COMMON SENSE AND SIMPLICITY IN
EMPIRICAL INDUSTRIAL ORGANIZATION

Ariel Pakes

Common Sense and Simplicity in Empirical Industrial Organization
Ariel Pakes
NBER Working Paper No. 10154
December 2003
JEL No. C1, L0

ABSTRACT

This paper is a revised version of a keynote address delivered at the inaugural International Industrial Organization Conference in Boston, April 2003. I argue that new econometric tools have facilitated the estimation of models with realistic theoretical underpinnings, and because of this, have made empirical I.O. much more useful. The tools solve computational problems thereby allowing us to make the relationship between the economic model and the estimating equations transparent. This, in turn, enables us to utilize the available data more effectively. It also facilitates robustness analysis and clarifies the assumptions needed to analyze the causes of past events and/or make predictions of the likely impacts of future policy or environmental changes. The paper provides examples illustrating the value of simulation for the estimation of demand systems and of semiparametrics for the estimation of entry models.

Ariel Pakes
Department of Economics
Littauer 117
Harvard University
Cambridge, MA 02138
and NBER
ariel@ariel.fas.harvard.edu

1   Introduction

What I want to argue is that advances in econometrics, in combination with at least the basics of a theoretical structure, and greatly facilitated by the parallel and interrelated growth of better data and computing facilities, have made

sensible things simple

and that, as a result, empirical industrial organization has become much more useful.

I will argue this through two examples. They are designed to show how new econometric tools enable us to dramatically improve our estimates of objects which are of fundamental importance to the analysis of market outcomes: demand functions and entry costs. The econometric tools used in these examples, simulation and semiparametrics, are also the two tools that have been most intensively used of late in empirical industrial organization.

1.1   Background Notes

Why give a talk with this title? My motivation comes from the need for an antidote to the disparaging comments that I have, of late, been hearing on the use of "complex" econometric tools, particularly on the use of econometrics in conjunction with tools from economic theory.

Many of us would agree that most good empirical work starts out with simple correlations of one form or the other. First and foremost the correlations provide a summary of what the model has to explain. They often also provide a basis for determining which among alternative more detailed structures is more appropriate.

Probably the most important role of the new econometric tools in empirical I.O. has been to enable us to use more sensible theoretical structures to explain these correlations. In this context the theoretical structures play two roles. First they make it possible to make the link between market outcomes and the decisions taken by economic units. Less appreciated, however, is that the theory also makes it possible to utilize the available data in a "sensible" way. It is this latter point that the examples are designed to illustrate.

The examples show that were we to use a realistic theoretical structure to interpret either demand or entry data, and, in addition, limit ourselves to standard econometric tools, we would run into rather immense computational problems. The new econometric tools alleviate these computational problems. Moreover they do so in a way the makes the relationship between the economic model and the estimating equations transparent. It is this transparency that makes it easy to see where the model might be lacking and how it might be improved.

Though the relationship between the estimating equations and the economic model is typically easy to understand, the resulting parameter estimates do not necessarily have a simple relationship to the underlying raw data moments. The parameter estimates do, however, describe the primitives that determine the behavior of the underlying economic agents. As a result their reasonableness can often be judged by consistency with the other information available on these primitives. Moreover, under the assumptions of the model, they are the parameters needed to either (i) evaluate the causal impacts of past events or (ii) analyze the counterfactuals needed for the evaluation of possible future policies. Of course if the assumptions are wrong then so might be the analysis. However one of the advantages of writing down the model is that it clarifies the assumptions needed for the analysis, and this allows us to engage in sensible robustness analysis. The robustness analysis is typically greatly facilitated by the transparency of the estimation procedure and the power of modern computers.

Put differently, the emphasis on integrating theory and the power of the modern econometric techniques allows us to focus the empirical work on determining

- the primitives generating the returns to actions (the demand and cost systems, the rules which determine the institutional environment, etc.) and

- appropriate behavioral assumptions (interpreted broadly enough to include the choice of equilibrium).

We then evaluate alternative explanations of past phenomena by working with the implications of these more primitive constructs, and we evaluate counterfactuals by making the changes in either the primitives or the behavioral assumptions that the counterfactuals imply.

Of course the "real world" is complex and we will never get the model exactly "right". That, however, is also a rather naive goal. The question is not whether a paper has gotten it "right" but rather whether the paper has provided a more meaningful approximation than the next best alternative. Firms are going to use data to help make decisions, agencies are going to use it to help determine policies, and academics are going to use it to interpret market outcomes, whether we like it or not. The only question is whether we can improve on how this is being done.

The fact that the model is not exactly right does, however, imply that we should not be using standard test statistics to either accept or reject a given model, or to choose between alternatives.[1] Our success is better evaluated by demonstrating the ability of the "structural" model to make sense of aspects of the data that the simpler models could not. Similarly "stopping rules" for model selection should be based on trading off the incremental complexity resulting from relaxing assumptions against any improvements in the quality of the analysis the more complex model delivers.

1.2 The Examples

I first consider the use of simulation to overcome aggregation problems in estimation.[2] Simulation allows us to base the analysis on microeconomic models but base the estimation on the market outcomes we more typically observe and are primarily interested in. In this context I will focus on the estimation of market level demand systems. The economic model here is any demand model that allows consumers' choices to depend on their attributes (i.e. on their income, family size, location of residence, and so on). What simulation does is

---

[1]Though sometimes it does make sense to employ them as summary statistics to be compared to the analogous summary statistics for other models (either computed on the same data or on other data sets of *similar* size).

[2]This use of simulation dates back to Pakes (1986).

allow us to sum up over the demands generated by this model and the *actual* distribution of consumer attributes in the market of interest. This generates a prediction for demand whose relationship to the micro model and the data are transparent. The estimator is obtained by simply finding the value of the model's parameter vector that makes these predictions as close as possible to the observed market demands. This is an application that has benefits which I think are so obvious that I can suffice with a simple summary of what we have gained by looking at demand in this way.

