + (\mathcal{E} - e_t)\log[1 - F^{\kappa}\{\widehat{VE}_t(\theta) \mid \theta\}].$$

2. 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 (or a pseudo minimum χ^2 estimator), and

3. 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 entry and exit rates predicted by the model for different values of θ.[14]

II. The estimated transition probabilities. We consider using

1. the empirical Markov matrix defined above and

2. computing estimates of the entry and exit probabilities at each z at each location (say by maximum likelihood), and then substituting these probabilities into the multinomial formula to generate the implied Markov matrix for each (n, z) (this is our "structural" estimate of the transition matrix).

[12] 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.

[13] 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 $p(z' \mid z)$ 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.

[14] 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.

III. The continuation and entry values. We consider using

1. the discounted sum of future profits given above, and
2. substituting the profit function and estimates of the transition probabilities into the contraction mapping defining a single agent's value function (equation (2)) and computing that mapping 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. These transition probabilities are then used to produce new first-stage estimates of continuation and entry values either using the matrix inversion in (III1) or the nested fixed point in (III2).[15] 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 no guarantee that they converge to an outcome that is consistent with our assumptions on the choice of equilibrium.

□ **Alternative objective functions.** Because we use estimated rather than actual exit and entry probabilities, 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 ensure that the Markov chain generated by the model is finite, see Assumption 1.2), one of the parameters to be estimated defines a point of support of the likelihood, which invalidates the standard regularity conditions for maximum likelihood.

So the usual limiting arguments in favor of maximum likelihood do not apply here. Moreover, there are two arguments that 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, because 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 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 (I2) is, like the log likelihood, likely to have problems due to the nonlinearities it induces. To see this, let s index states, $\hat{F}_s^\kappa(\theta) = F^\kappa(\widehat{VE}_s(\theta))$ 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

[15] Indeed, the estimator in Aguirregabiria and Mira (2007) noted in our introduction is, in our terminology, a pseudo–maximum-likelihood estimator as in (I1) that uses the structural transition matrices in (II2) and the nested fixed point to compute entry and continuation values as in (III2), and they consider a multistep version of their estimator that 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).

$$
\begin{aligned}
&[\hat{q}_e(s) - \hat{F}_s^{\kappa}(\theta)]\frac{\partial \hat{F}_s^{\kappa}(\theta)}{\partial \theta} \\
&\quad \equiv [\hat{q}_e(s) - F_s^{\kappa}(\theta)]\frac{\partial \hat{F}_s^{\kappa}(\theta)}{\partial \theta} + [F_s^{\kappa}(\theta) - E\hat{F}_s^{\kappa}(\theta)]\frac{\partial \hat{F}_s^{\kappa}(\theta)}{\partial \theta} + [E\hat{F}_s^{\kappa}(\theta) - \hat{F}_s^{\kappa}(\theta)]\frac{\partial \hat{F}_s^{\kappa}(\theta)}{\partial \theta},
\end{aligned}
$$

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, because $\hat{F}^{\kappa}(\cdot)$ is a nonlinear function of the first-stage estimation error and $\hat{F}^{\kappa}(\cdot)$ and $\partial \hat{F}^{\kappa}(\cdot)/\partial\theta$ are constructed from the same estimates of $\widehat{VE}$ (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., $\partial \hat{F}_s^{\kappa}(\theta)/\partial\theta$), which is correlated with the error in the estimate of the probability. Thus, if we replaced $\partial \hat{F}_s^{\kappa}(\theta)/\partial\theta$ 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.[16] The simpler method-of-moments estimator in (I3) is a case in point in which the "instrument" becomes "one" for all observations.

□ **Empirical vs. structural transition matrices.** 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 (because we do not observe transitions from these states), and the imputation errors will affect our estimators.[17] Second, to go from the binomial probabilities to the probabilities needed for the transitions from n to n' requires the computation of a complex probability distribution 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 III2). This increases the computational burden of the estimator, and the increase will be larger the larger the number of state variables in the model.

