Behavioral and Descriptive Forms of Choice Models[*]

Ariel Pakes[†]

April 1, 2014

Abstract

Empirical work on choice models, especially on relatively new topics or data sets, often starts with descriptive, or what is colloquially referred to as "reduced form", results. Our descriptive form formalizes this process. It is derived from the underlying behavioral model, has an interpretation in terms of fit, and can sometimes be used to quantify biases in agents' expectations. We consider estimators for the descriptive form of discrete choice models with interacting agents that take account of approximation errors as well as unobservable sources of endogeneity. We conclude with an investigation of descriptive forms for two period entry models.

Running Head: Forms of Choice Models

1   Introduction.

Empirical work on choice models, especially work on relatively new topics or data sets, frequently starts with descriptive, or what is often colloquially referred to as "reduced form", results. Authors often rightly stress that the parameters obtained from this part of the analysis are not meant to have a behavioral interpretation. Rather the descriptive results are viewed as suggestive of relationships that any appropriate model should generate.

This paper begins by providing a simple formalization of this process. It results in what I call the "descriptive form" of a choice model[1]. Though the descriptive form is derived from a behavioral model, its parameters will not have a behavioral interpretation. On the other hand the parameters of the descriptive form will have an interpretation in terms of fit. This will endow the descriptive form with a meaning in the sense that we will understand what differences in its value imply. It will also make them comparable across data sets making it easier to tell where and why modeling differences

---

[1] Initially I called this form an "interpretable reduced form". I changed my terminology so as not to conflict with other usage of the term "reduced form". However the adjective "interpretable", is the relevant distinction between the colloquial usage of term reduced form and what I am proposing here.

are required.

Section 3 of the paper considers the relationship between the descriptive form and the behavioral (or "structural") form of a choice model. This clarifies a number of conceptual points. Perhaps most important among them is the relationship between the expectation operator an agent uses in determining which course of action to take and the expectation operator emanating from the data generating process. We show that sometimes we can test whether the conditional mean of an agent's expectations are unbiased (conditional on a subset of the variables the agent knew at the time the choice was made). The test does not require us to specify any aspect of the behavioral model other than that the agent makes the choice that maximizes the agent's perceived value. In particular it does not require us to specify the contents of agents' information sets or their impact on the agent's perceptions (or "priors"); aspects of the specification the researcher typically knows little about. When the test fails, that is when we find that the agent's expectations are biased, the test results quantify the bias in the agent's expectations as functions of the variables that the test conditions

on.

The general discussion applies to both single agent and multiple agent problems, but the last three sections of the paper focus on the analysis of discrete choice models with interacting agents. These are discrete choice models where the outcome of one agent's choice depends on the actions of other agents. Both our theoretical and econometric models for interacting agent problems are at a relatively early stage of development. So these are cases where the descriptive form is particularly relevant.

The econometric treatment of interacting agent discrete choice models has been a topic of discussion for some time (see, for e.g., Heckman's, (1978), discussion of the "coherence" problem). Importantly, Tamer (2003) noted the connection of this literature to the phenomena of multiple equilibria that had been discussed extensively in the economic theory literature. The econometric analysis of models with multiple equilibria often has to deal with the fact that the empirical researcher does not have the information needed to focus on a particular equilibrium. We note that a simple set estimator, introduced by Pakes and Porter (2013), for the parameters of

the descriptive form of these models has desirable properties and compare it to other estimators recently introduced for this problem. Our estimator allows for the approximation errors generated by the construction of the descriptive form as well as the endogeneity correction that prior econometric research in this area has focused on.

This part of the paper contains an example. The example starts by computing a sequential Markov Perfect equilibria to a market. The equilibrium values for alternative actions and the optimal policies are then used to compute its true descriptive form. Proceeding in this way makes it clear why two period models designed to summarize the data on the impact of the number and type of competitors on the value of being active need to allow for both a structural error that is correlated with the right hand side variables of interest and for approximation errors which are uncorrelated (or mean independent) of the right hand side variables by construction. We then show that the estimator introduced in Pakes and Porter (2013) is able to accommodate both types of errors under quite weak assumptions. The paper concludes by using simulated data from the computed equilib-

rium in a Monte Carlo study of the performance of alternative estimators for the descriptive form of this model.

2   Behavioral and Descriptive Forms.

The behavioral form. We can not investigate the relationship between statistics that summarize the implications of an appropriate behavioral model and that model without specifying the behavioral model. So we start with a "high level" set of assumptions on the behavioral model.

Our agent choses a $d \in \mathcal{D}$, and we are after a summary of how variables in the agent's environment affect its choice. I will focus on the case where $\# \mathcal{D}$ is finite (discrete choice)[2]. $V(d, \omega)$ will denote the value that would be generated by the decision for any $d \in \mathcal{D}$ and state of the world $\omega \in \Omega$. $\mathcal{J} \in J$ will represent the agent's information set at the time the decision is made and since the decision is made prior to the realization of the outcome, $J \subset \Omega$. $\mathcal{E}[\cdot|\mathcal{J}]$ will provide the agent's expectation conditional on that

---

[2]The notion and use of the descriptive form do not depend on the form of the choice set, but implementation details do. In particular special issues arise in using it for models with discrete choice sets.

information set[3]. Then

$$V(d, \mathcal{I}) = \mathcal{E}[V(d, \omega)|\mathcal{I}], \qquad (1)$$

is the agent's perception of the value of the outcome from choosing "d" at the time the choice is made.

Notice that $V(d, \mathcal{I})$ needs to be specified for values of d that the agent does not chose. To describe the determinants of the agent's choice we will have to describe the determinants of the perceived values of the actions not taken (i.e. for counterfactual choices or values for actions "off the equilibrium path"). Finally we assume

$$d(\mathcal{I}) = argmax_{d \in \mathcal{D}} V(d, \mathcal{I}), \qquad (2)$$

so we are assuming, as in Savage (1956), that agent's decisions maximize the expected (discounted) returns from their perceptions of the values generated by different outcomes.

___
[3]I will assume that $V(\cdot)$ is finite everywhere and with the addition of assumptions introduced below this will imply that all expectations exist.