The second example is a bit more complex. It shows how the use of semiparametrics can simplify the empirical analysis of models that require decision rules that are difficult to compute.[3] I use the example of entry models, variants of which have been used in empirical industrial organization for some time now. In this context I am going to try and convince you of two issues. First I argue that the early "structural" entry models that were taken to data, models which used blatantly *unrealistic* assumptions, turned out to be *quite useful*. Second I am going to argue that semiparametrics can be used to provide models which *both* (i) get around an important subset of those unrealistic assumptions and (ii) are simple to estimate.

The difficulty with the early entry models is that they are two period models; firms earn profits in the second period and choose whether to enter in the first. Semiparametrics allows us to base the analysis on a truly dynamic environment without having to compute equilibrium strategies for a dynamic game. More precisely incumbents' decisions on whether to exit and potential entrants' decisions on whether to enter depend on the expected discounted values of the profits from continuing and entering, respectively. It is these values which the equilibrium solves for. What semiparametrics does is obtain a first stage estimate of these values by averaging over the realized discounted cash flows for the incumbents in the data who did continue, and of the potential entrants who did enter. Given these first stage estimates of continuation and entry values, the parameters of the entry and exit distribution

---

[3]This use of semiparametrics dates back to Hotz and Miller (1992) for single agent problems, and Olley and Pakes (1996) for multiple agent problems.

can be estimated from any of a number of simple algorithms.

Again the economic model is transparent; agents enter (continue) if the expected discounted value of entering (continuing) is greater than the cost of entry (the sell-off value). It is the computation of the expected discounted values that is complex. What semiparametrics does is provide a simple way of approximating them that is grounded in what actually happened.

2   Simulation and Demand Systems

Not long ago graduate lectures on demand systems were largely based on representative agent models. Applications, on the other hand, typically used market level data: they would regress quantity purchased on (average) income and prices. There were theoretical papers which investigated the properties of market level demand systems obtained by explicitly aggregating up from micro models of consumer choices (including a seminal paper by Houthakker,1955). However we could not use their results to structure estimation on market level data without imposing unrealistic *a priori* assumptions on the distribution of income and "preferences" (or its determinants like size, age, location, etc.).

What simulation methodology has done is enabled us to aggregate up from the *observed* distribution of consumer characteristics and any functional form that we might think relevant. That is we allow different consumers to have different income, age, family size, or location of residence. We then formulate a demand system which is conditional on the consumer's characteristics and a vector of parameters which determines the relationship between those characteristics and preferences over products (or over product characteristics). To estimate those parameters from market level data we simply

- draw vectors of consumer characteristics from the distribution of those characteristics in the market of interest (in the U.S., say from the March CPS),

- determine the choice that each of the households drawn would make for a given value

5

of the parameter vector,

- aggregate those choices into a prediction for aggregate demand conditional on the parameter vector, and

- employ a search routine that finds the value of that parameter vector which makes these aggregate quantities as close as possible to the observed market level demands.

What do we gain by aggregating from a micro model?

We gain an increase in both (i) the depth of our understanding of how "aggregate preferences" are formed and (ii) the precision of our parameter estimates. Not only does this enable us to be more detailed and precise in our analysis of industrial organization issues, it also enables us to use them to analyze a host of distributional issues of distinct interest to related fields (examples include tax incidence and voting patterns).

For example we all believe (and virtually all empirical work indicates) that the impact of price depends on income. Our micro model will therefore imply that the price elasticity of a given good depends on the density of the income distribution among the income/demographic groups attracted to that good. So if the income distribution differed across regional markets, and we used an aggregate framework to analyze demand, we would require different price coefficients for each market.

Table I provides some data on the distribution of the income distribution across U.S. counties (there are about three thousand counties in the U.S.). It is clear that the income distribution differs markedly across these "markets". The standard deviation of the fraction of households in our nine income groups varies between twenty and a hundred percent of their means (with an average across these groups of fifty three percent of the mean). Counties with more households tend to have a larger fraction of their populations in the highest income group, and a smaller fraction in the lowest.

Table I: Cross County Differences in Household Income

| Income Group (thousands) | Fraction of U.S. Population in Income Group | Statistics for Counties' | | Correlation of Fraction with Mean County Income |
| | | Distribution of Fraction Over Counties | | |
		Mean	Std. Dev.	
0-20	0.226	0.289	0.104	-0.845
20-35	0.194	0.225	0.035	-0.637
35-50	0.164	0.174	0.028	0.147
50-75	0.193	0.175	0.045	0.737
75-100	0.101	0.072	0.033	0.901
100-125	0.052	0.030	0.020	0.912
125-150	0.025	0.013	0.011	0.868
150-200	0.022	0.010	0.010	0.862
200 +	0.024	0.012	0.010	0.800

It is pretty clear, then, that if we rely on the aggregate demand framework we are likely to require different price coefficients in different markets. If we based our estimates on an underlying micro model, on the other hand, we could get price effects that differ in a sensible way across markets without needing to resort to parameterizing each market separately. Moreover, unlike the aggregate framework, the micro model would allow us to make sensible predictions for price elasticities in locations where the good has not yet been marketed. Of course getting sensible price effects is a prerequisite for getting sensible *markups* (since markups are closely tied to the inverse of the price elasticity of demand). Moreover getting sensible predictions for markups and demand are the prerequisites for getting sensible incentives for *product development*, and so on.

Relatedly we often study differentiated product markets in which goods differ in their quality attributes and prices reflect these quality differences. The last column of Table I makes it clear that intermarket differences in the density of consumers in any given income range have only a modest empirical relationship to mean income. Consequently aggregate

frameworks are going to find it difficult to account for, say, differences in demand for mid-priced cars between two regions with similar mean incomes; one with mostly poor and a few very rich households, and one in which all household incomes are near the mean.

Of course one could stick with the aggregate model and add more detailed features of the income distribution as "right hand side variables". However were we to go this route we would quickly move to a model with too many parameters to estimate, particularly once we begin interacting the percentiles of the income distribution with demographic and location characteristics of the population. Indeed there is a very real sense in which the major advantage of moving to the micro model is that it enables us to use economics to provide a sensible and empirically useful way of constraining the impacts of the joint distribution of consumer attributes on demand.

