Econometric Tools for Analyzing Market Outcomes.

Daniel Ackerberg, C. Lanier Benkard,
Steven Berry, and Ariel Pakes.*

June 14, 2006

Abstract

This paper outlines recently developed techniques for estimating the primitives needed to empirically analyze equilibrium interactions and their implications in oligopolistic markets. It is divided into an introduction and three sections; a section on estimating demand functions, a section on estimating production functions, and a section on estimating "dynamic" parameters (parameters estimated through their implications on the choice of controls which determine the distribution of future profits).

The introduction provides an overview of how these primitives are used in typical I.O. applications, and explains how the individual sections are structured. The topics of the three sections have all been addressed in prior literature. Consequently each section begins with a review of the problems I.O. researchers encountered in using the prior approaches. The sections then continue with a fairly detailed explanation of the recent techniques and their relationship to the problems with the prior approaches. Hopefully the detail is rich enough to enable the reader to actually program up a version of the techniques and use them to analyze data. We conclude each section with a brief discussion of some of the problems with the more recent techniques. Here the emphasis is on when those problems are likely to be particularly important, and on recent research designed to overcome them when they are.

Keywords; Demand Estimation, Production Function Estimation, Dynamic Estimation, Strategic Interactions, Equilibrium Outcomes.

JEL Classification; C1, C3, C5, C7, L1, L4, L5.

*The authors are from the UCLA, Stanford University, Yale University, and Harvard University, respectively, and thank their NSF grants for financial support. This paper was prepared for the *Handbook of Econometrics*, J.J. Heckman (ed.)

Recent complementary developments in computing power, data availability, and econometric technique have lead to rather dramatic changes in the way we do empirical analysis of market interactions. This paper reviews a subset of the econometric techniques that have been developed. The first section considers developments in the estimation of demand systems, the second considers developments in the estimation of production functions, and the third is on dynamic estimation, in particular on estimating the costs of investment decisions (where investments are broadly interpreted as any decision which affects future, as well as perhaps current, profits).

These are three of the primitives that are typically needed to analyze market interactions in imperfectly competitive industries. To actually do the analysis, that is to actually unravel the causes of historical events or predict the impact of possible policy changes, we need more information than is contained in these three primitives. We would also need to know the appropriate notion of equilibrium for the market being analyzed, and provide a method of selecting among equilibria if more than one of them were consistent with our primitives and the equilibrium assumptions. Though we will sometimes use familiar notions of equilibrium to develop our estimators, this paper does not explicitly consider either the testing of alternative equilibrium assumptions or the issue of how one selects among multiple equilibria. These are challenging tasks which the profession is just now turning to.

For each of the three primitives we do analyze, we begin with a brief review of the dominant analytic frameworks circa 1990 and an explanation of why those frameworks did not suffice for the needs of modern Industrial Organization. We then move on to recent developments. Our goal here is to explain how to use the recently developed techniques and to help the reader identify problems that might arise when they are used. Each of the three sections have a different concluding subsection.

There have been a number of recent papers which push the demand estimation literature in different directions, so we conclude that section with a brief review of those articles and why one might be interested in them. The section on production function concludes with a discussion of the problems with the approach we outline, and some suggestions for overcoming them (much of this material is new). The section on the costs of investments, which is our section on "dynamics", is largely a summary and integration of articles that are still in various stages of peer review; so we conclude here with some caveats to the new approaches.

We end this introduction with an indication of the ways Industrial Organization makes use of the developments outlined in each of the three sections of the paper. This should direct the researcher who is motivated by particular substantive issues to the appropriate section of the paper. Each section is self-contained, so the reader ought to be able to read any one of them in isolation.

Demand systems are used in several contexts. First demand systems are *the* major tool for comparative static analysis of any change in a market that does not have an immediate impact on costs (examples include the likely effects of mergers, tax changes, etc.). The static analysis of the change usually assumes a mode of competition (almost always either Nash in prices or in quantities) and either has cost data, or more frequently estimates costs from the first order conditions for a Nash equilibrium. For example in a Nash pricing (or Bertrand) equilibrium with single products firm, price equals marginal cost plus a markup. The markup can be computed as a function of the estimated demand parameters, so marginal costs can be estimated as price minus this markup. Given marginal costs, demand, and the Nash pricing assumption the analyst can compute an equilibrium under post change conditions (after the tax or the merger). Assuming the computed equilibrium is the equilibrium that would be selected, this generates the predictions for market outcomes after the change. If the analyst uses the pre-change data on prices to estimate costs, the only primitive required for this analysis is the demand function and the ownership pattern of the competing products (which is usually observed).

A second use of demand systems is to analyze the effect of either price changes or new goods on consumer welfare. This is particularly important for the analysis of markets that are either wholly or partially regulated (water, telecommunications, electricity, postage, medicare and medicaid, ...). In this context we should keep in mind that many regulatory decisions are either motivated by non-market factors (such as equity considerations), or are politically sensitive (i.e. usually either the regulators or those who appointed them are elected). As a result the analyst often is requested to provide a distribution of predicted demand and welfare impacts across demographic, income and location groups. For this reason a "representative agent" demand system simply will not do.

The use of demand systems to analyze welfare changes is also important in several other contexts. The "exact" consumer price index is a transform of the demand system. Thus ideally we would be using demand systems to construct price indices also (and there is some attempt by the BLS research staff to construct experimental indexes in this way). Similarly the social returns to (either publicly or privately funded) research or infrastructure investments are often also measured with the help of demand systems.

Yet a third way in which demand systems are important to the analysis of I.O. problems is that some of them can be used to approximate the likely returns to potential new products. Demand systems are therefore an integral part of the analysis of product placement decisions, and more generally, for the analysis of the dynamic responses to any policy or environmental change. Finally the way in which tastes are formed, and the impacts of advertising on that process, are problems of fundamental interest to I.O. Unfortunately these are topics we will not address in the demand section of this paper. Our only consolation is the hope is that the techniques summarized here will open windows that lead to a deeper understanding of these phenomena.

Production or cost functions are a second primitive needed for comparative static analysis. However partly because product specific cost data are not available for many markets, the direct estimation of cost functions has not been an active area of research lately. There are exceptions, notably some illuminating studies of learning by doing (see Benkard (2000) and the literature cited there), but not many of them.

What has changed in the past decade and a half is that researchers have gained access to a large number of plant (sometimes firm) level data sets on production inputs and outputs (usually the market value of outputs rather than some measure of the physical quantity of the output). This data, often from various census offices, has stimulated renewed interest in production function estimation and the analysis of productivity. The data sets are typically (though not always) panels, and the availability of the data has focused attention on a particular set of substantive and technical issues.

Substantively, there has been a renewal of interest in measuring productivity and gauging how some of the major changes in the economic environment that we have witnessed over the past few decades affect it. This includes studies of the productivity impacts of; deregulation, changes in tariff barriers, privatization, and broad changes in the institutional environment (e.g. changes in the legal system, in health care delivery, etc.). The micro data has enabled this literature to distinguish between the impacts of these changes on two sources of growth in aggregate productivity; (i) growth in the productivity of individual establishments, and (ii) growth in industry productivity resulting from a reallocating more of the output to the more productive establishments (both among continuing incumbents, and between exitors and new entrants). Interestingly, the prior literature on productivity was also divided in this way. One part focused on the impacts of investments, in particular of research and infrastructure investments, on the productive efficiency of plants. The other focused on the allocative efficiency of different market structures and the impacts of alternative policies on that allocation (in particular of merger and monopoly policy).

From an estimation point of view, the availability of large firm or plant level panels and the desire to use them to analyze the impacts of major changes in the environment has renewed interest in the analysis of simultaneity (endogeneity of inputs) and selection (endogeneity of attrition) on parameter estimates. The data made clear that there are both; (i) large differences in measured "productivity" across plants (no matter how one measures productivity) and that these differences are serially correlated (and hence likely to effect input choices), and (ii) large sample attrition and addition rates in these panels (see Dunne, Roberts and Samuelson (1988) and Davis and Haltwinger (1992) for some of the original work on U.S. manufacturing data). Moreover, the changes in the economic environment that we typically analyze had different impacts on different firms. Not surprisingly, the firms that were positively impacted by the changes tended to have disproportionate growth in their inputs, while those that it affected negatively tended to exhibit falling input demand, and not infrequently, to exit.

The traditional corrections for both simultaneity and selection, corrections based largely on simple statistical models (e.g., use of fixed effect and related estimators for simultaneity, and the use of the propensity score for selection) were simply not rich enough to account for the impacts of such major environmental changes. So the literature turned to simultaneity and selection corrections based on economic models of input and exit choices. The section of this chapter on production functions deals largely with these latter models. We first review the new procedures emphasizing the assumptions they use, and then provide suggestions for amending the estimators for cases where those assumptions are suspect.

The last section of the paper deals explicitly with dynamic models. Despite a blossoming empirical literature on the empirical analysis of static equilibrium models, there has been very little empirical work based on dynamic equilibrium models to date. The I.O. literature's focus on static settings came about not because dynamics were thought to be unimportant to the outcomes of interest. Indeed it is easy to take any one of the changes typically analyzed in static models and make the argument that the dynamic implications of the change might well overturn their static effects. Moreover, there was a reasonable amount of agreement among applied researchers that the notion of Markov Perfect equilibrium provided a rich enough framework for the analysis of dynamics in oligopolistic settings.

The problem was that even given this framework the empirical analysis of the dynamic consequences of the changes being examined was seen as too difficult a task to undertake. In particular, while some of the parameters needed to use the Markov Perfect framework to analyze dynamic games could be estimated without imposing the dynamic equilibrium conditions, some could not. Moreover until very recently the only available methods for estimating these remaining parameters were extremely burdensome, in terms of both computation and researcher time.

The computational complexity resulted from the need to compute the continuation values to the dynamic game in order to estimate the model. The direct way of obtaining continuation values was to compute them as the fixed point to a functional equation, a high order computational problem. Parameter values were inferred from observed behavior by computing the fixed point that determines continuation values at different trial parameter values, and then searching for the parameter value that makes the behavior implied by the continuation values "as close as possible" to the observed data. This "nested fixed point" algorithm is extremely computationally burdensome; the continuation values need to be computed many times and each time they are computed we need to solve the fixed point.

A recent literature in industrial organization has developed techniques that substantially reduce the computational and programming burdens of using the implications of dynamic games to estimate the parameters needed for subsequent applied analysis. That literature requires some strong assumptions, but delivers estimating equations which have simple intuitive explanations and are easy to implement.

Essentially the alternative techniques deliver different semiparametric estimates of continuation values. Conditional on a value of the parameter vector, these estimated continuation values are treated as the true continuation values and used to determine optimal policies (these can be entry and exit policies, investments of various forms, or bidding strategies in dynamic auctions). The parameters are estimated by matching the policies that are predicted in this way to the policies that are observed in the data. Note that this process makes heavy use of nonparametric techniques; nonparametric estimates of either policies or values must be estimated at every state observed in the data. Not surprisingly then Monte Carlo evidence indicates that the small sample properties of the estimators can be quite important in data sets of the size we currently use. This, in turn, both generates preferences for some semiparametric estimators over others, and makes obvious a need for small sample bias correction procedures which, for the most part, have yet to be developed. We now move on to the body of the paper.

1 Demand Systems.

Demand systems are probably the most basic tool of empirical Industrial Organization. They summarize the demand preferences that determines the incentives facing producers. As a result some form of demand system has to be estimated before one can proceed with a detailed empirical analysis of pricing (and/or production) decisions, and, consequently of the profits and consumer welfare likely to be generated by the introduction of new goods.

Not long ago graduate lectures on demand systems were largely based on "representative agent" models in "product" space (i.e. the agent's utility was defined on the product per se rather than on the characteristics of the product). There were a number of problems with this form of analysis that made if difficult to apply in the context of I.O. problems. We begin with an overview of those problems, and the "solutions" that have been proposed to deal with them.

Heterogeneous Agents and Simulation.

First almost all estimated demand system were based on 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.) across consuming units.

Simulation estimators, which Pakes (1986) introduced for precisely this problem, i.e. to enable one to use a micro behavioral model with heterogeneity among agents to structure the empirical analysis of aggregate data, have changed what is feasible in this respect. We can now 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, and/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 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.

The ability to obtain aggregate demand from a distribution of household preferences has had at least two important impacts on demand analysis. First it has allowed us to use the same framework to study demand in different markets, or in the same market at different points in time. A representative agent framework might generate a reasonable approximation to a demand surface in a particular market. However there are often large differences in the distribution of income and other demographic characteristics across markets, and these in turn make an approximation which fits well in one market do poorly in others.

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 variance being especially large in the high income groups (the groups which purchase a disproportionate share of goods sold).

Table I: Cross County Differences in Household Income*

Income Group (thousands)	Fraction of U.S. Population in Income Group	Distribution of Fraction Over Counties	
		Mean	Std. Dev.
0-20	0.226	0.289	0.104
20-35	0.194	0.225	0.035
35-50	0.164	0.174	0.028
50-75	0.193	0.175	0.045
75-100	0.101	0.072	0.033
100-125	0.052	0.030	0.020
125-150	0.025	0.013	0.011
150-200	0.022	0.010	0.010
200 +	0.024	0.012	0.010

* From Pakes (2004, *Review of Industrial Organization*) "Common Sense and Simplicity in Empirical Industrial Organization."

A heterogenous agent demand model with an interaction between price and income uses the available information on differences in the distribution of income to combine the information from different markets. This both enables us to obtain more precise parameter estimates, and provides a tool for making predictions of likely outcomes in new markets.

The second aspect of the heterogenous agent based systems that is intensively used is its ability to analyze the distributional impacts of policies or environmental changes that effect prices and/or the goods marketed. These distributional effects are often of primary concern to both policy makers and to the study of related fields (e.g. the study of voting patterns in political economy, or the study of tax incidence in public finance).

The Too Many Parameters and New Goods Problems.

There were at least two other problems that appeared repeatedly when we used the earlier models of demand to analyze Industrial Organization problems. They are both a direct result of positing preferences directly on products, rather than on the characteristics of products.

1. Many of the markets we wanted to analyze contained a large number of goods that are substitutes for one another. As a result when we tried to estimate demand systems in product space we quickly ran into the "too many parameters problem". Even a (log) linear demand system in product space for J products requires estimates of on the order of J^2 parameters (J price and one income coefficient in the demand for every one of the J products). This was often just too many parameters to estimate with the available data.
2. Demand systems in product space do not enable the researcher to analyze demand for new goods prior to their introduction.

Gorman's polar forms (Gorman (1959)) for multi-level budgeting were an ingenious attempt to mitigate the too many parameter problem. However they required assumptions which were often unrealistic for the problem at hand. Indeed typically the grouping procedures used empirically paid little attention to accommodating Gorman's conditions. Rather they were determined by the policy issue of interest. As a result one would see demand systems for the same good estimated in very different ways with results that bore no relationship to each other[1]. Moreover the reduction in parameters obtained from

[1]For example it was not uncommon to see automobile demand systems that grouped goods into imports and domestically produced in studies where the issue of interest involved tariffs of some form,

multilevel budgeting was not sharp enough to enable the kind of flexibility needed for many I.O. applications (though it was for some, see for e.g. Hausman (1996) and the literature cited there).

The new goods problem was central to the dynamics of analyzing market outcomes. That is in order to get any sort of idea of the incentives for entry in differentiated product markets, we need to be able to know something about the demand for a good which had not yet been introduced. This is simply beyond the realm of what product based demand systems can do. On the other hand entry is one of the basic dynamic adjustment mechanisms in Industrial Organization, and it is hard to think of say, the likely price effects of a merger[2], or the longer run effects of an increase in gas prices, without some way of evaluating the impacts of those events on the likelihood of entry.

The rest of this section of the paper will be based on models of demand that posit preferences on the characteristics of products rather than on products themselves. We do not, however, want to leave the reader with the impression that demand systems in product based, in particular product space models that allow for consumer heterogeneity, should not be used. If one is analyzing a market with a small number of products, and if the issue of interest does not require an analysis of the potential for entry, then it may well be preferable to use a product space system. Indeed all we do when we move to characteristic space is to place restrictions on the demand systems which could, at least in principle, be obtained from product space models. On the other hand these restrictions provide a way of circumventing the "too many parameter" and "new goods" problems which has turned out to be quite useful.

1.1 Characteristic Space: The Issues.

In characteristics space models

- Products are bundles of characteristics.
- Preferences are defined on those characteristics.

and alternatively by gas mileage in studies where the issue of interest was environmental or otherwise related to fuel consumption. Also Gorman's results were of the "if and only if" variety; one of his two sets of conditions were necessary if one is to use multi-level budgeting. For more detail on multi-level budgeting see Deaton and Muellbauer (1980).

[2]Not surprisingly, then, directly after explaining how they will analyze the price effects of mergers among incumbent firms, the U.S. merger guidelines (DOJ (1992)) remind the reader that the outcome of the analysis might be modified by an analysis of the likelihood of entry. Though they distinguish between different types of potential entrants, their guidelines for evaluating the possibility of entry remain distinctly more *ad hoc* then the procedures for analyzing the initial price changes.

- Each consumer chooses a bundle that maximizes its utility. Consumers have different relative preferences (usually just marginal preferences) for different characteristics, and hence make different choices.
- Simulation is used to obtain aggregate demand.

Note first that in these models the number of parameters required to determine aggregate demand is *independent* of the number of products per se; all we require is the joint distribution of preferences over the characteristics. For example if there were five important characteristics, and preferences over them distributed joint normally, twenty parameters would determine the own and cross price elasticities for all products (no matter the number of those products). Second, once we estimate those parameters, if we specify a new good as a different bundle of characteristics then the bundles currently in existence, we can predict the outcomes that would result from the entry of the new good by simply giving each consumer an expanded choice set, one that includes the old and the new good, and recomputing demand in exactly the same way as it was originally computed[3].

Having stated that, at least in principle, the characteristic space based systems solve both the too many parameter and the new goods problems, we should now provide some caveats. First what the system does is restrict preferences: it only allows two products to be similar to one another through similarities in their characteristics. Below we will introduce unmeasured characteristics into the analysis, but the extent to which unmeasured characteristics have been used to pick up similarities in tastes for different products is very limited. As a result if the researcher does not have measures of the characteristics that consumers care about when making their purchase decisions, the characteristic based models are unlikely to provide a very useful guide to which products are good substitutes for one another. Moreover it is these substitution patterns that determine pricing incentives in most I.O. models (and as a result profit margins and the incentives to produce new goods).

As for new goods, there is a very real sense in which characteristic based systems can only provide adequate predictions for goods that are not too “new”. That is, if we formed the set of all tuples of characteristics which were convex combinations of the characteristics of existing products, and considered a new product whose characteristics are outside of this set, then we would not expect the estimated system to be able to provide much information regarding preferences for the new good, as we would be “trying

[3]This assumes that there are no product specific unobservables. As noted below, it is typically important to allow for such unobservables when analyzing demand for consumer products, and once one allows for them we need to account for them in our predictions of demand for new goods. For an example see Berry, Levinsohn and Pakes, 2004.

to predict behavior outside of the sample". Moreover many of the most successful product introductions are successful precisely because they consist of a tuple of characteristics that is very different than any of the characteristic bundles that had been available before it was marketed (think, for example, of the lap top computer, or the Mazda Miata[4]).

Some Background.

The theoretical and econometric groundwork for characteristic based demand systems dates back at least to the seminal work of Lancaster (1971) and McFadden (1974, 1981)[5]. Applications of the Lancaster/McFadden framework however, increased significantly after Berry, Levinsohn, and Pakes (1995) showed how to circumvent two problems that had made it difficult to apply the early generation of characteristic based models in I.O. contexts.

The problems were that

1. the early generation of models used functional forms which restricted cross and own price elasticities in ways which brought into question the usefulness of the whole exercise,
2. the early generation of models did not allow for unobserved product characteristics.

The second problem was first formulated in a clear way by Berry (1994), and is particularly important when studying demand for consumer goods. Typically these goods are differentiated in many ways. As a result even if we measured all the relevant characteristics we could not expect to obtain precise estimates of their impacts. One solution is to put in the "important" differentiating characteristics *and* an unobservable, say ξ, which picks up the aggregate effect of the multitude of characteristics that are being omitted. Of course, to the extent that producers know ξ when they set prices (and recall ξ represents the effect of characteristics that are known to consumers), goods that have high values for ξ will be priced higher in any reasonable notion of equilibrium.

[4]For more detail on just how our predictions would fail in this case see Pakes (1995).

[5]Actually characteristics based models have a much longer history in I.O. dating back at least to Hotelling's (1929) classic article, but the I.O. work on characteristic based models focused more on their implications for product placement rather than on estimating demand systems per se. Characteristic based models also had a history in the price index literature as a loose rational for the use of hedonic price indices; see Court (1939), Griliches (1961), and the discussion of the relationship between hedonics and I.O. equilibrium models in Pakes (2004).

This produces an analogue to the standard simultaneous equation problem in estimating demand systems in the older demand literature; i.e. prices are correlated with the disturbance term. However in the literature on characteristics based demand systems the unobservable is buried deep inside a highly non-linear set of equations, and hence it was not obvious how to proceed. Berry (1994) shows that there is a unique value for the vector of unobservables that makes the predicted shares exactly equal to the observed shares. Berry, Levinsohn, and Pakes (1995; henceforth BLP) provide a contraction mapping which transforms the demand system into a system of equations that is linear in these unobservables. The contraction mapping is easy to compute, and once we have a system which is linear in the disturbances we can again use instruments, or any of the other techniques used in more traditional endogeneity problems, to overcome this "simultaneity problem."

The first problem, that is the use of functional forms which restricted elasticities in unacceptable ways, manifested itself differently in different models and data sets. The theoretical I.O. literature focussed on the nature competition when there was one dimension of product competition. This could either be a "vertical" or quality dimension as in Shaked and Sutton (1982) or a horizontal dimension, as in Salop (1979) (and in Hotelling's (1929) classic work). Bresnahan (1981), in his study of the automobile demand and prices, was the first to bring this class of models to data. One (of several) conclusions of the paper was that a one-dimensional source of differentiation among products simply was not rich enough to provide a realistic picture of demand: in particular it implied that a particular good only had a non-zero cross price elasticity with its two immediate neighbors (for products at a corner of the quality space, there was only one neighbor).

McFadden himself was quick to point out the "IIA" (or independence of irrelevant alternatives) problem of the logit model he used. The simplest logit model, and the one that had been primarily used when only aggregate data was available (data on quantities, prices, and product characteristics), has the utility of the i^{th} consumer for the j^{th} product defined as

$$U_{i,j} = x_j\beta + \epsilon_{i,j}$$

where the x_j are the characteristics of product j (including the unobserved characteristic and price) and the $\{\epsilon_{i,j}\}$ are independent (across both j for a given i and across i for a given j) identically distributed random variables[6]. Thus $x_j\beta$ is the mean utility of

[6]In the pure logit, they have a double exponential distribution. Though this assumption was initially quite important, it is neither essential for the argument that follows, nor of as much importance for current applied work. Its original importance was due to the fact that it implied that the integral that

product j and $\epsilon_{i,j}$ is the individual specific deviation from that mean.

There is a rather extreme form of the "IIA" problem in the demand generated by this model. The model implies that the distribution of a consumer's preferences over products *other than* the product it bought, does not depend on the product it bought. One can show that this implies the following.

- Two agents who buy different products are equally likely to switch to a particular third product should the price of their product rise. As a result two goods with the same shares have the same cross price elasticities with any other good (cross price elasticities are a multiple of $s_j s_k$, where s_j is the share of good j). Since both very high quality goods with high prices and very low quality goods with low prices have low shares, this implications is inconsistent with basic intuition.

- since there is no systematic difference in the price sensitivities of consumers attracted to the different goods, own price derivatives only depends on shares $(\partial s/\partial p) = -s(1-s)$. This implies that two goods with same share must have the same markup in a single product firm "Nash in prices" equilibrium, and once again luxury and low quality goods can easily have the same shares.

No data will ever change these implications of the two models. If your estimates do not satisfy them, there is a programming error, and if your estimates do satisfy them, we are unlikely to believe the results.

A way of ameliorating this problem is to allow the coefficients on x to be individual-specific. Then, when we increase the price of one good the consumers who leave that good have very particular preferences, they were consumers who preferred the x's of that good. Consequently they will tend to switch to another good with similar x's generating exactly the kind of substitution patterns that we expect to see. Similarly, now consumers who chose high priced cars will tend to be consumers who care less about price. Consequently less of them will substitute from the good they purchase for any given price increase, a fact which will generate lower price elasticities and a tendency for higher markups on those goods.

This intuition also makes it clear how the IIA problem was ameliorated in the few studies which had micro data (data which matched individual characteristics to the products those individuals chose), and used it to estimate a micro choice model which was then

---

determined aggregate demand had a closed form, a feature which receded in importance as computers and simulation techniques improved.

explicitly aggregated into an aggregate demand system. The micro choice model interacted *observed* individual and product characteristics, essentially producing individual specific β's in the logit model above. The IIA problem would then be ameliorated to the extent that the individual characteristic data captured the differences in preferences for different x-characteristics across households. Unfortunately many of the factors that determine different households preferences for different characteristics are typically not observed in our data sets, so without allowing for unobserved as well observed sources of differences in the β, estimates of demand systems typically retain many reflections of the IIA problem as noted above; see, in particular Berry, Levinsohn, and Pakes (2004; henceforth MicroBLP) and the literature cited there.

The difficulty with allowing for individual specific coefficients on product characteristics in the aggregate studies was that once we allowed for them the integral determining aggregate shares was not analytic. This lead to a computational problem; it was difficult to find the shares predicted by the model conditional on the model's parameter vector. This, in turn, made it difficult, if not impossible, to compute an estimator with desirable properties. Similarly in micro studies the difficulty with allowing for unobserved individual specific characteristics that determined the sensitivity of individuals to different product characteristics was that once we allowed for them the integral determining individual probabilities was not analytic. The literature circumvented these problems as did Pakes (1986), i.e. by substituting simulation for integration, and then worried explicitly about the impact of the simulation error on the properties of the estimators (see Berry, Linton, and Pakes (2004) and the discussion below).

1.2 Characteristic Space; Details of a Simple Model.

The simplest characteristic based models assumes that each consumer buys at most one unit of one of the differentiated goods. The utility from consuming good j depends on the characteristics of good j, as well as on the tastes (interpreted broadly enough to include income and demographic characteristics) of the household. Heterogenous households have different tastes and so may choose different products.

The utility of consumer (or household) i for good j in market (or time period) t if it purchases the j^{th} good is

$$u_{ijt} = U(\tilde{x}_{jt}, \xi_{jt}, z_{it}, \nu_{it}, y_{it} - p_{jt}, \theta), \tag{1}$$

where $\tilde{x}_{jt}$ is a K-dimensional vector of observed product characteristics other than price,

p_{jt} is the price of the product, $\xi_{j,t}$ represents product characteristics unobserved to the econometrician, z_{it} and ν_{it} are vectors of observed and unobserved (to the econometrician) sources of differences in consumer tastes, y_{it} is the consumer's income, and θ is a vector of parameters to be estimated. When we discuss decisions within a single market, we will often drop the t subscript.

Note that the "partial equilibrium" nature of the problem is incorporated into the model by letting utility depend on the money available to spend outside of this market $(y_i - p_j)$. In many applications, the expenditure in other markets is not explicitly modelled. Instead, y_i is subsumed into either ν_i or z_i and utility is modelled as depending explicitly on price, so that utility is

$$u_{ij} = U(\tilde{x}_j, \xi_j, z_i, \nu_i, p_j, \theta). \tag{2}$$

The consumer chooses one of j products and also has the $j = 0$ choice of not buying any of the goods (i.e., choosing the "outside option".) Denote the utility of outside good as

$$u_{i0} = U(\tilde{x}_0, \xi_0, z_i, \nu_i, \theta), \tag{3}$$

where $\tilde{x}_0$ could either be a vector of "characteristics" of the outside good, or else could be an indicator for the outside good that shifts the functional form of U (because the outside good may be difficult to place in the same space of product characteristics as the "inside" goods.) The existence of the outside option allows us to model aggregate demand for the market's products; in particular it allows market demand to decline if all within-market prices rise.

The consumer makes the choice that gives the highest utility. The probability of that product j is chosen is then the probability that the unobservables ν are such that

$$u_{ij} > u_{ir}, \ \forall r \neq j. \tag{4}$$

The demand system for the industry's products is obtained by using the distribution of the (z_i, ν_i) to sum up over the values for these variables that satisfy the above condition in the market of interest.

Note that, at least with sufficient information on the distribution of the (z_i, ν_i), the same model can be applied when; only market level data are available, when we have micro data which matches individuals to the choices they make, when we have strata samples or information on the total purchases of particular strata, or with any combination of the above types of data. In principal at least, this should make it easy to compare different studies on the same market, or to use information from one study in another.

Henceforth we work with the linear case of the model in equations (2 and 3). Letting $x_j = (\tilde{x}_j, p_j)$, that model can be written as

$$U_{ij} = \Sigma_k x_{jk} \theta_{ik} + \xi_j + \epsilon_{ij}, \tag{5}$$

with

$$\theta_{ik} = \overline{\theta}_k + \theta_k^o{\prime} z_i + \theta_k^u{\prime} \nu_i,$$

where the "**o**" and "**u**" superscripts designate the interactions of the product characteristic coefficients with the observed and the unobserved individual attributes, and it is understood that $x_{i0} \equiv 1$.

We have not written down the equation for $U_{i,0}$, i.e. for the outside alternative, because we can add an individual specific constant term to each choice without changing the order of preferences over goods. This implies we need a normalization and we chose $U_{i,0} = 0$ (that is we subtract $U_{i,0}$ from each choice). Though this is notationally convenient we should keep in mind that the utilities from the various choices are now actually the differences in utility between the choice of the particular good and the outside alternative[7].

Note also that we assume a single unobservable product characteristic, i.e. $\xi_j \in \mathcal{R}$, and its coefficient does not vary across consumers. That is, if there are multiple unobservable characteristics then we are assuming they can be collapsed into a single index whose form does not vary over consumers. This constraint is likely to be more binding were we to have data that contained multiple choices per person (see, for example Heckman and Snyder (1997))[8]. Keep in mind, however, that any reasonable notion of equilibrium would make p_j depend on ξ_j (as well as on the other product characteristics).

The only part of the specification in (5) we have not explained are the $\{\epsilon_{ij}\}$. They represent unobserved sources of variation that are independent across individuals for a

[7] We could also multiply each utility by positive constant without changing the order, but we use this normalization up by assuming that the $\epsilon_{i,j}$ are i.i.d. extreme value deviates, see below.

[8] Attempts we have seen to model a random coefficient on the ξ have lead to results which indicate that there was no need for one, see Das, Olley, and Pakes (1996).

given product, and across products for a given individual. In many situations it is hard to think of such sources of variation, and as a result one might want to do away with the $\{\epsilon_{ij}\}$. We show below that it is possible to do so, and that the model without the $\{\epsilon_{ij}\}$ has a number of desirable properties. On the other hand it is computationally convenient to keep the $\{\epsilon_{ij}\}$, and the model without them is a limiting case of the model with them (see below), so we start with the model in (5). As is customary in the literature, we will assume that the $\{\epsilon_{ij}\}$ are i.i.d. with the double exponential distribution.

Substituting the equation which determines $\theta_{i,k}$ into the utility function in (5) we have

$$U_{ij} = \delta_j + \Sigma_{kr} x_{jk} z_{ir} \theta^o_{rk} + \Sigma_{kl} x_{jk} \nu_{il} \theta^u_{kl} + \epsilon_{ij}, \tag{6}$$

where

$$\delta_j = \Sigma_k x_{jk} \overline{\theta}_k + \xi_j.$$

Note that the model has two types of interaction terms between product and consumer characteristics; (i) interactions between observed consumer characteristics (the z_i) and product characteristics (i.e. $\Sigma_{kr} x_{jk} z_{ir} \theta^o_{rk}.$), and interactions between unobserved consumer characteristics (the ν_i) and product characteristics (i.e. $\Sigma_{kl} x_{jk} \nu_{il} \theta^u_{kl}.$) It is these interactions which generate reasonable own and cross price elasticities (i.e. they are designed to do away with the IIA problem).

1.3 Steps in Estimation: Product Level Data.

There are many instances in which use of the model in (6) might be problematic, and we come back to a discussion of them below. Before doing so, however, we want to consider how to estimate that model. The appropriate estimation technique depends on the data available and the market being modelled. We begin with the familiar case where only product level demand data is available, and where we can assume that we have available a set of variables w that satisfies $E[\xi|w] = 0$. This enables us to construct instruments to separate out the effect of ξ from that of x in determining shares. The next section considers additional sources of information, and shows how the additional sources of information can be used to help estimate the parameters of the problem. In the section that follows we come back to the "identifying" assumption, $E[\xi|w] = 0$, consider the instruments it suggests, and discuss alternatives.

When we only have product level data all individual characteristics are unobserved, i.e. $z_i \equiv 0$. Typically some of the unobserved individual characteristics, the ν_i will have a known distribution (e.g. income), while some will not. For those that do not we assume that distribution up to a parameter to be estimated, and subsume those parameters into the utility function specification (for example assume a normal distribution and subsume the mean in $\overline{\theta}_k$ and the standard deviation in θ_k^u). The resultant known joint distribution of unobserved characteristics is denoted by $f_\nu(\cdot)$. We now describe the estimation procedure.

The first two steps of this procedure are designed to obtain an estimate of $\xi(\cdot)$ as a function of θ. We then require an identifying assumption that states that at $\theta = \theta_0$, the true value of θ, the distribution of $\xi(\cdot;\theta)$ obeys some restriction. The third step is a standard method of moments step that finds the value of θ that makes the distribution of the estimated $\xi(\cdot,\theta)$ obey that restriction to the extent possible.

Step I. We first find an approximation to the aggregate shares conditional on a particular value of (δ,θ). As noted by McFadden (1974) the logit assumption implies that, when we condition on the ν_i, we can find the choice probabilities implied by the model in (6) analytically. Consequently the aggregate shares are given by

$$\sigma_j(\theta,\delta) = \int \frac{exp[\delta_j + \Sigma_{kl} x_{jk}\nu_{il}\theta^u_{kl}]}{1+\sum_q exp[\delta_q + \Sigma_{kl} x_{qk}\nu_{il}\theta^u_{kl}]} f(\nu)\, d(\nu). \tag{7}$$

Typically this integral is intractable. Consequently we follow Pakes (1986) and use simulation to obtain an approximation of it. I.e. we take ns pseudo random draws from $f_\nu(\cdot)$ and compute

$$\sigma_j(\theta,\delta,P^{ns}) = \sum_{r=1}^{ns} \frac{exp[\delta_j + \Sigma_{kl} x_{jk}\nu_{ilr}\theta^u_{kl}]}{1+\sum_q exp[\delta_q + \Sigma_{kl} x_{qk}\nu_{ilr}\theta^u_{kl}]}. \tag{8}$$

where P^{ns} denotes the empirical distribution of the simulation draws. Note that the use of simulation introduces simulation error. The variance of this error decreases with ns but for given ns can be made smaller by using importance sampling or other variance reduction techniques (for a good introduction to these techniques see Rubinstein (1981)). Below we come back to the question of how the simulation error effects the precision of the parameter estimates.