The descriptive form. Our goal is to summarize the impact of different variables on $V(d, \mathcal{J})$ using statistics that both have an intrinsic meaning and can be compared across data sets. If $E(\cdot|\cdot)$ provides the conditional expectations generated by the true data generating process (or DGP), our descriptive summary is given by $g_d(x)$ where

$$g_d(x) \equiv E[V(d, \mathcal{J})|x]. \tag{3}$$

$g_d(x)$ is our descriptive form. Notice that it has the properties that $\forall d \in D$

$$V(d, \mathcal{J}_i) = g_d(x) + \epsilon_d, \;\; \text{with} \;\; E[\epsilon_d|x] = 0, \;\; \text{and} \;\; \epsilon_d \text{ measurable } \mathcal{J}. \tag{4}$$

$g_d(\cdot)$ has a familiar statistical interpretation. If it is the conditional mean (or regression function) of the agent's evaluations, then our descriptive form is the minimum mean square error predictor of those evaluations. Often researchers restrict $g_d(x)$ to be linear, in which case the descriptive form is the best linear predictor of those evaluations[4].

---

[4]Linearity and parsimony in the choice of x often enables us to focus in on characteristics of $V(d, \mathcal{J}_i)$ that can easily be kept in the mind of the researcher and the subsequent readers of the paper. On the other hand most of the technical detail in the paper would go through if, instead, we used a non-parametric estimate of $E[V(d, \mathcal{J}_i)|x]$, it would just be much harder to interpret the output of the estimation routine.

Note that, as described here, the descriptive form provides a summary of the correlates of the agent's perceptions of the value of a choice rather than the DGP's expected value of the choice. We investigate the relationship between these two values below. One advantage of distinguishing the agent's perceptions from the expectations generated by the DGP is that it will allow us to compare perceptions to outcomes.

There are three properties of the descriptive form that I want to emphasize as they underlie much of the discussion below.

- First

$$E[V(d,\omega)|\mathcal{I}] \neq \mathcal{E}[V(d,\omega)|\mathcal{I}] = V(d,\mathcal{I}),$$

  i.e. the agent's expectations need not equal the averages generated by the DGP, and the difference may be of interest.

- Second, the disturbances, i.e. the $\{\epsilon_d\}_{d\in\mathcal{D}}$, enter the descriptive form in an *additively separable* way and are either mean independent or orthogonal to the included variables *by construction.*

- These disturbances are known to the agent when the choice is made (i.e. $\{\epsilon_d\}_{d\in\mathcal{D}} \in \mathcal{I}$); it is only the analyst that does not know their

10

values.

Finaly keep in mind that when we apply the expectation operator $E[V(\cdot)|x]$ we average over agents with the same x in a *given* data set. Different data sets can generate different descriptive forms for many reasons; the distribution of information sets, agents' perceptions given those information sets, or the relationship of either to x may differ across environments. If $g_d(x)$ does differ across two data sets it will typically be of interest to understand why. Just as the function $g_d(x)$ might provide clues as to how the behavior in a given data set varies with x, differences in $g_d(x)$ across data sets might provide clues as to how different environments impact behavior. Of course to obtain a coherent explanation of behavioral differences we will need to understand what determines $V(\cdot)$. The descriptive form is only meant as a tool to this end. It provides a summary of what an appropriate model for $V(\cdot)$ must rationalize.

Familiar Examples. There is quite a bit of empirical work that can best be understood as presenting results on the descriptive form of a behavioral model, rather than on the behavioral model per se.

Perhaps the most obvious example is the use of two period models as a framework for summarizing relationships in the data before going on to a more realistic, though typically far more complicated, sequential models of behavior. A familiar example, and one we return to below, is the two-period entry model used in Industrial Organization. Here the goal is typically to summarize the impact of the number and types of firms on the value of being active (for a review of the literature, see Berry and Tamer, 2007). An important more recent strand in the I.O. literature has used two period models to document how different covariates are related to the profit splits in vertical markets (see, for e.g. Ho, 2009; Crawford and Yurukoglu, 2012; and Grennan, 2013).

Another example where descriptive models have been of extensive use are choice models where the opinions of multiple agents go into the decision making process in an unknown way. Examples include choices that affect more than one member of a family (see, chapter 5 in Browning, et. al., forthcoming) and hospital choices (determined by doctors patients and insurers; see Ho and Pakes, forthcoming). A familiar goal here is to deter-

mine the relationship between features of the environment (e.g.s; sources of income in family choice and features of insurance contracts in hospital choice) and the choice made.

Descriptive models are often also used in single agent decision making problems. This is particularly true when there are dynamic implications of the choice (for example the choice of durable, network, experience or storable goods) and/or when there is extensive uncertainty and the researcher does not have a lot of information on the contents of the agents' information sets or the form of its prior distributions (e.g. the choice of a place to shop at when the prices and goods at different outlets are uncertain).

In all of these cases it is possible to formulate and estimate a behavioral model. Typically however, with the data and prior information available, that model would require restrictive assumptions and permutations of those assumptions would require considerable programing and computational time. As a result the researcher often starts with summaries of the data which illuminate the aspects of the problem that the behavioral

model should capture. The challenge then becomes to construct a model that both rationalizes the dominant features of the summary statistics and attaches magnitudes to effects of interest. The descriptive form I am proposing is a way of formalizing the first step and examining the implications of that formalization. Its advantage is that the coefficients estimated from it have an interpretation and can be compared across data sets.

Descriptive forms and approximations to behavioral models. The descriptive form is not meant to be an approximation to a behavioral model. The distinction lies in the properties of the disturbance and the related issue of the interpretation of the estimates. To see this consider the issues that would arise were we to write down a behavioral model, linearize it, and take the linear form to data in an attempt to get the "first order effect" of a particular variable on an outcome. Typically the first question that would arise would be the relationship between the included controls and the disturbance term; in particular do we need instruments? The descriptive form, on the other hand, has a mean independent residual by construction.

If we want a descriptive form that conditions on a factor in the envi-

ronment that is not captured by the covariates in the data, then we have to create an observable that captures what we want to condtion on. For example a panel data study might want to condition on individual specific constant terms in order to generate a description of the variables that are related to changes in an agent's behavior over time. However if we are to add structure to a descriptive form we must be explicit about exactly the structure we are adding; that is we need to know what we are conditioning on to understand what we are describing. In practice, the distinction between a descriptive form and an approximation to a behavioral model has lead to some confusion. I will illustrate with the development of two-period entry models in Industrial Organization (though many of the analytic frameworks used in labor and/or public finance are subject to similar criticisms).

The two-period entry models are attempts to describe the factors that are related to the value of being active in a market. These models focused on the impact of the number of different types of competitors on that value as this was a particular concern of the anti-trust authorities. Early stud-

ies that used the number of active firms as a determinant of the value of being active often found the number of active firms was positively related to that value. This because more profitable markets induced more active firms and researchers could not condition on all the determinants of market profitability. Though a two period model is clearly a descriptive and not a behavioral model, the early entry models did not describe what the researchers wanted to describe.