Use of a framework constructed from an underlying micro model of demand and explicit aggregation also implicitly solves a number of other outstanding problems in demand estimation, a few of which are now discussed.

2.1 Integrating Alternative Data Sets

Since demand systems based on a micro model use data on the distribution of consumer characteristics, they use more of the information available then do demand systems based on an aggregate framework. Perhaps more important, however, is the fact that the micro models allow us to integrate information from alternative data sets in the same framework. That is use of a micro framework allows us to analyze micro choice data, data which matches the consuming unit to the good it chooses, within exactly the same framework used to analyze the market level data (and a similar comment applies to analyzing data at any intermediate level of aggregation, for e.g. data on differences in demand patterns among different income and/or demographic groups).

There are at least two important implications of this fact. First, it allows us to do away with the embarrassment of having mutually inconsistent sets of parameters for the same

problem. Second by using all the available data in one coherent framework we can obtain much sharper estimates of the objects of interest. In this context it is important to remember that prices and products (or product characteristics) typically vary over time but not in the cross section, while the characteristics of households (income, demographics, and location of residence) vary greatly over households but have a joint distribution which is pretty stable over time. Thus our ability to use the information at different levels of aggregation in one internally consistent estimation framework can be the key to getting precise estimates of demand patterns.

2.2 Gains to Variety and Dynamics

Use of a micro based model allows us to get a start on the difficult problem of analyzing the gains to variety. The aggregate frameworks used in macroeconomics and trade either to evaluate the gains from variety or to work out the implications of those gains, are totally driven by functional form assumptions that have not been verified empirically. The empirical problem of measuring the gains to variety is a problem of evaluating the welfare gains from new products. Fundamentally, though standard tools may find an empirical approximation to the welfare gained by a consumer who purchased the new good at one observed price but not at another, the tools cannot determine the inframarginal benefits accumulated by those consumers who purchased the good at all observed prices without strong functional form assumptions.

On the other hand if differences in variety can be broken down into the differences in the characteristics of the goods being marketed, and households with different attributes differ in their relative preferences over the product characteristics, then a micro model can provide a rather direct measure of variety gains. The micro model would use the difference in household attributes in the population and differences in the quality of products over time to estimate of the importance of the interactions between household attributes and product characteristics, and then use those estimates to build a measure of welfare gains. A simple

example of evaluating the impact of "horizontal" variety differences is the analysis of the cost of travel time that we can do by analyzing the impacts of the interaction of the locational distribution of stores (or airports) with consumer residences (or work places) in determining purchasing decisions. Any sensible analysis of the impacts of "zoning" laws would require such a framework. An example which is a bit more complex, but central to issues in growth, trade, and health care is the analysis of the gains from new improved "qualities" of products. This would require us to be able to measure quality dimensions (as, arguably is the case for computers, or hospitals), and the consumer characteristics that determine their usefulness (like the need for computer speed and/or storage, or the degree of sickness).

Finaly the nature of the tradeoff between the increase in demand that would accrue to a "horizontally" differentiated variety of a good, versus the demand increase that would be generated by an increase in the quality dimensions of an existing product, is often a fundamental tradeoff that determines the nature of the dynamic model relevant for the analysis of different industries (see, for example, Sutton, 1991). As a result were we to look for a "test" of situations in which one or the other dynamic model might be relevant that did not depend on some *a priori* classification, we would have to estimate a demand model that allowed for differences in tastes, and then see if differences in the primitives estimated are consistent with the theory's predictions for differences in the features of markets.

2.3   Distributional Issues and Political Economy

Lastly, use of a micro model of demand provides us with a basis for analyzing the distributional impacts of different policies. This feeds directly into the analysis of regulation and of political economy more generally. That is, to analyze most of the regulatory and political economy issues of interest, we need the entire distribution of benefits and costs from different policies (and not just market level summary statistics).

3 "Structural" Entry Models

Entry and exit are a key part of the dynamics of market adjustments. In particular we usually think that it is the size and nature of entry costs that allow one (or a group of) firms to maintain a dominant position in a profitable market.

Unfortunately most cost data are proprietary (and hence difficult for researchers to access). This is the reason we often have to resort to inferring marginal costs from their implications on an equilibrium pricing equation (a procedure whose validity requires detailed assumptions on both demand and on the nature of equilibrium). If we consider data on marginal costs to be rare, then we should think of data on many aspects of sunk costs as virtually nonexistent. The decision of a firm on whether to exit is determined by whether its continuation value is greater than its sell-off value, and the latter is often associated with factors as hard to measure as "goodwill" or the value of the firm's building and equipment in its "second best" alternative employment. The potential entrants' sunk costs can be largely determined by the time and effort required to formulate the idea to be marketed, or by an individual entrepreneur's cost in accessing startup capital and/or the requisite permissions from a local administration. As a result we have to infer the extent of sunk costs from other variables whose behavior depends on them.

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

3.1 Early Entry Models

Largely as a result of these computational problems the early empirical analysis of entry and exit used, as a modelling framework, two-period entry games. The two-period game framework is problematic in many respects. Most importantly it makes little sense unless sunk costs are absent (or at least are very small relative to single period profits). This is an assumption which is hard to swallow, particularly since much of our interest in entry models stems from a desire to investigate the origins and implications of sunk costs.

Despite this I would argue that the early papers were extremely useful for at least two reasons

- the structures they used provided a convenient framework for organizing empirical facts,

- the use of the structures forced the researchers to face up to questions which reappear in more realistic (truly dynamic) entry models, and the two period models provided an environment in which to explore them.

Indeed this is a subfield of industrial organization in which later authors built rather directly from earlier results.