□ **Matrix inversions vs. nested fixed points.** The nested fixed-point algorithm finds its estimates of continuation values by computing the contraction mapping

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

where

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

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

At least in cases where the matrix inversion formula in (9) is available, the nested fixed-point calculation will increase the computational burden of the estimator. When we use the matrix

[16] The argument here is very similar to the argument against using nonlinear 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).

[17] 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 that 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.

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.[18] This extra computational burden grows exponentially in the number of state variables (or locations) and will increase in the number of parameters to be estimated (as then the search routine will typically require more function evaluations). The extra burden will be compounded when we use the nested fixed point with structural probabilities, as then we will have 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 estimates of the continuation and entry values.

5. Monte Carlo

■ We begin with estimates of entry fees and sell-off values in first a one-location and then a two-location model.[19] Our focus is on the computational burden and parameter distributions generated by the alternative estimators.

For each model, we compute equilibrium policies using a variant of the Pakes and McGuire (1994) algorithm (we shut down the investment decision in that algorithm). These policies are then used to simulate a long time series of data. The data sets used for estimation are constructed by initiating C separate time series of length T from the approximate ergodic distribution of states generated by this simulation. 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.

Much of the earlier work on entry focused on the relationship between market size and the number of active firms in relatively isolated markets or, in the multiple-location case, between market size and the relationship between the number of firms in the different "locations" (see Bresnahan and Reiss, 1987; Mazzeo, 2002; Seim, 2005). There are about 250 isolated markets in South Dakota, and we chose the support of our population variable to mimic the range of population sizes in these markets. The actual population-growth rates in these markets were serially correlated, so we allowed for serially correlated growth rates in our sample also. Thus, the exogenous variables (i.e., the z in our notation) in the Monte Carlo consist of two variables; a level and a growth rate for market size.

In each case, we started out with six sample designs, a time dimension (T) of 5 or 15, and the number of markets or the cross-sectional dimension (our C) of 250, 500, or 1000. A seventh design of $C = 50$ and $T = 5$ was suggested by a referee. It gives us a chance to examine the problems that arise in data sets that are quite small by recent standards.[20]

□ **Our single-location example.** This example uses a Cournot model with linear demand to determine output and profits conditional on (n, z). z changes shift the demand curve, as does market size in Bresnahan and Reiss (1987). Appropriate choice of parameter values gives us

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

and we assume that $z = log[Z]$ is a second-order Markov process, so that there is persistence in growth rates, i.e., $z_{t+1} = z_t + g_{t+1}$, where g, the growth rate, is a first-order Markov process. Thus, the state variables for the dynamic problem are (n, g, z).

[18] If we assume the sell-off value distributes i.i.d. type II extreme value as in Aguirregabiria and Mira (2007), the integral over ϕ has an analytic form, which makes each calculation of the fixed point easier.

[19] That is, we assume the profit function was estimated elsewhere. 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).

[20] Mazzeo (2002) has 492 interstate exits as his markets, while Seim (2005) has data from 151 city groups with an average of 21 related markets per city group.

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 ensures that there will be no entry when there is a sufficient number of incumbents. The sell-off value is distributed exponentially with parameter σ.

In our parameterization, the maximum number of firms ever active was nineteen, z took on 45 values and we allowed three growth rates $(.05, 0, -.05)$.[21] This implies that the number of distinct (n, g, z) vectors possible is about 2700.

Monte Carlo results. 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. Its first row 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 fit the pseudo-likelihood; and if $OF = 2$, we fit 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 estimate entry and exit probabilities and use them and the binomial formula to compute structural transition probabilities. If $PR = 0$, we use the empirical transition probabilities (or $\tilde{M}$ above). The VF row indicates how the first-stage estimates of $(VC(\cdot), VE(\cdot))$ are computed. If $VF = 1$, they are computed via a nested fixed point; and if $VF = 0$, they are found by a single matrix inversion at the start of each run.