The reaction to these results included a new generation of two period entry models that worried about "endogeneity". An oft-cited starting point is Bresnahan and Reiss (1991). Their model did not allow for any heterogeneity among firms but subsequent work incorporated varying degrees of heterogeneity, culminating in the recent work of Ciliberto and Tamer (2009) which allows for arbitrary heterogeneity among firms. All these models recognized the problem created by unobserved determinants of market profitability. Different models endogenized the errors in different ways but they all went to the opposite extreme; they assumed that the two-period model, when combined with a simple (usually linear) functional form, was the true

value of being active.

We come back to this in section 4 where we compare these two period behavioral models to descriptive models for the value of being active. Both formulations try to account for unobservable determinants of market profitability. The conceptual differences are that; (i) the two period behavioral models do not allow for the approximation error which emanates from a descriptive form (the $\{\epsilon_d\}$ in equation 4), and (ii) the behavioral models do not specify the covariance structure of disturbances a priori, and hence are not able to be explicit about what one is conditioning on when describing the estimated effects. In addition there are fairly significant computational differences between the two approaches.

There are two remaining parts of the paper. First we clarify the relationship between the behavioral and descriptive form of a model. This is a bit of a digression from the main purpose of this paper which is to define the descriptive form and provide an estimator for it in discrete choice settings. However the digression does result in some useful insights. In particular we show that it is sometimes possible to examine properties of

the agent's expectation operator without specifying the behavioral model that underlies the agent's decisions. In particular we will be able to test whether the agent's expectations are unbiased, and if they are biased we can consider what that bias is a function of. We then return to econometric issues that arise in estimating the descriptive form of discrete choice decision problem.

3 Digression: Behavioral vs Descriptive Forms.

One advantage of having explicit definitions for both the descriptive and behavioral form is that it allows us to examine the relationship between them. This section explores that relationship.

The descriptive form depends on the expectations emanating from the data generating process, and the behavioral form depends on the agent's expectation operator. So the relationship between them will depend on the relationship between the agent's expectation of the returns from its decisions (our $V(d, \mathcal{J})$) and the expectation of the actual realizations of those returns (of $V(d, \omega)$) resulting from the DGP. Recalling that $E[\cdot|\cdot]$ is

18

our notation for the conditional expectations generated by the DGP, since $\mathcal{J} \subset \Omega$ we can write

$$V(d, \omega) = E[V(d, \omega)|\mathcal{J}] + \xi_d, \text{ with } E[\xi_d|\mathcal{J}] = 0.$$

Further for any x we can write

$$E[V(d, \omega)|\mathcal{J}] = f_d(x) + u_d \text{ with } E[u_d|x] = 0.$$

So

$$V(d, \omega) = f_d(x) + u_d + \xi_d, \tag{5}$$

with

$$\xi_d \in \Omega, \text{ and } E[\xi_d|\mathcal{J}] = 0, \quad \text{while} \quad u_d \in \mathcal{J}, \text{ and } E[u_d|x] = 0.$$

The behavioral form depends on the agent's expectations conditional on $\mathcal{J}$ and the descriptive form depends on the averages emanating from the DGP conditional on the researchers choice of explanatory variables or x. Hence the relationship between them depends on whether or not $x \in \mathcal{J}$.

If $x \notin \mathcal{J}$, or x was not known to the agent when its decision was made, then the decision could not have been a function of the realization of x. As a result $g_d(x) = E[V(d, \mathcal{J})|x]$ neither provides an interpretable description of

the agent's choice behavior (in the sense defined above), nor the behavioral form of the choice equation[5]. So we focus on the case in which x is known when d is chosen, or $x \in \mathcal{J}$.

Assume $x \in \mathcal{J}$. Choices are made based on agent's perceptions of the likely outcomes from those choices. So the descriptive form of the choice model is derived from agent's beliefs on likely outcomes. The actual outcomes depends on the DGP which may or may not be consistent with those beliefs. We now consider what we can learn about the relationship between perceptions and the actual process generating the data by comparing the descriptive form of the choice equation to the outcome equation. Note that this comparison does not depend on the actual form of the true behavioral model (the specification for information sets, priors, and functional forms for primitives), but rather just on the notion that agents make choices based on their own perceptions of what is best for them.

We compare the information obtained from estimating the descriptive

---

[5]The actual outcome, $V(\omega, d)$, can depend on an $x \notin \mathcal{J}$, but to use that information in estimation in a consistent way requires additional assumptions on how expectations are formed, and our goal in this section is to see what we can learn from the data about how those expectations are formed.

form of the choice model in equation (4), with the information that can be obtained from the outcome, or equation (5). Recall that the descriptive form of the choice model is $\mathcal{E}[V(d, \Omega)|\mathcal{I}] = g_d(x) + \epsilon_d$, $\epsilon_d \in \mathcal{I}$, and $E[\epsilon_d|x] = 0$, while the outcomes are given by $V(d, \omega) = f_d(x) + u_d + \xi_d$, $u_d \in \mathcal{I}$, and $E[u_d|x] = 0$. Since $E[\xi_d|x, \epsilon] = 0$, the relationship between between the expected value generated by the DGP and the agent's expectations (which determines choices) depends on the relationship between

- $f(x) \equiv [f_1(x), \ldots, f_D(x)]$ to $g(x) \equiv [g_1(x), \ldots, g_D(x)]$, and of

- $\epsilon \equiv [\epsilon_1, \ldots, \epsilon_D]$ to $u \equiv [u_1, \ldots, u_D]$.

Definitions: Weak and Strong Rationality.

I will say an agent is weakly rational w.r.t. an $x \in \mathcal{I}$ if

$$f(x) = g(x), \quad (a.e.\ x) \tag{6}$$

and I will say an agent is strongly rational if

$$f(x) = g(x) \quad \text{and} \quad \epsilon(\mathcal{I}) = u(\mathcal{I}), \quad (a.e.\ \mathcal{I}). \quad \spadesuit \tag{7}$$

If an agent is weakly rational and $x \in \mathcal{I}$ then the descriptive model, which is always an unbiased predictor of the agent's expectations, is also an

21

unbiased predictor of the outcomes. If the agent is strongly rational, it has unbiased expectations *and* its decisions are optimal in the sense that they maximize the expected values emanating from the DGP. Next we consider what can be learned on the relationship of perceptions to realizations when both choice and outcome data are available[6].