Step II. Let the vector of observed shares be $s^n = [s^n_1, \ldots, s^n_J]$, where n denotes the size of the sample from which these shares are calculated (which is often very large). Step

II finds the unique values of δ that makes the predicted shares for a given θ and set of simulation draws equal to s^n. BLP show that iterating on the system of equations

$$\delta_j^k(\theta) = \delta_j^{k-1}(\theta) + ln[s_j^n] - ln[\sigma_j(\theta, \delta^{k-1}, P^{ns})] \tag{9}$$

leads to the unique δ that makes $\sigma_j(\theta, \delta, P^{ns}) = s_j^n$ for all j[9].

Call the fixed point obtained from the iterations $\delta(\theta, s^n, P^{ns})$. The model in (6) then implies that

$$\xi_j(\theta, s^n, P^{ns}) = \delta(\theta, s^n, P^{ns}) - \Sigma_k x_{jk}\overline{\theta}_k. \tag{10}$$

I.e. we have solved for the $\{\xi_j\}$ as a function of the parameters, the data, and our simulation draws.

"Identification." An identifying restriction for our model will be a restriction on the distribution of the true ξ, the ξ obtained when we evaluate the above equation at $n = ns = \infty$, that will only be satisfied by $\xi_j(\theta, s^\infty, P^\infty)$ when $\theta = \theta_0$ (but not at other values of θ). Different restrictions may well be appropriate in different applied cases, and we come back to a discussion of possible restrictions below. For now, however, we illustrate by assuming we have a set of instruments, say w that satisfy $E[\xi(\theta_0)|w] = 0$. In that case the third and final step of the algorithm is as follows.

Step III. Interact $\xi_j(\theta, s^n, P^{ns})$ with function of w and find that value of θ that makes the sample moments as close as possible to zero. I.e. minimize $\|G_{J,n,ns}(\theta)\|$ where

$$G_{J,n,ns}(\theta) = \sum_j \xi_j(\theta, s^n, P^{ns}) f_j(w). \tag{11}$$

[9]Note that one has to recompute the shares at the "new" δ at each iteration. The system of equations is a mapping from possible values of δ into itself. BLP prove that the mapping is a contraction mapping with modulus less that one. The iterations therefore converge geometrically to the unique fixed point of the system.

Formal conditions for the consistency and asymptotic normality of this estimator are given in Berry, Linton, and Pakes (2004), and provided one accounts for simulation and sampling error in the estimate of the objective function, standard approximations to the limit distribution work (see, for e.g. Pakes and Pollard, 1989). A few of the properties of this limit distribution are discussed below. For now we want only to note that there is an analytic form for the $\overline{\theta}$ parameters conditional on the θ^u; i.e. for the given θ^u the solution for $\overline{\theta}$ is given by the standard instrumental variable formula. So the nonlinear search is only over θ^u.

1.4 Additional Sources of Information on Demand Parameters.

Often we find that there is not enough information in product level demand data to estimate the entire distribution of preferences with sufficient precision. This should not be surprising given that we are trying to estimate a whole distribution of preferences from just aggregate choice probabilities. Other than functional form, the information that is available for this purpose comes from differences in choice sets across markets or time periods (this allows you to sweep out preferences for given characteristics), and differences in preferences across markets or over time for a fixed choice set (the preferences differences are usually associated with known differences in demographic characteristics). The literature has added information in two ways. One is to add an equilibrium assumption and work out its implications for the estimation of demand parameters, the other is to add data. We now consider each of these in turn.

1.4.1 Adding the Pricing Equation.

There is a long tradition in economics of estimating "hedonic" or reduced form equations for price against product characteristics in differentiated product markets (see, in particular Court (1939) and Griliches (1961)). Part of the reason those equations were considered so useful, useful enough to be incorporated as correction procedures in the construction of most countries' Consumer Price Indices, was that they typically had quite high R^2's[10]. Indeed, at least in the cross section, the standard pricing equations estimated by I.O. economists have produced quite good fits (i.e. just as the model predicts, goods with similar characteristics tend to sell for similar prices, and goods in parts of the characteristic space with lots of competitors tend to sell for lower prices). Perhaps it is not surprising then that when the pricing system is added to the demand system

[10]For a recent discussion of the relationship between hedonic regressions and pricing equations with special emphasis on implications for the use of hedonics in the CPI, see Pakes (2004).

the precision of the demand parameters estimates tends to improve noticeably (see, for e.g. BLP).

Adding the pricing system from an oligopoly model to the demand system and estimating the parameters of two systems jointly is the analogue of adding the supply equation to the demand equation in a perfectly competitive model and estimating the parameters of those systems jointly. So it should not be surprising that the empirical oligopoly literature itself started by estimating the pricing and demand systems jointly (see Bresnahan (1981)). On the other hand there is a cost of using the pricing equation. It requires two additional assumptions; (i) an assumption on the nature of equilibrium, and (ii) an assumption on the cost function.

The controversial assumption is the equilibrium assumption. Though there has been some empirical work that tries a subset of the alternative equilibrium assumptions and sees how they fit the data (see for e.g. Berry, Levinsohn, and Pakes (1999) or Nevo (2001)), almost all of it has assumed static profit maximization, no uncertainty, and that one side of the transaction has the power to set prices while the other can only decide whether and what to buy conditional on those prices. There are many situations in which we should expect current prices to depend on likely future profits (e.g.'s include any situation in which demand or cost tomorrow depends on current sales, and/or where there are collusive possibilities; for more discussion see the last section of this paper). Additionally there are many situations, particularly in markets where vertical relationships are important, where there are a small number of sellers facing a small number of buyers; situations where we do not expect one side to be able to dictate prices to another (for an attempt to handle these situations see Pakes, Porter, Ho, and Ishii (2006)).

On the other hand many (though not all) of the implications of the results that are of interest will require the pricing assumption anyway, so there might be an argument for using it directly in estimation. Moreover, as we have noted, the cross-sectional distribution of prices is often quite well approximated by our simple assumptions, and, partly as a result, use of those assumptions is often quite helpful in sorting out the relevance of alternative values of θ.

We work with a Nash in prices, or Bertrand, assumption. Assume that marginal cost, to be denoted by mc, is log linear in a set of observables r_{kj} and a disturbance which determines productivity or ω_j, i.e.

$$ln[mc_j] = \sum r_{k,j}\theta^c_k + \omega_j. \qquad (12)$$

r will typically include product characteristics, input prices and, possibly the quantity produced (if there are non-constant returns to scale). As a result our demand and cost disturbances (i.e. ξ and ω) will typically be mean independent of some of the components of r but not of others. Also we might expect a positive correlation between ξ and ω since goods with a higher unobserved quality might well cost more to produce.

Since we characteristically deal with multiproduct firms, and our equilibrium assumption is that each firm sets each of its price to maximize the profits from all of its products conditional on the prices set by its competitors, we need notation for the set of products owned by firm f, say $\mathcal{J}_f$. Then the Nash condition is that firms set each of their prices to maximize $\sum_{j\in J_f}(p_j - C_j(\cdot))Ms_j(\cdot)$, where C_j is total costs. This implies that for $j = 1,\ldots,J$

$$\sigma_j(\cdot) + \sum_{l\in J_f}(p_l - mc_l)M\frac{\partial\sigma_l(\cdot)}{\partial p_j} = 0. \tag{13}$$

Note that we have added a system of J equations (one for each price) and $R = dim(r)$ parameters to the demand system. So provided $J > R$ we have added degrees of freedom.

To incorporate the information in (13) and (12) into the estimation algorithm rewrite the first order condition as $s + (p - mc)\Delta = 0$, where $\Delta_{i,j}$ is nonzero for elements of a row that are owned by the same firm as the row good. Then

$$p - mc = \Delta^{-1}\sigma(\cdot).$$

Now substitute from (12) to obtain the cost disturbance as

$$ln(p - \Delta^{-1}\sigma) - r'\theta^c = \omega(\theta), \tag{14}$$

and impose the restrictions that

$$Ef_j(w)\omega_j(\theta) = 0 \;\; at \;\; \theta = \theta_0.$$

We add the empirical analogues of these moments to the demand side moments in (11) and proceed as in any method of moments estimation algorithm. This entails one additional computational step. Before we added the pricing system every time we evaluated a θ we had to simulate demand and do the contraction mapping for that θ. Now we also have to calculate the markups for that θ.

1.4.2 Adding Micro Data.

There are a number of types of micro data that might be available. Sometimes we have surveys that match individual characteristics to a product chosen by the individual. Less frequently the survey also provides information on the consumer's second choice (see for e.g. MicroBLP), or is a panel which follows multiple choices of the same consuming unit over time. Alternatively we might not have the original survey's individual choice data, but only summary statistics that provide information on the joint distribution of consumer and product characteristics (for a good example of this see Petrin's, 2002, use of Consumer Expenditure Survey moments in his study of the benefits to the introduction of the minivan). We should note that many of the micro data sets are choice based samples, and the empirical model should be built with this in mind (see, for e.g. MicroBLP (2004); for more on the literature on choice based sampling see Manski and Lerman (1977) and Imbens and Lancaster (1994)).

Since the model in (6) is a model of individual choice, it contains all the detail needed to incorporate the micro data into the estimation algorithm. Thus the probability of an individual with observed characteristics z_i choosing good j given (θ, δ) is given by

$$Pr(j|z_i, \theta, \delta) = \int_\nu \frac{exp[\delta_j + \Sigma_{kl} x_{jk} z_{il} \theta^o_{kl} + \Sigma_{kl} x_{jk} \nu_{il} \theta^u_{kl}]}{1 + \sum_q exp[\delta_q + \Sigma_{kl} x_{qk} z_{il} \theta^o_{kl} + \Sigma_{kl} x_{jk} \nu_{il} \theta^u_{kl}]} f(\nu) d(\nu). \tag{15}$$

What Can be Learned from Micro Data. Assume temporarily that we can actually compute the probabilities in (15) analytically. Then we can use maximum likelihood to estimate (θ^o, θ^u). These estimates *do not* depend on any restrictions on the distribution of ξ. I.e. by estimating free δ_j coefficients, we are allowing for a free set of ξ_j.

On the other hand recall that

$$\delta_j = \Sigma_k x_{jk} \overline{\theta}_k + \xi_j.$$

So we cannot analyze many of the implications of the model (including own and cross price elasticities) without a further assumption which enables us to separate out the effect of ξ from the effect of the x on δ (i.e. without the identifying assumption referred to above). The availability of micro data, then, *does not* solve the simultaneity problem.

In particular it does not enable us to separate out the effect of price from unobservable characteristics in determining aggregate demand. On the other hand there are a few implications of the model that can be analyzed from just the estimates of $(\delta, \theta^o, \theta^u)$. In particular estimates of consumer surplus from the products currently marketed (and hence "ideal" consumer price indices) depend only on these parameters, and hence do not require the additional identifying assumption.

Now say we wanted to use the data to estimate $\overline{\theta}$. In order to do so we need a further restriction so assume, as before, that we have instruments w, and can provide instrumental variable estimates of the $\overline{\theta}$. The number of observations for the instrumental variable regressions is the number of products. That is, at least if we chose to estimate (θ^o, θ^u) without imposing any constraints on the distribution of ξ, the precision of the estimates of $\overline{\theta}$ will depend only on the richness of the product level data. Moreover IV regressions from a single cross-section of products in a given market are not likely to produce very precise results; in particular there is likely to be very little independent variance in prices. Since additional market level data is often widely available, this argues for integrating it with the micro data, and doing an integrated analysis of the two data sources.

One more conceptual point on survey data. What the survey data adds is information on the joint distribution of observed product and consumer attributes. We would expect this to be very helpful in estimating θ^o, the parameters that determine the interactions between z and x. There is a sense in which it also provides information on θ^u, but that information is likely to be much less precise. That is we can analyze the variance in purchases among individuals with the same choice set and the same value of z and use that, together with the i.i.d structure of the ϵ, to try and sort out the variance-covariance of the ν. However this requires estimates of variances conditional on z, and in practice such estimates are often quite imprecise. This is another reason for augmenting cross-sectional survey data with aggregate data on multiple markets (or time periods) in an integrated estimation routine; then the observed variance in z could determine the θ^o and differences in choice sets could help sweep out the impact of the θ^u parameter.

When the data does have second choice information, or when we observe the same consuming unit purchasing more than one product, there is likely to be much more direct information on θ^u. This because the correlation between the $x - intensity$ of the first choice and the second choice of a given individual is a function of both θ^o and the θ^u terms, and the θ^o terms should be able to be estimated from only the first choice data. A similar comment can be made for repeated choices, at least provided the utility function of the consuming unit does not change from choice to choice.

Table 2 illustrates some of these points. It is taken from MicroBLP where the data consisted of a single cross sectional survey of households, and the market level data from the same year. The survey contained information on household income, the number of adults, the number of children, the age (of the head) of household, and whether their residence was rural, urban, or suburban (and all of these were used in the estimation). That study had particularly rich information on vehicle preferences, as each household reported its second as well as its first best choice.

Table 2 provides the best price substitutes for selected models from demand systems for automobiles that were estimated in four different ways; (i) the full model allows for both the z_i and the ν_i (i.e. for interactions between both observed and unobserved individual characteristics and product characteristics), (ii) the logit models that allow for only the z_i, and (iii) the $\sigma' s$ only model allows for only the ν_i. The most important point to note is that without allowing for the ν_i there is a clear IIA problem. The prevalence of the Caravan and the FS pickups when we use the logit estimates (the models without the ν_i) is a result of them being the vehicles with the largest market shares and the apparent absence of the observed factors which cause different households to prefer different product characteristics differentially. Comparing to column (iv) it is clear that the extent of preference heterogeneity caused by household attributes not in our data is large. MicroBLP also notes that when they tried to estimate the full model without the second choice information their estimates of the θ^u parameters were very imprecise; too imprecise to present. However when they added the second choice data they obtained both rather precise estimates of the contributions of the unobserved factors and substitution patterns that made quite a bit of sense. Finally we note that the fact that there was only a single year's worth of data made the estimates of $\overline{\theta}$ quite imprecise, and the paper uses other sources of information to estimate those parameters.

Computational and Estimation Issues: Micro Data. There are a number of choices to make here. At least in principal we could (i) estimate $(\theta^o, \theta^u, \delta)$ pointwise, or (ii) make an assumption on the distribution of ξ (eg. $E[\xi|w] = 0$), and estimate $(\theta^o, \theta^u, \overline{\theta})$ instead of $(\theta^o, \theta^u, \delta)$. However the fact that ξ is a determinant of price, and price is in the x vector, makes it difficult to operationalize (ii). To do so it seems that one would have to make an assumption on the primitive distribution of ξ, solve out for equilibrium prices conditional on (θ, ξ, x), substitute that solution into the choice probabilities in (15), and then use simulation to integrate out the ξ and ν in the formula for those probabilities. This both involves additional assumptions and is extremely demanding computationally. The first procedure also has the advantage that its estimates of (θ^o, θ^u) are independent of the identifying restriction use to separate out the effect of ξ from the effect of x on $\overline{\theta}$.

Assume that we do estimate $(\theta^o, \theta^u, \delta)$. If there are a large number of products or J, this will be a large dimensional search (recall that there are J components of δ), and large dimensional searches are difficult computationally. One way to overcome this problem is

**Table 2: Price Substitutes for Selected Vehicles,
A Comparison Among Models*.**

Vehicle	Full Model	Logit 1^{st}	Logit 1^{st} & 2^{nd}	Sigma Only
Metro	Tercel	Caravan	Ford FS PU	Civic
Cavalier	Escort	Caravan	Ford FS PU	Escort
Escort	Tempo	Caravan	Ford FS PU	Ranger
Corolla	Escort	Caravan	Ford FS PU	Civic
Sentra	Civic	Caravan	Ford FS PU	Civic
Accord	Camry	Caravan	Ford FS PU	Camry
Taurus	Accord	Caravan	Ford FS PU	Accord
Legend	Town Car	Caravan	Ford FS PU	LinTnc
Seville	Deville	Caravan	Ford FS PU	Deville
Lex LS400	MB 300	Econovan	Ford FS PU	Seville
Caravan	Voyager	Voyager	Voyager	Voyager
Quest	Aerostar	Caravan	Caravan	Aerostar
G Cherokee	Explorer	Caravan	Chv FS PU	Explorer
Trooper	Explorer	Caravan	Chv FS PU	Rodeo
GMC FS PU	Chv FS PU	Caravan	Chv FS PU	Chv FS PU
Toyota PU	Ranger	Caravan	Chv FS PU	Ranger
Econovan	Dodge Van	Caravan	Ford FS PU	Dodge Van

From Berry, Levinsohn, and Pakes (2004, *JPE*)"Estimating Differentiated Product Models from a Combination of Micro and Macro Data."

to use the aggregate data to estimate δ conditional on θ from the contraction mapping in (9), and restrict the non-linear search to searching for (θ^o, θ^u).

Finally since the probabilities in (15) are not analytic, either they, or some transform of them (like the score), will have to be simulated. There is now quite a bit of work on simulating the probabilities of a random coefficient logit model (see Train (2003) and the literature cited there). Here we only want to remind the reader that in the applications we have in mind it is likely to be difficult to use the log (or a related) function of the simulated probabilities in the objective function. Recall that if $p^{ns}(\theta)$ is the simulated probability, and $p^{ns}(\theta) = p(\theta) + e^{ns}$, where e^{ns} is a zero mean simulation error, then

$$log[p^{ns}(\theta) \approx log[p(\theta)] + \frac{e^{ns}}{p(\theta)} - \frac{(e^{ns})^2}{2 \times p(\theta)^2}.$$

So if the simulated probabilities are based on ns independent simulation draws each of which has variance $V(p(\theta))$ the *bias* in the estimate of the log probability will be approximately

$$Elog[p^{ns}(\theta)] - log[p(\theta)] \approx -\frac{1}{2 \times ns \times p(\theta)},$$

and ns must be large relative to $p(\theta)$ for this bias to go away (this uses the fact that $Var(p^{ns}(\theta)) \approx p(\theta)/ns$).

In many Industrial Organization problems the majority of the population do not purchase the good in a given period, and the probabilities of the inside goods are formed by distributing the remainder of the population among a very large number of goods. For example in MicroBLP's auto example, only ten per cent of household's purchase a car in the survey year, and that ten percent is distributed among more than two hundred models of cars. So it was common to have probabilities on the order of 10^{-4}. It should not be a surprise then that they chose to fit moments which were linear functions of the error in estimating the probabilities (they fit the covariances of car characteristics and household characteristics predicted by the model to those in the data) rather than maximizing a simulated likelihood.

1.4.3 Identifying Restrictions.

Recall that the source of the endogeneity problem in the demand estimates is the correlation of the product specific unobservable, our ξ, with some of the observable characteristics of the product; in particular we are worried about a correlation of ξ with price. The contraction mapping in (9) is helpful in this respect as it delivers ξ as a linear function of observables. As a result, any of the standard ways of solving endogeneity problems in linear models can be employed here.

The most familiar way of dealing with endogeneity problems in linear models is to use instruments. The question then becomes what is an appropriate instrument for $x's$ in the demand system, a question which has been discussed extensively in the context of perfectly competitive models of supply and demand. As in those models cost shifters that are excluded from demand and uncorrelated with the demand error are available as instruments. The familiar problem here is that input prices typically do not vary much; at least not within a single market. There are a couple of important exceptions. One is when production takes place in different locations even though the products are all sold in one market (as is common when investigating trade related issues, see Berry, Levinsohn, and Pakes (1999)). Another is when a subset of the $x's$ are exogenous, the cost factors are differentially related to different $x's$, and the $x-$intensity of different product varies. In this case interactions between the cost factors and those x's should be useful instruments.

In addition to cost instruments, Nevo (2001) uses an idea from Hausman (1996) market-equilibrium version of the AIDS model, applied to a time-series/cross-section panel of geographically dispersed set of markets. The underlying assumption is that demand shocks are not correlated across markets while cost shocks are correlated across markets. The prices of goods in other markets then become instruments for the price of goods in a given market. Nevo (2001) studies breakfast cereals and so sources of common cost shocks include changes in input prices; sources of common demand shocks (which are ruled out) include national advertising campaigns.

In oligopoly markets prices typically sell at a markup over marginal cost. So if the product's own $(\tilde{x}_j, r_j)$'s are used as instruments, then so might the $(\tilde{x}_{-j}, r_{-j})$ of other products, giving us a lot of potential instruments. Moreover if price setting models like the one in equation (13) are appropriate (and recall that they often have a lot of explanatory power), the impact of the (x_{-j}, r_{-j}) on p_j will depend on whether the product's are owned by the same or by different firms. This type of reasoning dates back at least to Bresnahan (1987), who notes the empirical importance of the idea that markups will be lower in "crowded" parts of the product space and that they will be higher when "nearby" products are owned by the same firm. BLP and Berry, Levinsohn, and Pakes (1999) rely on this sort of argument to propose the use of functions of rivals' observed product characteristics, and of the ownership structure of products, as instruments. Relatedly exogenous changes in competitive conditions across markets are also candidate instruments (say due to the size of the market, or heterogeneity in entry costs).

It is difficult to specify *a priori* how to make optimal use of the product characteristics to predict markups. Both BLP and Berry, Levinsohn, and Pakes (1999) try approximations to the "optimal instrument" formula suggested by Chamberlain (1986). This assumes

$$E[\xi_j|\tilde{x}_j, \tilde{x}_{-j}, r_j, r_{-j}] = E[\omega_j|\tilde{x}_j, \tilde{x}_{-j}, r_j, r_{-j}] = 0,$$

homoscedasticity, and ignores the within market dependence induced by the market interactions. Chamberlain's results then imply that the optimal instrument for our problem is the derivative of these expectations with respect to the parameter vector.

In our context this will be a difficult to compute function of all the product characteristics. BLP tries to approximate this function "non-parametrically" using the exchangeable basis provided in Pakes (1993). Berry, Levinsohn, and Pakes (1999), try an alternative approximation which is more direct, but also more computationally burdensome. They use a first-stage estimate of the parameter vector, θ, to recalculate equilibrium prices with all values of $\xi = \omega = 0$. They then compute the derivative of ξ and ω with respect to θ at the first stage estimate of θ and the new equilibrium prices, and use it as an instrument. I.e. instead of evaluating the mean of the derivative they evaluate the derivative at the mean of the disturbance vector. Note that the instrument is then a function only of exogenous variables, and so results in consistent estimators (even though they are not quite efficient).

So far we have assumed mean independence of the unobservable characteristics, and, as noted, there are plausible reason to believe that product characteristics themselves are correlated with ξ. After all the product design team has at least some control over the level of ξ, and the costs and benefits of producing different levels of the unobservable characteristics might well vary with the observed characteristics of the product. One possible solution would be to completely model the choice of product characteristics, as in the dynamic models considered later in this chapter.

That said since p is typically not as hard to adjust as the other product characteristics, the relationship between ξ and $\tilde{x}$ does not seem to be nearly as direct as that between ξ and p (which is the reason it is often ignored; just as it was in traditional models of demand and supply). So one might be willing to make some reduced form assumption which allows us to proceed without all the detail of a dynamic game. In particular one might try to use changes in demand over time, or across markets, for the same good to control for the influence of unobserved product characteristics.

For example, suppose that we observe demand for the same product over time. It might be reasonable to suppose that the product characteristics are correlated with the unobservable in the year of product introduction. However one might also argue that any changes in the level of unobserved characteristics over time are due to changes in either perception of the product or in customer service that have little to do with the initial x choices. So if t_0 were the date of introduction of the good we might assume that

$$\xi_{j,t} = \xi_{j,t_0} + \eta_{j,t+1}. \tag{16}$$

where $\eta_{j,t+1}$ is mean independent of the observed characteristics of all products. Alternatively we could assume that $\xi_{j,t}$ followed a first order Markov process with only ξ_{j,t_0}, and not the increments in the process, correlated with observed characteristics.

Relatedly if the data contains sales of the same product in many markets one could think of restrictions on how the unobservable for a single product changes across markets. The most straightforward example of this is to require ξ to be the same across markets. This is quite a powerful restriction, and one might question it on the basis of differences in the distribution of consumer preferences across markets that impact on their estimated $\xi's$. A weaker assumption would be that the difference between ξ's for the same product across markets is uncorrelated with the observed x. Similarly, some products within a market may differ only by the the addition of some optional features and we could restrict the way that ξ changes across products that vary only in their options.

1.5 Problems With the Framework.

We have motivated our discussion on demand estimation by noting how the recent literature dealt with the problems that arose in using representative agent models in product space. There are many senses, however, in which the framework outlined above can be too restrictive for particular problems. This section reviews some of the more obvious of them. The impact of these problems depend upon the market one is analyzing and the issues one is focusing on. Also, at least partial solutions to some of these problems are available, and we will direct the reader to them where we can. In large part, however, this section is an outline of agendas for future research on demand estimation for I.O. problems.

We begin with multiple choice and/or dynamics, and then come back to the problem in the static discrete choice model considered above. Most empirical studies simply ignore issues related to multiple choices and/or dynamics. The hope is that the estimated demand system is still the best currently available approximation for analyzing the question of interest. To us the surprising part of the results of those studies is that the framework seems to provide a "reasonable" approximation to substitution patterns, and even more surprisingly, a reasonable approximation to pricing patterns. This despite the fact that we know that consumers' demands and the market equilibrium outcomes are products of much more complicated processes then those we model. Even so, as will become clear presently, there are a number of issues of importance to I.O. which can not be studied empirically without a more detailed understanding of multiple choice and/or the dynamic aspects of demand.

Multiple Units of Demand.

There are many situations for which a model based on the choice of either one or zero units of a good does not match reality[11]. Models for choosing a finite number of units from a set of substitute goods require a specification for the utility from multiple units. Then, at least in principle, we are back to a discrete choice for "tuples" of goods. However to maintain tractability when the number of units can grow large the specification is likely to require constraints which cut down the choice set by implying that some choices are dominated by others (otherwise the size of the choice set grows as J^C, where J is the number of products and C is the maximum number of purchases).

One example of the use of such constraints is Hendel's (1999) two-stage multiple-unit / multiple good framework for the demand of a firm for computers. He simplifies the problem by imagining that the firm faces a random, discrete number of tasks. For each task, it chooses only one type (brand) of computer and, according to the random size of the tasks, a number of computers to purchase. This explicitly accounts for decisions to purchase multiple units of multiple kinds of goods.

Gentzkow (2004) considers a problem with a small number of goods, but where there are a small number of choices. In that study of online and print newspapers, some of the goods are potentially complements, and this requires a different set of modifications. Moreover, as Gentzkow shows the determination of whether goods are in fact complements or substitutes interacts with the issue of the form of consumer heterogeneity in subtle ways reminiscent of the initial condition problems in panel data estimation (see Heckman (1981).

[11] Dubin and McFadden (1984) provide an earlier example with one discrete choice and one continuous choice.

A related problem involves continuous choice over multiple goods. If all goods are purchased in some positive amount by every consumer, then a traditional continuous demand approach, equating marginal rates of substitution across all goods, is appropriate. But many real-world consumer data problems involve a large number of goods with many zero purchase decisions and many positive purchase decisions. Chan (2002) considers the Kuhn-Tucker version of the traditional continuous choice problem to study soft drink purchases.

Dynamic Demand.

Yet another set of problems arises when the demand for the good is inherently dynamic, as occurs with either durable, storable, or experience goods. Models which are appropriate for dynamic demand estimation can become quite complex; they require forward looking consumers whose behavior depends on the likely distribution of future (as well as current) offerings. Moreover in a complete model these future offerings would, in turn, depend on producer's perceptions of consumer demand. A number of new studies make simplifying assumptions which allow them to make some headway.

Both Hendel and Nevo (2002) and Erdem, Keane, and Imai (2003) consider a problem of durable good demand in an explicitly dynamic framework. They consider shopping decisions when consumer's are allowed to store purchases, and use a reduced form assumption on the process generating prices. It has been clear to I.O. economists for some time that we are going to have to model intertemporal substitution of this form in order to understand "sales" in retail markets (see Sobel (1984)).

Two problems in this kind of study are that the rate of consumption (inventory reduction) at home is typically not observed and the dimension of the state space (which involves both the current price vector, which predicts future prices, and also the vector of household inventories of different brands.) In these models new purchases are added to a single-index of home inventories, with different brands of product receiving different utility weights in the inventory stock. This single index of inventories reduces the dimensionality of the state space. Another simplifying assumption is that unobserved household consumption follows a simple rule.

Esteban and Shum (2002) consider a model of durable automobile purchases. They assume a used-car market with zero transaction costs. The zero transaction costs imply that the joint distribution of past choices and consumer characteristics are not a state variable of the problem. Under these assumptions they are able to derive empirical implications about the dynamic pricing problem of the durable goods manufacturer (in determining current price the manufacturer has to worry about future aggregate supply of the used goods). Many, if not most, manufacturing goods are durable.

Studies of demand for advertised experience goods include Erdem and Keane (1996), Ackerberg (2003), and Crawford and Shum (2005). All of these papers feature Bayesian consumers who learn both from experience and from adverting. This lead to a fairly complex dynamic programming problems for the consumer. The studies largely ignore the firm's endogenous pricing and advertising decisions.

Problems With the Static Discrete Choice Specification.

There are also aspects of the static discrete choice specification of the model outlined above whose flexibility, and/or implications, are not yet well understood. One such issue is whether the second derivatives of the demand function are very flexibly estimated. This will determine whether two goods are strategic substitutes or strategic complements, and hence has implications for the analysis of the structure of strategic interaction, and appears to be largely unexplored in the current literature. More generally there are a host of questions on what we can learn non-parametrically about the structure of demand from different kinds of data that we have not touched on here (for a discussion of some of them, see Matzkin's contribution to this volume).

A second such issue concerns the role of the i.i.d. "idiosyncratic match values", the $\epsilon_{ij}' s$, in the models above. These are added to the model largely for computational convenience; they do not seem to match any omitted causal demand determinant. Moreover the presence of the ϵ_{ij} has implications. They imply that each product is "born" with a distribution of consumer tastes whose conditional distribution, conditional on the tastes for other products, has support that ranges from minus to plus infinity. This implies that every conceivable product, no matter its characteristics and price, will have a strictly positive (though perhaps quite tiny) expected market share.

Given the standard ϵ_{ij}'s, each product will also have a positive cross-price effect with every other product: competition is never completely local. Perhaps most problematic, it also implies that if we define a consumer by a (z, ν) combination, every consumer's utility will grow without bound as we increase the number of products – *regardless* of the characteristics or prices of the new products that are introduced. As a result there is a worry about the ability of the model in (6) to provide an adequate approximation to the benefits from introducing new goods[12].

[12] We hasten to note that estimating the consumer surplus generated by new products is an extremely difficult task in any framework. This because we typically do not have data on the demand for new products at prices that are high enough to enable us to estimate the reservation prices of a large fraction of consumers. The characteristic based demand model does use slightly more information in its estimation

To investigate these issues more fully, Berry and Pakes (2005) consider a "pure characteristic" model of demand. That model is exactly the model in equation (6) once we omit the the ϵ_{ij} terms. They consider the analytic properties of the model, then provide an estimation algorithm for it and explore its computational properties, and finally provide Monte Carlo evidence on its performance. Song (2004) has used this model to evaluate the gains from new semiconductor chips. The pure characteristics model is somewhat more computationally burdensome then the model in equation (6), largely because the equation for solving for δ for that model (the analogue to equation 9) is not necessarily a contraction with modulus less than one. On the other hand its shares are easier to simulate to sufficient accuracy. However the jury is still out on the major question; the question of whether the pure characteristic model tends to provide a better approximation to the consumer surplus gains from new goods then the model with the ϵ_{ij}.

Berry and Pakes (2005) and Bajari and Benkard (2005) discuss two different versions of the "pure characteristics" model with "no ϵ"s. Berry and Pakes (2005) consider a discrete choice version of the model, with a utility function of

$$u_{ij} = x_j \beta_i - \alpha_i p_j + \xi_j, \tag{17}$$

where β_i and α_i are random coefficients associated with consumer i's tastes for characteristics and price of product j. Berry and Pakes suggest a BLP-style estimation algorithm.

In contrast, Bajari and Benkard (2005) obtain an estimate of the unobservable demand component, ξ_j, from the pricing side of the model rather than the demand side. The argument is that in a "pure characteristics" model, prices must be strictly increasing in ξ conditional on other x's. Following on recent econometric literature, they show that a monotonic transformation of the ξ can be obtained from data on prices and x's. This transformed ξ is then used in the demand-side analysis to control for unobserved characteristics. Note, however, that consistency of this approach relies on asymptotics in the number of products, and further requires the assumption that products enter the market in such a way that eventually they "fill up" the product space (i.e., for every product, it is assumed that eventually there will be other products whose observed characteristics are arbitrarily close to those of the given product). In practice it is clear that the approach requires data with many products per market, but there has not been enough experience to date to know what "many" means in this context.

of consumer surplus gains than do demand models in product space, since it uses the price variance for products with similar characteristics. However the results are still not terribly robust. Petrin (2002) for example, reports large differences in consumer surplus gains from differences in specifications and data sources.

1.6 Econometric Details.

This subsection summarizes results from Berry, Linton, and Pakes (2004) who provide limit theorems for the parameter estimates from differentiated product models. The actual form of the limit distributions depends on the type of data and type of model. We will focus on the case where only one cross section of market level data is available. Our purpose is to give the reader some indication of how the various estimation errors that have been introduced are likely to effect the parameter estimates, and this is the simplest environment in which to show that[13].

Recall that the objective function minimized in the estimation algorithm or equation (11) is a norm of

$$G_J(\theta, s^n, P^{ns}) = \frac{1}{J}\sum_{j=1}^{J} \xi_j(\theta, \mathrm{s}^{\mathrm{n}}, P^{ns}) f_j(w).$$

The ξ_j are defined implicitly as the solution to the system

$$\mathrm{s}^{\mathrm{n}}_{\mathrm{j}} = \sigma_j(\xi, x; \theta, P^{ns}),$$

where $\sigma(\cdot)$ is defined in (8), the w satisfy $E[\xi|w,\theta_0] = 0$, s^n is the observed vector of market shares, and P^{ns} is notation for the vector of simulation draws used to compute the market shares predicted by the model.

The objective function, $\|G_J(\theta,\, s^n,\, P^{ns})\|$, has a distribution determined by three independent sources of randomness: randomness generated from the draws on the product characteristics (both observed and unobserved, in the full model these are vectors $\{\xi, \tilde{x}, r, \omega\}$), randomness generated from the sampling distribution of s^n, and that generated from the simulated distribution P^{ns}. Analogously there are three dimensions in which our sample can grow: as n, as ns, and as J grow large.