The two-period structure did make sense as an "organizer" of empirical facts. This is easiest to see in environments where the the only aspect of profitability that changes over time are idiosyncratic realizations of the sell-off and entry values of the different incumbents and potential entrants. Then the variables that determine the second period profit function in a two period game are the same as the "publicly observed" variables that determine the value of entry in a truly dynamic game. Of course sensibility demands that in most empirical examples we modify the two-period model's "profit" function. In reality what the two period models analyzed were differences in the number of active firms across markets and these differences depend on lagged, as well as on the current, values of the determinants of profitability. Moreover when we try to interpret the parameter estimates we have to keep

in mind that firms are trading off the current sunk cost of entry against a continuation value which incorporates the expected discounted value of a stream of future, and not just current, profits. These caveats makes it harder both to be precise about the implications of the parameter estimates and to do counterfactuals, but the two-period models use as an organizer of empirical facts is, I think, beyond doubt.

Bresnahan and Reiss (1987, 1990) focused on a model with identical firms. That is, all firms were assumed to have both the same sunk cost of entry *and* the same continuation values, but they noted the extensions that were required to allow for heterogeneity in sunk entry costs. Berry (1992) showed how simulation could be used to compute the implications of the model that allowed for differences in sunk costs, and based his empirical results on that model. The two-period identical firm model of Bresnahan and Reiss was primarily used as a framework to organize facts on how differences in the characteristics of regional markets (largely population) impacted the number of active firms. When differences in sunk costs were added they provided an avenue which could be used to explore how differences among firms might affect their entry possibilities.

However, there were conceptual problems that had to be solved before one could actually do the entry analysis. In particular once we allow for interfirm differences there is more than one possible equilibrium outcome that is consistent with the model and a given value for its parameter vector. This made it impossible to apply standard estimation techniques for models which predict discrete outcomes (like entry or exit). A standard model implies a unique outcome conditional on values for the parameter vector, the observed determinants of the decision (the x's), and the unobserved determinants (the ϵ's). To find the probability of the observed action conditional on the parameter vector, the econometrician adds up the probabilities for the ϵ that would lead to the action actually chosen. These probabilities are used to form the likelihood of that parameter value (or some other method of moments criteria function). Estimation consists of repeating this procedure whenever needed while searching for the value of the parameter vector that maximizes the objective function.

With nonuniqueness, the researcher cannot compute *the* equilibrium outcome implied by the parameters and then search for the value of the parameter vector that makes the data as close as possible to the implications of the theory. The insight used in the early entry literature to circumvent this problem was that, provided only sunk costs (and not continuation values) differ among firms, then the number of firms are uniquely determined even though their identities are not. They then proceed by developing an estimator which finds the value of the parameter vector which makes the theoretical prediction for the number of firms as close as possible to the number in the data.[4]

Subsequent work by Siem (2002) and Mazzeo (2002) extends the analysis further by developing estimation techniques for two period models that allow for differences in continuation values (more precisely differences associated with different entry "locations"). From an empirical point of view what this did is to allow us to investigate how continuation values of similar (but not identical) entry locations are affected by the number of firms operating in the neighboring locations. This is particularly important to the study of retail trade where location choice, defined broadly enough to include both physical location and store size, is central to the issues we want to analyze. From a modelling point of view the papers suggested two other ways to ameliorate (if not overcome) the uniqueness issue, both of which use more of the underlying detail in the data: add noise to the system so that one agent does not know the precise action of its potential competitors, and/or sequentially order decisions.

I review this material not just to provide a setting for what I am about to present, but also to make a point. We have made progress both in characterizing the determinants of entry, and in providing appropriate frameworks for subsequent analysis, by working with a patently unrealistic "structural" model. On the other hand we have not yet gotten to a framework which is rich enough to enable us to provide a realistic analysis of entry costs,

---

[4]A similar problem (i.e. nonuniqueness) arises in the analysis of pricing equilibria and a similar solution is typically used there. In the pricing game all equilibria must satisfy the first order conditions for a Nash equilibrium in prices, so we estimate off of those conditions. Note also that though this entry literature derives consistent and asymptotically normal parameter estimates, those estimates need not be efficient. An alternative estimator which makes use of more of the available information is provided in Tamer (2003).

which, recall, was a major reason for our interest in the study of entry in the first place. To do that we need to explicitly model the tradeoff between entry costs and the perceived value of entry (not the one period profit function).

My goal here is to go one step further and show how semiparametrics can provide an easy way of making the transition from two-period to truly dynamic models, and hence to more realistic inferences about entry costs and exit fees. Most of this part of the talk is taken from Pakes, Ostrovsky and Berry (2003), a paper I will refer to simply as POB, and I thank my coauthors for permission to use this material.[5]

3.2   A Semiparametric Entry Model

I will focus on the simple case where there is only one entry location and the same number of potential entrants in each period (POB shows how to extend the model and deal with a finite number of entry locations and a random number of potential entrants in every period). The issues here are a bit more detailed, so I will need some notation.

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 discrete state Markov process, and assume that there is a one-period profit function that is determined by these variables, say $\pi(n, z; \theta)$, where θ is a parameter vector to be estimated. An incumbent chooses to exit if current profits plus the discounted selloff value is greater than profits plus the discounted continuation value. So if ϕ is the sell-off (or exit) value and $0 < \delta < 1$ is the discount rate, the "Bellman" equation for the value of an incumbent is

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

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

---

[5]Similar ideas can be used to simplify empirical analysis that requires equilibrium decision rules in other complex contexts (for an example of an application to electric utility auctions see Pakes, Porter, and Wolfram, 2003)

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 (of future $V(\cdot)$), or

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

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

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

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

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

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

where

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

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

The potential entrant enters if

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

where κ is its sunk cost of entry.

3.2.1 Assumptions and Their Implications

In addition to technical "regularity" conditions, it is assumed that entry and exit decisions are made simultaneously at the beginning of the period. The following assumptions are also made.

1. There are a fixed number of potential entrants in each period (denoted by $\mathcal{E}$), and the distribution over

   - the sunk costs of entry, say $F^{\kappa}(r|\beta)$, which has a lower bound of $\underline{\kappa} > 0$, and

   - the returns to exiting, say $F^{\phi}(\cdot|\beta)$, which are assumed nonnegative,

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

2. Entrant's and incumbent's 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).

These assumptions are *truly restrictive*. They imply, for example, that there are no serially correlated state variables that are observed to the agents and not to the econometrician. However they (or simple generalizations of them that allow for multiple locations) are less restrictive than the assumptions used in any of the two-period models that have been taken to data to date. Moreover, as we will show below, these assumptions lead to an estimator

of continuation and entry values that has a transparent relationship to objects in the data, and therefore has both a great deal of intuitive appeal and is easy to work with.