3.1 Tests of rationality when $\mathcal{J}$ is unknown.

It is easiest to explain the issues that arise in quantifying agents' biases if we assume the researcher can observe the outcome up to an error which is mean zero conditional on x. This is often the case when the outcomes of interest are measured in monetary units. Below we generalize and allow for outcomes that are only observed up to both a mean zero error and a parameter vector that needs to be estimated. This would typically be the case if outcomes were measured in "utils" and agents' choices maximize their evaluations of the likely utility from their actions.

---

[6]One could study how perceptions are formed more directly by eliciting information on beliefs. For a review of the literature on eliciting beliefs, see Manski (2004). Hendren (2013), considers the case where the elicited beliefs might not be the same as the beliefs the agent basis its actions on.

Most of the empirical analysis of outcome equations assumes that all the researcher can construct are error prone measures of the realizations of the outcome from the decision actually made, our $V(d = d(\mathcal{J}), \omega)$. So we start with this assumption and show how it leads directly to the problem of separating out selection effects from the effects of interest. Then we allow the researcher access to error prone measures of outcomes for all possible decisions, for $\{V(d, \omega)\}_{d \in \mathcal{D}}$. When this is possible we can circumvent the selection problem and investigate biases in agents' perceptions directly.

Three preliminary points. Throughout we assume that the researcher does not know either the agents' information sets, or their priors given those information sets. Second, this section of the paper assumes that we can use the descriptive form of the choice model to estimate $g(x)$ but not the distribution ϵ conditional on x. The next section of the paper will provide a method of partially identifying $g(x)$ in discrete choice models (with or without interacting agents). Finally we note that the rest of this section can allow for measurement error in our measure of realizations, or of $V(d, \omega)$, provided it is mean zero conditional on the x used in the

descriptive model. For notational simplicity, instead of adding a separate term for measurement error, I simply reinterpret the ξ term in equation (5) to be the sum of the expectational and the measurement error.

Testing when all we observe is $V(d = d(\mathfrak{I}), \omega)$**.** Then the only regression functions that the outcomes generate is

$$E[V(d, \omega)|x, d = d(\mathfrak{I})] = \tag{8}$$

$$f_{d(\mathfrak{I})}(x) + E\left[u_{d(\mathfrak{I})}|\epsilon_{d(\mathfrak{I})} \geq \max_{q \neq d(\mathfrak{I})} (g_q(x) - g_{d(\mathfrak{I})}(x) + \epsilon_q)\right]$$

$$\equiv f_{d(\mathfrak{I})}(x) + r_{d(\mathfrak{I})}(x).$$

$r_{d(\mathfrak{I})}(x)$ is a "selection" term. It is the expectation of $u_{d(\mathfrak{I})}$ conditional on the choice made; that is conditional on $g_{d(\mathfrak{I})}(x) + \epsilon_{d(\mathfrak{I})} = \max_q[g_q(x) + \epsilon_q]$. It arises because the ϵ that determines the decision and the u that determines the outcome conditional on that decision will not, in general, be distributed independently of one another.

Generally for $r_{d(\mathfrak{I})}(x) \equiv 0$, that is for the regression function for the observed choice to uncover the mean outcome for a randomly drawn agent with characteristics x, the disturbance in the outcome equation would have

to be mean independent of the disturbance in the choice equation. Put differently the unlikely condition is that all variables that the agent uses to predict outcomes that are not in the econometricians data set are not, in fact, predictors of the mean outcome. In our context this is a rather negative result as it implies that as long as there is some variable that effects outcomes that the agent can, but the econometrician can not, condition on, we should expect a difference between the expectation of the function determining choices and the expectation of the function of the chosen alternative (and this, even if the agent is strongly rational). [7]

There is a large literature on correcting for selection in different contexts. Perhaps most closely related to the discussion here is the literature on estimating "treatment effects", a literature which focuses on the relationship between the outcome equation and the selection equation in models with

---

[7] Of course the function $E[V(d,\omega)|x,d = d(\mathcal{J})] = f_{d(\mathcal{J})}(x) + r_{d(\mathcal{J})}(x)$ may be of interest per se. It tells us the average value of the outcome for individuals with characteristics x who chose decision d. For example consider an immigration model, let $d = 1$ be the decision to leave one's native land, and let the outcome of interest be a measure of future income. Then $f_{d=1}(x) + r_{d=1}(x)$ would provide the average of the income measure for individuals with characteristic x who do migrate. The only point I am making is that $f_{d=1}(x) + r_{d=1}(x)$ does not provide the impact of migrating on a randomly chosen individual with characteristics x.

binomial choice (see for e.g. Imbens and Angrist, 1994, and Heckman and Vytlacil, 1999). Reviewing that literature would be a paper in itself, so I am going to suffice with a suggestion for one way of circumventing selection problems in multinomial choice models. When it is available it will generate a rather direct test for weak rationality and a quantification of biases when weak rationality fails.

Testing when we can construct an unbiased approximation to $\{V(d,\omega)\}_{d \in \mathcal{D}}$. There is an interesting dichotomy in the empirical literature. When we formulate choice models we assume we can write down an approximation to the agent's expectation of the value that would be generated were it to chose any of the possible alternatives (those that are chosen and those that are not). In contrast we seldom (if ever) analyze the outcome function for the alternatives not chosen. The counterfactual specifications needed for the choice model are for the returns the agent expected to obtain. To construct counterfactuals for the outcome equation we require a specification for the expectation of the actual returns from counterfactual choices. Given that agent's expectations are inherently un-

observed, it is often easier to specify the actual returns from counterfactual choices than to specify the agents' expected returns from those choices. The rest of this section examines possibilities when we can construct, possibly error prone, measures of outcomes for counterfactual choices.

I distinguish two cases. The first is when we can measure the value of the outcome directly (as when outcome values are measured in dollars). In the second case we only know what aspects of the outcome are valued, but not the specification of the function that aggregates these different aspects of the outcome into "utility". In these cases we assume we can measure the valued aspects of the outcome for counterfactual choices. As we illustrate below the assumptions and data needed to construct the counterfactual outcome, or (in the second case) the valued aspects of the outcome for counterfactual choices, are similar to the assumptions and data needed to construct the value of these objects for the choices actually made[8].

We begin with the case in which we can construct the values of counterfactual outcomes. When we can assess that value for counterfactual choices we can regress $V(d, \omega)$ on x for every realization of $\omega \in \Omega$. Since

---

[8]Though both of these will typically be harder for dynamic (in contrast to static) choice problems.

$E[u_d|x] = E[\xi_d|x] = 0$, this generates a consistent estimate of the $f_d(x)$ defined in equation (5). The bias in an agent's estimate of $V(d,\omega)$, as a function of x, which we label $b_d(x)$, will then be

$$b_d(x) \equiv E[V(d,\omega) - g_d(x)|x], \tag{9}$$

and if the agent is weakly rational then $b_d(x) \equiv 0$. So when we can construct $V(d,\omega)$ for all d we can test for weak rationality, and when it is not accepted we can measure how agent's biases are related to x.