The limit theorems allow different rates of growth for each dimension. Throughout we take pathwise limits, i.e., we write $n(J)$ and $ns(J)$, let $J \to \infty$, and note that our assumptions imply $n(J), ns(J) \to \infty$ at some specified rate. Note also that both s^n and $\sigma(\xi, \theta, P)$ take values in R^J, where J is one of the dimensions that we let grow in our limiting arguments. This is an unusual feature of the econometric model and causes complications in the limiting arguments. As will become obvious sampling error (error in

[13]Cases in which there is data from many regional markets but the same goods are sold in each of them will still have to deal with limits as the number of products grow large; it is just that then we might also want to let the number of markets increase as we increase the number of products. Also in cases with regional markets the computational problems we highlight will be even more severe, as then we will have to compute ξ separately in each different market.

s^n) plays an analogous role to simulation error (error in P^{ns}), so for notational simplicity assume that n is sufficiently large that we do not need to worry about sampling error. When there is no sampling (simulation) error we set n (ns) equal to zero.

We need to find an approximation for the objective function which allows us to separate out the roles of the three sources of error. To this end write

$$\xi(\theta,\, s^0,\, P^{ns}) = \xi(\theta,\, s^0,\, P^0) + \left\{\xi(\theta, s^0,\, P^{ns}) - \xi(\theta,\, s^0,\, P^0)\right\}. \tag{18}$$

The function $\sigma(\xi,\, \theta,\, P)$ is differentiable in ξ, and its derivative has an inverse, say

$$H^{-1}(\xi, \theta, P) = \left\{\frac{\partial\sigma(\xi,\, \theta,\, P)}{\partial\xi'}\right\}^{-1}.$$

Abbreviate $\sigma_o(\theta, s, P) = \sigma(\xi(s, \theta, P), \theta, P)$ and $H_o(\theta, s, P) = H(\xi(s, \theta, P), \theta, P)$, and let

$$\sigma(\xi, P^{ns}, \theta) = \sigma(\xi, P^0, \theta) + \varepsilon^{ns}(\theta).$$

Then from the fact that we obtain ξ from $\sigma(\cdot) = \sigma(\xi, P^0, \theta) + \varepsilon^{ns}(\theta)$ it follows that

$$\xi(\theta,\, s^0,\, P^{ns}) = \xi(\theta,\, s^0,\, P^0) + H_o^{-1}(\theta,\, s^0,\, P^0)\left\{\varepsilon^{ns}(\theta)\right\} + r(\theta,\, s^n,\, P^{ns})$$

where $r(\theta,\, s^n,\, P^{ns})$ is a remainder term. Substituting into (18)

$$G_J(\theta,\, s^n,\, P^{ns}) = G_J(\theta,\, s^0,\, P^0) + \frac{1}{J}z'H_o^{-1}(\theta,\, s^0,\, P^0)\left\{-\varepsilon^{ns}(\theta)\right\} + \frac{1}{J}z'r(\theta,\, s^n,\, P^{ns}).$$

The limit theorems in Berry, Linton, and Pakes (2004) work from this representation of $G_J(\theta, s^n, P^{ns})$. To prove consistency they provide conditions which insure that: i) the second and third terms in this equation converge to zero in probability uniformly in θ, and ii) an estimator which minimized $\|G_J(\theta, s^0, P^0)\|$ over $\theta \in \Theta$ would lead to a consistent estimator of θ^0.

Asymptotic normality requires, in addition, local regularity conditions of standard form, and a limiting distribution for $H_o^{-1}(\theta,\, s^0,\, P^0)\left\{-\varepsilon^{ns}(\theta)\right\}$. The rate needed for this limit distribution depends on how the elements of the $J \times J$ matrix $H_o^{-1}(\theta, s^0, P^0)$ grow, as J gets large. It is easiest to illustrate the issues that can arise here by going back to the simple logit model.

In that model; $u_{i,j} = \delta_j + \epsilon_{i,j}$, with the $\{\epsilon_{i,j}\}$ distributed i.i.d. type II extreme value, and $\delta_j = x_j\overline{\theta} + \xi_j$. Familiar arguments show that $\sigma_j = exp[\delta_j]/(1 + \sum_q exp[\delta_q])$, while

$\sigma_0 = 1/(1 + \sum_q exp[\delta_q])$. In this case the solution to the contraction mapping in (9) is analytic and

$$\xi_j(\theta, s^o, P^o) = (ln[s_j^o] - ln[s_0^o]) - x_j\beta.$$

Thus in this simple case

$$\frac{\partial \xi}{\partial s_j}|_{s^o} = \frac{1}{s_j^o}.$$

Now consider how randomness effects the estimate of $\xi_j(\theta)$. In the simple logit model the only source of randomness is in the sampling distribution of s^n. That is we observe the purchases of only a finite random sample of consumers. Letting their shares be s^n we have, $s^n - s^o = \epsilon^n$. The first order impact of this randomness on the value of our objective function at any θ will be given by

$$H_o^{-1}(\theta, s^0) \times \epsilon_n = \frac{\partial \xi}{\partial s}_{|s=s^0} \times \epsilon^n.$$

This contains expressions like $\epsilon_j^n \frac{1}{s_j^o}$. In the logit model as $J \to \infty$, $s_j^o \to 0$. So as J grows large the impact of any given sampling error grows without bound.

A similar argument holds for the estimator of BLP's model, only in this more complicated model there are two sources or randomness whose impacts increase as J grows large, sampling error and simulation error. Consequently Berry, Linton, and Pakes show that to obtain an asymptotically normal estimator of the parameter vector from this model both n and ns must grow at rate J^2. Note the similarity here to the reason that simulation error is likely to make use of maximum likelihood techniques with survey data computationally demanding; i.e. the impact of the simulation error on the objective function increases as the actual shares get smaller. The computational implication here is that for data sets with large J one will have to use many simulation draws, and large samples of purchasers, before one can expect to obtain an accurate estimator whose distribution is approximated well by a normal with finite variance.

Interestingly, this is not the case for the pure characteristic model discussed in the last subsection. We will not provide the argument here but Berry, Linton, and Pakes (2004) show that in that model both n and ns need only grow at rate J (and depending on the pricing equilibrium, sometimes slower rates will do), for the normal limit distribution to be appropriate. This gives the pure characteristic model a computational advantage in calculating shares, though, as noted above, it is harder to compute the analogue of the contraction mapping in (9) for the pure characteristics model, so it can still be computationally demanding.

1.7 Concluding Remark.

The last decade has seen a rather dramatic change in the way I.O. researchers analyze demand systems. There now is a reasonably substantial body of academic research using the new techniques, and it seems to indicate that, at least for many situations, they allow us to get better approximations to substitution patterns and the likely demand for new goods then had been possible previously. Perhaps not surprisingly then, the techniques have been picked up, to varying extents, by; the consulting community, various government offices, and even by a part of the business community. On the other hand, as we have tried to emphasize, there are empirically important issues and data sets that the new techniques are not able to analyze – at least not without substantial further developments. We welcome those developments. Moreover we hope that they will not be judged by any absolute criteria but rather by the simple test of whether they allow for improvements in our ability to empirically analyze one or more issue of substantive interest.

2 Production Functions.

As noted in the introduction, the advent of new micro data sets on the inputs and outputs from the production process has generated a renewed interest in the estimation of production functions and their use in the analysis of productivity. We begin this section by reviewing the basic simultaneity and selection issues that the recent literature on production function estimation has faced. We then consider the traditional solutions to these issues, pointing out why those solutions are not likely to be terribly helpful in our context.

Next we introduce an approach based on explicit models of input choices and exit decisions that was first introduced in a paper by Olley and Pakes (1996). Our presentation of the Olley-Pakes model will stress the assumptions they used which either we, or others before us, see as questionable (at least in certain environments). These include assumptions on; the timing of input choices, the cost of changing the levels of different inputs over time, the process by which productivity evolves over time, and the relationship of investment to that process. The rest of the section focuses on ways of testing these assumptions, and details recently proposed modifications to the estimation procedure which might be used when they seem appropriate.

2.1 Basic Econometric Endogeneity Issues.

We can illustrate all issues that will concern us with simple Cobb Douglas production technology

$$Y_j = A_j K_j^{\beta_k} L_j^{\beta_l}$$

with one output (Y_j) and two inputs; capital (K_j) and labor (L_j). A_j represents the Hicksian neutral efficiency level of firm j, which is unobserved by the econometrician[14].

Taking natural logs results in a linear equation

$$y_j = \beta_0 + \beta_k k_j + \beta_l l_j + \epsilon_j \tag{19}$$

where lowercase symbols represent natural logs of variables and $\ln(A_j) = \beta_0 + \epsilon_j$. The constant term β_0 can be interpreted as the mean efficiency level across firms, while ϵ_j is the deviation from that mean for firm j. ϵ_j might represent innate technology or management differences between firms, measurement errors in output, or unobserved sources of variance output caused by weather, machine breakdowns, labor problems, etc.

We have known since Marshak and Andrews (1944) that direct OLS estimation of (19) is problematic. The problem is that the right hand side variables, capital and labor, are generally chosen by the firm. If the firm has knowledge of its ϵ_j (or some part of ϵ_j) when making these input choices, the choices will likely be correlated with ϵ_j. For example, suppose that firms operate in perfectly competitive input and output markets (w_j, r_j, and p_j being the prices of labor, capital, and output respectively), that capital is a fixed input, that firms perfectly observe ϵ_j before choosing labor, and that firms' current choices of labor only impact current profits and have no affect on future profits. Then the firm's optimal short-run choice of labor input is given by

$$L_j = \left[\frac{p_j}{w_j} \beta_l e^{\beta_0 + \epsilon_j} K_j^{\beta_k} \right]^{\frac{1}{1-\beta_l}}. \tag{20}$$

[14]The methods discussed in this chapter are equally applicable to many other production functions. As we shall see the major requirements will be that variable inputs have positive cross-partials with productivity, and that the value of the firm is increasing in fixed inputs.

Since choice of L_j (and thus l_j) depends directly on ϵ_j, OLS will generate biased coefficient estimates. In more general models, firms' choices of K_j will also typically be correlated with ϵ_j[15].

There is a second, less well documented, endogeneity problem often inherent in OLS estimation of (19). Firm level datasets usually have a considerable level of attrition. For example, over a wide range of manufacturing industries, Dunne, Roberts, and Samuelson (1988) find exit rates higher than 30% between 5 year census pairs. In applied work, one only has data on firms prior to exiting. If firms have some knowledge of ϵ_j prior to exiting, the firms that continue to produce will have ϵ_j draws from a selected sample, and the selection criteria will be partially determined by the other fixed inputs. Again as a simple example, suppose that firms are monopolies that are exogenously endowed with different fixed levels of capital. Firms then observe ϵ_j, decide whether to exit or not, and choose labor and produce if they have not exited. Also for simplicity suppose that after production firms disappear, so that the firms have no dynamic considerations. Firms in this situation will have an exit rule of the following form:

$$\chi(\epsilon_j, K_j; p_j, w_j, \beta) = 0 \text{ (or exit)} \quad iff \quad \Pi(\epsilon_j, K_j; p_j, w_j, \beta) < \Psi$$

where β is the set of parameters $(\beta_0, \beta_l, \beta_k)$ and Ψ is the non-negative selloff value of the firm. Π is the argmax (over the variable input labor) of variable profits. This condition states that firms exit if variable profits are not at least as high as the selloff value of the firm[16] .

The key point is that this exit condition will generate correlation between ϵ_j and K_j *conditional* on being in the dataset (i.e. on not exiting). In the Cobb-Douglas case, both ϵ_j and K_j positively impact variable profits. As a result, selection will generate *negative* correlation between ϵ_j and K_j, since firms with higher K_j will be able to withstand lower ϵ_j without exiting. Thus, even if K_j is exogenous in the sense that it is uncorrelated with ϵ_j in the entire population of *potentially* active firms, selection can generate negative correlation in one's sample.

[15]Empirical results have lead practitioners to conclude that most often the bias imparted on the labor coefficient α is larger than the bias imparted on the capital coefficient β. This is consistent with models of input choice where labor is more easily adjustable than capital (i.e. labor is a "more variable" input than capital). The intuition here is that because it is more quickly adjustable, labor is more highly correlated with ϵ_j .

[16]This is a very simple example of an exit rule. More realistic models of exit would be dynamic in nature and distinguish between fixed and sunk costs; see the discussion below.

2.2 Traditional Solutions.

As is often the case, the two traditional solutions to these endogeneity problems are instrumental variables and fixed effects. Before discussing these approaches, we make two slight changes to our basic model. First, to explicitly consider the use of longitudinal panel data, we index our variables by time t. Second, to be precise about where exactly the endogeneity problems are coming from, we divide the unobservable ϵ_{jt} into two components, ω_{jt} and η_{jt}, i.e.

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \omega_{jt} + \eta_{jt}. \tag{21}$$

The η_{jt} here are intended to represent unobservables that are *not* observed (or predictable) by the firm before input and exit decisions at time t. As such, they will not be correlated with these choices of inputs or exit behavior. On the other hand, we do allow the possibility that ω_{jt} is observed (or predictable) by firms when they choose inputs and make exit decisions. Intuitively, ω_{jt} might represent factors like managerial ability at a firm, expected down-time due to machine breakdowns or strikes, or the expected rainfall at a farm's location. η_{jt} might represent deviations from expected breakdown rates in a particular year or deviations from expected rainfall at a farm. Another valid interpretation of η_{jt} is that it is classical measurement error in y_{jt} that is uncorrelated with inputs and exit decisions. The basic point here is that we have consolidated our endogeneity problems into ω_{jt}. η_{jt} is not a concern in that regard. We will often refer to ω_{jt} as the firms "unobserved productivity".

2.2.1 Instrumental Variables.

Instrumental variables approaches rely on finding appropriate instruments - variables that are correlated with the endogenous explanatory variables but do not enter the production function and are uncorrelated with the production function residuals. Fortunately, the economics of production suggests some natural instruments. Examining input demand functions (such as (20)) suggests that input prices (r_{jt} and w_{jt}) directly influence choices of inputs. In addition, these prices do not directly enter the production function. The last necessary condition is that the input prices need to be uncorrelated with ω_{jt}. Whether this is the case depends on the competitive nature of the input markets that the firm is operating in. If input markets are perfectly competitive, then input prices should be uncorrelated with ω_{jt} since the firm has no impact on market prices. This is the primary assumption necessary to validate input price instruments. Note why things break down when firms have market power in input markets. If this is the case, input prices will be a function of the quantity purchased inputs, which will generally depend on ω_{jt}.[17]

While using input prices as instruments may make sense theoretically, the IV approach has not been uniformly successful in practice. We believe there are at least four reasons for this. First input prices are often not reported by firms, and when firms do report the labor cost variable, i.e. w_{jt}, is often reported in a way that makes it difficult to use. Labor costs are typically reported as average wage per worker (or average wage per hour of labor). Optimally, we would want this variable to measure differences in exogenous labor market conditions faced by firms. Unfortunately, it may also pick up some component of unmeasured worker quality. Suppose we as econometricians do not observe worker quality, and that some firms employ higher quality workers than others. Presumably, the firms with higher quality workers must pay higher average wages. The problem here is that unobserved worker quality will enter the production function through the unobservable ω_{jt}. As a result, ω_{jt} will likely be positively correlated with observed wages w_{jt}, invalidating use of w_{jt} as an instrument.

Second, to use prices such as r_{jt} and w_{jt} as instruments requires econometrically helpful variation in these variables. While input prices clearly change over time, such time variation is not helpful when one wants to allow flexible effects of time in the production function (e.g. allowing β_0 to be a flexible function of t). One generally needs significant variation in r_{jt} and w_{jt} *across firms* to identify production function coefficients. This can be a problem as we often tend to think of input markets as being fairly national in scope. One might not expect, for example, the price of capital or labor market conditions to vary that much between states. Summarizing, to use the IV approach one 1) has to observe significant variation in input prices across firms in the data, and 2) believe that this variation is due primarily to differences in exogenous input market conditions, *not* due to differences in unobserved input quality.

A third problem with IV is that it relies fairly strongly on an assumption that ω_{jt} evolves exogenously over time, i.e. firms do not choose an input that effects the evolution of ω_{jt}. Allowing ω_{jt} to be affected by *chosen* inputs that we do not control for is very problematic econometrically for the IV approach, for then it would be hard to imagine finding valid instruments for observed input choices. One would need to find variables that affect one input choice but that do not affect other input choices. In general this

[17] Another possible instrument is output prices, as long as the firm operates in competitive output markets. These instruments have been used less frequently, presumably because input markets are thought to be more likely to be competitive. Other related instruments are variables that shift either the demand for output or the supply of inputs. While these types of instruments are typically harder to come by, one can argue that they are valid regardless of the competitive nature of input or output markets.

will be hard to do, since individual input choices typically depend on all input prices.

Finally, the IV approach only addresses endogeneity of input choice, not endogenous exit. Endogenous exit will tend to invalidate the direct use of input prices as instruments. The reason for this is that it is probable that the exit decision will be based in part on input prices. For example, we might expect that firms who face higher input prices to be more likely to exit (i.e. would exit at a higher ω_{jt}). This is likely to generate positive correlation between the instruments and the residuals in the the production function. While direct application of IV in this situation is problematic, it is possible that one could combine the population orthogonality assumptions with a selection model (e.g. Gronau (1974), Heckman (1974, 1976, 1979)) to generate a consistent estimator of the production function parameters.

2.2.2 Fixed Effects.

A second traditional approach to dealing with production function endogeneity issues is fixed effects estimation. In fact, fixed effects estimators were introduced to economics in the production function context (Hoch (1962), Mundlak (1961)). Fixed effects approaches make explicit use of firm panel data. The basic assumption behind fixed effects estimation is that unobserved productivity ω_{jt} is *constant* over time, i.e.

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \omega_j + \eta_{jt}. \tag{22}$$

This allows one to consistently estimate production function parameters using either mean differencing, first differencing, or least squares dummy variables estimation techniques. First differencing, for example, leads to

$$y_{jt} - y_{jt-1} = \beta_k(k_{jt} - k_{jt-1}) + \beta_l(l_{jt} - l_{jt-1}) + (\eta_{jt} - \eta_{jt-1}). \tag{23}$$

Given the assumption that the η_{jt}'s are uncorrelated with input choices $\forall t$,[18] this equa-

[18]The assumption that η_{jt}'s are uncorrelated with input choices (and possibly entry/exit decisions) at all time periods t is often described as a "strict" exogeneity assumption. One can often estimate these fixed effects models under weaker, "sequential" exogeneity assumptions, i.e. that η_{jt}'s are uncorrelated with input choices at all time periods $\leq t$. See Wooldridge (2002) for a discussion of these issues.

tion can be consistently estimated by OLS.[19] Note that this approach simultaneously solves the selection problem of endogenous exit, at least if exit decisions are determined by the time invariant ω_j (and not by the η_{jt}'s). While fixed effects approaches are fairly straightforward and have certainly been used in practice, they have not been judged to be all that successful at solving endogeneity problems in production functions either. Again, there are a number of reasons why this may be the case.

First, it is clearly a strong assumption that ω_j is constant over time. This is especially true given the longer time frames for which panel data is now becoming available. In addition, researchers are often interested in studying periods of data containing major economic environmental changes (e.g. deregulation, privatization, trade policy changes, . . .). Typically these changes affect different firms' productivities differently, and those firms that the change impacts positively will be more likely to increase their inputs and less likely to exit[20].

A second potential problem with fixed effects estimators is that when there is measurement error in inputs, fixed effects can actually generate worse estimates than standard level (OLS) estimators. Griliches and Hausman (1986) note that when inputs are more serially correlated over time than is input measurement error, differencing can lower the signal to noise ratio in the explanatory variables.[21] This can generate higher biases in fixed effects estimators than in OLS estimators, even if ω_j is constant over time and correlated with the explanatory variables[22].

Lastly, fixed effects estimators simply have not performed well in practice. One often gets unreasonably low estimates of capital coefficients[23]. Even one of the seminal papers, Hoch (1962), for example, finds estimates of returns to scale around 0.6 - almost certainly an unrealistically low number. Another empirical finding that appears to contradict the fixed effect assumption concerns the comparison of fixed effects estimates on balanced

[19] Note that generic OLS standard errors are wrong because the residuals will be correlated across observations.

[20] The restriction that ω_j is constant over time is one that has been relaxed (in parametric ways) in the dynamic panel data literature, e.g. Chamberlain (1984), Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1999). For example, these methods can allow ω_{jt} to be composed of a fixed effect plus an AR(1) process.

[21] By signal to noise ratio, Griliches and Hausman mean the variance in an observed explanatory variable due to true variance in the variable, vs. variance in the observed explanatory variable due to measurement error. This signal to noise ratio is inversely related to the bias induced by measurement error.

[22] Note that in this case (i.e. when there is measurement error in inputs), both fixed effects and OLS estimators are biased. Also, note that the more structural approaches discussed later in this chapter are likely also prone to this critique.

[23] "Unreasonable" is clearly not a completely precise statement here. We are referring to cases where the estimated capital coefficient is considerably below capital's cost share or where returns to scale are extremely low.

panels (containing only observations for firms appearing throughout the sample) to those on the full panel. As mentioned above, if ω_j is constant over time, fixed effects estimation completely addresses selection and input endogeneity problems. As a result, one should obtain similar fixed effects estimates whether one uses the balanced sample or the full sample. Olley and Pakes (1996), for example, finds very large differences in these two estimates, suggesting that the fixed effects assumption is invalid. That said, whether or not one takes fixed effects estimates as serious estimates of structural production function parameters, the fixed effect decomposition of variation into within and between components often provides a useful reduced form look at a dataset.

2.3 The Olley and Pakes (1996) Approach.

A recent paper by Olley and Pakes (1996) (henceforth OP) takes a different approach to solving both the simultaneity and selection problems inherent in production function estimation. Their empirical context is that of telecommunications equipment producers (using data from the U.S. Census Bureau's longitudinal research database). The basic empirical goal is to measure the impact of deregulation and the breakup of AT&T on measures of plant level productivity. Our focus is on the OP methodology for addressing the endogeneity problems rather than the actual empirical results.

As we work through the OP approach, it is useful to keep in mind three types of assumptions that will be important in the approach. First there are assumptions on timing and the dynamic nature of inputs. Timing refers to the point in time when inputs are chosen by the firm relative to when they are utilized in production. "Dynamic nature" refers to whether the input choices of the current period effect the cost of input use in future periods; if it does not the input is labelled non-dynamic and if it does the input is labelled as dynamic (and its current value becomes a "state variable" in the problem). Second, there will be a *scalar unobservable* assumption. This assumption limits the dimensionality of the econometric unobservables that impact firm behavior. Third, there will be a *strict monotonicity* assumption on the investment demand function - basically that investment level is strictly monotonic in the scalar unobservable (at least for firms whose investment level is strictly positive). We will see that this last assumption can be generated by more basic assumptions on economic primitives. While some of these assumptions can be relaxed in various ways, we delay that discussion until the next subsection.

Lastly, note that we focus on how to use the OP methodology in practice. We do not address the higher level technical aspects of the methodology, e.g. semi-parametric consistency proofs and alternative standard error derivations for their two-step estimators. For discussion of these issues, e.g. see Pakes and Olley (1996) and the literature they cite.

One might also look at Wooldridge (2004), who presents a concise, one-step, formulation of the OP approach for which standard error derivations are more straightforward[24]. This one-step approach may also be more efficient than the standard OP methodology.

The rest of this section discusses in detail the workings of the OP methodology. We start by describing a simple, bare bones, version of the model and methodology that ignores potential selection problems. We then move on to the full OP model, which does address selection. Lastly, we discuss caveats and extensions of the OP procedure.

2.3.1 The Model.

The OP approach considers firms operating through discrete time, making production choices to maximize the present discounted value (PDV) of current and future profits. The environment is as follows. First, the assumed production function is similar to (21), with an additional input a_{jt}

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + \beta_l l_{jt} + \omega_{jt} + \eta_{jt} \tag{24}$$

the natural log of the age (in years) of a plant. The interest in the age coefficient stems from a desire to separate out cohort from selection effects in determining the impact of age of plant on productivity.

Second, unobserved productivity ω_{jt} is assumed to follow an exogenous first order Markov process. Formally,

$$p\left(\omega_{jt+1} | \left\{\omega_{j\tau}\right\}_{\tau=0}^{t}, I_{jt}\right) = p\left(\omega_{jt+1} | \omega_{jt}\right) \tag{25}$$

where I_{jt} is the firms entire information set at time t. This is simultaneously an *econometric* assumption on unobservables *and* an *economic* assumption on how firms form their perceptions on (i.e. learn about) the evolution of their productivity over time. Specifically, a firm in period t, having just observed ω_{jt}, infers that the distribution of ω_{jt+1} is given by $p\left(\omega_{jt+1}|\omega_{jt}\right)$. Firms thus operate through time, realizing the value of

[24] Though Woolridge deals with input endogeneity, he does not explicitly consider the selection issue. However similar ideas can be used when one needs to incorporate selection corrections.

ω_{jt} at period t and forming expectations of future ω_j's using $p\left(\omega_{jt+1}|\omega_{jt}\right)$. Note that this first-order Markov assumption encompasses the fixed effects assumption where ω_{jt} is fixed over time (i.e. $\omega_{jt} = \omega_j$). OP also assume that $p\left(\omega_{jt+1}|\omega_{jt}\right)$ is stochastically increasing in ω_{jt}. Intuitively, this means that a firm with a higher ω_{jt} today has a "better" distribution of ω_{jt+1} tomorrow (and in the more distant future). Lastly, note that the ω_{jt} process is assumed to be a time-homogeneous Markov process, i.e. p is not indexed by t[25].

Third, capital is assumed to be accumulated by firms through a deterministic dynamic investment process, specifically

$$k_{jt} = (1 - \delta)k_{jt-1} + i_{jt-1}$$

Here we will assume that i_{jt-1} is chosen by the firm at period $t - 1$. That is we are assuming that the capital that the firm uses in period t was actually decided upon at period $t - 1$; so it takes a full production period for new capital to be ordered, received, and installed by firms[26]. Note that capital is a fixed (rather than variable) input.

Lastly, OP specify single period profits as

$$\pi(k_{jt}, a_{jt}, \omega_{jt}, \Delta_t) - c(i_{jt}, \Delta_t).$$

Note that labor l_{jt} is not explicitly in this profit function - the reason is that labor is assumed to be a variable and non-dynamic input. It is variable in that (unlike capital), l_{jt} is chosen at period t, the period it actually gets used (and thus it can be a function of ω_{jt}). It is non-dynamic in the sense that (again, unlike capital) current choice of labor has no impact on the future (i.e. it is not a state variable). This non-dynamic assumption rules out, for example, fixed hiring or firing costs of labor. We discuss relaxing this assumption in section 2.4. For now $\pi(k_{jt}, a_{jt}, \omega_{jt}, \Delta_t)$ can be interpreted as a "conditional" profit function - conditional on the optimal static choice of labor input.

[25]This assumption is not as strong as it might seem, as, e.g. one can easily allow average productivity to vary across time by indexing β_0 by t, i.e. β_{0t}. The assumption can also be relaxed in some cases, i.e. allowing $p_t\left(\omega_{jt+1}|\omega_{jt}\right)$ to be indexed by t.

[26]We note that there is a long literature on trying to determine the distributed lag which translates investment expenditures into a productive capital stock (see for e.g. Pakes and Griliches (1984) and the literature cited there), and one could incorporate different assumptions on this distributed lag into the OP framework (for some examples see Ackerberg et. al (2005)). OP themselves also tried allowing current investment to determine current capital, but settled on the specification used here.

Note also that both $\pi(\cdot)$ and $c(\cdot)$ depend on Δ_t, which represents the economic environment that firms face at a particular point in time. Δ_t could capture input prices, characteristics of the output market, or industry characteristics like the current distribution of the states of firms operating in the industry. The OP formulation allows all these factors to change over time, although they are assumed constant across firms in a given time period. Including market structure in the state space allows some of the competitive richness of the Markov-perfect dynamic oligopoly models of Ericson and Pakes (1995)[27].

Given this economic environment, a firm's maximization problem can be described by the following Bellman equation:

$$V(k_{jt}, a_{jt}, \omega_{jt}, \Delta_t) =$$
$$\max \left\{ \begin{array}{c} \Phi(k_{jt}, a_{jt}, \omega_{jt}, \Delta_t), \\ \max_{i_{jt} \geq 0} \{\pi(k_{jt}, a_{jt}, \omega_{jt}, \Delta_t) - c(i_{jt}, \Delta_t) + \beta E\left[V(k_{jt+1}, a_{jt+1}, \omega_{jt+1}, \Delta_{t+1}) | k_{jt}, a_{jt}, \omega_{jt}, \Delta_t, i_{jt}\right]\} \end{array} \right\}$$

k_{jt}, a_{jt} and ω_{jt} are sufficient to describe the firm specific component of the state space because labor is not a dynamic variable and because $(k_{jt}, a_{jt}, \omega_{jt})$ (and the control i_{it}) are sufficient to describe firms perceived distributions over future $(k_{jt+1}, a_{jt+1}, \omega_{jt+1})$.

The Bellman equation explicitly considers two decisions of firms. First is the exit decision - note that $\Phi(k_{jt}, a_{jt}, \omega_{jt}, \Delta_t)$ represents the sell off value of the firm. Second is the investment decision i_{jt}, which solves the inner maximization problem. Under appropriate assumptions[28], we can write the optimal exit decision rule as

$$\chi_{jt} = \begin{cases} 1 \quad \text{(continue)} & \text{if } \omega_{jt} \geq \overline{\omega}(k_{jt}, a_{jt}, \Delta_t) = \overline{\omega}_t(k_{jt}, a_{jt}) \\ 0 \qquad \text{(exit)} & \text{otherwise,} \end{cases} \tag{26}$$

[27] See Gowrisankaran (1995), Doraszelski and Satterthwaite (2004), and the third section of this chapter for more discussion of such equilibria.

[28] Other than assuming that an equilibria exists, the main assumption here is that the difference in profits between continuing and exiting is *increasing* in ω_{jt}. Given that ω_{jt} positively affects current profits and that the distribution $p(\omega_{jt+1}|\omega_{jt})$ is stochastically increasing in ω_{jt}, the value of continuing is clearly increasing in ω_{jt}. Thus as long as $\Phi(k_{jt}, \omega_{jt}, \Delta_t)$ either doesn't depend on ω_{jt}, decreases in ω_{jt}, or doesn't increase *too fast* in ω_{jt}, this will be satisfied. Note that to get the specific selection bias discussed in Section 1.1 above (i.e. k_{jt} negatively correlated with ω_{jt}), we also need the difference in returns between continuing and exiting to be *increasing* in k_{jt}.

and the investment demand function as

$$i_{jt} = i(k_{jt}, a_{jt}, \omega_{jt}, \Delta_t) = i_t(k_{jt}, a_{jt}, \omega_{jt}). \tag{27}$$

Note the slight change in notation - we are now representing the dependence on Δ_t through the subscript t. See Pakes (1994) for a discussion of conditions under which this investment demand function is strictly increasing in ω_{jt} in the region where $i_{jt} > 0$. That is, conditional on k_{jt} and a_{jt}, firms with higher ω_{jt} optimally invest more. This is an intuitive result - because $p(\omega_{jt+1}|\omega_{jt})$ is assumed stochastically increasing in ω_{jt} , ω_{jt} positively impacts the distribution of all future $\omega_{j\tau}$'s. Since $\omega_{j\tau}$'s positively impact the marginal product of capital in future periods τ, current investment demand should increase. The importance of this *strict monotonicity* condition will be apparent momentarily.

2.3.2 Controlling for Endogeneity of Input Choice.

Given the setup of the model, we can now proceed with the OP estimation strategy. We first focus on dealing only with the endogeneity of input choice, i.e. we assume there are no selection problems due to exit. We will also assume for now that investment levels are always positive, i.e. $i_{jt} > 0, \forall(j, t)$. Later we will relax both these assumptions.

Given that (27) is strictly monotonic in ω_{jt}, it can be inverted to generate

$$\omega_{jt} = h_t(k_{jt}, a_{jt}, i_{jt}) \tag{28}$$

Intuitively, this says that conditional on a firm's levels of k_{jt} and a_{jt}, its choice of investment i_{jt} "tells" us what its ω_{jt} must be. Note that the ability to "invert" out ω_{jt} depends not only on the strict monotonicity in ω_{jt}, but also the fact that ω_{jt} is the *only* unobservable in the investment equation.

This is the *scalar unobservable* assumption mentioned earlier. This, for example, means that there can be no unobserved differences in investment prices across firms[29], no other

[29]Recall that changes in the price of investment over time are permitted as they are picked up by the function h through its dependence on t.

state variables that the econometrician does not observe, and no unobserved separate factors that affect investment but not production. It also prohibits ω_{jt} from following higher than a first order Markov process.[30] We discuss both tests for this assumption and the possibilities for relaxing it in section 2.4.

Substituting (28) into the production function (24) gives

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + \beta_l l_{jt} + h_t(k_{jt}, a_{jt}, i_{jt}) + \eta_{jt}. \tag{29}$$

The first stage of OP involves estimating (29) using semi-parametric methods that treat the inverse investment function $h_t(k_{jt}, a_{jt}, i_{jt})$ non-parametrically. Note the advantages of treating $h_t(k_{jt}, a_{jt}, i_{jt})$ non-parametrically. $i_t(\cdot)$ (and thus its inverse $h_t(\cdot)$) are complicated functions that depend on all the primitives of the model (e.g. demand functions, the specification of sunk costs, the form of conduct in the industry, etc.). These functions are also solutions to a potentially very complicated dynamic game. The OP non-parametric approach therefore avoids both the necessity of specifying these primitives, and the computational burden that would be necessary to formally compute $h_t(\cdot)$.

Given the non-parametric treatment of $h_t(k_{jt}, a_{jt}, i_{jt})$, it is clear that β_0, β_k and β_a cannot be identified using (29). If, for example, $h_t(k_{jt}, a_{jt}, i_{jt})$ is treated as a polynomial in k_{jt}, a_{jt} and i_{jt}, the polynomial will be colinear with the constant, k_{jt}, and a_{jt} terms. Thus, we combine these terms into $\phi_t(k_{jt}, a_{jt}, i_{jt})$, i.e.

$$y_{jt} = \beta_l l_{jt} + \phi_t(k_{jt}, a_{jt}, i_{jt}) + \eta_{jt}. \tag{30}$$

Representing ϕ_t with a high order polynomial in k_{jt}, a_{jt} and i_{jt} (an alternative would be to use kernel methods, e.g. Robinson (1988)) and allowing a different ϕ_t for each time period, OP estimate this equation to recover an estimate of the labor coefficient $\widehat{\beta_l}$. To summarize this first stage, the scalar unobservable and monotonicity assumptions essentially allow us to "observe" the unobserved ω_{jt} - this eliminates the input endogeneity problem in estimating the labor coefficient. Note that it is important here that labor is assumed to be a non-dynamic input - if labor had dynamic implications, it would enter the state space, and thus the investment function and ϕ_t. As a result, β_l would not be

[30] If, for example, ω_{jt} followed a second order process, both ω_{jt} and ω_{jt-1} would enter the state space and the investment decision. With two unobservables in the investment function, it would not be possible to invert out ω_{jt} in the current model.

identified in this first stage. Again, this is an assumption that can potentially be relaxed - see section 2.4.