Panel A of Table 1 provides the results when $C = 1000$ and $T = 15$.[22] These results indicate that, with this design and a big enough sample, all estimators work "reasonably" well, though, even in this sized panel, there is some indication that $OF = 0$ estimators of σ are preferred to those from the other objective functions.[23] When we decrease the number of observations to $T = 5$, $C = 250$ (panel B) and $T = 15$, $C = 500$ (panel C), we learn more about the estimators' performance.

All the estimates of a have an upward bias. An upward bias in a implies a downward bias in the estimate of $1/a$. For the $OF = 1$ or pseudo–maximum-likelihood 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 that generate entry. In almost all observations, this derivative is positive, and that accounts for the positive deviations in those estimators. Note, however, that though there is a bias problem in $\hat{a}$ in small samples, it is still estimated quite precisely; and once we increase the sample size, the bias problem disappears

[21] 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.

[22] 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; (ii) 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, we sometimes get to terminal conditions that 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.

[23] The poor performance of the $\hat{\sigma}$ 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.

TABLE 1 One Location

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												
$a = .3$	.30	.30	.30	.30	.33	.31	.31	.31	.30	.30	.30	.30
SD(a)	.00	.00	.00	.00	.03	.01	.01	.00	.00	.00	.00	.00
$\sigma = .75$	.75	.75	.75	.75	.74	.72	.74	.74	.83	.86	.75	.75
SD(σ)	.01	.01	.01	.01	.02	.07	.01	.07	.56	1.00	.01	.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												
$a = .3$	.37	.37	.38	.37	.40	.37	.38	.37	.36	.36	.37	.37
SD(a)	.03	.03	.03	.03	.05	.03	.02	.02	.03	.03	.03	.03
$\sigma = .75$	.77	.75	.76	.74	.75	.71	.74	.71	9.30	1.85	5.02	.78
SD(σ)	.04	.04	.04	.04	.04	.07	.04	.03	27.15	7.20	19.92	.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	1.3	.1	32.9	.5	8.8	.5	207.4	.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												
$a = .3$	.32	.32	.32	.32	.35	.33	.33	.32	.31	.32	.32	.32
SD(a)	.01	.01	.01	.01	.04	.01	.01	.01	.01	.01	.01	.01
$\sigma = .75$	.75	.75	.75	.74	.74	.72	.74	.73	1.28	.97	.97	.75
SD(σ)	.02	.02	.02	.02	.02	.02	.02	.02	6.29	1.16	4.44	.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	2.2	.1	34.1	1.6	17.5	1.5	240.8	.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

Note: 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 transition probabilities. C and T are the cross sectional and time dimensions of the panel and R is the number of Monte Carlo repetitions.

rather rapidly. The $OF = 0$ estimates of σ are "right on", and the $OF = 1$ estimates are close, but the $OF = 2$ estimates of σ can be problematic.

No matter the sample design, the $OF = 0$ (or simple method of moments) estimators had both smaller biases *and* smaller variances than those of either the pseudo–maximum-likelihood ($OF = 1$) or the pseudo minimum χ^2 ($OF = 2$) estimators. This illustrates the impact of estimation error in $(\widehat{VC}, \widehat{VE})$ on the relative performance of the alternative objective functions. At $T = 15$, there is an average of six observed transitions per state visited, and at $T = 5$, that average is just 3. Perhaps what is most striking is how well the $OF = 0$ estimator handles this estimation error. Of course, each observed transition averages over all sample points reached from the state to which it transited, so *each* observed transition is implicitly *averaging* over observed sample paths from the point to which it transited.