An example might be helpful here. Assume we have estimated a demand system and either know or have estimated costs for a market consisting of a number of retail outlets. Say the outlets (e.g. gas stations) need to make a simultaneous decision on which of a finite set of prices to set for the next period; so $d_i \in \{p_1, \ldots, p_D\}$, and assume that the goal is to set prices to maximize current profits. The demand given the price depends on the prices of neighboring gas stations; if there are N competitors it depends on $d_{-i} \in \{p_1, \ldots, p_D\}^{N-1}$. The data contains what the agents did do.

We begin by analyzing the actual choice of prices as a function of x's known at the time the decision was made (for e.g. as a function of sum-

mary statistics calculated from past sales and prices). This gives us the $g(x)$ in equation (3). Since the game is a simultaneous move game, when constructing counterfactual profits for a given agent we can fix the agent's competitors' prices at the prices they actually chose, and evaluate the demand and cost system at alternative prices for the one agent whose counterfactual profits we are constructing. This gives us the profits the agent would have had were a particular p chosen for those who did not chose that p as well as for those who did. We then regress those profits on the same variables that were used to analyze the decision. This gives us the $f(x)$ in equation (5). This can then be compared to $g(x)$ to determine whether the agents were "weakly rational" and investigate the biases in the agents' expectations if they are not.

Things are slightly more complicated and require more assumptions when we only know what aspects of the outcome are valued, but not the function which aggregates them into a single outcome index, as then we need to also estimate the aggregator function. Here we assume that the aggregator function is common across agents and belongs to a paramet-

ric family whose parameters we estimate along with the $g(x)$ and $f(x)$ functions in equations (3) and (5). A second example illustrates this.

Consider a set of agents who wants to chose an outlet to minimize the cost of buying a particular good. We know that this cost is a function of travel time and the purchase price of the good; properties of the possible choices that may or may not be known by the agent at the time it makes its decision. The first step of the analysis is the same as above; we analyze outlet choice as a function of x's known to the agent at the time the decision was made (e.g.; distance to the outlet or past prices of goods bought at the outlet).

To analyze all (including counterfactual) outcomes we need to measure each agent's travel time and purchase price were it to shop at each outlet in the choice set (up to conditional mean zero errors). Next we run linear regressions of travel time and purchase price for each possible choice on the same x's we used to analyze the choice of outlet; call these vectors of functions (one for each outlet) $f^p(x)$ and $f^t(x)$ respectively.

Now we come to the extra step in this case. Unlike in the previous

case here we need to assume a function which aggregates travel times and purchase prices into an outcome value. Say we are willing to assume that the agent perceives cost to be a linear function of the travel time and the purchase price of the good. Then the expected cost of outlet d conditional on the x's used to analyze choices is $\alpha^p f_d^p(x) + \alpha^t f_d^t(x)$ for $d = 1, \ldots, D$ and some (α^p, α^t). We chose estimates of the (α^p, α^t) that minimize a metric in bias; so there is a sense in which our bias estimates are the minimal estimates of bias one could get from the data and our assumptions. For example we first find the (α^p, α^t) that

$$min_{\alpha^p,\alpha^t} \sum_{i,d \in \mathcal{D}} \left(\alpha^p f_d^p(x_i) + \alpha^t f_d^t(x_i) - g_d(x_i) \right)^2,$$

say $(\hat{\alpha}^p, \hat{\alpha}^t)$. Then our estimate of the bias in the perceived cost of outlet d as a function of x is

$$b_d(x) = \hat{\alpha}^p f_d^p(x) + \hat{\alpha}^t f_d^t(x) - g_d(x).$$

In both these cases the analysis of bias is intuitive; it compares the expectation of the actual outcome equation conditional on x to the expectation of the function that determines the agent's choices conditional on x.

If these differ, and the agent is maximizing the expected outcome conditional on its own beliefs, the agent's expectations are biased. Importantly the test *does not* require us to specify either

- the agent's information set (our $\mathcal{J}$), or

- the priors the agent uses to form $\mathcal{E}[V(\cdot)|\mathcal{J}]$.

Both of these are objects the researcher typically knows little about.

Moreover, once we obtain an estimate $b_d(x)$ we can back out an estimate of $r_d(x)$ from equation (8), and use this, together with what can be learned about the ϵ disturbances from the descriptive form of the choice equation, to provide a deeper understanding of the relationship between the agent's beliefs and actual outcomes (for e.g. if $b_d(x) \equiv 0$ we might ask if the agent is strongly rational). I leave this to subsequent research and now return to the issue of constructing the descriptive form for discrete choice models.

4   Descriptive Forms for Entry Games.

This section of the paper constructs descriptive forms for entry games. Its purpose is to illustrate how to construct, estimate, and interpret descriptive

forms for interacting agent discrete choice problems. Special emphasis will be given to the "endogeneity" problem briefly introduced above. That is we will be interested in a summary of the impact on the number and type of firms on the value of being active conditional on unobserved aspects of market profitability.

We start by computing Markov Perfect equilibria for two entry games (one where location is endogenous and one where it is not). Next the computed values from these games are used to generate the true descriptive forms for each model. This will make it clear that the worry about endogeneity that was the impetus behind the modern entry literature in Industrial Organization is justified. We then simulate data using the estimated policies and use that data in a Monte Carlo study of alternative estimators for the descriptive form.

4.1 The Computed Equilibria and the True Descriptive Form.

We compute equilibrium policies for two different Ericson-Pakes (1995) style Markov Perfect dynamic models. In each computed equilibria firms

differed in two dimensions; their location (the east or west side of town) and the quality of the product they market. For a given model we computed many different equilibria that differed only in market size. So the possible states of the game, which we index by σ, consists of a counting measure for the number of active firms at the different quality-location combinations and market size. Consumers also differed in two dimensions; their location, and their sensitivity to price. Both were assumed to distribute uniformly, and the west side of town was made somewhat more sensitive to price (or poorer) than the east.

The difference between the two models analyzed is that in one quality was endogenous in that firms invest in developing it (as in Goertler and Gordon, 2011, or Benkard, 2004) and firms were randomly assigned a location in which they could (but need not) enter (so location decisions where exogenous). In the second model quality was set exogenously but the firms could chose a location in which to enter (as in Seim, 2006, or Mazzeo, 2002).