This model is a special case of the model in Ericson and Pakes (1995) and so has an equilibrium, but there may be more than one of them. Each equilibrium generates a finite state Markov chain in (n, z) couples: i.e. the distribution of possible (n, z)'s in the next period depends only on the current (n, z) (and not on either prior history, or time itself). Indeed one can go a bit further and note 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). Thus, for example, there will be an $\overline{n}$ such that, provided the current n is lower than it, we will never observe an $n > \overline{n}$. The market is simply not profitable enough to induce entry if there are $\overline{n}$ incumbents, so states with $n > \overline{n}$ are not in $\mathcal{R}$. However all states in $\mathcal{R}$ "communicate" with each other, and will eventually be visited many times.

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

To see that the data can be used to pick out the equilibrium, note that (i) the agents only condition their perceptions of the behavior of their competitors on the publicly available information (on (n, z)) and (ii) precisely the same information is available to the econometrician. Moreover in equilibrium the realized distribution of entrants and exitors from each state must be consistent with these perceived distributions.

Now recall that the data will eventually wander into the recurrent subset of points, and once in that subset will visit each point in it repeatedly. As the sample gets large we obtain an empirical distribution of entrants and exitors from each (n, z), and by the law of large numbers that distribution will converge to the distribution which generated it (almost

surely). As noted this must be the distribution the agents use to form their perceptions, so we have just identified the perceived distributions needed for agents to make their decisions.

Given those perceived distributions equations (1) and (2) generate a unique best response for each incumbent and potential entrant. This is just the familiar statement that reaction functions are generically unique, and can be proven using Blackwell's theorem for single agent dynamic programs. Since there is only one policy that is consistent with both the data and our equilibrium assumptions at each $(n, z) \in \mathcal{R}$, and once we are in the set $\mathcal{R}$ we stay there forever, there is a unique equilibrium for any subgame starting from any (n, z) couple in $\mathcal{R}$ (a set which can be identified from the data).[6] We now provide semiparametric estimators for the continuation and entry values generated by that equilibrium.

3.2.2 Equilibrium Perceptions and Continuation Values

In equilibrium the perceptions of potential entrants and incumbents of the likelihood of entry and exit by their competitors must be consistent with the distribution of entry and exit actually generated by incumbent and potential entrant behavior. This observation leads directly to two semiparametric estimates of continuation values at each (n, z) observed at least once.

The first estimate simply averages the realized continuation values of all firms who did continue when $(n_t, z_t) = (n, z)$. Since agents perceptions of probabilities of reaching different states must be consistent with the actual probabilities of reaching the different states, the sample average of discounted future values will converge to expected continuation value we are after. This estimator is particularly simple, so I will provide detail on it below.

The second estimator, though also intuitive, is computationally more complex, so the reader who wants more detail should see POB and the literature referred to there. Briefly, if

---

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

we knew $p(e, x|n, z, \chi = 1)$ we could compute the fixed point which defines the continuation value in equations (1) and (2) for any value of θ. What the second procedure does is substitute the empirical distributions of entrants and exitors from each state for $p(e, x|n, z, \chi = 1)$ and compute the fixed point implied by these empirical distributions. Equations (1) and (2) define a contraction mapping, so solving the associated fixed point problem is not too difficult, but it must be solved every time we need to evaluate θ in the estimation algorithm. In the terminology of Rust (1994) this is a "nested fixed point" estimation algorithm.[7]

The next subsection provides a brief description of how to obtain the estimates based on sample averages. The reader who is not interested in these details should be able to skip directly to the remark at the end of this subsection, and continue from there.

3.2.3 Using Sample Averages to Form Continuation Values

To form this estimate we need estimates of (i) the distribution of future states given the current state and (ii) the realized value at those future states.

We begin with the realized values at the future states. An incumbent at t who continues to $t + 1$ receives that period's profits plus a choice between the continuation value at $t + 1$ and a draw from the distribution of sell-off values. The incumbent chooses the sell-off value if and only if it is greater than the continuation value.

Let functions indexed by t refer to the function evaluated at (n_t, z_t), and let the probability of exit be given by

$$g_{t+1}^x \equiv Pr\{\phi > VC_{t+1}(\theta)\}. \tag{4}$$

Then given how the incumbent chooses, the continuation value of firms who did continue

---

[7]Aguirreberia and Mira (2003), in independent work, provide a closely related estimator. Instead of using the empirical transition probabilities for (n, z) couples, they use the empirical exit (and for entrants, entry) probabilities to generate these transitions. That is, since all incumbents are identical, a consistent estimate of the probability that x of n incumbents exit is given by the binomial formula with an exit probability estimated by the fraction of incumbents that exit. A similar procedure can be used to generate entry probabilities. Aguirreberia and Mira (2003) construct a consistent estimate of $p(e, x|n, z, \chi = 1)$ from these binomial probabilities. They assume that the random entry and exit fees are i.i.d. draws from an extreme value distribution. This is not necessary but does ease the computational burden of their nested fixed point estimator. POB compare these and other alternative estimators.

from period t conditional on the (publicly available) information available in period $t + 1$, say $\widetilde{VCR}_t(\theta)$, is given by

$$\widetilde{VCR}_t(\theta) = \pi_{t+1}(\theta) + \delta(1 - g^x_{t+1})VC_{t+1}(\theta) + \delta g^x_{t+1}E[\phi|\phi > VC_{t+1}(\theta)] \tag{5}$$

where $E[\phi|\phi > VC_{t+1}(\theta)]$ is the expected sell-off value conditional on exiting.