The second stage of OP identifies the capital and age coefficients β_k and β_a. First, note that the first stage provides an estimate, $\widehat{\phi}_{jt}$, of the term

$$\phi_t(k_{jt}, a_{jt}, i_{jt}) = \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + \omega_{jt}.$$

If one uses a polynomial approximation to $\phi_t(k_{jt}, a_{jt}, i_{jt})$, $\widehat{\phi}_{jt}$ is just the estimated sum of the polynomial terms for a particular (k_{jt}, a_{jt}, i_{jt}) pair. This means that given a particular set of parameters $(\beta_0, \beta_k, \beta_a)$, we have an estimate of ω_{jt} for all j and t

$$\widehat{\omega}_{jt}(\beta_0, \beta_k, \beta_a) = \widehat{\phi}_{jt} - \beta_0 - \beta_k k_{jt} - \beta_a a_{jt}. \tag{31}$$

Next decompose ω_{jt} into its conditional expectation given the information known by the firm at $t-1$ (denote this by I_{jt-1}) and a residual, i.e.

$$\begin{aligned} \omega_{jt} &= E\left[\omega_{jt} | I_{jt-1}\right] + \xi_{jt} \\ &= E\left[\omega_{jt} | \omega_{jt-1}\right] + \xi_{jt} \\ &= g(\omega_{jt-1}) + \xi_{jt} \end{aligned} \tag{32}$$

for some function g. The second line follows from the assumption that ω_{jt} follows an exogenous first order Markov process. By construction, ξ_{jt} is uncorrelated with I_{jt-1} One can think of ξ_{jt} as the innovation in the ω process between $t-1$ and t that is unexpected to firms. The important thing is that given the information structure of the model, this innovation ξ_{jt} is *by definition* uncorrelated with k_{jt} and a_{jt}. The reason is that k_{jt} and a_{jt} are functions of only the information set at time $t-1$. Intuitively, since k_{jt} was actually decided on at time $t-1$ (from the investment decision i_{jt-1}), it cannot be correlated with unexpected innovations in the ω process that occurred *after* $t-1$. Lastly, note that since the stochastic process generating ω_{jt} has been assumed constant over time, the g function need not be indexed by t.[31]

[31] Were we to allow $p\left(\omega_{jt+1} | \omega_{jt}\right)$ to vary across time, we would simply index g by t.

Next, consider rewriting the production function as:

$$y_{jt} - \beta_l l_{jt} = \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + \omega_{jt} + \eta_{jt}. \tag{33}$$

Substituting in both (32) and (31) results in:

$$\begin{aligned} y_{jt} - \beta_l l_{jt} &= \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + g(\omega_{jt-1}) + \xi_{jt} + \eta_{jt} & (34) \\ &= \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + g(\phi_{jt-1} - \beta_0 - \beta_k k_{jt-1} - \beta_a a_{jt-1}) + \xi_{jt} + \eta_{jt} \\ &= \beta_k k_{jt} + \beta_a a_{jt} + \tilde{g}(\phi_{jt-1} - \beta_k k_{jt-1} - \beta_a a_{jt-1}) + \xi_{jt} + \eta_{jt} & (35) \end{aligned}$$

where $\tilde{g}$ encompasses both occurrences of β_0 in the previous line. The key point in (34) is that, as argued above, the residual $\xi_{jt} + \eta_{jt}$ is uncorrelated with all the right-hand side variables.

We do not observe β_l or ϕ_{jt-1}, but we do have estimates of them from the first stage. Substituting $\hat{\beta}_l$ and $\hat{\phi}_{jt-1}$ for their values in the equation above, and treating $\tilde{g}$ non-parametrically we obtain $\sqrt{n}$ consistent estimates of β_k and β_a. If one uses polynomials to approximate $\tilde{g}$, NLLS can be used for estimation[32].

Alternatively one can adapt the suggestion in Wooldridge (2004) to combine both stages into a single set of moments and estimate in one step. This should be more efficient than the OP approach (as it uses the information in the covariances of the disturbances, and any cross equation restrictions). The moment condition in this case is

$$E \left[\begin{array}{ccc} \eta_{jt} & \otimes & f_1(k_{jt}, a_{jt}, i_{jt}, l_{jt}) \\ (\xi_{jt} + \eta_{jt}) & \otimes & f_2(k_{jt}, a_{jt}, k_{jt-1}, a_{jt-1}, i_{jt-1}) \end{array} \right] = 0$$

[32] An alternative way to construct a moment condition to estimate (34) is as follows (see Ackerberg, Caves, and Frazer (2004)). Given β_k and β_a, construct $\widehat{\omega}_{jt} = \widehat{\phi}_{jt} - \beta_k k_{jt} - \beta_a a_{jt}$ $\forall t$. Non-parametrically regress $\widehat{\omega}_{jt}$ on $\widehat{\omega}_{jt-1}$ to construct estimated residuals $\widehat{\xi}_{jt}$ (note that if using polynomial approximation, this can be done using linear methods (since β_k and β_a are given)). Construct a moment condition interacting $\widehat{\xi}_{jt}$ with k_{jt} and a_{jt}. Estimation then involves searching over (β_k, β_a) space to make this moment close to zero.

where f_1 and f_2 are vector valued instrument functions, and $\otimes$ is the Kronecker product operator. Appropriate choices for f_1 and f_2 lead to moments similar to those used by OP. Note that there is a different set of conditioning variables for the moment in η_{jt} than that in the moment for $\xi_{jt} + \eta_{jt}$ (since l_{jt} can be correlated with ξ_{jt})[33].

2.3.3 Controlling for Endogenous Selection.

Next we relax the assumption that there is no endogenous exit. Firms now exit according to the exit rule given in (26). A first important observation is that the first stage of the OP procedure is not affected by selection. The reason is that by construction, η_{jt}, the residual in the first stage equation (30), represents unobservables that are not observed (or predictable) by the firm before input and exit decisions. Thus there is no selection problem in estimating (30). Intuitively, the fact that in the first stage we are able to completely proxy ω_{jt} means that we can control for both endogenous input choice and endogenous exit.

In contrast, the second stage estimation procedure is affected by endogenous exit. Examining (34), note that the residual contains not only η_{jt}, but ξ_{jt}. Since the firms exit decision in period t depends directly on ω_{jt} (see (26)), the exit decision will be correlated with ξ_{jt}, a component of ω_{jt}[34].

We now correct for the selection. Starting from (33), take the expectation of both sides conditional on both the information at $t-1$ and on $\chi_{jt} = 1$ (i.e. being in the dataset at t). This results in

$$\begin{aligned} E\left[y_{jt} - \beta_l l_{jt} | I_{jt-1}, \chi_{jt} = 1\right] &= E\left[\beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + \omega_{jt} + \eta_{jt} | I_{jt-1}, \chi_{jt} = 1\right] \\ &= \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + E\left[\omega_{jt} | I_{jt-1}, \chi_{jt} = 1\right] \qquad (36) \end{aligned}$$

[33] As Wooldridge notes, one can add further lags of variables to these instrument functions, increasing the number of moments; though more lags will not be able to be used on the observations for the initial years.

[34] This correlation relies on OP allowing firms to know the realization of ξ_{jt} before making the exit decision. Otherwise exit would not cause a selection problem. The longer the time period between observations the more serious the selection problem is likely to be. This point comes out clearly in OP's comparison of results based on their "balanced" panel (a data set constructed only from the observations of plants that were active throughout the sample period), to results from their full panel (a panel which keeps the observations on exiting firms until the year they exit and uses observations on new startups from the year they enter). Selection seemed a far larger problem in the balanced than in the full panel.

The second line follows because 1) k_{jt} and a_{jt} are known at $t-1$, and 2) η_{jt} is by definition uncorrelated with either I_{jt-1} or exit at t. Focusing on the last term, we have

$$\begin{aligned} E\left[\omega_{jt}|I_{jt-1},\chi_{it}=1\right] &= E\left[\omega_{jt}|I_{jt-1},\omega_{jt}\geq\overline{\omega}_t(k_{jt},a_{jt})\right] && (37) \\ &= \int_{\overline{\omega}_t(k_{jt},a_{jt})}^{\infty}\omega_{jt}\frac{p(\omega_{jt}|\omega_{jt-1})}{\int_{\overline{\omega}_t(k_{jt},a_{jt})}^{\infty}p(\omega_{jt}|\omega_{jt-1})d\omega_{jt}}d\omega_{jt} && (38) \\ &= g(\omega_{jt-1},\overline{\omega}_t(k_{jt},a_{jt})) \end{aligned}$$

The first equality follows from the exit rule. The second and third equalities follows from the exogenous first order Markov process assumption on the ω_{jt} process.

While we do know ω_{jt-1} conditional on the parameters (from (31)), we do not directly observe $\overline{\omega}_t(k_{jt},a_{jt})$. Modelling $\overline{\omega}_t(k_{jt},a_{jt})$ as a non-parametric function of k_{jt} and a_{jt} might be a possibility, but this would hinder identification of β_k and β_a due to collinearity problems. What we can do is try to control for $\overline{\omega}_t(k_{jt},a_{jt})$ using data on observed exit. Recall that our exit rule is given by

$$\chi_{jt}=\begin{cases} 1 \ \text{(continue)} \\ 0 \ \text{(exit)} \end{cases} \quad \text{according as} \quad \omega_{jt} \gtreqless \overline{\omega}_t(k_{jt},a_{jt}). \qquad (39)$$

This means that the probability of being in the data (at period t) conditional on the information known at $t-1$ is:

$$\Pr(\chi_{jt}=1|I_{jt-1})=\Pr(\omega_{jt}\geq\overline{\omega}_t(k_{jt},a_{jt})|I_{jt-1}) \qquad (40)$$

$$=\Pr(\chi_{jt}=1|\omega_{jt-1},\overline{\omega}_t(k_{jt},a_{jt}))=\widetilde{\widetilde{\varphi}}_t(\omega_{jt-1},\overline{\omega}_t(k_{jt},a_{jt}))$$

$$=\widetilde{\varphi}_t(\omega_{jt-1},k_{jt},a_{jt})=\varphi_t(i_{jt-1},k_{jt-1},a_{jt-1})=P_{jt}$$

The second to last equality holds because of (28), and the fact that k_{jt} and a_{jt} are deterministic functions of i_{jt-1},k_{jt-1}, and a_{jt-1}.

Equation (40) can be estimated non-parametrically, i.e. modelling the probability of surviving in t as a non-parametric function of i_{jt-1},k_{jt-1}, and a_{jt-1}. OP do this in two alternative ways - first using a probit model with a 4th order polynomial in $(i_{jt-1}, k_{jt-1}, a_{jt-1})$ as the latent index, second using kernel methods. For a plant characterized by $(i_{jt-1}, k_{jt-1}, a_{jt-1})$, these estimates allow us to generate a consistent estimate of the probability of the plant surviving to period t $(\widehat{P}_{jt})$.

Next, note that as long as the density of ω_{jt} given ω_{jt-1} is positive in an area around $\overline{\omega}_t(k_{jt}, a_{jt})$, (40) can be inverted to write $\overline{\omega}_t(k_{jt}, a_{jt})$ as a function of ω_{jt-1} and P_{jt}[35], i.e.

$$\overline{\omega}_t(k_{jt}, a_{jt}) = f(\omega_{jt-1}, P_{jt}). \tag{41}$$

Substituting (41) into (37) and (36), and using (31) gives us

$$\begin{aligned} E\left[y_{jt} - \beta_l l_{jt} | I_{jt-1}, \chi_{jt} = 1\right] &= \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + g(\omega_{jt-1}, f(\omega_{jt-1}, P_{jt}) \\ &= \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + g'(\omega_{jt-1}, P_{jt}) \\ &= \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + g'(\phi_{jt-1} - \beta_0 - \beta_k k_{jt-1} - \beta_a a_{jt-1}, P_{jt}) \end{aligned} \tag{42}$$

This is similar to (36), only differing in the additional P_{jt} term in the non-parametric g' function. P_{jt} controls for the impact of selection on the expectation of ω_{jt} - i.e. firms with lower survival probabilities who *do in fact* survive to t likely have higher ω_{jt}'s than those with higher survival probabilities.

Equation (42) implies that we can write

$$\begin{aligned} y_{jt} - \beta_l l_{jt} &= \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + g'(\phi_{jt-1} - \beta_0 - \beta_k k_{jt-1} - \beta_a a_{jt-1}, P_{jt}) + \zeta_{jt} \\ &= \beta_k k_{jt} + \beta_a a_{jt} + \widetilde{g}(\phi_{jt-1} - \beta_k k_{jt-1} - \beta_a a_{jt-1}, P_{jt}) + \zeta_{jt} + \eta_{jt} \end{aligned} \tag{43}$$

where, as in (34), the two β_0 terms have been encompassed into the non-parametric function $\widetilde{g}$. By construction the residual in this equation satisfies $E\left[\zeta_{jt} + \eta_{jt} | I_{jt-1}, \chi_{jt} = 1\right] =$

[35]Formally, (40) implies that $P_{jt} = \widetilde{\widetilde{\varphi}}_t(\omega_{jt-1}, \overline{\omega}_{jt})$. With positive density of ω_{jt} around $\overline{\omega}_{jt}$, $\widetilde{\widetilde{\varphi}}_t$ is strictly monotonic in $\overline{\omega}_{jt}$, so this can be inverted to generate (41).

0. Substituting $\hat{P}_{jt}$, $\hat{\phi}_{jt}$ and $\hat{\beta}_l$ for P_{jt}, ϕ_{jt} and β_l, (43) can also be estimated with NLLS, approximating $\widetilde{g}$ with either a polynomial or a kernel [36].

In this estimation procedure information on β_k and β_a is obtained by comparing labor productivities of firms with the same ω_{jt-1} and P_{jt} but different k_{jt} and a_{jt}. In addition, since the functions $\varphi_t(\cdot)$ and $P_t(\cdot)$ vary across t with changes in industry conditions (while $g(\cdot)$ is assumed constant over time), it also uses information from variation in inputs across firms in *different* time periods that have the same ω_{jt-1} and P_{jt}.

In the selection literature, $\widehat{P}_{jt}$ is referred to as the propensity score - for discussion of these techniques, see, e.g. Heckman (1974, 1978, 1979), Rosenbaum and Rubin (1983), Heckman and Robb (1985), and Ahn and Powell (1993). An important difference between OP and this selection literature is that controlling for the propensity score is not sufficient for OP's model; they require a control for both ω_{jt-1} and for $\overline{\omega}_{jt-1}$.

A number of recent papers have applied the OP procedure successfully. As an example consider Table 3, which displays the results from the food processing industry in Pavcnik (2002) - this is the first out of the eight industries in her table 2. Comparing the OLS to the OP estimates, we see the changes that we should expect. Returns to scale decrease (consistent with positive correlation between unobserved productivity and input use), with the coefficients on the more variable inputs accounting for all of the fall (consistent with this correlation being more pronounced for the variable inputs). Consistent with selection, the capital coefficient rises moving from OLS to OP. The fixed effects estimates are the most difficult to understand, as they generate a coefficient for capital near zero, and an estimate of economies of scale below .9. These results are indicative of those for the other industries in Pavcnik's table 2. The average of the returns to scale estimate across industries when estimated by OLS is 1.13, when estimated by OP it is 1.09, and when estimated by Fixed Effects it is .87. The average of the capital coefficients across industries from OLS is .066, from OP .085, and from Fixed Effects only .021 (with two industries generating negative capital coefficients).

Olley and Pakes (1995) themselves compare their estimates to estimates obtained using OLS and fixed effect on both a balanced panel (a panel constructed only from firms that were operating during the entire fifteen year sample period) and from the full sample (constructed by keeping firms that eventually exit until the year prior to their exit and introducing new entrants as they appear). The difference between the balanced sample estimators and OP estimators on the full sample are truly dramatic, and those between the OLS and fixed effect estimators on the full sample and the OP estimators are similar to those reported above (though somewhat larger in absolute value). In both papers, the OP estimator generates standard errors for the labor coefficient that are not too different

[36] OP try both the kernel and a polynomial with only minor differences in results.

Table 3: Production Function Estimates from Pavcnik (2002)*

	OLS	Fixed Effects	Olley-Pakes
Unskilled Labor	0.178 (0.006)	0.210 (0.010)	0.153 (0.007)
Skilled Labor	0.131 (0.006)	0.029 (0.007)	0.098 (0.009)
Materials	0.763 (0.004)	0.646 (0.007)	0.735 (0.008)
Capital	0.052 (0.003)	0.014 (0.006)	0.079 (0.034)

* From Pavcnik, N. (2002) "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants," *The Review of Economic Studies* 69, January, pp. 245-76

then those generated by OLS, but, as might be expected, standard errors for the capital coefficient do increase (though much less so in the OP results than in Pavcnik's).

2.3.4 Zero Investment Levels.

For simplicity, we assumed above that investment levels for all observations were non-zero. This allowed us to assume that the investment equation was strictly monotonic in ω_{jt} everywhere (and hence could be inverted to recover ω_{it} for every observation). Observations with zero investment call into question the strict monotonicity assumption. However, the OP procedure actually only requires investment to be strictly monotonic in ω_{jt} for a *known* subset of the data. OP themselves take that subset to be all observations with $i_t > 0$, i.e. they simply do not use the observations where investment equals 0.

Even with this selected sample, first stage estimation of (29) is consistent. Since ω_{jt} is being completely proxied for, the only unobservable is η_{jt}, which is by assumption uncorrelated with labor input and with the selection condition $i_{it} > 0$. Second stage estimation of (43) is also consistent when OP discard the data where $i_{jt-1} = 0$ ($\widehat{\phi}_{jt-1} - \beta_0 - \beta_k k_{jt-1} - \beta_a a_{jt-1}$ is not computable when $i_{jt-1} = 0$). The reason is that the error term in (43) is by construction uncorrelated with the information set I_{jt-1}, which contains the investment level i_{jt-1}. In other words, conditioning on $i_{jt-1} = 0$ doesn't say anything about the unobservable ζ_{jt}.

While the OP procedure can accommodate zero investment levels, this accommodation is not without costs. In particular, there is likely to be an efficiency loss from discarding the subset of data where $i_{jt} > 0$. Levinsohn and Petrin (2003) (henceforth LP) suggest an alternative estimation routine whose primary motivation is to eliminate this efficiency loss. They start by noting that in many datasets, particularly those from developing countries, the set of observations with zero investment can be quite large. For example, in LP's dataset on Chilean plants more than 50% of the observations have zero investment (note that in OP's U.S. plant data, this proportion is much less, $\approx$ 8%). To avoid a potentially large efficiency loss, LP suggest using variables other than investment to proxy for the unobserved ω_{jt}. In particular LP focus on firms' choices of intermediate inputs (e.g. electricity, fuels, and/or materials) - these are rarely zero.[37]

Consider the production function

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \beta_m m_{jt} + \omega_{jt} + \eta_{jt} \tag{44}$$

with additional input m_{jt} (e.g. materials). LP assume that like labor, m_{jt} is a variable (i.e. chosen at t), non-dynamic input, and consider the following material demand equation

$$m_{jt} = m_t(k_{jt}, \omega_{jt}). \tag{45}$$

As with the OP investment equation, the demand equation is indexed by t to allow, e.g. input prices, market structure, and demand conditions to vary across time[38]. LP state conditions under which this demand equation is monotonic in ω_{jt}. Given this monotonicity, estimation proceeds analogously to OP. First, (45) is inverted to give:

$$\omega_{jt} = h_t(k_{jt}, m_{jt}). \tag{46}$$

[37] An alternative to LP might be to augment the original OP procedure with a more complete model of investment and/or distributional assumptions on ω, allowing one to utilize the zero investment observations.

[38] Given that materials are a static choice (in contrast to dynamic investment), one might be more willing to make parametric assumptions on this input demand function (since it depends on fewer primitives, e.g. it does not depend on expectations about the future). However, there are caveats of such an approach, see section 2.4.1

Next, (46) is substituted into (44) to give:

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \beta_m m_{jt} + h_t(k_{jt}, m_{jt}) + \eta_{jt} \tag{47}$$

Treating the h_t function non-parametrically results in the following estimating equation

$$y_{jt} = \beta_l l_{jt} + \phi_t(k_{jt}, m_{jt}) + \eta_{jt} \tag{48}$$

where β_k and β_m are not separately identified from the non-parametric term. As in OP, the first stage of LP involves estimating (48) to obtain $\widehat{\beta_l}$ and $\widehat{\phi_{jt}}$. The second stage of LP again proceeds following OP, the main difference being that the parameter on the intermediate input, β_m, still needs to be estimated. Moving the labor term to the left hand side and using (32) gives:[39]

$$\widetilde{y}_{jt} = \beta_k k_{jt} + \beta_m m_{jt} + \widetilde{g}(\phi_{jt-1} - \beta_k k_{jt-1} - \beta_m m_{jt-1}) + \xi_{jt} + \zeta_{jt}, \tag{49}$$

and nonparametric estimates of ϕ_{jt} and of $\tilde{g}(\cdot)$ are used in estimation.

Note that since k_{jt} is assumed decided at $t-1$, it is orthogonal to the residual, $\xi_{jt} + \eta_{jt}$. However, since m_{jt} is a variable input, it is clearly not orthogonal to ζ_{jt}, the innovation component of ω_{jt}. LP address this by using m_{jt-1} as an instrument for m_{jt} in estimation of (49). In their application LP find biases that are generally consistent with those predicted by OP, but some differences in actual magnitudes of coefficients.

2.4 Extensions and Discussion of OP.

The OP model was designed to produce estimates of production function coefficients which are not subject to biases due to simultaneity and selection problems generated by the endogeneity of input demands and exit decisions. We begin this section with a

[39]While the LP procedure does not formally address selection, they note that their procedure could be extended to control for it in the same way as OP.

test of whether the coefficient estimates obtained using OP's assumptions are robust to different sources of misspecification.

There are a variety of reasons why this test could fail and the rest of this subsection considers some of the more likely candidates. Each time a source of possible misspecification in OP's assumption is introduced, we consider modifications to their estimation techniques which produce consistent estimates of production function coefficients under that misspecification. This is in keeping with our belief that different modifications are likely to be appropriate for different industries and data sets. Though the extended models may well be of more general interest, as they typically will produce richer dynamics with more detailed policy implications, we limit ourselves to considering their implications for estimating production function coefficients.

In this context we first investigate relaxing assumptions on the dynamic implications of inputs (e.g. that labor choices today have no dynamic implications) and on the timing of input choices. We then investigate the potential for relaxing the scalar unobservable assumptions of OP. Most of the discussion regarding the timing and dynamic implications of inputs is based on Ackerberg, Caves, and Fraser (2004) (ACF) (also see Buettner (2004a) for some related ideas), while much of the discussion on non-scalar unobservables is taken from Ackerberg and Pakes (2005). We also briefly discuss two recent contributions by Buettner (2004b) and Greenstreet (2005).

2.4.1 A Test of Olley and Pakes' Assumptions.

This subsection combines results from section 4.1 in OP with results from ACF. Broadly speaking, there are two questionable implications of the assumptions used in OP that are central to their estimation strategy. First there is the implication that, conditional on capital and age, there is a one to one mapping between investment and productivity (we give reasons for doubting this implication below). Second there is the direct assumption that the choice of labor has no dynamic implications; i.e. that labor is not a state variable in the dynamic problem.

Focusing on the second assumption first, assume instead that there are significant hiring or firing costs for labor, or that labor contracts are long term (as in, for example, unionized industries). In these cases, current labor input choices have dynamic implications, labor becomes a state variable in the dynamic problem, and equation (28) becomes

$$\omega_{jt} = h_t(k_{jt}, l_{jt}, a_{jt}, i_{jt}). \tag{50}$$

Now the labor coefficient will not be identified in the first stage; i.e. from equation (34)

- the first stage can not separate out the impact of labor on production, or β_l, from its impact on the $h(\cdot)$ function.

ACF point out that under these assumptions β_l can still be identified from the second stage. To see this note that the second stage is now

$$y_{jt} = \beta_l l_{jt} + \beta_k k_{jt} + \beta_a a_{jt} + \widetilde{g}(\phi_{jt-1} - \beta_l l_{jt-1} - \beta_k k_{jt-1} - \beta_a a_{jt-1}) + \xi_{jt} + \eta_{jt}. \quad (51)$$

After substituting $\widehat{\phi}_{jt-1}$ for ϕ_{jt-1}, we can estimate the production function parameters using a semiparametric GMM procedure related to the above. Note, however, that if we maintain the rest of OP's assumptions, then l_{jt} differs from k_{jt} in that labor can adjust to within period variation in productivity. This implies that unlike k_{jt}, l_{jt} can be correlated with ξ_{jt}. As a result we need to use an "instrument" for l_{jt} when estimating equation (51). A fairly obvious instrument is l_{jt-1}. Since l_{jt-1} was decided on at $t-1$, it is uncorrelated with ξ_{jt}, and l_{jt} and l_{jt-1} are typically highly correlated. With this modification, estimation can proceed as before using, say, a polynomial or kernel approximation to $\widetilde{g}$.

Note that even though the first stage does not directly identify any of the parameters of the model in this procedure, we still need the first stage to generate estimates of $\widehat{\phi}_{jt-1}$. Indeed we still need (an extended version) of the assumptions that generates the first stage equation. Before we needed the assumption that conditional on values for (k_{jt}, a_{jt}) there was a one to one map between productivity and investment. Now we need the assumption that conditional on values of (k_{jt}, a_{jt}, l_{jt}) there is a one to one map between productivity and investment.

In fact equation (51) is closely related to the test for the inversion proposed in Olley and Pakes (1995). Recall that they assume that labor is not a dynamic input. In that case when they subtract their first stage estimate $\hat{\beta}_l$ times l from both sides of their second stage equation they obtain

$$y_{jt} - \hat{\beta}_l l_{jt} = (\beta_l - \hat{\beta}_l) l_{jt} + \beta_k k_{jt} + \beta_a a_{jt} + \widetilde{g}(\hat{\phi}_{jt-1} - \beta_k k_{jt-1} - \beta_a a_{jt-1}) + \xi_{jt} + \eta_{jt}, \quad (52)$$

which is an equation with over identifying restrictions[40].

In particular the term $(\beta_l - \hat{\beta}_l) l$ in equation (52) should be zero if the inversion which lead to the estimate of the labor coefficient was a good approximation to reality. Further the inversion implies that what we must subtract from our estimate of ϕ_{jt-1} to obtain lagged productivity is determined by the contribution of (k_{jt-1}, a_{jt-1}) to production of y_{jt-1},

[40]We have omitted a term that results from substituting $\hat{\phi}_{jt-1}$ for the true ϕ_{jt-1} in this equation. The additional term's impact on the parameter estimates is $o_p(1/\sqrt{J})$, and so does not effect their limit distributions.

i.e. by (β_k, β_a). These coefficients also determine the contribution of (k_{jt}, a_{jt}) to y_{jt} given ω_{jt-1}. If our inversion were seriously in error we would expect that $\phi_{jt-1} - \beta_k k_{jt-1} - \beta_a a_{jt-1}$ would not be perfectly correlated with ω_{jt-1} and as a result there would be a residual component of productivity we are not controlling for. Provided there was an endogeneity problem in the first place, this residual should be correlated with (k_{jt}, a_{jt}). Thus OP allow the coefficients of (k_{jt}, a_{jt}) to differ from those of (k_{jt}, a_{jt}) in equation (52), use $(l_{j,t-1}, k_{jt}, a_{jt})$ and powers and lags of these variables as instruments, and then test whether; $\hat{\beta}_l - \beta_l = 0$, and whether the coefficients on the current lagged values of k and a are equal[41].

As noted when the current labor choice has dynamic implications the first stage estimate of β_l obtained by OP is inconsistent (regardless of whether the inversion is correct). However even if labor is dynamic equation (51) does still generate over identifying restrictions; the existence of the inversion implies that the current and lagged values of (l, k, a) should enter in this equation with the same factors of proportionality. I.e. if the inversion is correct then what we must subtract from our estimate of ϕ_{jt-1} to obtain lagged productivity is determined by the contribution of $(l_{jt-1}, k_{jt-1}, a_{jt-1})$ to production of y_{jt-1}, i.e. by $(\beta_l.\beta_k, \beta_a)$. These coefficients also determine the contribution of (l_{jt}, k_{jt}, a_{jt}) to y_{jt} given ω_{jt-1}. That is if we were to estimate

$$y_{jt} = \beta_l^* l_{jt} + \beta_k^* k_{jt} + \beta_a^* a_{jt} + \widetilde{g}(\phi_{jt-1} - \beta_l l_{jt-1} - \beta_k k_{jt-1} - \beta_a a_{jt-1}) + \xi_{jt} + \eta_{jt}, \quad (53)$$

and then test whether $(\beta_l^*, \beta_k^*, \beta_a^*) = (\beta_k, \beta_l, \beta_a)$, with a large enough data set we should reject the null of equality if assumptions which underlie the inversion are wrong. Given the additional parameters one will need additional instruments to estimate this specification. Natural instruments would be those used in OP, i.e. $(l_{j,t-1}, k_{jt}, a_{jt})$ and powers and lags of these variables.

Two other points about the test. First, the OP test conditions on the fact that labor is variable (i.e. it is not a state variable) and endogenous (current labor is correlated with ξ_{jt}), and then tests whether the inversion is a good approximation. We could have proceeded by conditioning on the inversion and then test one or both of the assumptions that; labor is dynamic and/or labor choices are fixed prior to the realization of ξ_t. We would do this by estimating the two equations in ACF simultaneously and then imposing constraints. The constraint to be tested in asking whether labor can be treated as a non-dynamic (or variable) input is whether $\phi(l_{jt}, i_{jt}, k_{jt}, a_{jt}) = \beta_l l_{jt} + \phi(i_{jt}, k_{jt}, a_{jt})$. To test whether labor is endogenous in the sense that it can react to ξ_{jt} we estimate the system once using l_{jt-1} as an instrument for l_{jt} in equation (51) and once using l_{jt} as an instrument on itself. Exactly what it makes sense to condition on (and what to test for) is likely to depend on the characteristics of the industry being studied. Alternatively we

[41]Note that we can not use both current and lagged values of a_{jt} as instruments for the two are collinear. We could, however, use different functions of a_{jt} as additional instruments.

could improve the power of the omnibus test in equation (51) by estimating the first stage in ACF simultaneously with this equation and then asking $(\beta_l^*, \beta_k^*, \beta_a^*) = (\beta_l, \beta_k, \beta_a)$. If that is accepted we could then try the additional (nested) constraints implied by an assumption that labor is not endogeous[42].

Finally a word of caution on the usefulness of these tests. First we have made no attempt to look at the power of these tests. Though OP find very precise estimates of differences in coefficients from (53), their data seems to deliver more precise estimates than many other data sets (see, for e.g. ACF). Second it is important to realize that the test that $(\beta_l^*, \beta_k^*, \beta_a^*) = (\beta_k, \beta_l, \beta_a)$ is designed to ask the limited question of whether making our approximations greatly hinders our ability to obtain reasonable production function coefficient. As a result we are using the difference in these coefficients, normalized by the variance-covariance of those differences, as our metric for "reasonableness". There are other metrics possible, one of which would be to have some prior knowledge of the characteristics of the industry the researcher is working on (and we expect these results to vary by industry). Further there may well be independent reasons for interest in the timing of input decisions or in our invertibility assumption (see the discussion below), and a test result that our approximations do not do terrible harm to production function estimates does not imply that they would do little harm in the analysis of other issues (for e.g. in the analysis of the response of labor hiring to a change in demand, or in the response of investment to an infrastructure change which increases productivity).

2.4.2 Relaxing Assumptions on Inputs.

This subsection assumes that there is an inversion from productivity to investment conditional on the state variables of the problem, and investigates questions regarding the nature of the input demands given this inversion. ACF note that there are two dimensions along which we can classify inputs in this context, and the two dimensions have different implications for the properties of alternative estimators. First inputs can either be variable (correlated with ξ_{jt}) or fixed (uncorrelated with ξ_{jt}). Second the inputs can either be dynamic, i.e. be state variables in the dynamic problem and hence conditioned on in the relationship between productivity and investment, or static. So if we generalize,

[42] Note that both these ideas; that one can allow labor to have dynamic effects, and that some of the assumptions behind these procedures are testable - are related to the the dynamic panel literature cited above (e.g. Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1999)) in that further lags of inputs are typically used as instruments. If one were willing to assume that the $\eta_{j,t}$ are independently distributed across time then the residuals should be uncorrelated with past values of output also. However if η_{jt} represented serially correlated measurement error in the observations on y_t then the η_{jt} may be serially correlated, and we could not expect a zero correlation between past output and the disturbance from (53). ACF flesh out the distinction between their methods and the dynamic panel literature further.

and allow for inputs of each of the four implicit types we have

$$y_{jt} = \beta^{vs} X_{jt}^{vs} + \beta^{vd} X_{jt}^{vd} + \beta^{fs} X_{jt}^{fs} + \beta^{fd} X_{jt}^{fd} + \omega_{jt} + \eta_{jt} \tag{54}$$

where the input acronyms correspond to these dimensions, e.g. X_{jt}^{vs} represent variable, non-dynamic inputs, while X_{jt}^{fd} represent fixed, dynamic inputs, and so on.

The various coefficients can be identified in different ways. β^{vs}, like labor in the original OP framework, can be identified either in the first stage, or in the second stage using X_{jt-1}^{vs}as an instrument (because X_{jt}^{vs} is variable and thus potentially correlated with ξ_{jt}, it cannot be used as an instrument in the second stage). β^{fd}, like capital in the original OP framework, cannot be identified in the first stage, but it can be identified in the second stage using either X_{jt}^{fd} or X_{jt-1}^{fd} (or both) as instruments. β^{vd}, the coefficients on the inputs that are variable and dynamic, also cannot be identified in the first stage, but can be identified in the second stage using X_{jt-1}^{vd} as an instrument. Lastly, β^{fs} can be identified either in the first stage or in the second stage using either X_{jt}^{fs} or X_{jt-1}^{fs} (or both) as instruments.