The differences in the distributions generated by the alternative $OF = 0$ estimators are small, but 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 for which we observe data on transitions, and the $\#(\widehat{p})$ row, which is the number of states for which we need to compute transition probabilities). 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

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	.00	1.00	.50	.00	1.00	.50
a = .3	.47	.28	.28	.37	.28	.28
SD(a)	.08	.03	.03	.03	.01	.01
σ = .75	.82	.71	.72	.77	.71	.72
SD(σ)	.11	.08	.08	.04	.04	.04
Average % Bias, % Standard Error, and % MSE of *VE* and *VC* Estimates						
VE bias	.03	.20	.18	.01	.16	.14
VE "se"	.21	.15	.18	.13	.06	.07
VE "$\sqrt{MSE}$"	.21	.29	.29	.13	.24	.22
VC bias	.02	.05	.04	.01	.03	.03
VC "se"	.24	.11	.13	.15	.05	.06
VC "$\sqrt{MSE}$"	.25	.18	.18	.15	.16	.15

Note: 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.

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.[24] 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 the results indicate a clear preference for the simple method-of-moments estimators that uses empirical transitions probabilities (i.e., $OF = PR = 0$). 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 that uses $OF = PR = 0$ but the nested fixed point for calculating values ($VF = 1$).[25] Note that neither of these two estimators is particularly computationally burdensome. The compute time with $T = 5$ and $C = 250$ was about 10 seconds, and it was only about a half a minute on our largest sample.

We now move to Table 2, which provides estimates from our smallest samples. Given the results in Table 1, we only use the $OF = PR = VF = 0$ estimators in this table. However, because the samples are 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.

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 than do the

[24] 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 80% of the states, the imputation does not seem to cause a noticeable bias in the estimates.

[25] As noted, were we doing a more complex problem or estimating more parameters, we would see more of a computational preference for the $OF = PR = VF = 0$ over the $OF = PR = 0$ but $VF = 1$ estimator. On the other hand, there does seem to be a slight improvement in the statistical performance of the estimator with $VF = 1$ over $VF = 0$ in the smaller samples.

nonsmoothed 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 nonsmoothed estimators.[26]

□ **Our two-location example.** This 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, he can choose either the low- or high-quality good. Consumers are differentiated by their price coefficient (meant to mimic their marginal utility of income), and the inverse of that coefficient (which should be increasing in income) distributes exponentially. The model generates demand for the low- and high-quality options, respectively, as

$$Q_1 = M\left(e^{-\lambda\frac{p_1}{\delta_1}} - e^{-\lambda\frac{p_2-p_1}{\delta_2-\delta_1}}\right) \quad \text{and} \quad Q_2 = M\left(e^{-\lambda\frac{p_2-p_1}{\delta_2-\delta_1}}\right),$$

provided $(p_2 - p_1)/(\delta_2 - \delta_1) > p_1/\delta_1$ (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 is 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 $c_2/c_1 > \delta_2/\delta_1$, as this guarantees positive equilibrium quantities.

We assume 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. Because κ_1 and κ_2 reflect differences in a given firm'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 for the same firm of building the high-quality motel in a given market is larger than its cost of building the low-quality motel, or that $\kappa_2 > \kappa_1$ with probability one. In particular, we assume the cost of entry into the low-quality product distributes as does the entry cost in the one-location model (equation (17)) and 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,[27] so there is a two-parameter entry distribution (a_1, a_2). Exit fees are i.i.d., exponential with parameters (σ_1, σ_2) in the two locations.

Monte Carlo results. We begin with a summary of the results from a number of runs on large samples that were reported in earlier versions of this paper, as they allowed us to rule out certain estimators. Turning first to the pseudo–maximum-likelihood estimators, they typically failed in the sense that they could only find zero values for the likelihood. The reason this occurred was that the first stage produced a $\widehat{VE}_1(\cdot) > \widehat{VE}_2(\cdot)$ for all parameter values when there was entry in location 2. Because 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 did happen is $VE_1(\cdot) < VE_2(\cdot)$; i.e., the disturbances in the first-stage estimates have reversed the order of the two entry values. Because the pseudo-likelihood does not recognize the possibilities generated by first-stage estimation error, it can record a probability of zero for events that do happen. If this occurs even for *only one* state, the pseudo–maximum-likelihood estimate will fail.[28] Clearly, we should not be using pseudo–maximum-likelihood estimates, at least not

[26] 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 that deserves more research. In particular, there are a number of alternative estimation procedures that 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.