The easiest case is when sell-off values are exponentially distributed with parameter σ, because a truncated exponential is just another exponential with its origin shifted to the truncation point.[8] In this case, then

$$E[\phi|\phi > VC_{t+1}(\theta)] = \sigma + VC_{t+1}(\theta). \tag{6}$$

Now make two substitutions in the equation for $\widetilde{VCR}_t(\theta)$: (i) substitute this expression for $E[\phi|\phi > V_{t+1}(\theta)]$ and (ii) substitute an estimate of the exit probability obtained by averaging over the fraction who exit at the different periods in the data when $(n, z) = (n_t, z_t)$, say $\tilde{g}^x_t$, for the unobserved g^x_t. Rearranging terms we obtain

$$\widetilde{VCR}_t(\theta) = \pi_{t+1}(\theta) + \delta VC_{t+1}(\theta) + \delta\sigma\tilde{g}^x_{t+1}. \tag{7}$$

Note that the expectation of $\widetilde{VCR}_t(\theta)$ is equal to the expected continuation value of all those firms who continue. That is not quite $VC_t(\theta)$, since in forming the expectation which determines whether it should continue the incumbent conditions on itself continuing (while in the realizations measured by $\widetilde{VCR}_t(\theta)$ only a fraction continue). As shown in POB the ratio of the probabilities observed in the data, to the probabilities that the incumbent uses is consistently estimated by $w_t = [1 - \tilde{g}^x_t]/[1 - (x_t/n_t)]$, where $\tilde{g}^x_t$ is, as before, the fraction of incumbents who exit when $(n, z) = (n_t, z_t)$.

---

[8]Virtually any other distributional assumption could be used here, but most others (with the exception of the uniform) would result in a somewhat more computational complex estimation algorithm. On the other hand the shape restrictions of the exponential seems reasonable for sell-off values.

So if we let

$$VCR_{t+1} = w_t \widetilde{VCR}_t(\theta)$$

then

$$E[VCR_t(\theta)] = VC_t(\theta)$$

the unknown continuation value as defined in equation (1).

The sample average of the observations on $VCR_t(\theta)$ computed as an average over those periods where (n_t, z_t) was a particular value of (n, z), say (n^*, z^*), converges to the population expectation of the continuation value for (n^*, z^*), that is to the $VC(n^*, z^*; \theta)$ we are after. More precisely if we let $T(n, z)$ be the set of periods when there was n firms active and the state variable had value z, and $\#T(n, z)$ be the number of such periods, and define

$$\widetilde{VC}(n, z) \equiv \frac{1}{\#T(n, z)} \sum_{t \in T(n,z)} w_t \widetilde{VCR}_t(\theta), \tag{8}$$

then if $\rightarrow_P$ is read as converges in probability, we have

$$\widetilde{VC}(n, z) \rightarrow_P VC(n, z)$$

provided $\#T(n, z) \rightarrow \infty$.

Finally we need to connect the estimates of the continuation values from the different (n, z) to each other. Here I will need matrix notation. Arrange the $\widetilde{VC}(n, z; \theta)$ of the (n, z) combinations observed in the data into the vector $\widetilde{VC}(\theta)$, let the matrix of empirical transition probabilities from one observed (n, z) combination to another (weighted as above) be $\tilde{M}$, and gather the average fraction of firms exiting from each state into the vector $\tilde{g}^x$. Then, if we substitute our expression for $\widetilde{VCR}(\theta)$ from equation (7) into the expression for $\widetilde{VC}(n, z)$ in equation (8) and rearrange we have

$$\widetilde{VC}(\theta) \equiv \tilde{M} \left[\pi(\theta) + \delta\sigma\tilde{g}^x\right] + \delta\tilde{M}\, VC(\theta).$$

22

This gives us our estimates of continuation values in terms of objects that we know (at least up to the parameters to be estimated), and the continuation values itself. We now iteratively substitute our estimate of continuation values for the continuation values that appear in this expression. Doing this once we obtain

$$\widetilde{VC}(\theta) \equiv \tilde{M}\left[\pi(\theta) + \delta\sigma\tilde{g}^x\right] + \delta\tilde{M}\left[\pi(\theta) + \delta\sigma\tilde{g}^x\right] + \delta^2\tilde{M}^2 VC(\theta),$$

and if we continue this process of substitution we get

$$
\begin{aligned}
\widetilde{VC}(\theta) &= \sum_{\tau=0}^{\infty} \delta^{\tau+1} M^{\tau}\left[\pi(\theta) + \delta\sigma\tilde{g}^x\right] \\
&= [I - \delta\tilde{M}]^{-1}\tilde{M}[\pi(\theta) + \delta\sigma\tilde{g}^x],
\end{aligned}
$$

as a consistent estimator for the unknown continuation values.

Note that $\widetilde{VC}(\theta)$ is just the discounted value of the returns of the incumbents who did continue (adjusted to account for the fact that the incumbent conditions on itself continuing). This is the sense in which our estimator of continuation values is transparent and likely to have empirical content. We expect the actual average of realized continuation values to be close to the true expected continuation value.

3.2.4 Remark on Ease of Computation

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

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

for a data matrix $\tilde{A} = [I - \delta\tilde{M}]^{-1}\tilde{M}$ and a data vector $\tilde{a} = \delta[I - \delta\tilde{M}]^{-1}\tilde{g}^x$. So if profits were linear functions of θ, the first stage estimates of continuation values are also.

POB shows (and it is easy to verify) that consistent estimates of entry values, say $\widetilde{VE}(\theta)$, can be obtained as

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

where

$$\tilde{B} \equiv \tilde{M}^e + \delta\tilde{M}^e\tilde{A}, \quad \tilde{b} \equiv \delta\tilde{M}^e\tilde{a} + \delta\tilde{M}^e\tilde{g}^x,$$

and $\tilde{M}^e$ is the Markov transition matrix formed after weighting the observed transitions with $w_t^e = (e_t/\mathcal{E})/\tilde{g}_t^e$, where $\tilde{g}_t^e$ is an estimate of the entry probability obtained by averaging over the fraction who enter at the different periods in the data when $(n, z) = (n_t, z_t)$.

3.2.5 Estimation Methods

We estimate models by finding the value of the model's parameter vector that make the predictions of the model as close as possible to the observed data.