Note also that if we have any static or fixed inputs we have over identifying restrictions.[43] This over identification can potentially be useful in testing some of the timing assumptions. For example, suppose one starts by treating capital as a fixed, dynamic input. One could then estimate the second stage using both k_{it-1} and k_{it} as instruments, an over identified model. In the GMM context, one could test this over identification with a J-test (Hansen(1982)). Since k_{it} is a valid instrument only when capital is truly fixed (yet k_{it-1} is a valid instrument regardless) rejection of the specification might be interpreted as evidence that capital is not a completely fixed input. Consistent estimation could then proceed using only k_{it-1} as an instrument. Again, the Wooldridge (2004) framework makes combining these multiple sources of identification and/or testing very convenient.

ACF also look deeper into the various assumptions on inputs. They note that, under the assumption that l_{jt} is a variable input, for it to have the independent variance needed to estimate our first stage equation (30), there must be a variable, say z_{jt}, that impacts firms' choices of l_{jt} but that does not impact choices of investment at t. This variable z_{jt} must also have some variance that is independent of ω_{jt} and k_{jt}. If this were not the

[43] In all these cases, further lags (prior to $t-1$) of the inputs can be used as instruments and thus as over identifying restrictions, although it is not clear how much extra information is in these additional moment conditions, and one will not be able to use these additional lags in the initial time periods.

case, e.g. if $l_{jt} = f_t(\omega_{jt}, k_{jt})$, then one can show that l_{jt} is perfectly collinear with the nonparametric function in equation (30), implying that one cannot estimate β_l from that equation. Note that the variable z_{jt} does not need to be observed by the econometrician.

Thus, to proceed as OP do we need the demand function for labor to be

$$l_{jt} = f_t(\omega_{jt}, k_{jt}, z_{jt})$$

where z_{jt} are additional factors that affect demand for labor (or more generally, demand for the variable inputs) with non-zero conditional variance (conditional on ω_{jt}, k_{jt}). Note that the z_{jt} *cannot* be serially correlated. If this were the case, then z_{jt} would become part of the state space, influence i_{jt}, and one would not be able to do the inversion[44]

Even with this restriction, there are at least two possible z_{jt}'s here; i.i.d. firm specific input price shocks, and i.i.d. random draws to the environment that cause differences in the variance of η_{jt} over time (since the profit function is a convex function of η the variance in this variable will affect labor demand). The latter could be associated with upcoming union negotiations, the likelihood of machine break downs due to age of equipment, or the approach of maintenance periods. One problem with the i.i.d. input price shock story is that it is somewhat at odds with the assumptions that all other components of prices are constant across firms and that the other unobservables (ω_{jt}) in the model are serially correlated over time.

ACF provide two additional ways of overcoming this problem. First they note that if one weakens OP's timing assumptions slightly, one can still identify l_{jt} in the first stage. Their observation also reopens an avenue of research on the timing of input decisions which dates back at least to Nadiri and Rosen (1974). Suppose that l_{jt} is actually not a perfectly variable input, and is chosen at some point in time between periods $t-1$ and the completion of period t. Denote this point in time as $t-b$, where $0 < b < 1$. Suppose that ω evolves between the subperiods $t-1$, $t-b$, and t according to a first order Markov process, i.e.

$$p\left(\omega_{jt}|I_{jt-b}\right) = p\left(\omega_{jt}|\omega_{jt-b}\right), \quad \textit{and} \quad p\left(\omega_{jt-b}|I_{jt-1}\right) = p\left(\omega_{jt-b}|\omega_{jt-1}\right).$$

In this case, labor input is not a function of ω_{jt}, but of ω_{jt-b}, i.e.

[44]Note also that observing z_{jt} would not help in this serially correlated case. While one would now be able to do the inversion, z_{jt} would enter the non-parametric function, again generating perfect collinearity.

$$l_{jt} = f_t(\omega_{jt-b}, k_{jt}).$$

Since ω_{jt-b} cannot generally be written as a function of k_{jt}, a_{jt}, and i_{jt}, l_{jt} will *not* generally be collinear with the non-parametric term in (30), allowing the equation to be identified. The movement of ω between $t-b$ and t is what breaks the collinearity problem between l_{jt} and the non-parametric function. The second alternative suggested by ACF avoids this collinearity problem by abandoning the first stage identification of the labor coefficient. Instead, they suggest identifying the labor coefficient in the second stage using l_{jt-1} as an instrument.

Importantly, ACF argue that this collinearity problem is more severe when using the LP procedure. They contend that it is considerably harder to tell a believable story in which the assumptions of LP hold and where l_{jt} varies independently of the non-parametric function in (48). The reason for this is that it is hard to think of a variable z_{jt} that would affect a firms' labor choices but not their material input choices (either directly or indirectly through the labor choice).[45] ACF suggest a couple of procedures as alternatives to LP.[46] The first, based on the discussion above, again involves simply identifying the labor coefficient in the second stage. This can be done using either l_{jt} or l_{jt-1} to form an orthogonality condition, depending on what one wants to assume about the timing of the labor choice. Moreover, it can also be done in a manner that is also consistent with labor having dynamic effects. The second procedure is more complicated and involves sequentially inverting the value of ω_{jt} at each point in time at which inputs are chosen. While this procedure depends on independence (rather than mean independence) assumptions on innovations in ω_{jt}, it has the added advantage of allowing one to infer something about the point in time that labor is chosen. Bond and Soderbom (2005) make a somewhat related point regarding collinearity. They argue that in a Cobb-Douglas context where input prices are constant across firms, it is hard if not impossible to identify coefficients on inputs that are perfectly variable and have no dynamic effects. This is important for thinking about identification of coefficients on X_{jt}^{vs} in the above formulation.

[45] ACF note that one probably will not observe this perfect collinearity problem in practice (in the sense that the first stage procedure will actually produce an "estimate"). However, they point out that unless one is willing to make what they argue are extremely strong and unintuitive assumptions, the lack of perfect collinearity in practice must come entirely from misspecification in the LP model.

[46] An alternative approach to dealing with these collinearity problems might be to model the input demand functions (investment or materials) parametrically. If $g()$ is parametric, one doesn't necessarily have this collinearity problem. However, at least in the LP situation this does not guarantee identification. ACF show that in the Cobb-Douglas case, substituting in the implied parametric version of the material input function leads to an equation that cannot identify the labor coefficient.

2.4.3 Relaxing the Scalar Unobservable Assumption.

The assumption of a scalar unobserved state variable is another aspect of the OP approach that might be a source of concern. We begin with three reasons for worrying about this assumption and then provide a way of modifying the model to account for each of them. In each case we bring information on additional observables to bear on the problem. As a result, one way of looking at this section is as a set of robustness tests conducted by asking whether the additional observables affect the results.

Our three concerns in order of increasing difficulty are as follows. First productivity itself is a complex functions of many factors, and it may not be appropriate to assume that one can represent it as a first order Markov process. Second investment might well respond to demand factors that are independent of the firm's productivity. Then there is no longer a one to one mapping between investment and productivity given capital and age. Consequently we can not do the inversion in equation (28) underlying the first stage of the OP procedure. Finally, at least in some industries we often think of two sources of increments in productivity, one that results from the firm's own research investments, and one whose increments do not depend on the firm's behavior. A process formed from the sum of two different first order Markov processes is not generally a first order Markov process, and if one of those processes is "controlled" it may well be difficult to account for it in the same way as we can control for exogenous Markov processes.

First assume that productivity follows a second-order (rather than first-order) Markov process. This changes the investment demand equation to

$$i_{jt} = i_t(k_{jt}, a_{jt}, \omega_{jt}, \omega_{jt-1}). \tag{55}$$

Since there are the two unobservables $(\omega_{jt}, \omega_{jt-1})$ the investment equation cannot be inverted to obtain ω_{jt} as a function of observables, and the argument underlying the first stage of the OP process is no longer valid.

One possible solution to the estimation problem is through a second observed control of the firm. Suppose, for example, one observes firms' expenditures on another investment (advertising, expenditure on a distributor or repair network), say s_{it}[47]. Then we have the bivariate policy function

$$\begin{pmatrix} i_{jt} \\ s_{jt} \end{pmatrix} = \Upsilon_t(k_{jt}, a_{jt}, \omega_{jt}, \omega_{jt-1}).$$

If the bivariate function $\Upsilon_t \equiv (\Upsilon_{1,t}, \Upsilon_{2,t})$ is a bijection in $(\omega_{jt}, \omega_{jt-1})$ (i.e. it is onto),

[47] One can modify this argument to allow also for a second state variable, the stock of advertising or the size of the repair network, provided that stock is known up to a parameter to be estimated.

then it can be inverted in ω_{jt} to obtain:

$$\omega_{jt} = \Upsilon_t^{-1}(k_{jt}, a_{jt}, i_{jt}, s_{jt}).$$

Given this assumption the first stage proceeds as in OP, except with a higher dimensional non-parametric function to account for current productivity (it is a function of s_{jt} as well as (k_{jt}, a_{jt}, i_{jt})).

OP's second stage is modified to be

$$\widetilde{y}_{jt} = \beta_k k_{jt} + \beta_a a_{jt} + \widetilde{g}(\widehat{\phi}_{jt-1} - \beta_k k_{jt-1} - \beta_a a_{jt-1}, \widehat{\phi}_{jt-2} - \beta_k k_{jt-2} - \beta_a a_{jt-2}) + \xi_{jt} + \eta_{jt},$$

where $\widetilde{y}_{jt} = y_{jt} - \hat{\beta} l_{jt}$ and the $\widehat{\phi}_{jt}$ variables are obtained from the first stage estimates at $t-1$ and $t-2$. Note that since the conditional expectation of ω_{jt} given I_{jt-1} now depends on ω_{jt-2} as well as ω_{jt-1}, we need to use estimates of ϕ from two prior periods. The extension to control for selection as well is straightforward. Moreover, provided the number of observed control variables is at least equal to the order of the Markov process, higher order Markov processes can be handled in the same way.

We now move on to allow investment to depend on an unobservable demand shock that varies across firms, in addition to the (now first-order) ω_{jt} process. Suppose that the demand shock, μ_{jt}, also follows a first-order Markov process that is *independent* of the ω_{jt} process. Then the investment function will be a function of both unobservables, or $i_{jt} = i_t(k_{jt}, a_{jt}, \omega_{jt}, \mu_{jt})$. Again we will assume the existence of a second control and use it to allow us to substitute for ω_{jt} in the first stage of OP's procedure.

More precisely, assume we also observe the firms' pricing decisions, p_{jt}. At the risk of some notational confusion, again let the bivariate policy function determining (i_{jt}, p_{jt}) be labelled $\Upsilon(\cdot)$, and assume it is a bijection in (ω_{jt}, μ_{jt}) conditional on (k_{jt}, a_{jt}). Then it can be inverted to form

$$\omega_{jt} = \Upsilon_t^{-1}(k_{jt}, a_{jt}, i_{jt}, p_{jt}), \tag{56}$$

and one can proceed with the first stage of estimation as above.

For the second stage observe that since the μ_{jt} process is independent of the ω_{jt} process the firm's conditional expectation of ω_{jt} given I_{jt-1} only depends on ω_{jt-1}. Thus, the second stage is

$$\widetilde{y}_{jt} = \beta_k k_{jt} + \beta_a a_{jt} + \widetilde{g}(\widehat{\phi}_{jt-1} - \beta_k k_{jt-1} - \beta_a a_{jt-1}) + \xi_{jt} + \eta_{jt}. \tag{57}$$

Note that the demand shock, if an important determinant of i_{jt}, may help with the precision of our estimates, as it generates independent variance in $\widehat{\phi}$.

The estimation problem becomes more complicated if, for some reason, the two Markov processes are dependent. The problem is that in this case, the firm's conditional expectation of ω_{jt} given I_{jt-1} depends on *both* ω_{jt-1} and μ_{jt-1}. Then equation (57) will have to be amended to allow $\widetilde{g}(\cdot)$ to also depend on μ_{jt-1}. If we let

$$\mu_{jt-1} = \Upsilon^{-1}_{2,t-1}(k_{jt-1}, a_{jt-1}, i_{jt-1}, p_{jt-1}), \tag{58}$$

our second stage can then be written as

$$\widetilde{y}_{jt} = \beta_k k_{jt} + \beta_a a_{jt} + \widetilde{g}(\omega_{jt-1}, \mu_{jt-1}) + \xi_{jt} + \eta_{jt} \tag{59}$$

$$= \beta_k k_{jt} + \beta_a a_{jt} + \widetilde{g}(\phi_{jt-1} - \beta_k k_{jt-1} - \beta_a a_{jt-1}, \Upsilon^{-1}_{2,t-1}(k_{jt-1}, a_{jt-1}, i_{jt-1}, p_{jt-1})) + \xi_{jt} + \eta_{jt}.$$

Unfortunately, this equation cannot identify β_k and β_a since it requires us to condition on a nonparametric function of $(k_{jt-1}, i_{jt-1}, a_{jt-1})$. More formally, since $k_{jt} = (1-\delta)k_{jt-1} + i_{jt-1}$ (and $a_{jt} = a_{jt-1} + 1$), there is no remaining independent variance in (k_{jt}, a_{jt}) to be used to identify β_k and β_a.

To avoid this problem, we need an explicit ability to solve for or estimate μ_{jt-1}. This would generally require demand side data. For example the Berry, Levinsohn, and Pakes (1995) demand estimation procedure produces estimates of a set of "unobserved product characteristics" which might be used as the μ_{jt}'s. Of course once one brings in the demand side, there is other information that can often be brought to bear on the problem. For example, the production function estimates should imply estimates of marginal cost which, together with the demand system, would actually determine prices in a "static" Nash pricing equilibrium (see the first section of this chapter). We do not pursue this further here.

Finally we move to the case where there are two sources of productivity growth, one evolving as a controlled Markov process, and one as an exogenous Markov process. In this case the production function is written as

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + \beta_l l_{jt} + \omega^1_{jt} + \omega^2_{jt} + \eta_{jt},$$

where ω^1_{jt} is the controlled, and ω^2_{jt} is the exogenous, first-order Markov process.

Assume now that we have data on both *R&D* expenditures, say r_t, which is the input of the controlled process, and a "technology indicator" or T_t (like patents, or licensing fees) which is an output of the controlled process. As above, assume the policy functions for physical and R&D investment are a bijection, so we can write

$$\begin{aligned} \omega^1_{jt} &= \Upsilon^{-1}_{1t}(k_{jt}, a_{jt}, i_{jt}, r_{jt}) \\ \omega^2_{jt} &= \Upsilon^{-1}_{2t}(k_{jt}, a_{jt}, i_{jt}, r_{jt}). \end{aligned} \tag{60}$$

Now the first stage consists of using the technology indicator to isolate ω_{jt}^1. In other words, we assume

$$T_{jt} = \omega_{jt}^1 \theta + \eta_{2jt}, \tag{61}$$

where $\eta_{2,t}$ is mean independent of all the controls. We then substitute a nonparameteric function of $(k_{jt}, a_{jt}, i_{jt}, r_{jt})$ for ω_{jt}^1 in equation (61). This provides us with an estimate of $\omega_{jt}^1\theta$, say $\hat{\Upsilon}_{1tj}^{-1}$.

Our second stage mimics the first stage of OP except we treat $\hat{\Upsilon}_{1tj}^{-1}$ as an input. That is, we estimate

$$y_{jt} = \beta_l l_{jt} + \theta^{-1}\hat{\Upsilon}_{1tj}^{-1} + \phi(k_{jt}, a_{jt}, i_{jt}, r_{jt}) + \eta_{1jt}, \tag{62}$$

where

$$\phi(k_{jt}, a_{jt}, i_{jt}, r_{jt}) \equiv \beta_k k_{jt} + \beta_a a_{jt} + \omega_{jt}^2.$$

Then, without a selection correction, the third stage becomes

$$\tilde{y}_{jt} = \beta_a a_{jt} + \beta_k k_{jt} + \tilde{g}(\phi_{jt-1} - \beta_k k_{jt-1} - \beta_a a_{jt-1}, \hat{\Upsilon}_{1tj}^{-1}) + \xi_{jt} + \eta_{1jt},$$

Once again, we can modify this to allow for selection by using the propensity score as an additional determinant of $\tilde{g}(\cdot)$.

Buettner (2004b) explores a related extension to OP. While he only allows one unobserved state variable, he does allow the distribution of ω_{jt} to evolve endogenously over time, i.e. firms invest in R&D and these investments affect the distribution of ω_{jt} (conditional on ω_{jt-1}).[48] Unlike the above, Buettner does not assume that a "technology indicator" is observed. He develops a dynamic model with investments in R&D and physical capital that generates invertible policy functions such that the first stage of OP can be directly applied (and the labor coefficient can be estimated). However, second stage estimation is problematic, since the conditional expectation of ω_{jt} now depends on the full state vector through the choice of R&D. Furthermore, with the endogenous productivity process, he cannot rely on exogenous variation (such as changes in the economic environment over time) for identification. It remains to be seen whether this problem can be solved.

Greenstreet (2005) proposes and utilizes an alternative model/methodology that, while related to the above procedures, does not require the first stage inversion. This is a very

[48] Recall that "endogenous" evolution of ω_{jt} is problematic for IV approaches.

nice attribute since as a result, the procedure does not rely at all on the key scalar unobservable and monotonicity assumptions of the OP/LP/ACF procedures. Greenstreet achieves this by making a different assumption on firms' information sets. Specifically, instead of observing ω_{jt} and η_{jt} individually (after production at t), firms only ever observe the *sum* $\omega_{jt} + \eta_{jt}$. Because of this alternative informational assumption, the econometrician does not need the first-stage inversion to recreate the information set of the firms. While this does avoid the scalar unobservable and monotonicity assumptions, Greenstreet's approach still relies on similar timing assumptions, involves a slightly more complicated learning process than the above procedures (requiring Kalman filtering), and also generates some new initial conditions problems that require additional assumptions to solve.

2.5 Concluding Remark.

The increase in the availability of plant and/or firm level panels together with a desire to understand the efficiency implications of major environmental and policy changes has led to a renewed interest in productivity analysis. Most of this analysis is based on production function estimates, and the literature has found at least two empirical regularities. First, there are indeed large efficiency differences among firms and those differences are highly serially correlated. Second, at least in many environments, to obtain realistic production function estimates the researcher must account for the possibility of simultaneity and selection biases.

Put differently, to study either the changes in the allocative efficiency of production among firms of differing productivities, or the correlates of productivity growth within individual establishments, we first have to isolate the productivity variable itself. Since firms' responses to the changes in the environment being studied typically depend on how those changes impacted their productivity, movements in productivity can not be isolated from changes in input and exit choices without an explicit model of how those choices are made.

The appropriateness of different models of how these decisions are made will undoubtedly depend on the environment being studied. We have presented a number of alternatives, and discussed their properties. However this is an empirically driven subfield of estimation, and there are undoubtedly institutional settings where alternative frameworks might be better to use. It is not the precise framework that is important, but rather the fact that productivity studies must take explicit account of the fact that changes in productivity (or, if one prefers, sales for a given amount of inputs) in large part determine how firms respond to the changes being studied, and these must be taken into account in the estimation procedure.

3 Dynamic Estimation.

This chapter considers structural estimation of dynamic games. Despite a blossoming empirical literature on structural estimation of static equilibrium models, there has been relatively little empirical work to date on estimation of dynamic oligopoly problems. Four exceptions are Gowrisankaran and Town (1997), Benkard (2004), Jofre-Benet and Pesendorfer (2003), and Ryan (2005). The literature's focus on static settings came about not because dynamics were thought to be unimportant to market outcomes, but rather because empirical analysis of dynamic games was seen as too difficult. In particular, while some of the parameters needed to analyze dynamic games could be estimated without imposing the dynamic equilibrium conditions, some could not and, until very recently, the only available methods for estimating these remaining parameters were extremely burdensome, in terms of both computation time and researcher time.

This computational complexity resulted from the need to compute the continuation values to the dynamic game in order to estimate the model. The direct way of obtaining continuation values was to compute them as the fixed point to a functional equation, a high order computational problem. Parameter values were inferred from observed behavior by computing the fixed point that determines continuation values at different trial parameter values, and then searching for the parameter value that makes the behavior implied by the continuation values "as close as possible" to the observed behavior. This "nested fixed point" algorithm is extremely computationally burdensome because the continuation values need to be computed many times.

However, a recent literature in industrial organization (Aguirregabiria and Mira (2006), Bajari, Benkard, and Levin (2006), Jofre-Benet and Pesendorfer (2003), Pakes, Ostrovsky, and Berry (2006), and Pesendorfer and Schmidt-Dengler (2003)) has developed techniques that substantially reduce the computational and programming burdens of estimating dynamic games. This literature extends a basic idea that first appeared in the context of single agent problems in Hotz and Miller (1993). Hotz and Miller (1993) provided a set of assumptions under which one could obtain a nonparametric estimate of continuation values without ever computing the fixed point[49]. Rust (1994) suggests the extension of these ideas to the context of dynamic games. The recent literature in industrial organization has shown that, at least under a certain set of assumptions, these approaches can be extended to estimate continuation values in a wide variety of dynamic games, even in the presence of multiple equilibria.

[49]In related work Olley and Pakes (1996) use nonparametrics to get around the problem of computing the fixed point needed to obtain an agent's decision rule in a multiple agent framework; but they use the nonparametric estimates to control for unobservables and do not recover the implied estimates of continuation values.

This chapter summarizes the currently available techniques for estimating dynamic games, concentrating on this recent literature. The chapter proceeds as follows. We first outline the goals of the estimation procedure and consider what might be gained by modelling dynamics in an oligopoly situation. Then we present a general framework for dynamic oligopoly problems, with three simple examples from the recent literature. Next we overview existing estimation methods, providing details for the three examples. We conclude with a brief discussion of techniques available to ameliorate one (of many) outstanding problems; that of serially correlated unobserved state variables.

We note that there are at least two issues that appear in the literature and are not considered here. First we do not consider identification issues (at least not directly). Our feeling is that many of the parameters determining behavior in dynamic games can be estimated without ever computing an equilibrium, and those parameters that remain depend on the nature of the problem and data availability. Second, we do not consider "timing" games, such as those in Einav (2003) and in Schmidt-Dengler (2004). Our only excuse here is our focus on the evolution of market structure in oligopolies.

3.1 Why Are We Interested?

One contribution of the recent literature is that it provides a means of obtaining information about certain parameters that could not be obtained via other methods. For example, the sunk costs of entry and the sell-off values (or costs) associated with exit are key determinants in the dynamics of market adjustments to policy and environmental changes. Knowledge of the level of sunk costs is critical, for example, in a regulatory authority's decision of whether to approve a merger, or in the analysis of the likely impacts of changes in pension policy on shut down decisions. However, actual data on sunk costs are extremely rare. Besides being proprietary, and thus hard to access, sunk costs can also be very difficult to measure. Thus, in many cases the only option for learning the extent of sunk costs may be to infer them from equilibrium behavior using other variables that we can observe. Since sunk costs are only paid once upon entry, while firms may continue to operate for many periods, inferring the level of sunk costs from equilibrium behavior requires a dynamic framework. Similar arguments can be made regarding the parameters determining, among other diverse phenomena, the transaction costs of investments (including installment, delivery, and ordering costs), the costs of adjusting output rates or production mix, and the extent of learning-by-doing.

There are a number of other uses for techniques that enable us to empirically analyze dynamic games. For example, there are many industries in which an understanding of the nature of competition in prices (or quantities) requires a dynamic framework. In such cases, the empirical literature in industrial organization has often used static models to approximate behavior that the authors are well aware is inherently dynamic. For example, there has been much work on identifying and estimating the form of competition in markets (e.g., Bresnahan (1982, 1987), Lau (1982)). This literature typically compares a static Nash equilibrium with particular static "collusive" pricing schemes. In reality, the set of collusive pricing schemes that could be supported in equilibrium depends on the nature of the dynamic interactions (e.g., Abreu, Pearce and Stacchetti (1986), Green and Porter (1984), Rotemberg and Saloner (1985), Fershtman and Pakes (2000)). A related point is that static price or quantity setting models are known to be inappropriate when future costs depend directly on the quantity sold today, as in models with learning by doing or adjustment costs, and/or when future demand conditions depend on current quantities sold, as in models with durable goods, experience goods, the ability to hold inventory, and network externalities.

Similarly, most of the existing empirical literature on entry relies on two-period static models. While these models have proven very useful in organizing empirical facts, the two period game framework used makes little sense unless sunk costs are absent. Therefore, the results are not likely to be useful for the analysis of policy or environmental changes in a given market over time. This leaves us with an inability to analyze the dynamic implications of a host of policy issues, and there are many situations where dynamics may substantially alter the desirability of different policies. For example, Fershtman and Pakes (2000) show that because collusive behavior can help promote entry and investment, it can enhance consumer welfare. Similarly, a static analysis would typically suggest that mergers lower consumer welfare by increasing concentration, whereas a dynamic analysis might show that allowing mergers promotes entry, counterbalancing the static effects.

3.2 Framework.

This section outlines a framework for dynamic competition between oligopolistic competitors that encompasses many (but not all) applications in industrial organization. Examples that fit into the general framework include entry and exit decisions, dynamic pricing (network effects, learning-by-doing, or durable goods), dynamic auction games, collusion, and investments in capital stock, advertising, or research and development. The defining feature of the framework is that actions taken in a given period affect future payoffs, and future strategic interaction, by influencing only a set of *commonly observed* state variables. In particular we will assume that all agents have the same information to use in making their decisions, up to a set of disturbances that have only transitory effects on payoffs.

We use a discrete time infinite horizon model, so time is indexed by $t = 1, 2, ..., \infty$. At time t, prevailing conditions are summarized by a state, $\mathbf{s}_t \in S \subset \mathbb{R}^G$, that reflects aspects of the world relevant to the payoffs of the agents. Relevant state variables might include firms' production capacities, the characteristics of the products they produce, their technological progress up to time t, the current market shares, stocks of consumer loyalty, or simply the set of firms that are incumbent in the market. We assume that these state variables are commonly observed by the firms. Note that we have not yet specified which state variables are observed by the econometrician. This distinction will be made in the applications below.

Given the state $\mathbf{s}_t$ at date t, the firms simultaneously choose actions. Depending on the application, the firms' actions could include decisions about whether to enter or exit the market, investment or advertising levels, or choices about prices and quantities. Let $a_{it} \in A_i$ denote firm i's action at date t, and $\mathbf{a}_t = (a_{1t}, ..., a_{Nt})$ the vector of time t actions, where N_t is the number of incumbents in period t (entry and exit, and hence N_t, are endogenous in these models).

We also assume that before choosing its action each firm, i, observes a *private* shock $\nu_{it} \in \mathbb{R}$, drawn independently (both over time and across agents) from a distribution $G(\cdot|\mathbf{s}_t)$.[50] Private information might derive from variability in marginal costs of production that result, say, from machine breakdowns, or from the need for plant maintenance, or from variability in sunk costs of entry or exit. We let the vector of private shocks be $\nu_t = (\nu_{1t}, ..., \nu_{Nt})$.

In each period, each firm earns profits equal to $\pi_i(\mathbf{a}_t, \mathbf{s}_t, \nu_{it})$. Profits might include variable profits as well as any fixed or sunk costs, including the sunk cost of entry and the selloff value of the firm. Conditional on the current state, $\mathbf{s}_0$, and the current value of the firm's private shock, ν_{i0}, each firm is interested in maximizing its expected discounted sum of profits,

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t \pi_i(\mathbf{a}_t, \mathbf{s}_t, \nu_{it}) \,\middle|\, \mathbf{s}_0, \nu_{i0}\right], \tag{63}$$

where the expectation is taken over rival firms' actions in the current period as well as the future values of all state variables, the future values of the private shock, and all rivals' future actions. We assume firms have a common discount factor β.

The final aspect of the model is to specify the transitions between states. We assume that the state at date $t+1$, denoted $\mathbf{s}_{t+1}$, is drawn from a probability distribution $P(\mathbf{s}_{t+1}|\mathbf{s}_t, \mathbf{a}_t)$. The dependence of $P(\cdot|\mathbf{s}_t, \mathbf{a}_t)$ on the current period actions $\mathbf{a}_t$ reflects

[50] Here we assume that firm i's private shock is a single scalar variable. However, as will be seen in the examples below, there is no conceptual difficulty in allowing the shock to be multi-dimensional.

the fact that some time t decisions may affect future payoffs, as is clearly the case if the relevant decision being modelled is an entry/exit decision or a long-term investment. Of course, not all the state variables necessarily depend on past actions; for example, one component of the state could be a transitory *iid* shock that affects only the current payoffs, such as an *iid* shock to market demand.

Note that we have assumed that firms' private information does not influence state transitions directly (i.e., it only influences transitions through its impact on a_{it}). For example, incumbent firms care only about whether or not a potential entrant enters the market, and not what the entrant's sunk cost of entry was. On the other hand this assumption does rule out applications where firms' investment outcomes are their private information (e.g., Fershtman and Pakes (2005)).

We are interested in equilibrium behavior. Because the firms interact repeatedly and the horizon is infinite, there are likely to be many Nash, and even subgame perfect equilibria, possibly involving complex behavioral rules. For this reason, we focus on pure strategy Markov perfect equilibria (MPE).

In our context a Markov strategy for firm i describes the firm's behavior at time t as a function of the commonly observed state variables and firm i's private information at time t. Formally, it is a map, $\sigma_i : S \times \mathbb{R} \to A_i$. A profile of Markov strategies is a vector, $\sigma = (\sigma_1, ..., \sigma_n)$, where $\sigma : S \times \mathbb{R}^n \to A$. A Markov strategy profile, σ, is a MPE if there is no firm, i, and alternative Markov strategy, σ_i', such that firm i prefers the strategy σ_i' to the strategy σ_i given its opponents use the strategy profile σ_{-i}. That is, σ is a MPE if for all firms, i, all states, $\mathbf{s}$, and all Markov strategies, σ_i',

$$V_i(\mathbf{s}, \nu_i|\sigma_i, \sigma_{-i}) \geq V_i(\mathbf{s}, \nu_i|\sigma_i', \sigma_{-i}). \tag{64}$$

If behavior is given by a Markov profile σ, firm i's present discounted profits can be written in recursive form:

$$V_i(\mathbf{s}, \nu_i|\sigma) = \mathbb{E}_{\nu_{-i}}\left[\pi_i(\sigma(\mathbf{s}, \nu), \mathbf{s}, \nu_i) + \beta \int V_i(\mathbf{s}', \nu_i'|\sigma) dG(\nu_i'|\mathbf{s}') dP(\mathbf{s}'|\sigma(\mathbf{s}, \nu), \mathbf{s})\right]. \tag{65}$$

3.2.1 Some Preliminaries.

The framework above is a generalization of the Ericson and Pakes (1995) model. The existence proofs for that model that are available have incorporated additional assumptions to those listed above (see Gowrisankaran (1995), and Doraszelski and Satterthwaite (2004)). Typically, however, the algorithms available for computing an equilibrium do find an equilibrium even when the available sets of sufficient conditions for existence are not satisfied (i.e. the algorithm outputs policies and values that satisfy the fixed point conditions that define the equilibrium up to a precision determined by the researcher). There may, however, be more than one set of equilibrium policies (for an explicit example see Doraszelski and Satterwaite (2004)).

If the regularity conditions given in Ericson and Pakes (1995) are satisfied, each equilibrium generates a finite state Markov chain for the $\mathbf{s}_t$ process. That is, the vector of state variables can only take on a finite set of values, a set we will designate by $\mathcal{S}$, and the distribution of the future $\{\mathbf{s}_\tau\}_{\tau=t}^{\infty}$ conditional on all past history depends only on the current value of $\mathbf{s}_t$. Every sequence from this finite state Markov chain will, in finite time, wander into a subset of the states called a recurrent class or an $\mathcal{R} \subset \mathcal{S}$, and once in $\mathcal{R}$ will stay there forever. Every $s \in \mathcal{R}$ will be visited infinitely often[51].

Throughout we assume that agents' perceptions of the likely future states of their competitors depend only on $\mathbf{s}_t$ (i.e., we assume that $\mathbf{s}_t$ is a complete description of the state variables observed by the firms). As detailed by Pakes, Ostrovsky, and Berry (2006), this implies that there is only one equilibrium policy for each agent that is consistent with the data generating process; at least for all $\mathbf{s}_t \in \mathcal{R}$. To see this it suffices to note that since we visit each point in $\mathcal{R}$ infinitely often, we will be able to consistently estimate the distribution of future states of each firm's competitors given any $\mathbf{s}_t \in \mathcal{R}$. Given that distribution, each agent's best response problem is a single agent problem. Put differently, since reaction functions are generically unique, once the agent knows the distribution of its competitors' actions, its optimal policy is well defined. Thus, given the data generating process, policies are well defined functions of the parameters and the state variables. Consequently, standard estimation algorithms can be used to recover them.[52]

Finally, in all of the examples below we will assume that the discount factor, β, is one of the parameters that is known to the econometrician. It is a straightforward extension to estimate the discount parameter. However, our focus here is on obtaining estimates of parameters that we have little other information on.

[51] Formally, the dynamics of the model are described by a Markov matrix. Each row of the matrix provides the probability of transiting from a given $\mathbf{s}$ to each possible value of $\mathbf{s} \in \mathcal{S}$. Ericson and Pakes (1995) also provide conditions that imply that the Markov matrix is ergodic, that is there is only one possible $\mathcal{R}$.

[52] Note that if our data consists of a panel of markets, this implicitly assumes that, conditional on s_t, the policy rule (our σ) in one market is the same as in the other.

3.2.2 Examples.

The framework above is general enough to cover a wide variety of economic models. We provide three examples below. In general the objects that need to be recovered in the estimation are the period profit function, $\pi(\cdot)$, the transition probabilities, $P(\mathbf{s}_{t+1}|\mathbf{s}_t, \mathbf{a}_t)$, and the distribution of the private shocks, $G(\cdot|\mathbf{s})$.

Example 1: A Simple Model of Entry/Exit.

This example is based on Pakes, Ostrovsky, and Berry (2006). Let the state variables of the model be given by a pair, $s_t = (n_t, z_t)$, where n_t denotes the number of firms active at the beginning of each period, and z_t is a vector of profit shifters that evolve exogenously as a finite state. In the model, operating profits are determined solely by these variables. In any period, t, in which a firm is active it earns profits equal to

$$\tilde{\pi}(n_t, z_t; \theta).$$

The model focuses on entry and exit. In each period, each incumbent firm receives a random draw, denoted ϕ_{it}, determining the selloff value of the firm. The selloff values are assumed to be private information. However, their distribution is commonly known to the agents. The firm chooses to exit if the selloff value of the firm is greater than the expected discounted value of continuing in the market. Otherwise, the firm continues in the market.

Entry is described similarly. For ease of exposition, we assume that there are E potential entrants each period, where E is known to the agents.[53] Each period, each potential entrant firm receives a random draw, denoted κ_{it}, determining its sunk cost of entry. As above, the entry cost is private information, but its distribution is commonly known. The firm enters the market if the expected discounted value of entering is greater than the entry cost. Otherwise, the entrant stays out of the market and earns nothing.