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

[28] Because 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 that were

without an auxiliary procedure that ameliorates the problems caused by imprecise first-stage estimates, so we ignore pseudo–maximum-likelihood estimates in what follows.

Even on sample sizes as large as ($C = 1000, T = 15$), the estimators that used the pseudo minimum χ^2 objective function ($OF = 2$) performed very poorly. In the two-location model, there are fewer incumbents in each state (because 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 first-stage estimators were 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 4), makes those estimators problematic.

Table 3 presents the simple method-of-moments (or $OF = 0$) estimators. With samples of size $C = 1000$ and $T = 15$, all of the $OF = 0$ estimators did quite well. The estimates of σ that use structural probabilities (i.e., $OF = 0$, but $PR = 1$) do have slightly larger mean-square errors, which likely result from 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. However, there are stark differences in the computational burdens of the alternative $OF = 0$ estimators. Estimates that use structural probabilities and $VF = 0$ are 15 times as computationally burdensome as those that use empirical probabilities and $VF = 0$, and estimates that use the fixed point combined with the structural probabilities ($VF = PR = 1$) are 20 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 this would increase were we estimating more parameters).

Moving to the smaller samples, we note that, because 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 rapidly.

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;[29] so the estimators work most of the time even for our small samples, but an econometrician who uses them has to use some judgment. The outlier problem in small samples disappears when we use kernels (see below). 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. Because 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. Moreover, the computational burden of the estimators that use structural probabilities is 20 or more times that of the corresponding estimators that use the empirical probabilities.

Comparing the two estimators that use $OF = PR = 0$, we find that, in the smallest two samples, it is clear that we prefer the estimator that does the matrix inversion ($VF = 0$) to the estimator that uses the nested fixed point ($VF = 1$). However, by the time we get to a sample size

visited infrequently. Though trimming did help somewhat with the amount of times we could find a positive likelihood, it also biased the estimates and caused their variance to go up (in some cases dramatically).

[29] There were 9 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, in which a_2 was estimated to lie on the boundary of our search space. Excluding these outliers, the average estimates of a_2 are .51 and .52 respectively, with standard errors of .41 and .42.

TABLE 3 Two Locations, OF = 0, Different Sample Sizes

PR	0	0	1	1	0	0	1	1
VF	0	1	0	1	0	1	0	1
	T = 15, C = 1000, R = 50				**T = 15, C = 500, R = 500**			
$a_1 = .3$	.30	.30	.29	.29	.32	.32	.29	.29
SD(a_1)	.01	.01	.02	.02	.02	.02	.02	.02
$a_2 = .3$	.30	.30	.30	.30	.32	.32	.31	.31
SD(a_2)	.02	.02	.03	.02	.03	.03	.03	.03
$\sigma_1 = 1$	1.02	1.00	1.06	1.02	1.06	1.02	1.17	1.08
SD(σ_1)	.09	.07	.10	.07	.16	.11	.21	.13
$\sigma_2 = .5$	.51	.50	.52	.50	.53	.51	.56	.52
SD(σ_2)	.03	.03	.04	.03	.07	.04	.09	.05
t(setup)	47	47	920	920	26	26	730	730
t(search)	22	27	177	385	13	15	104	251
t(total)	69	75	1096	1305	39	41	834	981
$\#(n, g, z)$	595	595	595	595	531	531	531	531
$\#\hat{p}$	602	602	3463	3463	537	537	3160	3160
	T = 5, C = 500, R = 500				**T = 5, C = 250, R = 500**			
$a_1 = .3$	.34	.34	.28	.28	.39	.39	.28	.28
SD(a_1)	.04	.04	.03	.03	.11	.10	.04	.04
$a_2 = .3$	.39	.39	.31	.31	2.30	2.49	.34	.33
SD(a_2)	.08	.08	.06	.05	13.25	13.95	.66	.52
$\sigma_1 = 1$	1.10	1.36	1.74	1.57	1.35	3.52	2.34	2.37
SD(σ_1)	.73	2.33	.97	1.15	1.84	9.04	1.87	2.25
$\sigma_2 = .5$	.64	.81	.82	.71	.68	1.90	1.10	1.08
SD(σ_2)	.88	2.57	.55	.51	.85	4.45	.90	1.62
t(setup)	15	15	514	514	10	10	396	396
t(search)	9	11	85	210	8	9	71	210
t(total)	24	26	599	724	18	19	467	605
$\#(n, g, z)$	404	404	404	404	321	321	321	321
$\#\hat{p}$	427	427	2918	2918	347	347	2682	2682