In our case the model predicts that, conditional on (n, z), the number of entrants and exitors are independent draws from two binomial distributions. We have shown that a consistent estimate of the model's binomial probability of exit in a given period is

$$1 - F^\phi(\tilde{A}_t\pi(\theta) + \sigma\tilde{a}_t|\theta)$$

and of the binomial entry probability is

$$F^\kappa(\tilde{B}_t\pi(\theta) + \sigma\tilde{b}_t|\theta)$$

where the t subscript on (A, a, B, b) picks out the row of each at which $(n, z) = (n_t, z_t)$. Relatedly, consistent estimates of the model's predictions for the expected number of exitors from a state is just the number of incumbents at that state times the exit probability at that state (as defined above). Similarly the model's prediction for the expected number of entrants at that state is just the number of potential entrants times the entry probability at

24

that state.

Different estimators are obtained by using different measures of the distance between the theoretical predictions for the probabilities conditional on θ, and the empirical observations. For example the "pseudo" maximum likelihood estimator treats our estimated continuation and entry values as the true continuation and entry values (when in fact they are just estimates of the truth), forms the binomial likelihood of the observed number of entrants and exits conditional on these estimates, i.e.

$$
\begin{aligned}
& F^{\phi}(\tilde{A}_t \pi(\theta) + \sigma \tilde{a}_t | \ \theta)^{n_t - x_t} \left[1 - F^{\phi}(\tilde{A}_t \pi(\theta) + \sigma \tilde{a}_t | \theta)\right]^{x_t} \\
\times \quad & F^{\kappa}(\tilde{B}_t \pi(\theta) + \sigma \tilde{b}_t | \theta)^{e_t} \left[1 - F^{\kappa}(\tilde{B}_t \pi(\theta) + \sigma \tilde{b}_t | \theta)\right]^{\mathcal{E} - e_t},
\end{aligned}
$$

and then maximizes the sum of the logs of these likelihoods over periods.

Alternative method of moment estimators can be formed by minimizing a weighted average of either (i) the squared differences between the model's estimates of the average entry and exit rates and the entry and exit rates observed in the data (where the average is taken over observed periods) or (ii) the squared differences between the model's estimates of the average and entry and exit rates at the alternative observed (n, z) states (here we average over the observations at the same state, square the differences between the models predictions for the average at that state and the sample average at that state, and then minimize a weighted average of these differences over states).

All of these estimators will be consistent and have limiting normal distributions. They will, however, differ in both their variance, and in their computational complexity. POB contains a discussion of these issues. For now all I want to note is that there are good reasons to think that pseudo maximum likelihood will not do well in this context, so one should probably focus on the alternative method of moments estimators.

What I do want to stress here is the *simplicity* of all of these estimators. I.e. to obtain any of the estimators discussed above we need only minimize analytic function of the pa-

rameters. This makes the computational burden of these estimators as easy (or easier) than the estimators for any of the earlier entry models discussed in Section 3.1. Moreover, the procedure has the following additional benefits.

- It is easily generalized to models which allow entry into multiple locations (the number of entrants to the different locations are then draws from a multinominal rather than that of a binomial distribution), and to models with a random number of entrants (the number of entrants into the different locations are then a draw from a mixture of multinominals, so the estimating equation is still an analytic function of the parameters of the problem).

- Though it is true that all of the estimators are "two-step" estimators, and that their variance must take account of the fact that the "regressors" used in the second step were estimated, there is a simple "parametric boostrap" procedure which provides consistent estimators for the variances of the parameter estimates.[9]

3.2.6 Monte Carlo Evidence on Computational Burden

Table II provides Monte Carlo results that are designed to illustrate how this simplicity translates into ameliorating the computational burden from analyzing entry and exit problems. It considers two models: a one-location model roughly patterned after Bresnahan and Reiss' (1987) study, and a two-location model, patterned after Mazzeo's (2002) study (for

---

[9]The parametric bootstrap is obtained as follows. Let the estimate of the parameter vector be $\hat{\theta}$, and compute $\hat{VC}(\hat{\theta})$ and $\hat{VE}(\hat{\theta})$, the estimates of the continuation and entry values obtained by substituting the $\hat{\theta}$ into the equations for those values. Starting from the observed $(n_{t=0}, z_{t=0})$, randomly draw sell-off values from $F^\phi(\hat{\theta})$ and entry values from $F^\kappa(\hat{\theta})$ for each of the incumbents and potential entrants. Assume each incumbent exits if its sell-off value is greater than the component $\hat{VC}(\hat{\theta})$ that is associated with $(n_{t=0}, z_{t=0})$. Similarly the potential entrants enter if their randomly drawn entry cost was less than $\hat{VE}(\hat{\theta})$. Keep this predicted number of entrants and exitors in memory, and construct $n_{t=1}$ as the initial number of incumbents minus the exitors plus the entrants. Next randomly draw a $z_{t=1}$ from (a consistent estimate of) the Markov process generating z. This gives us $(n_{t=1}, z_{t=1})$, and we can determine exit and entry from this point in the same way as we did above. Continue in this way until we have a simulated sample with exactly the same number of periods (or of markets and periods) as in the original sample. Once we have this "pseudo random" sample, substitute it into the estimation algorithm and generate a new estimate of θ. Do this for a number of simulated samples, and calculate the variance-covariance of the estimates of θ obtained from the alternative pseudo random samples. This is a consistent estimate of the variance covariance matrix of $\hat{\theta}$.

more details see POB, 2003). In each case we used the algorithm introduced in Pakes and McGuire (1994) to compute an equilibrium for a fixed value of the parameter vector, and then used the equilibrium policy functions produced by that algorithm to generate a panel of "data". This data was then used to estimate the original parameters. In all cases we used a simple method of moments estimation algorithm: one that minimized the squared differences between the average (over periods) entry and exit probabilities predicted by the model and those in our data. The models were all computed on the desktop PC in my office.

Table 2: Monte Carlo Results

Number of Cross Sections	Average Bias (Percent of Parameter)	Average Standard Error (Percent of Parameter)	Computation Time (Seconds)
Single Location Model (Two Parameters)			
250	11%	7%	26
500	6.5%	2%	42
1000	4%	1%	60
Two Location Model with Random Number of Entrants (Four Parameters)			
250	34%	95%	80
500	8.5%	17%	109
1000	6%	11%	131

Notes: Results reported using a five-year time dimension, i.e., $T = 5$.