From the point of view of the exercise conducted below there are two

important differences between the two models. In the model in which location was endogenous we allowed for a market-specific fixed cost which differed across quality levels and time periods and will be assumed known to the econometrician. There was no such fixed cost in the market with endogenous quality, so the model with endogenous location will have an additional "instrument" in the analysis below. Second in the model with investment the discrete dynamic choice for both potential entrants and incumbents was "binomial" (they could either be active in the coming period or not)[9]. In the model with a location choice, there were three possible choices for potential entrants; they could enter west, enter east, or not enter.

The rest of the specification of the two models was standard. In both the prices and quantities marketed conditional on σ were determined by a Nash in prices (or Bertrand) equilibrium. In the model in which one could invest to improve the quality of the product marginal cost was assumed constant and the same across firms. In the model where quality was set

---

[9]In that model incumbents also chose an investment with stochastic outcomes. The outcomes conditional on the investment were modeled as in Pakes McGuire (1994).

exogenously, marginal costs increased with quality. The equilibrium profit vectors for every σ were fed into a Pakes and McGuire (1994) algorithm to determine dynamic policies that satisfied Markov Perfect equilibrium conditions. Information was symmetric except for the random draws determining entry costs and selloff values which were only known to the firm receiving them; though all firms knew the distribution from which they were drawn (as in Seim, 2006). Both the entry costs and selloff value distributions were exponential.

We computed equilibria for the model with endogenous quality but exogenous location for ten different market sizes. Each market could have up to six active firms. There were nine potential entrants in the model with exogenous quality but endogenous location (three at each quality level). For this model we computed equilibria for sixty different markets which differed by both the fixed costs assigned to the different quality levels, and market size[10]. So in the model with endogenous location choice we will have two sets of variables that generate permanent differences between markets;

---

[10]The fixed costs and market sizes were drawn from four independent mean zero normals. The market size had unit variance and the fixed cost had a variance of .25.

market size, and the fixed costs of different quality levels. In the model with endogenous location choices the only permanent difference between markets will be their market size.

To obtain the data set used in the Monte Carlo analysis we used the equilibrium policies to simulate a "cross-section" of markets. The markets for the model with investment are obtained as three thousand draws from the invariant distributions of each of the ten market sizes. The markets for the model with multinomial entry decisions were fifty draws from the invariant distribution of each the sixty markets.

The estimators we want to evaluate will all be set estimators derived from various moment inequality constraints. We chose this large a data set (approximately thirty thousand markets) to insure that the "confidence intervals" obtained from the dimension by dimension end points of the set estimators generated by the inequalities obtained from our data are close to the dimension by dimension end points of the "identified set" that would be obtained from an infinity of markets generated by our particular DGPs.

The True Descriptive Forms. We want to characterize the determinants of the the value of being an active firm and, as in the prior literature, we will be particularly interested in quantifying the effects of increased competition on that value[11]. We will also mimic prior research in that we will use linear functional forms for the descriptive form. In the Monte Carlo analysis that follows we will not assume that the researcher has access to all the relevant state variables. In particular we will assume the researcher does not know market size. As a result for each model we will present a true descriptive form that does, and that does not, condition on market size.

The absence of adequate controls for the differential profitability of different markets was a major reason for the development of empirical entry models which allowed for the endogeneity of the variables measuring the extent of competition. It was thought that a lack of data on either the demand or supply conditions in the market would generate a positive bias in the estimates of the impact of competing firms on the value of being

---

[11]The value being active of an incumbent in a given location differs from that of a potential entrant in the same location as the entrant must spend one period in setting up before it earns profits.

active, and the focus of the research was the effect of competing firms on that value. So the goal of the endogeneity correction was to analyze the effects of competition on firm value conditional on unobserved determinants of market profitability. By leaving out market size we insure that we will be faced with an analogous endogeneity problem.

We will also assume that the researcher only observed high or low quality. In the model with investment there were five possible quality levels so we assumed the researcher only knows whether q was in one of the three highest or the two lowest quality levels. In the model where quality is chosen at the outset there were three quality levels and we assume that the researcher only observes whether active firms are in one of the top two or not.

The researcher's goal is to characterize how the following variables affect the value of an active firm

- $n_{q,l}$, the number of firms in the same (q, l) slot as the firm in question,

- n_q, the number of competitors of the same quality as the firm in question,

- n_l, the number of competitors in the same location as the firm in question.

Let j index firms and (H_j, l_j) be indicators for high quality and a western location respectively. Then $X_j = (H_j, l_j, n_{l_j, q_j}, n_{l_j}, n_{q_j})$ is the set of variables the researcher observes on each firm and the market the firm is in. Note that X_j does not include market size, which we denote by S. However the vector of variables that the researcher actually wants to condition the firm's value on is $s \equiv (X_j, S)$. More formally let $V(q_j, l_j; \sigma)$ be the value of being active when a firm's location is l_j, its quality is q_j, and the market is in state σ (which recall consists of a counting measure for the number of active firms at the different quality-location combinations and market size). Then the researcher wants to estimate the coefficients in the projection of $V(q_j, l_j, \sigma)$ onto X_j conditional on s_σ. By construction

$$V(q_j, l_j, \sigma) = X_{j,\sigma} \beta_x + \sum_s \{S = s_\sigma\} \beta_s + \mu_{j,\sigma},$$

with, for each component of X, say X_k, $\sum_{j,\sigma} \mu_{j,\sigma} X_{j,k,\sigma} = 0$; while for each distinct value of s_σ, $\sum_j \mu_{j,\sigma} \{S = s_\sigma\} = 0$. Notice that $\mu_{j,\sigma}$ is likely to be

correlated across j for a given σ. So to complete the specification we define

$$\nu_\sigma = E[\mu_{j,\sigma}|\sigma] \quad \text{and} \quad \epsilon_{j,\sigma} = \mu_{j,\sigma} - \nu_\sigma$$

so that the $\epsilon_{j,\sigma}$ are mean independent of σ, or $E[\epsilon_{j,\sigma}|\sigma] = 0$. The resulting

descriptive model states that a firm will be active if and only if

$$V(q_j, l_j, \sigma) = X_{j,\sigma}\beta_x + \sum_s \{S = s_\sigma\}\beta_s + \nu_\sigma + \epsilon_{j,\sigma} \tag{10}$$

is greater than zero. Since the value of being active has both a disturbance

which varies across incumbents in the same market ($\epsilon_{j,\sigma}$), and a disturbance

that has the same value for all firms in that market (ν_σ), we should expect

correlation in the disturbances of potentially active firms in a given market.