Note: See the footnote to Table 1.

of $C = 500$ and $T = 15$, when both estimators are doing reasonably well, the nested fixed-point estimator seems to have marginally better performance.

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 4 presents the results. The kernels do away with the problem of extreme outliers that appears in the unsmoothed estimators. So 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. The iterated estimators used the $OF = 0$ but $PR = VF = 1$ estimators of θ to construct structural probabilities, then used the fixed point to get the continuation values generated by the structural transition probabilities and different values of θ, and then minimize $OF = 0$. The iterations seem to cause cycles of increasing amplitude in the estimators, and this reinforces the theoretical problems with iterated estimators (the iterations need not converge and because 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 iterations greatly increase the computational burden.

□ **Robustness to serially correlated unobservables.** The Monte Carlo results presented thus far investigated the distributional and computational properties of our estimators. In empirical work, we typically also worry about the specification of the model. An advantage of our estimator

TABLE 4 Kernel and Iterated Estimators for T = 5, C = 250

Two Locations, Kernel Estimates of *VE* and *VC*				
Bandwidth	0	0-No Bad Runs	1	.5
$a_1 = .3$	.39	.39	.18	.18
SD(a_1)	.11	.11	.03	.03
$a_2 = .3$	2.30	.51	.30	.29
SD(a_2)	13.25	.41	.15	.14
$\sigma_1 = 1$	1.35	1.35	.66	.73
SD(σ_1)	1.84	1.86	.08	.13
$\sigma_2 = .5$	.68	.67	.33	.35
SD(σ_2)	.85	.83	.03	.04
# runs	500	491	500	500
Iterated Estimates: From OF = 0, PR = 1, VF = 1; R = 50				
Iteration	1	2	3	4
$a_1 = .3$	.21	.28	.15	.26
SD(a_1)	.04	.03	.04	.04
$a_2 = .3$	.38	.29	.39	.30
SD(a_2)	.24	.06	.41	.08
$\sigma_1 = 1$	1.13	7.21	1.43	7.53
SD(σ_1)	.31	8.26	.97	5.97
$\sigma_2 = .5$	.54	2.53	.62	5.46
SD(σ_2)	.14	2.44	.21	8.07
t(setup)	351	351	349	352
t(search)	183	286	204	326
t(cumulative)	1088	1725	2278	2956

Note: See the footnotes to Tables 1 and 2.

is that its simplicity should facilitate an analysis of the impacts of different profit functions or distributional assumptions. However, the estimators presented thus far cannot account for serially correlated unobserved state variables.

In thinking about this problem, it is important to distinguish between cases in which we have measures of profits conditional on the state of the system and cases where no such information is available (so that the parameters 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, the presence of an unobserved serially correlated state variable is expected to lead to an upward bias on the (usually negative) effect of the number of incumbents on profits.[30] However, if there is an independent measure of profits, this bias does not occur and, as we now show, there is not an inconsistency in all the dynamic measures of interest.

For simplicity, consider our one-location model and assume the exogenous variable (z) is both serially correlated and unobserved by the econometrician. From equation (8), continuation values depend on (n, z). Use the invariant distribution of z given n, say $p(z \mid n)$, to form the average continuation value for a given n,

$$\widetilde{VC}(n) = \sum_z VC(n, z)p(z \mid n).$$

Because the empirical distribution of z conditional on n will converge to $p(z \mid n)$, our estimates

[30] Entry will be higher and exit will be lower when that unobserved state variable is higher. Because 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 is higher. Consequently, there will be a positive bias in the estimated relationship between the number of incumbents and profits.