To see how this model fits into the general framework, let $\chi_{it} = 1$ for any firm i that is active in the market in period t, and $\chi_{it} = 0$ otherwise. Note that we assume that when an incumbent firm exits $\chi_{it} = 0$ thereafter. In that case the period profit function is,

$$\pi_i(\mathbf{a}_t, \mathbf{s}_t, \nu_{it}) = \{\chi_{it} = 1\}\tilde{\pi}(n_t, z_t; \theta) + (\chi_{it} - \chi_{i,t-1})^{-}\phi_{it} - (\chi_{it} - \chi_{i,t-1})^{+}\kappa_{it},$$

[53]The extension to a random number of entrants is straightforward. See Pakes, Ostrovsky, and Berry (2006) for details.

where the notation $\{\chi_{it} = 1\}$ denotes an indicator function that is one if the firm is active and zero otherwise, the notation $f^+ \equiv \{f > 0\}f$, for any function f, and similarly $f^- \equiv \{f < 0\}|f|$. On the right hand side, χ represents firms' actions, a; n and z represent the states, $\mathbf{s}$; and ϕ and κ represent the private shocks, ν.

Note that while this model does not allow for observed heterogeneity among incumbent firms, this can be achieved by allowing for multiple entry locations. We consider this extension below. Note further that this model is a special case of the Ericson and Pakes (1995) model in which investment is not modelled. We add investment back to the model in the next example.

Example 2: An Investment Game with Entry and Exit.

This example is a straightforward extension of the Ericson and Pakes (1995) model due to Bajari, Benkard, and Levin (2006). Similarly to the above example, there are a set of incumbent firms competing in a market. Firms are heterogeneous, with differences across firms described by their state variables, s_{it}, which are commonly known. For ease of exposition, we will omit any other exogenous profit shifters from the set of state variables.

Each period, firms choose investment levels, $I_{it} \geq 0$, so as to improve their state the next period. Investment outcomes are random, and each firm's investment affects only its own state so that there are no investment spillovers. Therefore, each firm's state variable, s_{it}, evolves according to a process $Pr(s_{i,t+1}|s_{it}, I_{it})$.

Here are some examples of models that are consistent with this framework.

(i) Firms' state variables could represent (one or more dimensions of) product quality, where investment stochastically improves product quality.

(ii) Firms' state variables could represent the fraction of consumers who are aware of the firm's product, where investment is a form of advertising that increases awareness (e.g. Doraszelski and Markovich, 2004).

(iii) Firms' state variables could represent capital stock, where investment increases a firm's capital stock.

Firms earn profits by competing in a spot market. Because quantity and price are assumed not to influence the evolution of the state variables, they are determined in static equilibrium conditional on the current state. In any period, t, in which a firm is active in the market it earns profits equal to

$$q_{it}(\mathbf{s}_t, \mathbf{p}_t; \theta_1)\,(p_{it} - \text{mc}(s_{it}, q_{it}; \theta_2)) - C(I_{it}, \nu_{it}; \theta_3), \tag{66}$$

where q_{it} is quantity produced by firm i in period t, $\mathbf{p}_t$ is the vector of prices, mc is the marginal cost of production, ν_{it} represents a private shock to the cost of investment, $\theta = (\theta_1, \theta_2, \theta_3)$ is a parameter vector to be estimated, and we have assumed that the spot market equilibrium is Nash in prices.

The model also allows for entry and exit. Each period, each incumbent firm has the option of exiting the market and receiving a scrap value, Φ, which is the same for all firms (this differs from the prior example in which there is a distribution of exit costs). There is also one potential entrant each period with a random entry cost, κ_{it}.[54] The entrant enters if the expected discounted value of entering exceeds the entry cost. As above, the entry cost is assumed to be private information, but its distribution is commonly known.

Relative to the general framework above, current period returns are given by

$$\begin{aligned}\pi_i(\mathbf{a}_t, \mathbf{s}_t, \nu_{it}) =& \{\chi_{it} = 1\}\left[q_{it}(\mathbf{s}_t, \mathbf{p}_t; \theta_1)\,(p_{it} - \text{mc}(s_{it}, q_{it}; \theta_2)) - C(I_{it}, \nu_{it}; \theta_3)\right] + \\ & (\chi_{it} - \chi_{i,t-1})^{-}\Phi - (\chi_{it} - \chi_{i,t-1})^{+}\kappa_{it}.\end{aligned}$$

On the right hand side, prices (p), investment (I), and entry/exit (χ) are the actions (a), while the private shocks are the shock to investment (ν_{it}) and the entry cost (κ_{it}).

Example 3: A Repeated Auction Game with Capacity Constraints.

This example is based on Jofre-Benet and Pesendorfer (2003). In this example, a set of incumbent contracting firms compete in monthly procurement auctions. The auctions are heterogeneous because the contracts that come available each month are of differing size and scope. The firms bidding on the contracts are also heterogeneous as each has a different cost of completing each contract. In a given month, each firm also has a different backlog of contracts, which might affect its ability to take on new contracts.

Let z_t be the characteristics of the contract to be auctioned in month t, including both the contract size (in dollars), and the number of months required to complete the contract. We assume that z_t evolves exogenously as a finite state. Let $\omega_{i,t}$ be the backlog of work for firm i in period t and $\omega_t = (\omega_{1,t}, ..., \omega_{N,t})$ be the vector of backlogs. A firm's backlog

[54] It is straightforward to generalize the model to have a random number of potential entrants each period.

of work represents the remaining size in dollars, and the remaining number of days left until completion of each contract previously won by the firm. It therefore evolves deterministically depending on the current auction outcome according to the map,

$$\omega_{t+1} = \Gamma(\omega_t, z_t, j),$$

where j is the winner of the time t auction and the map Γ is known. The state variables of the model are $\mathbf{s}_t = (\omega_t, z_t)$. All states are assumed to be common knowledge.

Each firm also has a different cost, c_{it}, for each contract that is private information to the firm. Bidders' costs are drawn independently from a distribution $G(c_{it}|\omega_{it}, \omega_{-it}, z_t)$ that is commonly known.

In each period, each firm views its cost for the contract being offered and then chooses a bid, b_{it}. Each firm earns current profits equal to

$$\pi_i(\mathbf{a}_t, \mathbf{s}_t, \nu_{it}) = (b_{it} - c_{it})\{b_{it} \leq \min_j(b_{jt})\} \tag{67}$$

where the indicator function takes the value one if firm i submits the lowest bid and hence wins the auction (assume there are no ties). On the right hand side the bids (b_{jt}) are the action variables (a_t) and the costs c_{it} are the private shocks (ν_{it}).

Note that the state variables do not directly enter current profits in this model. However, the state variables influence all firms' costs and hence a firm's optimal bid depends on the current state both through its own costs directly and through the firm's beliefs about the distribution of rivals' bids. For the same reason, expected profits are also a function of the current state.

Note also that an important distinction between the investment model above and this example is that, in this example, each firm's choice variable (in this case, its bid) affects the evolution of all firms' states. In the investment model above, a firm's investment affects only the evolution of its own state. This distinction is important because many I.O. models share this feature. For example, models of dynamic pricing (learning by doing, network effects, or durable goods) would have this feature when firms compete in prices (though not if firms compete in quantities). Such models can be placed in the general EP framework we have been using, but to do so we need to adjust that framework to allow the control that affects the distribution of current profits (bids, quantities, or prices) to also have an impact on distribution of future states; see the discussion in Pakes (1998). We note that to our knowledge Jofre-Benet and Pesendorfer (2003) were the first to show that a two-step estimation approach was feasible in a dynamic game.

3.3 Alternative Estimation Approaches.

In order to conduct policy analysis in any of the economic models above, it is typically necessary to know all of the parameters of the model, including the profit function, the transition probabilities, and the distribution of the exogenous shocks. Often many of the parameters can be estimated "off line", that is, without needing to compute equilibria of the dynamic game. At one extreme here is Benkard's (2004) analysis of the commercial aircraft industry. He was able to obtain a large amount of cost data on sunk as well as marginal costs which, together with generally available information on demand, enabled him to estimate all the parameters he needed off line. Given these parameters he could focus on computing the dynamic implications of alternative policies.

However, such an approach is rarely possible. More typically, at least cost data are unavailable, either because they are proprietary and hence difficult for researchers to access, or because they are hard to measure. In static settings we often solve the problem of a lack of cost data by inferring marginal costs from their implications in an equilibrium pricing equation. A similar approach can be taken in this dynamic setting. However, characterizing the relationship between the data generating process and equilibrium play in the models above is complicated by the fact that the model involves repeated interactions.

Observed behavior in the model represents the solution to a maximization problem that involves both the profit function, which typically has a known parametric form, and the value function, which results from equilibrium play and therefore has unknown form. For example, the value of entering a market depends both on current profits, and expected future profits, which in turn depend on future entry and exit behavior. In order to describe the data generating process, then, we need the ability to compute the equilibrium continuation values.

Thus, conceptually, estimation of dynamic models can be separated into two main parts. The first part involves obtaining the continuation values for a given parameter value, θ. The second part is to use the continuation values obtained in the first part to maximize an objective function in the parameters, θ. Note that the continuation values must be obtained for many different values of θ in order to perform this maximization, and thus the first part is the source of most of the computational burden of the estimation. The key differences in estimation approaches lie in the details of how each of these two parts is performed.

3.3.1 The Nested Fixed Point Approach.

The nested fixed point approach is a logical extension of the method of Rust (1987) to games. The general idea is as follows:

1. Given a parameter vector, θ, compute an equilibrium to the game, $V(\mathbf{s};\theta)$, numerically, using the computer.
2. Use the computed values, $V(\mathbf{s};\theta)$, to evaluate an objective function based on the sample data.
3. Nest steps one and two in a search routine that finds the value of θ that maximizes the objective function.

A framework capable of computing equilibria to models like those above has existed for some time (Pakes and McGuire (1994)), and recent papers by Pakes and McGuire (2001), Doraszelski and Judd (2004), and Weintraub, Benkard, and Van Roy (2005) enable significant improvements in computational times, at least in some problems (for a discussion of these, and other alternatives, see Doraszelski and Pakes (2006)). All of these algorithms rely on similarities between the dynamic framework above and dynamic programming problems. The general idea of these algorithms is to start with an initial guess at the value function, $V^0(\mathbf{s};\theta)$, and substitute that into the right hand side of the Bellman equation (equation (65)). Then, at each state point and for each firm, solve the maximization equation on the right hand side of (65) yielding a new estimate of the value function, $V^1(\mathbf{s};\theta)$. This procedure is iterated until convergence is achieved, so that the new and old value functions are the same. Unlike single agent problems, in the context of a game, convergence of the algorithm is not guaranteed (the mapping is not a contraction) and, indeed, initial iterations will often seem to move away from equilibrium. However, in practice the algorithms typically converge and, once they do, the value functions obtained must represent an equilibrium.

An important feature of the nested fixed point algorithm is that the first step is performed without using any data. As a result, the value functions are obtained precisely; that is, they contain no sampling error. This lack of sampling error makes the second part of the algorithm, in which the parameters are estimated, straightforward.

On the other hand the algorithm is computationally burdensome. For models rich enough to use in empirical work, it is often difficult to compute an equilibrium even once, and in the nested fixed point algorithm it is necessary to compute an equilibrium once for each iteration of the maximization routine; implying that up to hundreds, if not thousands, of fixed points must be calculated. Moreover, setting up an efficient algorithm often requires a large amount of complex computer programming, creating a relatively large burden on researcher time. As a result there are very few examples in the literature where the nested fixed point algorithm has been applied to estimate parameters.

One exception is Gowrisankaran and Town (1997), who use a nested fixed point approach to apply a model similar to the investment model above to data for the hospital industry. In each iteration of the estimation they compute an equilibrium using the algorithm of Pakes and McGuire (1994). They then estimate the model using a GMM objective function that matches cross-sectional moments such as average revenue per hospital, average expenditures per hospital, average investment per hospital, and average number of hospitals of each type (nonprofit and for-profit) per market. The nested fixed point approach was feasible in their application because their model was parsimonious and there were never more than three hospitals in any market in the data.

Another difficulty with the nested fixed point algorithm arises from the fact that dynamic oligopoly models can admit more than one equilibria. While the assumptions given above in principle allow the researcher to use the data to pick out the correct equilibrium, actually achieving this selection using the nested fixed point algorithm is likely to be difficult. Moreover, equilibrium selection has to take place for every candidate value of the parameters to be estimated. Alternative sets of assumptions could be used to select different equilibria, but unless we were willing to assume "a priori" that equilibrium was unique, somehow we must investigate the issue of the relationship between the equilibrium computed in the algorithm, and that observed in the data.

3.3.2 Two-Step Approaches.

The biggest obstacle to implementing the nested fixed point algorithm in practice is the heavy computational burden that results from the need to compute equilibria for each trial parameter value. Fortunately, the recent literature (Aguirregabiria and Mira (2006), Bajari, Benkard, and Levin (2006), Jofre-Benet and Pesendorfer (2003), Pakes, Ostrovsky, and Berry (2006), and Pesendorfer and Schmidt-Dengler (2003)) has derived methods for estimating dynamic oligopoly models that impose the conditions of a dynamic equilibrium without requiring the ability to compute an equilibrium. The new literature sidesteps the equilibrium computation step by substituting nonparametric functions of the data for the continuation values in the game. These nonparametric estimates are in general much easier to compute than the fixed point calculations in the nested fixed point algorithm. As a result, these methods have substantially lower computational burden.

Below we outline five different two-step methods of estimating dynamic games. The overall approach is similar throughout, but each method does both the first and second steps of the estimation differently. To our knowledge, Hotz and Miller (1993) were the first to show that it was possible to estimate the continuation values in a dynamic programming problem nonparametrically instead of computing them. In a single agent dynamic discrete choice problem, Hotz and Miller showed that the agent's dynamic choice problem mimics a static discrete choice problem with the value functions replacing the mean utilities. Thus, the agent's continuation values can be obtained nonparametrically by first estimating the agent's choice probabilities at each state, and then inverting the choice problem to obtain the corresponding continuation values. This inversion is identical to the one commonly used in discrete choice demand estimation to obtain the mean utilities.

We begin our discussion of estimation by showing that if the game has only discrete actions, and there is one unobserved shock per action for each agent in the game, then under the information structure given in the general framework above, estimators very similar to those of Hotz and Miller (1993) can still be used (see also Aguirregabiria and Mira (2006)). Sticking with the single agent framework, Hotz et. al (1994) use estimated probabilities to simulate sample paths. They then calculate the discounted value of utility along these paths, average those values for the paths emanating from a given state, and use these averages as the continuation values at that state. The Bajari, Benkard, and Levin (2006) paper discussed below shows that related ideas can be used to incorporate continuous controls into estimation strategies for dynamic games.

Pakes, Ostrovsky, and Berry (2006) also consider dynamic discrete games but, instead of inverting the choice problem, they estimate the continuation values directly by computing (approximately) the average of the discounted values of future net cash flows that agents starting at a particular state actually earned in the data (at least up to the parameter vector of interest). Econometrically, they use a nonparametric estimate of the Markov transition probabilities that determine the evolution of the state of the system to form an analytic estimate of the probability weighted average of the discounted returns earned from different states. Given equilibrium play, these averages will converge to the true expected discounted value of future net cash flow, that is of the continuation values we are after.

Bajari, Benkard, and Levin (2006) instead begin by projecting the observed actions on the state variables to compute nonparametric estimates of the policy functions of each agent at each state. Then they use the estimated policies to simulate out the discounted values of future net cash flows. This procedure is computationally light even in models with large state spaces and is easily applied to models with continuous controls, such as investment, quantity, or price (including models with both discrete and continuous controls like the investment game above). Given equilibrium play, the continuation values obtained in this fashion will be consistent estimates of the continuation values actually perceived by the agents.

Berry and Pakes (2002) provide an alternative approach for estimating models with continuous controls that is likely to be useful when the dynamic environment is complex, but sales and investment data are available. They assume that current period net returns are observable up to a parameter vector to be estimated, but do not require that the state variables of the model be observed, or even specified (so it would not be possible to estimate policy functions conditional on those state variables as in Bajari, Benkard and Levin). They derive an estimating equation from the first order condition for the continuous control (investment in our example) by substituting observed profit streams for the expected profit streams, and noting that the difference must be orthogonal to information known at the time investment decisions are made.

Jofre-Benet and Pesendorfer (2003) provide an estimator for the dynamic auction model. They show that it is possible to derive an expression for the equilibrium continuation values in the auction game that involves only the bid distributions. Since bids are observed, the bid distributions can be recovered nonparametrically from the data and then substituted into these expressions. Provided that agents are bidding close to optimally, the continuation values obtained from this procedure will be consistent estimates of the continuation values perceived by the agents.

In many of the cases we consider several of the methods could be used in estimation. In these cases it is not currently known how the methods compare to one another on such dimensions as computational burden and econometric efficiency. Hybrid methods are also possible in which features of two or more of the approaches could be combined. We expect these issues to be sorted out in the future.

Finally, there are also some costs associated with the two-step approaches. First, because the continuation values are estimated rather than computed, they contain sampling error. This sampling error may be significant because these models often have state spaces that are large relative to the available data. As we will see below, this influences the properties of the second step estimators in important ways. To summarize, the choice of second stage estimation method will be influenced as much or more by a need to minimize small sample bias caused by error in the continuation value estimates as it is by the desire to obtain asymptotic efficiency.

Second, for the two step approaches to produce estimators with desirable properties the data must visit a subset of the points repeatedly. Formally the requirement for the limit properties of the estimators is that all states in some recurrent class $\mathcal{R} \subseteq \mathcal{S}$ be visited infinitely often. Moreover equilibrium strategies must be the same every time each point in $\mathcal{R}$ is visited. Whether or not this assumption is reasonable for the problem at hand depends on the nature of the available data and the institutional setting which generated it. If the data consists of a time series on one market then we would require stationarity of the process over time. There are different ways to fulfill this requirement in panels (i.e. when we follow a cross section of markets over time); one possibility is that the initial state in each market is a random draw from a long run ergodic distribution. Note that the nested fixed point approach has a weaker data requirement.

These costs must be weighed against the benefit that the two-step estimators eliminate most of the computational burden of the nested fixed point approach. Indeed, the entire two-step algorithm might well have less computational burden than one iteration of the nested fixed point algorithm.

3.4 A Starting Point: Hotz and Miller.

Because of the similarity of this section to the previous literature on single agent problems, we will keep this section short, concentrating mainly on extending Hotz and Miller to games. For more detail on the approach in single agent problems see Hotz and Miller (1993), Hotz et al. (1994), Magnac and Thesmar (2002), and Rust (1994). See also Aguirregabiria and Mira (2006) and Pesendorfer and Schmidt-Dengler (2003) for a discussion in the context of entry games.

The idea behind Hotz and Miller's estimation method for single agent problems is to set up a dynamic discrete choice problem such that it resembles a standard static discrete choice problem, with value functions taking the place of standard utility functions. This allows a two step approach in which a discrete choice model is used as a first step for recovering the value functions, and the parameters of the profit function are recovered in a second step once the value functions are known.

We make two simplifying assumptions that will assist in the exposition. First, we suppose that agents' current profits do not depend on rivals' actions (though they do depend on rival's states whose evolution depends on those actions). Second, we assume that the unobserved shocks are additive to profits. In that case, current profits are given by,

$$\pi_i(\mathbf{a}_t, \mathbf{s}_t, \nu_{it}) = \tilde{\pi}(a_{it}, \mathbf{s}_t) + \nu_{it}(a_{it}),$$

where ν_{it} is agent i's vector of profitability shocks and $\nu_{it}(a_{it})$ is the shock associated with agent i's action a_{it}.

The first assumption simplifies the agents' choice problem because, if agents' current profits depend on rivals' actions then, since rivals' actions depend on their own current shocks, in its own maximization problem each agent would have to integrate current profits over all rivals' current actions. This would not change the overall approach but would complicate the computations below (we would need to integrate over distributions of competitors outcomes to compute the expected profits; see Aguirregabiria and Mira, 2006, for a model in which profits do depend on rivals' actions). The second simplification, additive separability in the private shocks, is also not strictly required. If the error terms entered profits nonlinearly then we could rewrite the problem in terms of expected profits and an additively separable projection error and work with that framework instead. However, such an approach does have the unattractive property that it changes the interpretation of the profit function. Thus, it is typically the case that in practice people assume that the profit function has additive structural error terms.

With these assumptions the Bellman equation can be simplified to (suppressing the subscripts)

$$V(\mathbf{s}, \nu) = \max_a \left\{ \tilde{\pi}(a, \mathbf{s}) + \nu(a) + \beta \int V(\mathbf{s}', \nu') dG(\nu'|\mathbf{s}') dP(\mathbf{s}'|\mathbf{s}, a) \right\}. \tag{68}$$

Equation (68) represents a discrete choice problem in which the mean utilities are given by,

$$v_a(\mathbf{s}) = \tilde{\pi}(a, \mathbf{s}) + \beta \int V(\mathbf{s}', \nu') dG(\nu'|\mathbf{s}') dP(\mathbf{s}'|\mathbf{s}, a). \tag{69}$$

Thus, since the private shocks are independent across time and across agents, the choice probabilities for a given agent can be generated in the usual manner of a static discrete choice problem

$$Pr(a|\mathbf{s}) = Pr(v_a(\mathbf{s}) + \nu(a) \geq v_{a'}(\mathbf{s}) + \nu(a'), \forall a'). \tag{70}$$

Assuming that the data consists of a large sample of observations on states and actions, the probability of each action at each state, $Pr(a|\mathbf{s})$, can be recovered from the data. In that case, the left hand side of (70) is known, at least asymptotically. Let $P(\mathbf{s})$ be the vector of choice probabilities for all feasible actions. Hotz and Miller show that for any distribution of the private shocks there is always a transformation of the choice probabilities such that

$$v_a(\mathbf{s}) - v_1(\mathbf{s}) = Q_a(\mathbf{s}, P(\mathbf{s})). \tag{71}$$

That is, the differences in the choice specific value functions can be written as a function of the current state and the vector of choice probabilities. The transformation on the right hand side is the same as the inversion used in the discrete choice demand estimation literature. Berry (1994) proves that the solution is unique. Berry, Levinsohn, and Pakes (1995) provide a transformation from the data to the mean utilities which is a contraction, and hence enables the researcher to actually compute the mean utilities (for more discussion see the first part of this chapter).

In general, this transformation can be used to recover the normalized choice specific value functions, $v_a - v_1$, at each state, using the estimated choice probabilities. If the distribution of the private shocks is known, the mapping does not depend on any unknown parameters. For example, in the case of the logit,

$$Q_a(\mathbf{s}, P(\mathbf{s})) = \ln(Pr(a|\mathbf{s})) - \ln(Pr(a_1|\mathbf{s})). \tag{72}$$

However, in general the mapping may be a function of unknown parameters of the distribution of the private shocks.

Note that, as in static discrete choice models, only the value differences can be recovered nonparametrically. Thus, some further information is required to obtain the values themselves. This difficulty is not just a feature of this particular estimation approach, but comes from the underlying structure of the discrete choice framework, in which only utility differences are identified from the observed choices. One consequence of this is that, even if the discount factor and the distribution of private shocks are completely known, the profit function can not be recovered nonparametrically (see Magnac and Thesmar (2002) for a detailed proof and analysis for single agent dynamic discrete choice problems, and Pesendorfer and Schmidt-Dengler (2003) for results extended to dynamic discrete games). This feature is inherent to the dynamic discrete choice setup and carries through to the context of a dynamic discrete game. As noted earlier our feeling is that the appropriate resolution of identification issues, such as this one, is context specific and will not be discussed here.

To obtain the continuation values from the choice specific values we can use the fact that,

$$V(\mathbf{s}, \nu) = \max_a \{v_a(s) + \nu(a)\}. \tag{73}$$

Because the continuation values are obtained by inverting from the observed choice probabilities, the structure of the profit function has not yet been imposed on them, and they are not yet functions of the profit function parameters. In order to estimate the profit function parameters, Hotz and Miller iterate the Bellman equation once, inserting the estimated continuation values on the right hand side,

$$\widehat{V}(\mathbf{s}; \theta) = \int \max_a \left\{ \tilde{\pi}(a, \mathbf{s}; \theta) + \nu(a) + \beta \int \widehat{V}(\mathbf{s}', \nu') dG(\nu'|\mathbf{s}') dP(\mathbf{s}'|\mathbf{s}, a) \right\} dG(\nu|\mathbf{s}), \tag{74}$$

Note that for some distributions such as those of type GEV the integral on the right hand side has an analytic form. In other cases it can be simulated.

These new estimates of the continuation values contain the profit function parameters (θ) and can be used in an estimation algorithm to estimate θ. The way this is typically done is to compute new predicted choice probabilities, (70), based on the new continuation value estimates, $\widehat{V}(\mathbf{s};\theta)$. Then, these choice probabilities can be used to construct either a pseudo-likelihood or some other GMM objective function that matches the model's predictions to the observed choices.

As noted above, the nonparametric estimates of the continuation values and transition probabilities on the right hand side of (74) introduce estimation error into the second stage objective function nonlinearly. Hotz and Miller show that if this estimation error disappears quickly enough then the estimator obtained is consistent and asymptotically normal. However, there are other methods that may be preferable in this context to a pseudo likelihood. Because of the nonlinearity of the pseudo likelihood in the continuation values, estimation error in the continuation values causes increased small sample bias in the parameter estimates obtained using this method. We discuss methods that at least partially address this problem in the next section.

3.5 Dynamic Discrete Games: Entry and Exit.

In this section we consider estimation of the entry/exit game in example one using the methods of Pakes, Ostrovsky, and Berry (2006) (henceforth, POB). We maintain the assumption that all of the state variables, (n_t, z_t), are observed and that the number of entrants (e_t) and exitors (x_t) are also observed. Entry and exit costs are assumed not to be observed and are the objects of interest in the estimation. We discuss the possibilities for estimation when there are one or more unobserved state variables in section 3.8.1.

Consider first exit behavior. Redefining the value function from the start of a period, prior to the point at which the private scrap value is observed, the Bellman equation for incumbent firms is given by (t subscript suppressed),

$$V(n,z;\theta) = \tilde{\pi}(n,z;\theta) + \beta\mathbb{E}_{\phi}\left[\max\{\phi_i, VC(n,z;\theta)\}\right] \tag{75}$$

where VC denotes the continuation value of the firm, which equals

$$VC(n,z;\theta) \equiv \sum_{z',e,x} V(n+e-x, z';\theta)P(e,x|n,z,\chi=1)P(z'|z). \tag{76}$$

In the above equation, e and x denote the number of entering and exiting firms, and $P(e,x|n,z,\chi=1)$ denotes the incumbent's beliefs about the likely number of entrants and exitors starting from state (n, z) *conditional* on the incumbent itself continuing $(\chi = 1)$.

If the equilibrium continuation values, $VC(n, z; \theta)$, were known, then it would be straightforward to construct a likelihood function since the probability of exit is given by

$$Pr(i \text{ exits}|n, z; \theta) = Pr(\phi_i > VC(n, z; \theta)), \tag{77}$$

and is independent across firms. Thus, we need to find a simple way to construct the equilibrium continuation values using observed play.

The continuation values represent the expected discounted value of future profits conditional on the incumbent continuing. They are a function of the profit function, $\tilde{\pi}(n, z; \theta)$, which determines future profits at each state (n, z), and the processes determining the evolution of the state variables, n and z. The profit function is known up to the parameters, θ. Therefore, in order to construct the continuation values as a function of the parameters, we need only estimate the evolution of the number of firms, which is determined by entry and exit, and the evolution of the profit shifters, $P(z'|z)$. The easiest way to do this is to use their empirical counterparts. Starting from a certain state, to estimate the evolution of the number of firms we can use the actual evolution of the number of firms each time that state was observed in the data. Similarly, we can use the observed evolution of the profit shifters to estimate the process $P(z'|z)$. That way the estimated continuation values reflect, approximately, the actual profits of firms that were observed in the data. The next subsection outlines this process in detail.

3.5.1 Step 1: Estimating Continuation Values.

To facilitate estimation of the continuation values, it helps to rewrite the Bellman equation in terms of the continuation values, VC,

$$VC(n, z; \theta) = \sum_{n', z'} \left[\tilde{\pi}(n', z'; \theta) + \beta \mathbb{E}_\phi \left[\max\{\phi_i, VC(n', z'; \theta)\} \right] \right] P(n'|n, z, \chi = 1) P(z'|z), \tag{78}$$

where to shorten the notation we let $n' \equiv n + e - x$.

Next, rewrite (78) in vector form. Let $VC(\theta)$ be the $\#\mathcal{S} \times 1$ vector representing $VC(n, z; \theta)$ for every (n, z) pair, and define $\tilde{\pi}(\theta)$ similarly. Also let M^i be the $\#\mathcal{S} \times \#\mathcal{S}$ matrix whose (i, j) element is given by $P(n_j|n_i, z_i, \chi = 1)P(z_j|z_i)$. This is the matrix whose rows give us the equilibrium transition probabilities from a particular (n, z) to each other possible (n, z). Note that if we were not conditioning on $\chi = 1$ an unbiased estimate of the rows of this matrix could be obtained by simply counting up the fraction of transits from (n, z) that were to each other state. Since the continuation value the agent cares about is the continuation value should the agent continue, these estimates have to be modified for conditioning on $\chi = 1$, see the discussion below.

With this notation, (78) becomes,

$$VC(\theta) = M^i \tilde{\pi}(\theta) + \beta M^i \mathbb{E}_\phi \left[\max\{\phi_i, VC(\theta)\}\right]. \tag{79}$$

In this last equation, $\tilde{\pi}(\theta)$ is a known vector (up to θ). In a structural model the distribution of ϕ would also typically be known up to a parameter vector. Therefore, the only unknowns in the equation are M^i and $VC(\theta)$. If M^i were known, $VC(\theta)$ could be calculated as the solution to the set of equations, (79). We discuss the estimation of M^i below and turn first to the solution for $VC(\theta)$.

One of the insights of POB is that the expectations term on the right hand side of (79) can sometimes be simplified, making computation of $VC(\theta)$ simple. Expanding the expectations term at a single state (n, z) gives,

$$\begin{aligned}\mathbb{E}_\phi \left[\max\{\phi_i, VC(n,z;\theta)\}\right] =& Pr(\phi_i < VC(n,z;\theta)) * VC(n,z;\theta) + \\ & Pr(\phi_i > VC(n,z;\theta)) * \mathbb{E}_\phi \left[\phi_i | \phi_i > VC(n,z;\theta)\right] \\ =& (1 - p_x(n,z)) * VC(n,z;\theta) + p_x(n,z) * \mathbb{E}_\phi \left[\phi_i | \phi_i > VC(n,z;\theta)\right]\end{aligned}$$

where $p_x(n, z)$ is the probability of exit at state (n, z). Provided that the distribution of scrap values is log-concave, the above equation is a contraction mapping (see Heckman and Honore (1990)). In that case, given estimates of M and p^x, the equation can be solved for $VC(\cdot)$ in a straightforward manner. Moreover, when the distribution of scrap values is exponential, a distribution often thought to be reasonable on a priori grounds,

$$\mathbb{E}_\phi \left[\phi_i | \phi_i > VC(n,z;\theta)\right] = \sigma + VC(n,z;\theta),$$

where σ is the parameter of the exponential, and

$$\begin{aligned}\mathbb{E}_\phi \left[\max\{\phi_i, VC(n,z;\theta)\}\right] &= (1 - p_x(n,z)) * VC(n,z;\theta) + p_x(n,z) * [VC(n,z;\theta) + \sigma] \\ &= VC(n,z;\theta) + \sigma p_x(n,z).\end{aligned}$$

Substituting this expression into (79) and iterating gives,

$$\begin{aligned}VC(\theta) &= M^i[\tilde{\pi}(\theta) + \beta\sigma p_x] + (M^i)^2[\tilde{\pi}(\theta) + \beta\sigma p_x] + (M^i)^3 VC(\theta) + \dots \\ &= \sum_{\tau=1}^{\infty} (M^i)^\tau [\tilde{\pi}(\theta) + \beta\sigma p_x] \\ &= (I - \beta M^i)^{-1} M^i [\tilde{\pi}(\theta) + \beta\sigma p_x].\end{aligned} \tag{80}$$

The only thing that remains is to estimate M^i and p_x using the data. Both can be estimated in a variety of different ways, but the simplest approach, and the one supported by POB's Monte Carlo results, is to use their empirical counterparts. Let

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

be the set of periods in the data with the same state (n, z). Then, the empirical counterpart to p_x is

$$\widehat{p}_x(n, z) = \frac{1}{\#T(n, z)} \sum_{t \in T(n,z)} \frac{x_t}{n}.$$

Due to the Markov property, $\widehat{p}_x(n, z)$ is a sum of independent draws on the exit probability, and therefore it converges to $p_x(n, z)$ provided $\#T(n, z) \to \infty$.

Similarly, the matrix M^i can be estimated element-by-element using

$$\widehat{M}^i_{i,j} = \frac{\sum_{t \in T(n_i,z_i)} (n_i - x_t) 1\{(n_{t+1}, z_{t+1}) = (n_j, z_j)\}}{\sum_{t \in T(n_i,z_i)} (n_i - x_t)}.$$

This expression weights the actual observed transitions from (n_i, z_i) in different periods by the number of incumbents who actually continue in those periods. This weighting corrects the estimated transition probabilities for the fact that incumbents compute continuation values under the assumption that they will continue in the market.

Note that because this procedure uses empirical transition probabilities it never requires continuation values or transition probabilities from points not observed in the data. As a result there is no need to impute transition probabilities or continuation values for states not visited.[55] Since typical data sets will only contain a small fraction of the points in $\mathcal{S}$, this reduces computational burden significantly.

Substituting the estimated transition and exit probabilities into (80) provides an expression for the estimated continuation values,

$$\widehat{VC}(\theta, \sigma) = (I - \beta \widehat{M}^i)^{-1} \widehat{M}^i [\tilde{\pi}(\theta) + \beta \sigma \widehat{p}_x] \tag{81}$$

Note first that the estimates of continuation values using the expression in (81) are, approximately, the averages of the discounted values of the incumbents who did continue.[56] This makes the relationship between the data and the model transparent. Provided only

[55] Strictly speaking this is only true if the last period's state in the data was visited before. If it were not we would have to impute transition probabilities for it.

[56] This is only approximately true because the transitions for all firms that reached a state (n, z) are used to compute transitions for each firm, so information is pooled across firms in computing the continuation values.

that the specification of the profit function is correct, the actual average of realized continuation values should be close to the expected continuation values used by the agents in making their decisions.