Estimates of the Descriptive Form. The actual descriptive forms

generated by our simulated data for our two models are provided in Table

1. There are two sets of estimates for each model. One provides the

coefficients from the descriptive forms for the value of being active when

we do, and one when we do not, include the market size dummies. When

we condition on market size the estimates are similar across models; all

competitive effects are negative and highly significant, and the absolute

41

value of the effect of the number of firms in the same quality and location is noticeable larger than for firms that are marketing goods of different quality or are found in a different location[12]. Notice also that the standard errors are extremely small as expected from sample sizes this large.

The comparison of the results from when we do, and when we do not, condition on market size make it clear that applied researchers' worries about endogeneity are warranted. When we do not include the market size dummies the coefficients which represent competitive effects fall markedly in both models. The coefficient on the number of competitors in the same location and quality falls (in absolute value) by a third or more, that on the number of firms in the same location but different quality is divided by about three, and that on firms in a different location but the same quality goes from significantly negative to significantly positive (where significant can be interpreted in either a statistical or an economic sense). In addition the R^2 falls dramatically when we do not condition on market size. Keep

---

[12]The largest percentage difference between models is in the impact of the number of firms in the same location, which is less important in the model with endogenous location choices. This should not be surprising since the model with endogenous location choices allows for arbitrage across locations.

in mind that the goal of the estimators that use an endogeneity correction
is to obtain estimates of the parameters in columns (1) and (3) when the
data on market size is not available.

Table 1: Descriptive Forms For the Value of Being Active[*]

Model	Exogenous Location				Endogenous Location			
θ	With Mkt. Size		No Mkt. Size		With Mkt. Size		No Mkt. Size	
$n_{l,q}$	-0.73	(0.01)	-0.48	(0.01)	-1.04	(0.03)	-0.51	(0.01)
n_{-l}	-0.58	(0.01)	-0.17	(0.02)	-0.36	(0.04)	-0.13	(0.01)
n_{-q}	-0.07	(0.01)	0.18	(0.01)	-0.15	(0.01)	0.62	(0.05)
q	1.59	(0.01)	1.69	(0.02)	1.43	(0.06)	1.33	(0.06)
R^2	0.74	n.r.	0.28	n.r.	0.85	n.r.	0.41	n.r.

[*] The estimated descriptive forms included a location indicator. We have normalized its coefficient so as to be comparable to the estimators from the moment inequality estimators (which require a normalization, see below).

There is one more aspect of the results that is important to what follows.
Even when the market size dummies are included, about twenty percent of
the variance in the value of being active in the descriptive form is noise in

the sense that it is not correlated with the right hand side variables determining the competitive effects. The market size variable is a determinant of whether any agent wants to be active, and, as a result, is correlated with the competitive effects variables. So these results imply that, at least when we do not have the data on all variables that we want the descriptive form to condition on, the estimation procedure should allow for two types of unobservables. One that assumes the structure of the omitted variable that we want to condition on and is the cause of the endogeneity problem. The other unobservable is generated by the approximation involved in going from the behavioral to the descriptive form and is uncorrelated with the included right hand side variables by construction.

4.2 Estimation Algorithms.

As noted early work on entry models that did not allow the disturbance to be correlated with the variables measuring competitive effects often found unintuitively small (sometimes positive) effects of the number of competitors on the value of being active. These results were typically

attributed to omitted determinants of the profitability of markets. So researchers looked for an analytic framework which allowed the variables representing the effects of competition, i.e. the number of different types of firms, to be correlated with the disturbance.

An assortment of estimation algorithms that allowed for this correlation have been considered. The most detailed of them is the generalization of discrete choice introduced by Ciliberto and Tamer (2009) specifically for the entry problem (though the basic idea applies much more generally and largely dates back to Tamer, 2003). I labeled this model the generalized discrete choice (the GDC) model in Pakes (2010). It uses the inequality constraints generated by Nash equilibria, together with an ingenious estimation algorithm which is outlined below, to account for the endogeneity of right hand side variables.

I will also compare this to an estimation algorithm that explicitly treats our approximation to the value of being active as a descriptive form and hence allows for approximation error; albeit a descriptive form which conditions on an omitted, market-specific, profitability parameter. This esti-

mator, which is based on a theorem in Pakes and Porter (2013), is also relatively easy to implement.

Both estimation algorithms treat the estimation problem as a discrete choice problem. Write the value for each possible choice as

$$V(d, X_{j,m}, \sigma) = X_{j,\sigma}\beta_d + \sum_s \{S = s_\sigma\}\beta_{s,d} + \nu_{d,\sigma} + \epsilon_{d,j,\sigma}$$

where $d \in \mathcal{D}_j$, the choice set, and we use the normalization $\beta_d = \beta_{s,d} = \nu_{d,\sigma} = \epsilon_{d,j,\sigma} = 0$ for the choice of not to be active. In both models the incumbent choice is binomial (to be active or not) which is the same as the choice set for potential entrants in the model with exogenous location. In the model with endongenous location choices the potential entrant faces a trinomial choice; it can remain inactive, or chose to enter west or east.

Recall that the researcher does not observe σ; so we can not impose the constraints that would be implied if two markets had the same σ. Hence we simply index the markets (or different time periods in a given market) by m and rewrite the choice equation as

$$V(d, X_{j,m}, m) = X_{j,m}\beta_d + \sum_s \{S = s_m\}\beta_{s,d} + \nu_{d,m} + \epsilon_{d,j,m}, \qquad (11)$$

for $d \in \mathcal{D}_j$, $j = 1, \ldots n_m$, and $m = 1, \ldots M$. We assume that the appropriate limiting approximations for the properties of estimators has M growing large.

An Application of a Theorem in Pakes and Porter (2013). We begin with an estimator for the descriptive form. Recall that though we want a descriptive form that conditions on market characteristics, s_m is both unobserved and correlated with components of $X_{j,m}$. One way of treating this problem is to set $\lambda_{d,m} = \sum_s \{S = s_m\} \beta_{s,d} + \nu_{d,m}$ in equation (11), rewrite that equation as

$$V(d, X_{j,m}, m) = X_{j,m} \beta_d + \lambda_{d,m} + \epsilon_{d,j,m}, \tag{12}$$

and try to develop an estimator which allows for the $\lambda_{d,m}$.

Formally the problem is now a discrete choice problem with choice-specific fixed effects. We are interested in the β_d parameters but must contend with the fact that to estimate those parameters we need to condition on the $\{\lambda_{d,m}\}$; a set of terms which we expect to be correlated with the $X_{j,m}$ and whose cardinality grows with sample size. Also though the

47

model tells us that the $\epsilon_{d,j,m}$ are mean independent of the right hand side variables, it provides no further information about their distribution.