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 $p(z \mid n)$. Though knowledge of $\{\widetilde{VC}(n), \widetilde{VE}(n)\}_n$ is likely to be useful per se, to do a complete analysis of entry/exit responses, we also need the distribution of entry and exit costs.

One possibility is to fit the average entry and exit rate at each n to the rates predicted by $\{\widetilde{VC}(n), \widetilde{VE}(n)\}_n$ and the distributions of entry and exit costs. Because potential entrants and incumbents condition on the realization of z at the time they make their decisions and the probabilities of entry and exit are convex functions of the entry and continuation value, this estimation routine will generate inconsistent estimators, but there is a question of just how large the asymptotic bias would be. We did a Monte Carlo analysis to get some idea of the answer to this question.

Briefly, we computed the equilibrium of our two-location model with an additional serially correlated state variable, a variable that shifted entry and exit costs from period to period. We then used the equilibrium policies to simulate data from industries in which these four state variables determined behavior. Finally, we "pretended" that this data was generated from a model with only three state variables and used our estimators assuming that the *misspecified* three-state variable model generated the data. The results were encouraging. In particular, in large data sets, none of the parameter estimates were centered at a point more than 3% different then the true parameter value (for tabled results and more detail on the model, see the earlier version of this paper). We conclude that, when independent data on profits are available, their is reason to believe that the extent of the asymptotic bias caused by serially correlated unobservables is quite small.

6. Conclusions

■ This paper provides estimators for the parameters of discrete dynamic games that are easy to use and examines their properties. We begin by providing assumptions that ensure that there is a unique equilibrium associated with a 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 (incumbents) who entered at (continued from) particular 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 from a matrix inversion, which need only be done once at the beginning of the estimation run (whatever 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 were suggested and both their distributional and their computational properties were investigated. The results were unusually clear cut. The fact that the multinomial probabilities that go into the objective functions are not known functions of the data and parameter vector, but rather semiparametric estimates of those functions, generates a clear preference for the simplest method-of-moments objective function (our $OF = 0$ case, which is constructed by instrumenting state-specific differences between observed and empirical entry and exit rates with state-specific observables). Use of the empirical transition matrix, rather than multinomial probabilities computed from the observed entry and exit rates, minimizes both computational time and imputational errors. Finally, if available, continuation and entry values computed from the matrix inversion have a slight computational advantage.[31]

[31] The computational differences between the matrix inversion and the nested fixed-point algorithms will be accentuated in more complex models with more states and parameters. This because (i) the matrix inversion need only be done once, rather than repeatedly, in the estimation algorithm and, (ii) if we use empirical probabilities, the matrix to be inverted is likely to be "sparse," and hence relatively easy to invert.

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 substituting kernel averages over nearby entry and continuation values, for the pointwise estimates of these values, into the algorithm.

The good news is that the computational burdens of the preferred estimators were small enough for us to expect the effective barrier to the empirical analysis of discrete dynamic games to shift from being computational to being the richness of the data available to support the analysis. There remains the question of the extent of modeling complexity (increases in the number of distinct states) that can be supported by alternative data sets. There is reason for some optimism here. When we use empirical transition probabilities, the number of states grows, as does the size of the recurrent class, which in I.O. models tends to grow at a much smaller rate than the size of the state space per se (see Pakes and McGuire, 2001). Also, computerization is making larger data sets increasingly available.

References

AGUIRREGABIRIA, V. AND MIRA, P. "Sequential Estimation of Dynamic Discrete Games." *Econometrica*, Vol. 75 (2007), pp. 1–53.

BAJARI, P., BENKARD, L., AND LEVIN, J. "Estimating Dynamic Models of Imperfect Competition." *Econometrica*, forthcoming.