Second, note how easy it is to compute the estimated continuation values. If the discount factor is known, then,

$$\widehat{VC}(\theta,\sigma) = \tilde{A}\tilde{\pi}(\theta) + \tilde{a}\sigma \tag{82}$$

where $\tilde{A} = (I - \beta\widehat{M}^i)^{-1}\widehat{M}^i$ and $\tilde{a} = \beta(I - \beta\widehat{M}^i)^{-1}\widehat{p}_x$. Both $\tilde{A}$ and $\tilde{a}$ are independent of the parameter vector and can therefore be computed once and then held fixed in the second step of the estimation.

Finally, note that the parameters of the entry distribution do not enter into the calculation of the continuation values. The reason for this is that sunk costs of entry are paid only once at the time of entry. After that, the sunk costs distribution only affects profits indirectly through rival firms' entry decisions. Thus, all that matters for computing continuation values is the probability of entry, not the associated level of sunk costs. As a result the computational burden of the model does not depend in any major way on the form of the entry cost distribution, a fact which is particularly useful when we consider models with multiple entry locations below.

Entry behavior can be described similarly. A potential entrant enters the market if the expected discounted value of entering is greater than the entry cost, i.e., if χ^e is the indicator function which is one if the potential entrant enters and zero elsewhere

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

where

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

similarly to VC before. The main difference here is that the entrant is not active in the current period and therefore forms beliefs slightly differently from the incumbent.

The incumbent forms beliefs conditional on it remaining active. The entrant forms beliefs based on it *becoming* active. In vector form, the expression for the entrants' continuation values is

$$VE(\theta,\sigma) = M^e(\tilde{\pi} + \beta VC(\theta) + \beta p_x\sigma),$$

where the elements of M^e represent a potential entrant's beliefs about the distribution over tomorrow's states conditional on that entrant becoming active. An estimator for M^e that is analogous to the one above is given by

$$\widehat{M}^e_{i,j} = \frac{\sum_{t\in T(n_i,z_i)} e_t 1\{(n_{t+1},z_{t+1}) = (n_j,z_j)\}}{\sum_{t\in T(n_i,z_i)} e_t}.$$

Accordingly, a consistent estimator of $\widehat{V}E(\theta, \sigma)$ is given by,

$$\widehat{VE}(\theta, \sigma) = \tilde{B}\tilde{\pi}(\theta) + \tilde{b}\sigma \tag{83}$$

where $\tilde{B} = \widehat{M}^e(I + \beta\tilde{A})$, and $\tilde{b} = \beta\widehat{M}^e(\tilde{a} + \widehat{p}_x)$.

3.5.2 Step 2: Estimating the Structural Parameters.

If the continuation values (VE and VC) were known, any of a number of method of moments algorithms would provide consistent estimators of (θ, σ) and maximum likelihood would provide the efficient estimator. Since $\widehat{VE}$ and $\widehat{VC}$ are consistent estimators of the unknown continuation values, an obvious way to obtain a consistent estimator is to substitute them for VC and VE in any of these algorithms and proceed from there. For example, the implied "pseudo" maximum likelihood estimator would maximize

$$\begin{aligned} l(x_t, e_t|\theta, \sigma) = {} & (n_t - x_t)\log F^{\phi}\left[\widehat{VC}(n_t, z_t; \theta, \sigma)\right] + x_t \log\left[1 - F^{\phi}(\widehat{VC}(n_t, z_t; \theta, \sigma))\right] \\ & + e_t \log F^{\kappa}\left[\widehat{VE}(n_t, z_t; \theta, \sigma)\right] + (E - e_t)\log\left[1 - F^{\kappa}(\widehat{VE}(n_t, z_t; \theta, \sigma))\right], \end{aligned}$$

where F^{ϕ} is the distribution of scrap values and F^{κ} is the distribution of entry costs.

POB stress the importance of remembering that $\widehat{VE}$ and $\widehat{VC}$ contain sampling error. Though this sampling error does converge to zero with sample size, the fact that we have to estimate separate continuation values for each sample point means that, for standard sample sizes, the sampling error should not be ignored. This has implications both for the choice of estimators, and for how we compute standard errors for any given choice.

In this context there are two problems with the pseudo maximum likelihood estimator. First since it does not "recognize" that there is sampling error in the probabilities it uses, events can occur that the likelihood assigns zero probability to, no matter the value of (θ, σ) (even though the true probabilities of these events are nonzero; POB shows that this tends to occur in their two location model). If this happens even once in the data, the pseudo maximum likelihood estimator is not defined. Second, even if the pseudo likelihood is well defined, it's first order condition involves a function that is both highly nonlinear in, and highly sensitive to, the sampling error. The nonlinearity implies that the impact of the sampling error on the first order conditions will not average out over sample points. The sensitivity is seen by noting that the first order effect of the sampling error on the log likelihood will be determined by one over the probabilities of entry and exit, and these probabilities are typically quite small.

POB consider two alternatives to the likelihood approach. The first is a pseudo minimum χ^2 estimation algorithm that minimizes the sum of squares in the difference between the observed and predicted state specific entry and exit rates (i.e., the entry and exit rates for each observed (n, z) pair), where the predicted state specific entry and exit rates are given by

$$\begin{aligned} E[x_t|n_t, z_t] &= n_t * Pr(\phi_i > VC(n_t, z_t; \theta, \sigma)), \text{ and} \\ E[e_t|n_t, z_t] &= E * Pr(\kappa < VE(n_t, z_t; \theta, \sigma)). \end{aligned}$$

Their second estimator matches the overall entry and exit rates (across all observed state pairs) to those predicted by the model, or more generally takes a sum of squares in the differences between the predicted and actual entry and exit rates at different states multiplied by a *known* function of the state variables at those states.

They show that in finite samples the pseudo minimum χ^2 estimator has an extra bias term that reflects the sampling covariance between the estimated probability and its derivative with respect to the parameter vector, and their Monte Carlo evidence indicates that the extra bias term can have large effects. Thus they prefer the simplest method of moments algorithm and show that with moderately sized samples this estimator is both easy to calculate and performs quite well.

The second general point is that the variance of the second stage estimates, $(\widehat{\theta},\widehat{\sigma})$, depends on the variance of the first stage estimates.[57] It is possible to use standard semi-parametric formulae to obtain the asymptotic variance of the parameter estimates analytically. However these formula are somewhat complex and can be difficult to evaluate. Moreover there is little reason to do the calculation. Since we have a complete model and the computational burden of obtaining estimates is minimal it is relatively easy to obtain estimates of standard errors from a parametric bootstrap.

For an empirical example which uses these techniques see Dunne, Klimek, Roberts and Xu (2005). They estimate the parameters of a dynamic entry game from data on entry and exit of dentists and chiropractors in small towns. They first estimate the variable profit function (which depends on the number of active competitors) from observed data on revenues and costs. They then employ POB's method to provide estimates of the sunk costs of entry and of exit values. Their parameters could be used, for example, to predict the effect of a government subsidy intended to increase the number of medical service professionals in small towns.

[57]This follows from the fact that the derivative of the objective function with respect to the estimates of VC and VE are not conditional mean zero.

3.5.3 Multiple Entry Locations.

We now show how to generalize the model to allow for observed heterogeneity among incumbents. We do this by allowing entrants to choose from multiple entry locations. For ease of exposition, we will consider only two locations. However, expanding this to a larger number is straightforward.

Entrants have entry costs (κ_1, κ_2) in the first and second locations respectively, where entry costs are drawn from a distribution, $F^\kappa(\kappa_1, \kappa_2|\theta)$, that is independent over time and across agents. Note that we place no restrictions on F^κ so that entry costs of the same potential entrant at the different locations may be freely correlated. Once in a particular location, the entrant cannot switch locations, but can exit to receive an exit fee. Exit fees are an *iid* draw from the distribution $F_1^\phi(\cdot|\theta)$ if the incumbent is in location one, and an *iid* draw from $F_2^\phi(\cdot|\theta)$ if the incumbent is in location two.

The Bellman equation for an incumbent in the two location model is,

$$V_1(n_1, n_2, z; \theta) = \tilde{\pi}_1(n_1, n_2, z; \theta) + \beta \mathbb{E}_\phi \left[\max\{\phi_i, VC_1(n_1, n_2, z; \theta)\}\right],$$

where the subscript "1" indicates the value function for a firm in location one and the continuation values are,

$$VC_1(n_1, n_2, z; \theta) \equiv \sum_{z', e_1, e_2, x_1, x_2} V_1(n_1{+}e_1{-}x_1, n_2{+}e_2{-}x_2, z'; \theta) P(e_1, e_2, x_1, x_2|n_1, n_2, z, \chi = 1) P(z'|z).$$

Behavior of incumbent firms is identical to before, with the probability of exit given by (77) except using the new continuation values. However, because they have potentially different continuation values and different scrap values, firms in location one will in general behave differently than firms in location two.

Behavior of potential entrant firms is different from before because potential entrant firms now have three options. They can enter location one, enter location two, or not enter at all. A potential entrant will enter into location 1 if and only if it is a better alternative than both not entering anywhere, and entering into location 2, i.e. if,

$$\beta VE_1(n_1, n_2, z; \theta) \geq \kappa_1, \text{ and } \beta VE_1(n_1, n_2, z; \theta) - \kappa_1 \geq \beta VE_2(n_1, n_2, z; \theta) - \kappa_2.$$

The entry process therefore generates a multinomial distribution with probabilities derived from the inequalities above.

Firms' beliefs are now comprised of the probability of exit for an incumbent in location one, the probability of exit for an incumbent in location two, the three entry probabilities (location one, location two, or not at all), and the distribution determining the evolution of the profit shifters. In computing the continuation values we now have to consider all of these together.

Consistent estimates of a location-one incumbent's perceived transition probabilities from state (n_{i1}, n_{i2}, z_i) to (n_{j1}, n_{j2}, z_j) are obtained analogously to before using

$$\widehat{M}_{i,j}^{i,1} = \frac{\sum_{t \in T(n_{i1}, n_{i2}, z_i)} (n_{i1} - x_{t1}) 1\{(n_{t+1,1}, n_{t+1,2}, z_{t+1}) = (n_{j1}, n_{j2}, z_j)\}}{\sum_{t \in T(n_{i1}, n_{i2}, z_i)} (n_{i1} - x_{t1})}.$$

Similarly, estimates of a potential location-one entrant's perceived transition probabilities can be obtained using

$$\widehat{M}_{i,j}^{e,1} = \frac{1}{\#T(n_{i1}, n_{i2}, z_i)} \frac{\sum_{t \in T(n_{i1}, n_{i2}, z_i)} e_{t1} 1\{(n_{t+1,1}, n_{t+1,2}, z_{t+1}) = (n_{j1}, n_{j2}, z_j)\}}{\sum_{t \in T(n_{i1}, n_{i2}, z_i)} e_{t1}}.$$

As before these estimates are not exactly equal to the empirical frequency of state transitions but are a weighted average based on the fact that, when computing continuation values, an incumbent assumes it will continue, and a potential entrant assumes that it will enter.

As in the single location model, given the matrix inversion formula for continuation values, the computational burden of obtaining estimates for the parameters of the model is minimal. Indeed in their Monte Carlo results POB show that in two location models with relatively large data sets (on the order of 7500 observations) one finds estimates in under fifteen minutes on a five year old desktop computer. Most of that computation time is devoted to computing the Markov transition matrix and its inverse. The time required to compute the inverse can grow polynomially in the number of distinct states and, at least given market size, this typically increases with the number of locations. Whether it does or not depends on the structure of the matrix being inverted, and the way one computes the inverse. Models which only allow transitions to "near by" states, which are likely to dominate in I.O. applications, should not be as problematic in this respect.

Second, though the estimators remain consistent when the number of entry states is increased, their small sample properties may change. In particular, the estimates of the continuation values will become noisier in small samples and this is likely to cause increased small sample bias and variance in the second stage estimates. POB show that the use of smoothing techniques, such as those discussed in the next section, can be helpful in this context.

3.6 Models with Discrete and Continuous Controls: Investment Games.

In this section we consider Bajari, Benkard and Levin's (2006) (henceforth, BBL) estimation approach in the context of the investment model in example two. The main conceptual difference in BBL's general approach that separates it from the above methods is that, instead of estimating continuation values directly, BBL first estimate policy functions. Then, the estimated policy functions are used to simulate the continuation values. As noted earlier this is similar to the single agent approach of Hotz et. al. (1994), but BBL show that there are assumptions and techniques that allow the researcher to use this approach in a wide class of models, including models with both discrete and continuous controls such as investment models, some models of dynamic pricing, and dynamic auction problems. The assumptions used do carry with them some restrictions, and we will try to be clear on those restrictions below.

The presence of both discrete and continuous controls in the investment model affects both the first and second stage of the estimation. In particular, the second stage is augmented in order to incorporate information from firms' investment, as well as its entry/exit, choices. Additionally, when the stress is on investment we generally consider models with larger state spaces, and, as noted above, both computation of the estimates of continuation values, and the precision of those estimates, can become problematic. BBL introduce simulation techniques that, depending on the structure of the model, can cause a significant reduction in the computational burden of obtaining estimates of the continuation values. They also use techniques that smooth estimated continuation values across states to lower the mean square error of those estimates.

In the investment model from example two there are three policies (entry, exit, and investment) that are set in dynamic equilibrium, and one policy (price) that is set in static equilibrium. Since the pricing equilibrium is consistent with a large past literature on demand estimation, we will not consider estimation of the demand and marginal cost functions (θ_1 and θ_2) here as they would typically be estimated using existing methods. Instead, we will treat those parameters as if they were known and focus on estimation of the investment cost function (θ_3) and the entry and exit costs parameters.

We assume that all of the state variables, $\mathbf{s}$, are observed, as well as entry, exit, quantity, price, and investment levels. Entry and exit costs, the cost of investment function, and marginal costs are not observed. Note that it would be possible for some of the state variables to be unobserved as long as they could be recovered beforehand during estimation of the demand and cost systems. We discuss the issue of unobserved states further in section 3.8.1.

Let $\tilde{\pi}_i(\mathbf{s})$ represent the profits earned by firm i in the spot market equilibrium at state

s. Since the demand and marginal cost functions are assumed to be known, the function $\tilde{\pi}_i(\cdot)$ is also known, as the spot market equilibrium can be computed from these primitives.

Firms maximize the expected discounted value of profits. From the beginning of a period (prior to realization of the private shock), and for incumbent firms, this is

$$\mathbb{E}\sum_{t=0}^{\infty}\beta^t\left[\{\chi_{it}=1\}\left(\tilde{\pi}_i(\mathbf{s}_t)-C(I_{it},\nu_{it};\theta_3)\right)+(\chi_{it}-\chi_{i,t-1})^{-}\Phi\,\middle|\,\mathbf{s}_0\right], \tag{84}$$

where $\chi_{it}=1$ indicates that the incumbent continues in the market at period t and $\chi_{it}=0$ indicates that the incumbent exits, and it is understood that each exiting firm receives the same exit value and never operates thereafter. Note that unlike in the entry/exit example above, in this model we assume that the incumbent's choice of its discrete control (whether or not to exit) is not subject to a random cost shock.

For expositional (and computational) simplicity we will assume the following quadratic cost of investment function,

$$C(I,\nu;\theta_3)=\{I\geq 0\}(\theta_{3,0}+\theta_{3,1}I+\theta_{3,2}I^2+\theta_{3,3}\nu I). \tag{85}$$

The indicator function for $I\geq 0$ above allows for an adjustment cost that is incurred only if investment is nonzero. Zero investment is a phenomenon that is often observed and can easily result from either flatness of the value function reflecting low returns to investment (see Ericson and Pakes (1995)), or non convex investment costs (e.g., Caballero and Engle (1999)).

Potential entrant firms' expected discounted values are similar to (84), except that in the initial period they must pay a random entry cost, κ_{it}, in order to enter. We assume that entrants take one period to setup the firm and therefore do not earn profits in the spot market and do not invest until the subsequent period. For ease of exposition, we also assume that entrants always enter at the same initial state.[58]

3.6.1 Step 1: Estimating Continuation Values.

The goal of the first step of the estimation procedure is to compute the continuation values given by the expected discounted values in (84), under equilibrium strategies. These expected discounted values are functions of the profits earned at each state and the probability distributions determining future states and actions conditional on the starting state, $\mathbf{s}_0$.

[58]It is straightforward to allow entrants' initial state to be randomly determined.

BBL compute estimates of the continuation values by first estimating policies for each state, then using the estimated policies to simulate sample paths of industry states and actions, and then evaluating discounted profits on each sample path. In order to do this we need both the ability to simulate sample paths of states and actions, and the ability to evaluate profits along those paths given the states and actions in each period. Evaluating profits requires a (possibly flexible) description of the profit function and knowledge of the distribution of the private shocks, at least up to a parameter vector to be estimated. We treat these two as "primitives" of the dynamic model.

The evolution of the states depends on firms' entry, exit, and investment policies. BBL recover these policy functions using the observed data. In our investment model, private information is known to the firms before any actions are taken, so in MPE, strategies for investment, exit and entry are functions of both the states and this private information,

$$I(\mathbf{s}_t, \nu_{it}), \quad \chi(\mathbf{s}_t, \nu_{it}), \text{ and } \chi^e(\mathbf{s}_t, \kappa_{it}),$$

where χ is the exit policy function for incumbent firms, χ^e is the entry policy function and $\chi^e = 1$ indicates that the entrant enters the market. Since potential entrants cannot invest in the first period, entry strategies depend only on the random entry cost. Both the investment and exit strategies depend only on the shock to the marginal cost of investment.

Consider first exit and entry. The optimal exit policy has the form of a stopping rule

$$\chi_{i,t} = 1 \text{ iff } \nu_{it} \leq \bar{\nu}(\mathbf{s}_t).$$

All we require is a nonparametric estimate of the probability that $\chi = 1$ conditional on $\mathbf{s}_t$. Similarly, there a critical entry level of κ conditional on $\mathbf{s}_t$ that determines entry, and the entry policy is obtained as a nonparametric estimate of the probability of entry conditional on $\mathbf{s}_t$. In both cases we also have the restriction that the policies are exchangeable in rivals' states. In models with large state spaces, such that there are some states in the data with few or zero observations, it would typically be optimal to employ some smoothing in these estimates. In their Monte Carlo studies, BBL found that local linear methods worked well for this.

As far as investment is concerned, one can show that, conditional on a firm continuing in the market, investment is a (weakly) monotone function of ν_{it}, $I(\mathbf{s}_t, \nu_{it})$. Thus, if we knew the distribution of investment at each state, $F(I_{it}|\mathbf{s}_t)$, we could map the quantiles of ν into investment levels at each state. More precisely, the investment policy function is given by

$$F_{I|\mathbf{s}}^{-1}(G(\nu|\mathbf{s})).$$

The function G is a primitive of the model known up to a parameter vector, and the function F can be recovered nonparametrically.

There is an additional complication and that is that investment is not observed for firms that exit the market, which happens if $\nu_{it} > \bar{\nu}(\mathbf{s}_t)$. However, since both the exit and investment rules are monotonic in the shock, this is handled easily. Conditional on a firm continuing in the market, we observe the distribution of investment conditional on $\mathbf{s}$ that corresponds to $\nu_{it} \leq \bar{\nu}(\mathbf{s}_t)$. Therefore, if we first estimate the probability of exit at each state, and then recover the distribution of investment at each state conditional on staying in the market, then we have a complete description of the optimal exit and investment policy functions.[59] Note also that in the simulations below it is important that we maintain this link between exit and investment since one draw on the private shock to investment, ν_{it}, determines both policies.

If there was a second unobservable, say a random exit fee ϕ_{it}, then the exit decision would depend on both (ν_{it}, ϕ_{it}). The probability of exit could still be obtained as above, but the distribution of investment conditional on not exiting would depend on both ν_{it} and ϕ_{it}. Then, without further restrictions it would not be possible to invert the observed distribution of investment to obtain the policy decision as a function of ν conditional on $\mathbf{s}$ and not exiting.[60]

There also remains the question of how best to estimate the investment function, and this depends to some extent on its likely properties. Here it is important to keep in mind that investment is a complicated function of the primitives. Indeed the only restriction we have on its form is that it is exchangeable in the states of the competitors (which is already embodied in the definition of $\mathbf{s}$). Standard nonparametric approaches assume a certain amount of smoothness that is not necessarily guaranteed by the primitives of the model. The theoretical properties of the investment function in the EP model depend on the underlying properties of the family $\{P(s_{it+1}|I_{it}, s_{it})\}$. If conditional on s_{it} the points of support of this family do not depend on I_{it}[61], then by appropriate choice of primitives one can ensure that the investment function is smooth; see EP, the Monte Carlo evidence in BBL, and the generalizations of this in Dorazelski and Satterwaithe (2004). In their Monte Carlo studies BBL also found that local linear regression worked well for estimating the investment function.

Assume now that, for each state, we have consistent estimators of the entry probability, the exit probability, the investment distribution, and the distribution of future states.

[59]We cannot use the data to learn what an exiting firm would have invested had it stayed in the market, but it is not necessary to know this.

[60]Note that the problem here is that there is more than one error influencing the choice of investment. Therefore, a feasible alternative would be to allow a random exit cost but no shock to the marginal cost of investment.

[61]This assumption allows for stochastic outcomes to investment processes which is an assumption often made in Industrial Organization. However it does rule out the deterministic accumulation models traditionally used in growth theory.

This is all one needs to compute consistent estimates of the continuation values in (84) as a function of the parameters. To do so analytically, however, would involve high dimensional integration, so what BBL do is show how to extend the "forward simulation" idea in Hotz, Miller, Saunders and Smith (1994) to simplify the analysis of the more complex problems they deal with.

Starting from a given state, $\mathbf{s}_0$, one draw is taken on the shock to the marginal cost of investment for the firm of interest, ν_i. This draw determines the firm's investment and exit policies through the estimated policy functions above (i.e., the same draw determines the correct quantile for both investment and exit, as discussed above). These policies, along with the state and the value of the private shock, determine current profits as a function of the parameters. Draws are then taken for the investment shocks for the remaining incumbent firms, and one draw on the entry distribution is taken for the potential entrant. These draws, along with draws determining the outcomes of each firm's investment process, determine $\mathbf{s}_1$. The process is repeated to obtain one simulated path of states and the associated discounted stream of profits. Many such paths can be simulated to obtain an estimate of $V_i(\mathbf{s}_0)$. Consistency of the estimation algorithm requires that the number of simulated paths goes to infinity.

This forward simulation procedure is not too computationally burdensome, though one does have to hold one set of simulation draws in memory and use these same draws to evaluate the continuation values at the different values of θ tried in the estimation algorithm. Moreover, much of what computational burden remains disappears when we deal with models that are linear in the parameters. For example, suppose we consider the investment model above where the private shock to investment has a normal distribution, $\nu \sim N(0,1)$. (The investment shock is normalized to be standard normal without loss of generality because its mean and variance parameters are absorbed into the parameters $\theta_{3,0}$ and $\theta_{3,3}$.) Since all of the parameters enter the continuation values linearly, they can be factored out as follows,

$$\begin{aligned}
V_i(\mathbf{s}_0;\sigma_i,\sigma_{-i}) =& \mathbb{E}\sum_{t=0}^{\infty}\beta^t\{\chi_{it}=1\}\tilde{\pi}_i(\mathbf{s}_t) - \theta_{3,0}\mathbb{E}\sum_{t=0}^{\infty}\beta^t\{\chi_{it}=1\}\{I_{it}\geq 0\}- \\
& \theta_{3,1}\mathbb{E}\sum_{t=0}^{\infty}\beta^t\{\chi_{it}=1\}\{I_{it}\geq 0\}I_{it} - \theta_{3,2}\mathbb{E}\sum_{t=0}^{\infty}\beta^t\{\chi_{it}=1\}\{I_{it}\geq 0\}I_{it}^2- \\
& \theta_{3,3}\mathbb{E}\sum_{t=0}^{\infty}\beta^t\{\chi_{it}=1\}\{I_{it}\geq 0\}I_{it}\nu_{it} + \Psi\mathbb{E}\sum_{t=0}^{\infty}\beta^t(\chi_{it}-\chi_{i,t-1})^- \\
\equiv& W_i(\mathbf{s}_0;\sigma_i,\sigma_i')'\begin{pmatrix}1\\ \theta_3\\ \Psi\end{pmatrix},
\end{aligned} \tag{86}$$

where $W_i(\mathbf{s}_0, \sigma_i, \sigma_{-i})$ represents the expected discounted value terms above when i follows policy σ_i and rival firms follow policy σ_{-i}. The estimated continuation values are then computed by plugging in the estimated policies and simulating the expectations terms,

$$\widehat{V}_i(\mathbf{s}_0; \widehat{\sigma}_i, \widehat{\sigma}_i') = \widehat{W}_i(\mathbf{s}_0; \widehat{\sigma}_i, \widehat{\sigma}_i')' \begin{pmatrix} 1 \\ \theta_3 \\ \Psi \end{pmatrix}.$$

The key observation here is that if the model is linear in the parameters, then the parameters factor out of the continuation value calculations. In that case the W terms need only be computed once, and the continuation values at different values of the parameter vector can be obtained by multiplying two small dimensional vectors.

This simplification is an extension of the one used in the entry/exit example above, except here we exploit linearity in the investment cost parameters as well as the linearity in the period profits. Since the continuation values need to be calculated many times in the second step of the estimation, and since computing continuation values is the primary source of computational burden, such simplifications can lead to a substantial reduction in the overall computational burden of the estimator.

3.6.2 Step 2: Estimating the Structural Parameters.

As with the entry model above, once the continuation values have been estimated there are potentially many ways of estimating the structural parameters. The main difference is that now there is one continuous control variable (investment) in addition to the two discrete controls (entry/exit), and we want to use the information in the continuous control to help estimate the parameters. Accordingly all the issues that arose in the discussion of estimation of the entry/exit model are also relevant here. In particular there is error in the estimated continuation values that can contaminate the second stage estimates, so it is desirable to find a second step estimator that is close to linear in the estimated continuation values.

There are at least three possible estimators; (i) an inequality estimator that finds a value of the parameter vector that insures that the observed policies generate higher simulated continuation values than alternative policies (see below), (ii) a method of moments estimator that fits the mean of the policies implied by the simulated continuation values (i.e., at each state in the data you substitute the simulated continuation values into the right hand side of the Bellman equation (65) and solve for the optimal policy) to nonparametric estimates of the policies at each state, and (iii) a method of moments estimator that fits the nonparametric estimates of the distribution of the polices to the distribution of policies predicted by the simulated continuation values at each state. BBL provide

Monte Carlo evidence on the first two of these. Here we review the inequality estimator, that is the estimator found by satisfying the optimality inequalities (64) that define the MPE for the simulated values.

At the true values of the parameters, for all states, $\mathbf{s}_0$, all firms, i, and all alternatives, σ_i', it must be that

$$W_i(\mathbf{s}_0;\sigma_i,\sigma_{-i})'\begin{pmatrix}1\\ \theta_3\\ \Psi\end{pmatrix} \geq W_i(\mathbf{s}_0;\sigma_i',\sigma_{-i})'\begin{pmatrix}1\\ \theta_3\\ \Psi\end{pmatrix}.$$

Let x refer to a particular $(i,\mathbf{s}_0,\sigma')$ combination, such that x indexes inequalities, and let

$$g(x;\theta_3,\Psi) = (W_i(\mathbf{s}_0;\sigma_i,\sigma_{-i}) - W_i(\mathbf{s}_0;\sigma_i',\sigma_{-i}))'\begin{pmatrix}1\\ \theta_3\\ \Psi\end{pmatrix}.$$

Then it must be the case that $g(x;\theta_3,\Psi)\geq 0$ at the true values of the parameters for every x.

A natural thing to do in estimation would be to compute g at the estimated policies from the first stage and then find the values of the parameters that best satisfy the entire set of inequalities. However, when there are continuous controls this is difficult because there are too many possible alternative policies. Instead, BBL use simulation to choose a small subset of the inequalities to impose in estimation. The inequalities can be chosen according to any random rule that selects all of them asymptotically. However, it is important to remember that the exact rule used will influence efficiency. In their Monte Carlo studies, for investment alternatives BBL use policies of the form

$$I'(\mathbf{s}_t,\nu_{it}) = \widehat{I}(\mathbf{s}_t,\nu_{it}) + \epsilon'$$

where ϵ' is drawn from a normal distribution with mean zero and standard deviation chosen by the researcher. Alternative entry and exit policies were chosen similarly by shifting the cutoff rule by an amount ϵ' drawn from a normal distribution.

Suppose n_i inequalities are sampled, and let $\widehat{g}_{n_s}(x;\theta_3,\Psi)$ be a simulator for $g(x;\theta_3,\Psi)$ evaluated at the estimated policy functions, where n_s is the number of simulation draws used to simulate each W_i term. Then the inequality estimator minimizes the objective function

$$\frac{1}{n_I}\sum_{k=1}^{n_I} 1\{\widehat{g}_{n_s}(x_k;\theta_3,\Psi)<0\}\widehat{g}_{n_s}(x_k;\theta_3,\Psi)^2.$$

Because the estimator is computationally light, it is easy to choose (n_I,n_s) to be large enough that the simulation contributes nothing to the variance of the estimator. All of the variance comes from error in the estimation of the continuation values. BBL work out the asymptotic distribution of the estimator. However, the expression is difficult to evaluate and in practice the simplest way to compute standard errors is to use subsampling or a bootstrap.

The inequality estimator has several advantages. One is that it is very easy to implement even in complex models. It is conceptually simple and requires a minimum of computer programming, the main programming burden being the forward simulation routine. Additionally, the method can be used with very little alteration even if the model is only set-identified. In that case, all that is required is to use an alternative method for computing standard errors (see BBL, as well as Chernozhukov, Hong, and Tamer (2004) for details).

However, one potential disadvantage of the approach is that, similarly to the pseudo-likelihood methods shown in the examples above, the estimator is nonlinear in the first stage estimates, and therefore the estimates obtained are likely to contain small sample bias. For that reason, BBL also tested a natural alternative estimator based on a set of moment conditions that match the observed choice data. The general idea of this estimator is to substitute the estimated continuation values into the right hand side of the Bellman equation and then solve for an optimal policy rule conditional on those continuation values. This estimator is linear in the estimated continuation values, though those values are still nonlinear functions of the estimated policies. The expected value of the optimal policy is then matched against the average policy observed at each state in the data. In their Monte Carlo studies BBL found that this second estimator did help reduce small sample bias in the second stage estimates.

For an empirical example that uses these techniques see Ryan (2005). He estimates the parameters of a dynamic oligopoly model of US cement producers. In the first stage he estimates the static profits demand and cost parameters using demand data and a static equilibrium assumption. He also estimates the entry, exit, and investment policy functions using data on the set of firms operating in a panel of markets and their capacities. In the second stage he uses BBL's inequality estimator to estimate the sunk costs of entry and exit, as well as the adjustment costs of investment. He finds that the 1990 amendments to the Clean Air Act significantly raised the sunk costs of entry in the cement industry, and that a static analysis would have missed an associated welfare penalty to consumers.

3.6.3 An Alternative Approach.

Berry and Pakes (2002) provide an alternative approach for estimating models with continuous choice variables that uses quite different assumptions from POB or BBL. They assume that profits are observable up to a parameter vector to be estimated, but do not require that the state variables that determine current and expected future profits are observed, and do not even require the researcher to specify what those state variables are. In applications where the environment is complex, but sales and investment data are quite good, this alternative set of assumptions can be quite attractive.

Let the random variable τ_i refer to the period in which firm i exits the market. Then, firm i's continuation value in the investment game starting at state $\mathbf{s}_0$ is

$$V_i(\mathbf{s}_t) = \mathbb{E}\left[\sum_{r=t}^{\tau_i} \beta^{r-t} \left(\tilde{\pi}_i(\mathbf{s}_r) - C(\sigma(\mathbf{s}_r);\theta_3)\right) + \beta^{\tau_i - t}\Phi \,\middle|\, \mathbf{s}_t\right] \tag{87}$$

where σ is the equilibrium policy function. Note that we have assumed there is no private shock to investment; an assumption that is needed for the consistency of this estimator.

Berry and Pakes note that, if firms have rational expectations, then the actual discounted stream of profits earned by a given firm is an unbiased estimate of its expected discounted profits. Suppose that profits ($\tilde{\pi}_{it}$), investment (I_{it}), and exit (χ_{it}) are observed. Then the actual discounted sum of profits earned by the firm (corresponding to (87)) is

$$\widehat{V}_i(\mathbf{s}_t;\theta_3,\Phi) \equiv \sum_{r=t}^{\tau_i} \beta^{r-t} \left(\tilde{\pi}_{ir} - C(I_{ir};\theta_3)\right) + \beta^{\tau_i - t}\Phi, \tag{88}$$

where, in a slight abuse of notation, τ_i now refers to the actual period in which the firm exited. By rational expectations we have that, at the true values of the parameters,

$$\widehat{V}_i(\mathbf{s}_t;\theta_3,\Phi) = V_i(\mathbf{s}_t) + \epsilon_{it}$$

where $E[\epsilon_{it}|\mathbf{s}_t] = 0$.

A unique feature of the Berry and Pakes approach is that the estimated continuation values here are unbiased. However, in contrast to POB and BBL, Berry and Pakes (2002) do not have a first stage that provides consistent estimates of continuation values. Since the state variables are assumed not to be observed, there is no longer any way of identifying a set of data points that correspond to the same state vector. Thus, there is no way to average out across observations so as to obtain consistent estimates of the continuation values, as in POB and BBL.