Since the early papers studied isolated markets, we chose South Dakota and counted that there were approximately 250 relatively "isolated" towns in it. This became the smallest number of markets in any of our runs. Here we report results where the time dimension was five years (we have also tried fifteen, and the results are correspondingly better); i.e. we are assuming we have a five year panel data set. Our modification of the Bresnahan Reiss model

27

has entry costs drawn from a unimodal one parameter family, and sell-off values generated by a one parameter exponential. It is the parameters of these two distributions that we try to estimate (i.e. we assume the profit function is known).

The two location model is somewhat more complex. Here a random number of potential entrants appear in every period and each one of them receives a draw on a couple which determines the sunk cost of building a low and a high quality "motel", respectively. The draws are correlated (indeed, for any given individual the cost of building a high quality motel is always higher than the cost of building a low quality motel), and described by a two parameter family. The sell-off values from the two different locations are exponential with two different parameters. So there are four parameters we estimate for this model.

Each model has about five hundred "populated" states. In the one location model our procedures produce nonparametric estimates of entry and continuation values for each of them. The two location model produces separate entry and continuation values for each of the two locations for each state. The computation time column provides the sum of the time required to arrange the data into the form we use (largely the time required to compute the Markov transition matrices and their inverses) and the search time. It seems pretty clear that the estimators are not particularly computationally demanding. Estimation time for the one location model was only about a half a minute for the smallest sample and about one minute for the largest sample. This increases to just over two minutes for the large sample in the two location model. Of course in actual applications we might want to allow for more observable determinants of profitability or sunk costs, and this will increase the computational burden. Still it seems we have a long way we can go before computational burden becomes a problem.

Also, as one might expect, the larger the number of parameters we are estimating and the smaller is the sample, the larger the bias and the variances of the estimators. Still recall where we started: typically we have almost no knowledge of sunk costs. If we had obtained these estimates from real data on an industry of interest, even the most poorly behaved

28

estimators in the table would be providing a great deal of useful information.

4   A Final Note

I admit that the estimation frameworks outlined in these two examples are somewhat "complex". Moreover their complexity stems directly from the fact that their estimating equations are derived from a theoretical structure. To use the theory to guide estimation we had to transform either the model or the data. In the demand example we had to derive the implications of the model for aggregates over individuals. In the entry example we had to transform the data to approximate discounted values over time. However, in both cases the estimating equations were based on constructs which should have a transparent meaning to economists.

As a result I would argue that these relatively complex estimators are not only eminently "sensible", they are also about as direct a method of obtaining empirical information on either demand or on entry costs as is available for most industries. Moreover it is because the estimation equations are developed from an economic model that (i) it is often easy to consider robustness to model misspecification and (ii) there is an unambiguous and straightforward way to use the results for the investigation of past causal effects and/or subsequent policy analysis. It is, of course, true that the quality of this analysis is likely to depend on how good our assumptions were in approximating reality. What we should keep in mind, however, is that *all* causal analysis requires assumptions. Indeed all we have done is been "up front" about the assumptions that we need. This, in turn, provides a scientific basis for a debate on their relevance.

Of course this transparency is not as obvious in the early work that this paper is based upon. Indeed I think the lesson we have learned in empirical industrial organization is that complex and/or less realistic models are often stepping stones to developing empirical frameworks capable of a simpler, more direct analysis of the issues we are interested in. So

my hope is that you stay "tuned in" to the econometric developments in our field. We should be giving them a chance to complement the advances that we have seen over the last few decades in theory, computation, and data availability.

References

Aguirreberia, V., and P. Mira (2002) 'Identification and Estimation of Dynamic Discrete Games', Boson University mimeo.

Berry, S. (1992) 'Estimation of a Model of Entry in the Airline Industry', *Econometrica*, **60**, 889–917.

Brenahan, T., and P. Reiss (1987) 'Do Entry Conditions Vary Across Markets?', *Brookings Papers on Economic Activity, Microeconomics Annual*, **1**, 833–882.

Bresnahan, T. and P. Reiss (1990) 'Entry in Monopoly Markets', *Review of Economic Studies* **57**, 531–553.

Ericson, R., and A. Pakes (1995) 'Markov-Perfect Industry Dynamics: A Framework for Empirical Work', *Review of Economic Studies*, **62**, 53–82.

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

Hotz, J., and R. Miller (1993) 'Conditional Choice Probabilities and the Estimation of Dynamic Models', *The Review of Economic Studies*, **60**, 265–289.

Houthakker, H. S. (1955) 'The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis', *Review of Economic Studies*, **23**, 27–31.

Mazzeo, M. (2002) 'Product Choice and Oligopoly Market Structure', *RAND Journal of Economics*, **33**, 221–242.

Olley, S., and A. Pakes (1996) 'The Dynamics of Productivity in the Telecommunications Equipment Industry', *Econometrica*, **64**, 1263–1297.

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

Pakes, A., and P. McGuire (1994) 'Computing Markov-Perfect Nash Equilibria: Numerical Implications of a Dynamic Differentiated Product Model', *RAND Journal of Economics*, **25**, 555–589.

Pakes, A., and P. McGuire (2001) 'Stochastic Algorithms, Symmetric Markov Perfect Equilibria and The 'Curse' of Dimensionality', *Econometrica*, **69**, 1261–1281.

Pakes, A., M. Ostrovsky, and S. Berry (2003) 'Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry/Exit Examples)', Harvard University mimeo.

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

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

Seim, K. (2002) 'Geographic Differentiation and Firms' Entry Decisions: The Video Retail Industry', Stanford University Graduate School of Business mimeo.

Sutton, J. (1991) *Sunk Costs and Market Structure*. M.I.T. Press.

Tamer, E. (2003) 'Incomplete Simultaneous Discrete Response Model with Multiple Equilibria', *Review of Economic Studies* **70**, 147–165.