BERRY, S. "Estimation of a Model of Entry in the Airline Industry." *Econometrica*, Vol. 60 (1992), pp. 889–917.

——— AND TAMER, E. "Identification in Models of Oligopoly Entry." Working Paper, Yale University, 2006.

BRESNAHAN, T. AND REISS, P. "Do Entry Conditions Vary Across Markets." *Brookings Papers on Economic Activity, Microeconomics Annual*, Vol. 1 (1987), pp. 833–882.

——— AND ———. "Entry and Competition in Concentrated Markets." *Journal of Political Economy*, Vol. 99 (1991), pp. 977–1009.

DORASZELSKI, U. AND PAKES, A. "A Framework for Applied Dynamic Analysis in I.O." In M. Armstrong and R.H. Porter, eds., *The Handbook of Industrial Organization*. Amsterdam: North Holland, forthcoming.

——— AND SATTERTHWAITE, M. "Computable Markov-Perfect Industry Dynamics: Existence, Purification, and Multiplicity." Working Paper, Harvard University, 2007.

EINAV, L. "Not All Rivals Look Alike: Estimating an Equilibrium Model of the Release Date Timing Game." Working Paper, Stanford University, 2003.

ERICSON, R. AND PAKES, A. "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies*, Vol. 62 (1995), pp. 53–82.

FREEDMAN, D. *Markov Chains*. New York: Springer Verlag, 1983.

HECKMAN, J. AND HONORE, B. "The Empirical Content of the Roy Model." *Econometrica*, Vol. 58 (1990), pp. 1121–1149.

HOTZ, J. AND MILLER, R. "Conditional Choice Probabilities and the Estimation of Dynamic Models." *Review of Economic Studies*, Vol. 60 (1993), pp. 265–289.

LAFFONT, J.J., OSSARD, H., AND VUONG, Q. "Econometrics of First-Price Auctions." *Econometrica*, Vol. 63 (1995), pp. 953–980.

MAGNAC, T. AND THESMAR, D. "Identifying Dynamic Discrete Decision Processes." *Econometrica*, Vol. 70 (2002), pp. 801–816.

MASKIN, E. AND TIROLE, J. "A Theory of Dynamic Oligopoly, I and II." *Econometrica*, Vol. 56 (1988), pp. 549–599.

MAZZEO, M. "Product Choice and Oligopoly Market Structure." *RAND Journal of Economics*, Vol. 33 (2002), pp. 221–242.

MUTH, J. "Rational Expectations and the Theory of Price Movements." *Econometrica*, Vol. 29 (1961), pp. 315–335.

PAKES, A. "Econometrics and Theory in Applied I.O." Fisher-Schultz Lecture, The World Congress of The Econometric Society, London, 2005.

——— AND MCGUIRE, P. "Computing Markov-Perfect Nash Equilibria: Numerical Implications of a Dynamic Differentiated Product Model." *RAND Journal of Economics*, Vol. 25 (1994), pp. 555–589.

——— AND ———. "Stochastic Algorithms, Symmetric Markov Perfect Equilibria and The 'Curse' of Dimensionality." *Econometrica*, Vol. 69 (2001), pp. 1261–1281.

———, OSTROVSKY, M., AND BERRY, S. "Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry/Exit Examples)." Working Paper, Harvard University, 2005.

PESENDORFER, M. AND SCHMIDT-DENGLER, P. "Identification and Estimation of Dynamic Games." NBER Working Paper no. 9726, 2003.

RUST, J. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica*, Vol. 55 (1987), pp. 999–1033.

SCHMIDT-DENGLER, P. "The Timing of New Technology Adoption: The Case of MRI." Working Paper, London School of Economics, 2006.

SEIM, K. "An Empirical Model of Firm Entry with Endogenous Product-Type Choices." Working Paper, Stanford University, 2005.

STARR, A.W. AND HO, Y.C. "Nonzero-Sum Differential Games." *Journal of Optimization Theory and Applications,* Vol. 3 (1969), pp. 184–206.