Berry and Pakes get around the problem of having only unbiased, and not consistent estimates of continuation values, by using an estimating equation that has the error in the estimated continuation value entering linearly. More precisely, their estimating equation is derived from the first order condition for the firm's continuous control. Conditional on investment being strictly positive (a condition that is determined by the information available when the investment decision is made, and hence that is independent of the realization of ϵ_{it}), that first order condition is obtained by setting the derivative of (87) equal to zero. Using the cost of investment function after eliminating the i.i.d. shock to investment this gives us

$$\begin{aligned}
0 &= -\theta_{3,1} - 2 * \theta_{3,2} * I_{it} + \beta \sum_{\mathbf{s}_{t+1}} V_i(\mathbf{s}_{t+1}) \frac{\partial}{\partial I} P(\mathbf{s}_{t+1}|I_{it}, \mathbf{s}_t, \chi_{it} = 1) \\
&= -\theta_{3,1} - 2 * \theta_{3,2} * I_{it} + \beta \sum_{\mathbf{s}_{t+1}} V_i(\mathbf{s}_{t+1}) \frac{\frac{\partial}{\partial I} P(\mathbf{s}_{t+1}|I_{it}, \mathbf{s}_t, \chi_{it} = 1)}{P(\mathbf{s}_{t+1}|I_{it}, \mathbf{s}_t, \chi_{it} = 1)} P(\mathbf{s}_{t+1}|I_{it}, \mathbf{s}_t, \chi_{it} = 1) \\
&= -\theta_{3,1} - 2 * \theta_{3,2} * I_{it} + \beta E\left[V_i(\mathbf{s}_{t+1}) \frac{\partial \ln P(\mathbf{s}_{t+1}|I_{it}, \mathbf{s}_t, \chi_{it} = 1)}{\partial I} \,\middle|\, \mathbf{s}_t, I_{it}, \chi_{it} = 1 \right]
\end{aligned} \tag{89}$$

Adding and subtracting a term in $\widehat{V}(\mathbf{s}_{t+1}; \theta_3, \Psi)$ gives

$$0 = -\theta_{3,1} - 2 * \theta_{3,2} * I_{it} + \widehat{V}_i(\mathbf{s}_{t+1}; \theta_3, \Phi) \frac{\partial \ln P(\mathbf{s}_{t+1}|I_{it}, \mathbf{s}_t, \chi_{it} = 1)}{\partial I} + \eta_{it}(\theta_3, \Phi) \tag{90}$$

where we have defined

$$\begin{aligned}
\eta_{it}(\theta_3, \Phi) \equiv \beta E\left[V_i(\mathbf{s}_{t+1}) \frac{\partial \ln P(\mathbf{s}_{t+1}|I_{it}, \mathbf{s}_t, \chi_{it} = 1)}{\partial I} \,\middle|\, \mathbf{s}_t, I_{it}, \chi_{it} = 1 \right] \\
- \widehat{V}_i(\mathbf{s}_{t+1}; \theta_3, \Phi) \frac{\partial \ln P(\mathbf{s}_{t+1}|I_{it}, \mathbf{s}_t, \chi_{it} = 1)}{\partial I},
\end{aligned} \tag{91}$$

and consequently

$$E[\eta_{it}(\theta_3, \Phi)|\mathbf{s}_t] = 0, \tag{92}$$

at the true values of the parameters vector. Condition (92) follows from the twin facts that $V_i(\mathbf{s}_{t+1}) = \widehat{V}_i(\mathbf{s}_{t+1}; \theta_3, \Phi) - \epsilon_{i,t+1}$ and

$$E\left[\frac{\partial \ln P(\mathbf{s}_{t+1}|I_{it}, \mathbf{s}_t, \chi_{it} = 1)}{\partial I} \right] \epsilon_{i,t+1} = 0,$$

as the derivative is a function of information known at t. It follows that (90) provides a set of conditional moment restrictions that can be used as the basis for estimation.

There are a number of disadvantages of this approach. One that can potentially be corrected is that as presented in Berry and Pakes (2002) the algorithm does not incorporate the additional information in the data that comes from the choice of discrete controls (e.g. entry and exit), or from controls chosen to be at a corner of the choice set (e.g., $I_{i,t} = 0$). One could add a set of inequality constraints to the Berry-Pakes model to account for entry and exit and the $I_{i,t} = 0$ case. Also, as mentioned above, it is difficult to incorporate a shock to the cost of investment into this model.

However the major difference between this model and the other models discussed above is that Berry and Pakes do not need to specify and control for all the state variables in the dynamic system. This is an obvious advantage for complex problems. Of course, if we can not identify and control for all the state variables of the system, we can not make use of averaging techniques that enable us to use information on similar states to construct estimates of the policies and returns at a given state. In problems where the state variables are easy to identify and control for, averaging techniques can be very helpful in reducing variance. It remains to be seen if hybrids can be developed that make effective use of all of these techniques.

3.7 A Dynamic Auction Game.

In this section we consider estimation of the auction model in example three. This section closely follows Jofre-Benet and Pesendorfer (2003) (henceforth, JP). We assume that all bids, contract characteristics, and bidders' state variables are observed. A unique feature of the auction model is that the period payoff function is not a function of any unknown parameters. The goal of estimation, then, is to recover the distribution of bidders' privately known costs at each state.

Since the outcome of the auction affects not only current profits but also the firm's backlog, firms choose their bids so as to maximize the expected discounted value of future profits. Recall that z_t provides the characteristics of the contracts to be auctioned in month t and evolves as a Markov process, $\omega_{i,t}$ provides the backlog of work of firm i in period t, and if $\omega_t = (\omega_{1,t}, \ldots, \omega_{N,t})$, then $\omega_{t+1} = \Gamma(\omega_t, z_t, j)$ where j is the winning bidder.

It is convenient to write the maximization problem from the beginning of a period, prior to realization of the private shock and prior to realization of the contract characteristics. Then firms choose their bidding strategy so as to maximize the expected discounted sum

$$\mathbb{E}\left[\sum_{t=0}^{\infty}\beta^t(b_{it}-c_{it})\{b_{it}\leq\min_j(b_{jt})\}\,\bigg|\,\omega_0,z_{-1}\right], \tag{93}$$

where z_{-1} refers to last period's contract and where the expectation is defined over rival's bids in all periods as well as over the firm's own costs in all periods. Due to the Markov structure, this maximization problem can be written recursively,

$$\begin{aligned}V_i(\omega_t,z_{t-1}) = \int\int \max_{b_{it}}\Big[&(b_{it}-c_{it})Pr(i\text{ wins}|b_{it},\omega_t,z_t)+\\ &\beta\sum_{j=1}^{N}Pr(j\text{ wins}|b_{it},\omega_t,z_t)V_i(\Gamma(\omega_t,z_t,j),z_t)\Big]dF(c_{it}|\omega_t,z_t)dG(z_t|z_{t-1}).\end{aligned} \tag{94}$$

As is now common practice in the empirical auctions literature (Guerre, Perrigne, and Vuong (2000)), JP show that bidders' costs can be recovered by inverting the first order condition associated with the optimal bid. Let $G^i(\cdot|\omega_t,z_t)$ be the distribution of bids submitted by bidder i conditional on the state variables and $g^i(\cdot|\omega_t,z_t)$ be the density function. Let

$$h^i(\cdot|\omega_t,z_t)=\frac{g^i(\cdot|\omega_t,z_t)}{1-G^i(\cdot|\omega_t,z_t)}$$

denote the associated hazard function, and note that

$$\frac{\partial Pr(i\text{ wins}|b_{it},\omega_t,z_t)}{\partial b_{i,t}}=-\sum_{j\neq i}h^j(b_{it}|\omega_t,z_t)Pr(i\text{ wins}|b_{it},\omega_t,z_t)$$

while

$$\frac{\partial Pr(j\text{ wins}|b_{it},\omega_t,z_t)}{\partial b_{i,t}}=h^j(b_{it}|\omega_t,z_t)Pr(i\text{ wins}|b_{it},\omega_t,z_t).$$

Using these expressions, the first order condition for optimal bids yields the equation

$$b_{it}=c_{it}+\frac{1}{\sum_{j\neq i}h^j(b_{it}|\omega_t,z_t)}-\beta\sum_{j\neq i}\frac{h^j(b_{it}|\omega_t,z_t)}{\sum_{l\neq i}h^l(b_{it}|\omega_t,z_t)}\big[V_i(\Gamma(\omega_t,z_t,i),z_t)-V_i(\Gamma(\omega_t,z_t,j),z_t)\big]. \tag{95}$$

The optimal bid equals the cost plus a markup that has two terms. The first term reflects competition in the current auction. The second term accounts for the incremental effect on future profits of firm i winning today's auction.

Since the first order condition is strictly increasing in c it can be inverted to obtain

$$c=\phi(b|\omega_t,z_t), \tag{96}$$

where ϕ is a function of the observed bids, the hazard function of bids, h, the transition function, Γ, and the continuation values, V. The transition function is a known function. Since the bids, contract characteristics, and state variables are observed, the hazard function of bids can be obtained from the data. Thus, if the continuation values were known, then the relationship in (96) could be used to infer bidders' costs. Hence, as in the examples above, in order to estimate the parameters of the cost distribution we need first to obtain estimates of the continuation values.

3.7.1 Estimating Continuation Values.

In order to estimate the continuation values, JP note that the continuation values can be written as a function only of the distribution of bids. The easiest way to see this is to inspect (93). The expected discounted value involves terms in the probability of winning, which can be derived from the distribution of bids, and terms in the expected markup. Equation (95) shows that the optimal markup is a function of the distribution of bids and the continuation values. JP show that by combining these two equations it is possible to write the continuation values as a function only of the distribution of bids.

The derivation is long so we omit it here and instead refer readers to the appendix of JP. Proposition 1 in JP shows that equations (94) and (95) can be manipulated to obtain,

$$\begin{aligned} V_i(\omega_t, z_{t-1}) = \int \Bigg\{ \int & \frac{1}{\sum_{j \neq i} h^j(\cdot|\omega_t, z_t)} dG^{(i)}(\cdot|\omega_t, z_t) \\ & + \beta \sum_{j \neq i} \left[Pr(j \text{ wins}|\omega_t, z_t) + \int \frac{h^i(\cdot|\omega_t, z_t)}{\sum_{l \neq i} h^l(\cdot|\omega_t, z_t)} dG^{(j)}(\cdot|\omega_t, z_t) \right] \\ & \times V_i(\Gamma(\omega_t, z_t, j), z_t) \Bigg\} dG(z_t|z_{t-1}), \end{aligned} \tag{97}$$

where the notation

$$G^{(i)}(\cdot) = \Pi_{k \neq i} \left[1 - G^k(b|\omega_t, z_t) \right] g^i(b|\omega_t, z_t).$$

The terms in square brackets in the second line of (97) sum to one and therefore can be interpreted as transition probabilities. This interpretation leads to the following construction. Assume that the state space is discrete and let A_i be a vector with one element for each state representing the first term above,

$$A_i(s) = \int \int \frac{1}{\sum_{j \neq i} h^j(\cdot|\omega_t, z_t)} dG^{(i)}(\cdot|\omega_t, z_t) dG(z_t|z_{t-1}).$$

Next, construct the matrix M^i such that each element (k,l) reflects the transition probabilities above,

$$M^i_{k,l} = \begin{cases} \left[Pr(j \text{ wins}|\omega_k, z_l) + \int \frac{h^i(\cdot|\omega_k,z_l)}{\sum_{l\neq i} h^l(\cdot|\omega_k,z_l)} dG^{(j)}(\cdot|\omega_k, z_l)\right] Pr(z_l|z_k), & \text{if } \omega_l = \Gamma(\omega_k, z_l, j) \\ 0, & \text{otherwise.} \end{cases}$$

Then the value function can be expressed as,

$$V_i = [I - \beta M^i]^{-1} A_i \tag{98}$$

The matrices M^i and A_i can be estimated using estimates of the bid distribution.

3.7.2 Estimating the Cost Distribution.

Once the continuation values are known, estimating the cost distribution is straightforward. There is a relationship between the cost distribution and the bid distribution that is given by

$$\begin{aligned} F(c|\omega_t, z_t) =& G(b(c, \omega_t, z_t)|\omega_t, z_t) \\ =& G(\phi^{-1}(c|\omega_t, z_t)|\omega_t, z_t) \end{aligned}$$

(provided that ϕ is invertible). The function ϕ can be estimated using the first order condition, (95), and the estimated continuation values. The estimated ϕ can then be substituted into the estimated bid distribution in order to obtain an estimate of the cost distribution.

3.8 Outstanding Issues.

The literature on structural estimation of dynamic games is relatively recent. As a result our focus has been on reviewing assumptions and techniques that make it *feasible* to use the implications of dynamic games to make inferences on parameters of interest to I.O. We have paid little attention to a host of related issues including; the asymptotic efficiency of alternative estimators, the small sample properties of those estimators, identification in the absence of auxiliary information, and the likely validity of various assumptions.

It is not our attention to minimize any of these issues. Indeed we think it important to explore all of them, particularly the assumptions underlying the analysis. This includes the behavioral assumptions and the assumptions regarding the selection of equilibria, as well as more traditional assumptions on the properties of the unobservables in the model. The simple fact is that we have little to report on most of these issues. There is however one exception; problems that arise due to the presence of serially correlate unobserved state variables. Since this is an issue that has appeared in several related literatures, we do have some idea of how to deal with it in the context of estimating dynamic games.

3.8.1 Serially Correlated Unobserved State Variables.

In all of the examples above it is assumed that all of the states that are commonly known to the agents are also observed by the econometrician. In many empirical applications this assumption is questionable. For example, in many cases we might expect there to be an aggregate shock to profits that is known to all of the firms, but not controlled for by the econometrician. The models presented above can be modified to accommodate these shocks if they are *i.i.d.* over time. However we would often expect aggregate shocks to be serially correlated, just as most aggregate variables are. In that case, behavior in previous periods would depend on previous realizations of the unobserved states, leading to a correlation between today's values of the observed and unobserved states.

The statement of the problems caused by unobserved serially correlated state variables in dynamic models with discrete outcomes dates back at least to Heckman (1981). Pakes (1993) reviews three possible solutions to the problem: (i) solving for the unobserved states as a function of observables, (ii) simulating the model from a truly exogenous distribution of initial conditions, and (iii) using the ergodic distribution to model the long run relationship between the unobserved and observed states. With the advent of longer panels there is also the possibility of using techniques that allow one or more parameters to differ across markets in a panel of markets (say a market specific time invariant profit parameter, or a separate initial condition for each market), and then estimate those parameters pointwise.

The first case is quite promising in contexts where there is an observable continuous response to the unobservable state. Then conditional on the parameter vector, there is often a one to one relationship between the unobserved states and the observed states and controls. Several papers in the literature on static demand system estimation (Berry (1994), Berry, Levinsohn, and Pakes (1995), and Bajari and Benkard (2005)) have used such a condition to recover serially correlated unobserved product characteristics using data on quantities, prices, and observed product characteristics. Timmins (2002) uses a similar procedure to control for the initial conditions in a single agent dynamic control problem with unobserved state variables. Olley and Pakes (1996) use the implications of a dynamic Markov Perfect equilibrium model to recover a serially correlated productivity term. These methods could be used to recover the unobserved state variables prior to the dynamic estimation, and then the unobserved state variables could be treated as if they were observed in the dynamic estimation algorithm (at least up to estimation error).

Things become more difficult when the unobserved states are not recoverable in this way. In single-agent dynamic models, several papers (e.g., Pakes (1986) and Keane and Wolpin (1997)) have used exogenous initial conditions to solve the problem of serially correlated unobserved states. Starting with an exogenous initial distribution of states, the model can be used to simulate the relationship between the observed and unobserved states in future periods. However, while there may be many reasonable ways of modelling initial conditions for a single agent (be it a firm or an individual), such conditions tend to be less realistic for an industry, whose history is typically much longer than the available data.

The third case is perhaps more appropriate for industry studies. Instead of using an exogenous initial condition for the unobserved states at the time the the market starts up, we assume that the draws on the joint distribution of the unobserved states and the initial condition are draws from an invariant distribution. That distribution is then estimated along with the other parameters of the problem. The rational here is that if the markets in question have been in existence long enough, the joint distribution of the initial condition and the unobserved state will not depend on the early years of the industry's evolution. Rather it will depend only on the limiting structure of the Markov process generated by the nature of the dynamic game, a structure we can analyze.

Aguirregabiria, V. and P. Mira (2006) implement one version of this solution. They allow for an unobserved fixed effect which varies across markets and assume both that the fixed effect can only take on a finite number of values and that the transition probabilities for the observed exogenous variables are independent of the values of the fixed effect. They then solve for an invariant distribution of the state of the system and the fixed effect, form the conditional distribution of the initial condition given the fixed effect, and integrate out over possible values of the fixed effect. They report that allowing for the fixed effect has a noticeable impact on their empirical results.

Of course if one has a reasonably long panel of markets one should be able to estimate the fixed effect (or some other unobserved initial condition) pointwise; our fourth solution possibility. In that case continuation values could be estimated as described above but separately for each market in the data. The observations across markets could then be pooled together in the second stage in order to estimate the structural parameters that are the same across markets. This would lead to substantially higher estimation error in the continuation values, and one might want to think hard about estimators that would be designed to minimize the impact of these errors[62]. Some Monte Carlo work on just how long a panel is likely to be required for this procedure to be fruitful would be extremely helpful.

[62]Note that this process could also solve the multiple equilibrium problem that might exist across markets.

References.

Abreu, D., D. Pearce and E. Stacchetti (1986), “Optimal Cartel Equilibria with Imperfect Monitoring,” *Journal of Economic Theory*, vol. 39, no. 1, pp. 251-259.

Ackerberg, D. (2003), ”Advertising Learning and Customer Choice in Experience Good Markets: A Structural Empirical Examination,” *International Economic Review*, vol. 44, no. 3, pp. 1007-1040.

Ackerberg, D., K. Caves and G. Fraser (2004), ”Structural Identification of Production Functions,” mimeo, UCLA.

Ackerberg, D. and A. Pakes (2005), ”Notes on Relaxing the Scalar Unobservable Assumption in Olley and Pakes,” mimeo, Harvard University.

Aguirregabiria, V. and P. Mira (2006), “Sequential Estimation of Dynamic Games,” *Econometrica*, forthcoming.

Ahn, H., and J. Powell (1993), ”Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism,” *Journal of Econometrics*, vol. 58, no. 1-2, pp. 3-29.

Arellano, M. and S. Bond (1991), ”Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations,” *Review of Economic Studies*, vol. 58, pp. 277-97.

Arellano, M. and O. Bover (1995), ”Another Look at the Instrumental Variable Estimation of Error-Components Models,” *Journal of Econometrics*, vol. 68, no. 1, pp. 29-51.

Bajari, P. and L. Benkard (2005), ”Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach,” *Journal of Political Economy*, vol. 113, no. 6, pp. 1239-1276.

Bajari, P., L. Benkard and J. Levin (2006), “Estimating Dynamic Models of Imperfect Competition,” *Econometrica*, forthcoming.

Benkard, L. (2000), “Learning and Forgetting: The Dynamics of Aircraft Production,” *American Economic Review*, vol. 90, no. 4, pp. 1034-54.

Benkard, L. (2004), “A Dynamic Analysis of the Market for Wide-Bodied Commercial Aircraft,” *Review of Economic Studies*, vol. 71, no. 3, pp. 581-612.

Berry, S. (1994), ”Estimating Discrete Choice Models of Product Differentiation,” *RAND*, vol. 25, no. 2, pp. 242-262.

Berry, S., J. Levinsohn and A. Pakes (1995), ”Automobile Prices in Market Equilibrium,” *Econometrica*, vol. 63, no. 4, pp. 841-890.

Berry, S., J. Levinsohn and A. Pakes (1999, ”Voluntary Export Restraints on Automobiles: Evaluating a Trade Policy,” *American Economic Review*, vol. 89, no. 3, pp. 400-430.

Berry, S. and A. Pakes (2002), “Estimation from the Optimality Conditions for Dynamic Controls,” mimeo, Yale University.

Berry, S. and A. Pakes (2005), ”The Pure Characteristics Demand Model,” mimeo, Yale University.

Berry, S., J. Levinsohn and A. Pakes (2004), ”Estimating Differentiated Product Demand Systems from a Combination of Micro and Macro Data: The New Car Model,” *Journal of Political Economy*, vol. 112, no. 1, pp. 68-105.

Berry, S., O. Linton and A. Pakes (2004), ”Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems,” *Review of Economics Studies*, vol. 71, no. 3, pp. 613-654.

Blundell, R. and S. Bond (1999), ”GMM Estimation with Persistent Panel Data: An Application to Production Functions,” The Institute for Fiscal Studies, Working Paper Series No. W99/4.

Bresnahan, T. (1981), ”Departures from Marginal-Cost Pricing in the American Automobile Industry: Estimates for 1977-1978,” *Journal of Econometrics*, vol. 17, no. 2, pp. 201-227.

Bresnahan, T. (1982), “The Oligopoly Solution Concept is Identified,” *Economics Letters*, vol. 10, pp. 87-92.

Bresnahan, T. (1987), ”Competition and Collusion in the American Automobile Industry: The 1955 Price War,” *Journal of Industrial Economics*, vol. 35, no. 4, pp. 457-82.

Buettner, T. (2004a), ”Productivity Dynamics When Labor is a State Variable,” mimeo, LSE.

Buettner, T. (2004b), ”R&D and the Dynamics of Productivity,” mimeo, LSE.

Caballero, R. and E. M. R. A. Engel (1999), “Explaining Investment Dynamics in U.S. Manufacturing: A Generalized (S,s) Approach,” *Econometrica*, vol. 67, no. 4, pp. 783-826.

Chamberlain, G. (1984), ”Data Panel,” in K. Arrow and M. Intriligator eds. ”Handbook of Econometrics,” Elsevier Science, vol. 2, chp. 22, pp. 1247-1318.

Chamberlain, G. (1987), ”Asymptotic Efficiency in Estimation with Conditional Moment Restrictions,” *Journal of Econometrics*, vol. 34, no. 3, pp. 305-334.

Chan, T. (2002), ”Demand for Soft Drinks: Characteristics, Corners and Continuous Choice,” mimeo, Washington University, Marketing Department.

Chernozhukov, V., H. Hong and E. Tamer (2004), “Parameter Set Inference in a Class of Econometric Models,” mimeo, Princeton University.

Court, A. (1939), ”Hedonic Price Indexes with Automotive Examples,” in The Dynamics of Automobile Demand, General Motors Corporation, Detroit, pp. 99-117.

Crawford, G. and M. Shum (2005), ”Uncertainty and Learning in Pharmaceutical Demand,” *Econometrica*, forthcoming.

Das, M., S. Olley and A. Pakes (1996), ”The Evolution of the Market for Consumer Electronics,” mimeo, Harvard University.

Davis, S. and J. Haltwanger (1992), "Gross Job Creation, Gross Job Destruction, and Employment Reallocation" *Quarterly Journal of Economics*, vol. 107, no. 2, pp. 819-64.

Deaton, A. and J. Muellbauer (1980), "An Almost Ideal Demand System," *American Economic Review* vol. 70, no. 3, pp. 312-26.

Department of Justice (1992), "U.S. Merger Guidelines."

Doraszelski, U. and S. Markovich (2004), "Advertising Dynamics and Competitive Advantage," mimeo, Harvard University.

Doraszelski, U. and K. Judd (2004), "Solution Methods for Markov Perfect Nash Equilibria of Continuous Time, Finite State Stochastic Games," mimeo, Hoover Institution.

Doraszelski, U. and M. Satterthwaite (2004), "Foundations of Markov-Perfect Industry Dynamics: Existence, Purification, and Multiplicity," mimeo, Hoover Institution.

Doraszelski, U. and A. Pakes (2006); "A Framework for Dynamic Analysis in Applied I.O; Computation."

Dubin, J. and D. McFadden, 1984, "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, vol. 52, no. 2, pp. 345-362.

Dunne,T., S. Klimek, M.Roberts and Y. Xu (2005), "Entry and Exit in Geographic Markets", mimeo, Pennsylvania State Univeristy.

Dunne, T., M. Roberts and L. Samuelson (1988), "Patterns of Firm Entry and Exit in U.S. Manufacturing Industries," *Rand Journal of Economics*, vol. 19, no. 4, pp. 495-515.

Einav, L. (2003), "Not All Rivals Look Alike: Estimating an Equilibrium Model of the Release Date Timing Game", mimeo, Stanford University.

Erdem, T. and M. Keane (1996), "Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science*, vol. 15, no. 1, pp. 1-20.

Erdem. T., S. Imai and M. Keane (2003), "Brand and Quantity Choice Dynamics Under Price Uncertainty," *Quantitative Marketing and Economics*, vol. 1, no. 1, pp. 5-64.

Ericson, R. and A. Pakes (1995), "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies*, vol. 62, no. 1, pp. 53-83.

Esteban, S. and M. Shum (2002), "Durable Good Monopoly with Secondary Markets: The Case of Automobiles," mimeo, Johns Hopkins.

Fershtman, C. and A. Pakes (2000), "A Dynamic Oligopoly with Collusion and Price Wars," *RAND Journal of Economics*, vol. 31, vol. 2, pp. 207-236.

Fershtman, C. and A. Pakes (2005), "Finite State Dynamic Games with Asymmetric Information: A Computational Framework," mimeo, Harvard University.

Gentzkow, M. (2004), "Valuing New Goods in a Model with Complementarity: Online Newspapers," mimeo, Chicago GSB.

Gorman, W. (1959), "Separable Utility and Aggregation," *Econometrica*, vol. 27, pp. 469-481.

Gowrisankaran, G. (1995), "A Dynamic Analysis of Mergers," Ph.D. Dissertation, Department of Economics, Yale University.

Gowrisankaran, G. and R. Town (1997), "Dynamic Equilibrium in the Hospital Industry," *Journal of Economics and Management Strategy*, vol. 6, no. 1, pp. 45-74.

Green, E. and R. Porter (1984), "Non-cooperative Collusion Under Imperfect Price Information," *Econometrica*, vol. 52, pp. 87-100.

Greenstreet, D. (2005), "Exploiting Sequential Learning to Estimate Establishment-Level Productivity Dynamics and Decision Rules," mimeo, Univ. of Michigan.

Griliches, Z. (1961), "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change," in The Price Statistics of the Federal Government, NBER, New York.

Griliches, Z. and J. Hausman (1986), "Errors in Variables in Panel Data," *Journal of Econometrics*, vol. 31, no. 1, pp. 93-118.

Gronau, R. (1974), "Wage Comparisons-A Selectivity Bias," *Journal of Political Economy*, vol. 82, no. 6, pp. 1119-43.

Guerre, E., I. Perrigne and Q. Vuong (2000), "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, vol. 68, no. 3, pp. 525-574.

Hansen, L. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, vol. 50, no. 4, pp. 1029-1054.

Hausman, J. (1996), "Valuation of New Goods under Perfect and Imperfect Competition," in T. Bresnahan and R. Gordon eds. The Economics of New Goods, University of Chicago Press, Chicago, chapter 5, pp. 209-248.

Heckman, J. (1974), "Shadow Prices, Market Wages, and Labor Supply," *Econometrica*, vol. 42, no. 4, pp. 679-694.

Heckman, J. (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic & Social Measurement*, vol. 5, no. 4, pp. 475-492.

Heckman, J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, vol. 46, no. 4, pp. 931-959.

Heckman, J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, vol. 47, no. 1, pp. 153-161.

Heckman, J. (1981), "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time Discrete Data Stochastic Process," Chp. 4 in C. Manski and D. McFadden eds. Structural Analysis of Discrete Data and Econometric Applications, MIT press.

Heckman, J. and R. Robb (1985), "Alternative Methods for Evaluating the Impact of Interventions: An Overview," *Journal of Econometrics*, vol. 30, no. 1-2, pp. 239-267.

Heckman, J. and B. Honore (1990), “The Empirical Content of the Roy Model,” *Econometrica*, vol. 58, pp. 1121-1149.

Heckman, J. and C. Snyder (1997), “Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators,” *RAND Journal of Economics*, vol. 28(0), S142-S189.

Hendel, I. (1999), ”Estimating Multiple-Discrete Choice Models: An Application to Computerization Returns,” *Review of Economic Studies*, vol. 66, no. 2, pp. 423-446.

Hendel, I. and A. Nevo (2002), ”Sales and Consumer Inventory,” NBER working paper # 9048.

Hoch, I. (1962), ”Estimation of Production Parameters Combining Time-Series and Cross-Section Data,” *Econometrica*, vol. 30, no. 1, pp. 34-53.

Houthakker, H.S. (1955), ”The Pareto Distribution and the Cobb-Douglas Production Function,” *Review of Economic Studies*, vol. 23, no. 1, pp. 27-31.

Hotelling, H. (1929), ”Stability in Competition,” *Economic Journal*, vol. 39, pp. 41-57.

Hotz, V. and R. Miller (1993), “Conditional Choice Probabilities and the Estimation of Dynamic Models,” *Review of Economic Studies*, vol. 60, no. 3, pp. 497-529.

Hotz, V., R. Miller, S. Sanders and J. Smith (1993), “A Simulation Estimator for Dynamic Models of Discrete Choice,” *Review of Economic Studies*, vol. 60, pp. 397-429.

Imbens, G. and T. Lancaster (1994), ”Combining Micro and Macro Data in Microeconometric Models,” *The Review of Economic Studies*, vol. 61, no. 4, pp. 655-680.

Jofre-Bonet, M. and M. Pesendorfer (2003), “Estimation of a Dynamic Auction Game,” *Econometrica*, vol. 71, no. 5, pp. 1443-1489.

Judd, K. (1998), Numerical Methods in Economics, MIT Press, Cambridge, MA.

Keane, M. and K. Wolpin (1997), “The Career Decisions of Young Men,” *Journal of Political Economy*, vol. 105, no. 3, pp. 473-522.

Lancaster, K. (1971), Consumer Demand: A New Approach, Columbia University Press, New York.

Lau, L. J. (1982), “On Identifying the Degree of Competitiveness from Industry Price and Output Data,” *Economics Letters*, vol. 10, pp. 93-99.

Levinsohn, J. and A. Petrin (2003), ”Estimating Production Functions Using Inputs to Control for Unobservables,” *The Review of Economic Studies*, vol. 70, no. 2, pp. 317-341.

Magnac, T. and D. Thesmar (2002), “Identifying Dynamic Discrete Decision Processes,” *Econometrica*, vol. 70, no. 2, pp. 801-816.

Manski, C. and S. Lerman (1977), ”The Estimation of Choice Probabilities from Choice Based Samples,” *Econometrica*, vol. 45, no. 8, pp. 1977-88.

Marschak, J. and W.H. Andrews (1944), ”Random Simultaneous Equations and the Theory of Production,” *Econometica*, vol. 12, no. 3-4, pp. 143-205.

McFadden, D. (1974), ”Conditional Logit Analysis of Qualitative Choice Behavior,” in P. Zarembka eds. Frontiers of Econometrics, Academic Press, New York.

McFadden, D. (1981), ”Econometric Models of Probabilistic Choice,” in C. Manski and D. McFadden, eds. Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, MA.

Mundlak, Y. (1961), ”Empirical Production Function Free of Management Bias,” *Journal of Farm Economics*, vol. 43, no. 1, pp. 44-56.

Nadiri, N. and S. Rosen (1974), A Disequilibrium Model of Demand for Factors of Production, the National Bureau of Economic Research, New York.

Nevo, A. (2001), ”Measuring Market Power in the Ready-to-Eat Cereal Industry,” *Econometrica*, vol. 69, no. 2, pp. 307-342.

Olley, S. and A. Pakes (1996), "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, vol. 64, no. 6, pp. 1263-1298.

Pakes, A. and Z. Griliches (1984), " Estimating Distributed Lags in Short Panels with an Applicaton to the Specification of Depereciation Patterns and Capital Stock Constructs" *Review of Economic Studies*, vol. 51, pp. 243-262.

Pakes, A. (1986), "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, vol. 54, pp. 755-784.

Pakes, A. and D. Pollard (1989), "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, vol. 57, no. 5, pp. 1027-1057.

Pakes, A. (1994), "The Estimation of Dynamic Structural Models: Problems and Prospects, Part II. Mixed Continuous-Discrete Control Models and Market Interactions" J.J. Laffont and C. Sims eds, Advances in Econometrics: Proceedings of the 6th World Congress of the Econometric Society, Chapter 5.

Pakes, A. and P. McGuire (1994), "Computing Markov-Perfect Nash Equilibria: Numerical Implications of a Dynamic Differentiated Product Model," *Rand Journal of Economics*, vol. 25, no. 4, pp. 555-89.

Pakes, A. and S. Olley (1995), "A Limit Theorem for a Smooth Class of Semiparametric Estimators", *Journal of Econometrics* vol. 65, no.1, pp. 295-332.

Pakes, A. (1995), "Biases in the CPI," testimony before the Senate Finance Committee, the *United States Congress*.

Pakes, A. (1998), "A Framework for Dynamic Analysis in Applied I.O.," N.B.E.R. discussion paper No. 8024.

Pakes, A. and P. McGuire (2001), "Stochastic Approximation for Dynamic Models: Markov Perfect Equilibrium and the 'Curse' of Dimensionality," *Econometrica*, vol. 69, no. 5, pp. 1261-81.

Pakes, A. (2004), "Common Sense and Simplicity in Industrial Organization," *Review of Industrial Organization*, vol. 23, pp. 193-215.

Pakes, A., M. Ostrovsky and S. Berry (2006), “Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry/Exit Examples),” *RAND Journal of Economics*, forthcoming.

Pakes, A., J. Porter, K. Ho and J. Ishii (2006), ”Moment Inequalities and Their Application,” mimeo, Harvard University.

Pavcnik, N. (2002) ”Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants,” *Review of Economic Studies* January, vol. 69, pp. 245-76.

Pesendorfer, M. and P. Schmidt-Dengler (2003), “Identification and Estimation of Dynamic Games,” NBER working paper # 9726.

Petrin, A. (2002), ”Quantifying the Benefits of New Products: The Case of the Minivan,” *Journal of Political Economy*, vol. 110, no. 4, pp. 705-729.

Robinson, P. (1988), ”Root- N-Consistent Semiparametric Regression,” *Econometrica*, vol. 56, no. 4, pp. 931-954.

Rosenbaum, P. and D. Rubin (1983), ”The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika*, vol. 70, no. 1, pp. 41-55.

Rotemberg, J. and G. Saloner (1985), “A Supergame-Theoretic Model of Business Cycles and Price Wars During Booms,” *American Economic Review*, vol. 76, pp. 390-407.

Rubinstein, R. (1981), Simulation and the Monte Carlo Method, John Wiley and Sons, New York.

Rust, J. (1994), “Structural Estimation of Markov Decision Processes,” in R. F. Engle and D. L. McFadden (eds.), Handbook of Econometrics, Volume IV, Elsevier Science, Amsterdam.

Ryan, S. (2005), ”The Costs of Environmental Regulation in a Concentrated Industry,” mimeo, Duke University.

Salop, S. (1979), ”Monopolistic Competition with Outside Goods,” *RAND*, vol. 10, no. 1, pp. 141-156.

Shaked, A. and J. Sutton (1982), ”Relaxing Price Competition Through Product Differentiation,” *Review of Economic Studies*, vol. 49, no. 1, pp. 3-13.

Schmidt-Dengler, P. (2003), “The Timing of New Technology Adoption: The Case of MRI” mimeo, Yale University.

Sobel, J. (1984), “The Timing of Sales,” *The Review of Economic Studies,* vol. 51, no. 3, pp. 353-368.

Song, M. (2004), ”Measuring Consumer Welfare in the CPU Market: An Application of the Pure Characteristics Demand Model,” mimeo, Georgia Tech.

Timmins, C. (2002), “Measuring the Dynamic Efficiency Costs of Regulators’ Preferences: Municipal Water Utilities in the Arid West,” *Econometrica*, vol. 70, no. 2, pp. 603-629.

Train, K. (2003), Discrete Choice Models With Simulation, Cambridge University Press.

Weintraub, G., C.L. Benkard and B. Van Roy (2005), “Markov Perfect Industry Dynamics with Many Firms,” mimeo, Stanford University.

Wooldridge, J. (2002), Econometric Analysis of Cross Section and Panel Data, MIT Press, Cambridge, MA.

Wooldridge, J. (2004), ”On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables,” mimeo, Michigan State University.