We now introduce a theorem that allows us to bound the parameters of interest without imposing any constraints on the $\{\lambda_{d,m}\}$ or a particular distribution on the $[\epsilon_{1,j,m}, \ldots \epsilon_{D_j,j,m}]$; which can differ arbitrarily across choice sets and markets. On the other had it does require that all agents with the same choice set in the same market have the same distribution of the $[\epsilon_{1,j,m}, \ldots \epsilon_{D_j,j,m}]$.

Assumption 1. Assume each agent j in market m choses d to

$$\max_{d \in \{1,\ldots,\mathcal{D}_j\}} \left[X_{j,m}\beta_d + \lambda_{d,m} + \epsilon_{d,j,m} \right],$$

and that if agents (j,q) in market m face the same choice set (i.e $\mathcal{D}_j = \mathcal{D}_q$), then the distribution of

$$[\epsilon_{1,q,m}, \ldots, \epsilon_{D,q,m}] \text{ conditional on } [\{\lambda_{d,m}\}, X_{j,m}, X_q] \equiv \Omega(j,q)$$

is the same as the distribution of

$$[\epsilon_{1,j,m}, \ldots, \epsilon_{D,j,m}] \text{ conditional on } \Omega(j,q) \quad \spadesuit.$$

Note that the assumption on the error distribution does not restrict

the joint distribution of the disturbances across choices (either their correlation pattern or their marginals), and it allows these distributions to differ arbitrarily across choice sets and markets. Indeed our Assumption *almost* follows from the way we constructed the descriptive form as that construction guarantees that the disturbances are uncorrelated with $X_{j,m}$ and $\lambda_{d,m}$. "Almost" because that construction does not insure that the disturbances in the restrictive form satisfy the condition that two agents in the same market with the same choice set satisfy the exchangeability assumption above. So our statistical model does require a strengthening of the assumptions used to generate our economic model. The following theorem presents implications of this Assumption used in estimation.

Theorem: Pakes and Porter (2013). Let $r \in \{D_j, D_j - 1, \ldots, 2\}$ where D_j is the cardinality of $\mathcal{D}_j$, the choice set. If $\mathcal{D}_j = \mathcal{D}_q$, define $\delta^r_{j,q}(\beta; m)$ to be the rth largest value of $[X_{j,m,d} - X_{q,m,d}]\beta$. That is

$$\delta^{D_j}_{j,q}(\beta; m) = argmax_{d \in \mathcal{D}_j}[X_{j,m,d} - X_{q,m,d}]\beta_d,$$

while

$$\delta_{j,q}^{D_j-1}(\beta;m) = argmax_{\{d:d\neq\delta_{j,q}^{D_j}(\beta;m),d\in\mathcal{D}_j\}}[X_{j,m,d} - X_{q,m,d}]\beta_d$$

is the choice that yields the second highest value of that difference, and so

on. Then for $l = 0,\ldots,D-2$

$$Pr\Big\{ \sum_{r=0}^{l}\{d_q = \delta(j,q)^{D-r}(\beta_0;m)\} \,|X_{q,m},\lambda_{1,m},\ldots,\lambda_{D_j,m}\Big\} \quad \leq$$

$$Pr\Big\{ \sum_{r=0}^{l}\{d_j = \delta(j,q)^{D-r}(\beta_0;m)\} \,|X_{j,m},\lambda_{1,m},\ldots,\lambda_{D_j,m}\Big\}. \quad \spadesuit$$

The theorem provides $D-1$ conditional moment inequalities. It states

that at the true β_0 agent j's probability of each of the following events

conditional on $(X_{j,m}, X_{q.m})$ is at least as large as those events for agent q:

choosing $\delta_{j,q}^{D_j}(\beta;m)$; choosing one of the set of choices $\{\delta_{j,q}^{D_j}(\beta;m),\delta_{j,q}^{D_j-1}(\beta;m)\}$;

choosing one of the set of choices $\{\delta_{j,q}^{D_j}(\beta;m),\delta_{j,q}^{D_j-1}(\beta;m),\delta_{j,q}^{D_j-2}(\beta;m)\}$, and

so on. Consequently the theorem implies that if $f(\cdot)$ is any positive func-

tion of $(X_{j,m}, X_{q,m})$, then for any l we can form the unconditional moment

inequality

$$E\Big(\sum_{r=0}^{l}\Big[\{d_j = \delta(j,q)^{D-r}(\beta;m)\}-\{d_q = \delta(j,q)^{D-r}(\beta;m)\}\Big]f(X_{j,m}, X_{q,m})\Big) \geq 0,$$

(13)

and know that it is satisfied at $\beta = \beta_0$. That is values of β that do not

make inequalities positive are unlikely to be the true β_0.

These are inequalities we can take to data. Of course if β satisfies them,

then so will $\kappa \times \beta$ for any $\kappa > 0$. So we will have to normalize one coefficient

in the analysis below.

The GDC Estimator (Ciliberto and Tamer, 2009). The GDC es-

timator also considers the choice model in equation (12), that is

$$V(d, X_{j,m}, m) = X_{j,m}\beta_d + \lambda_{d,m} + \epsilon_{d,j,m},$$

where $\lambda_{d,m} = \sum_s\{S = s_m\}\beta_{s,d} + \nu_{d,m}$ and $\beta_d = \lambda_{d,m} = \epsilon_{d,j,m} = 0$ for the

choice of not begin active. However the GDC estimator makes different

assumptions on the $\{\lambda_{d,m}\}$ and the $\{\epsilon_{d,j,m}\}$ then the estimator discussed

above, and this leads to a different estimator for the parameters of interest.

The GDC estimator assumes that the data is generated by a Nash equi-

librium to the game in which both $\{\lambda_{d,m}\}$ and the $\{\epsilon_{d,j,m}\}$ are known to all the agents playing the game. As a result both of the unobservables are determinants of the $\{n_{l,q}, n_{-l}, n_{-q}\}$ components of $X_{j.m}$. This differs from the implications of the assumptions used to derive the descriptive form; their the $\{\lambda_{d,m}\}$ are correlated with the competitive effect variables but the $\{\epsilon_{d,j,m}\}$ represent approximation error and are uncorrelated with $\{n_{l,q}, n_{-l}, n_{-q}\}$ by construction. The GDC estimator also assumes a parametric distribution for both the $\{\lambda_{d,m}\}$ and the $\{\epsilon_{d,j,m}\}$; the descriptive form places no restriction on the $\{\lambda_{d,m}\}$ and uses the orthogonality condition implied by the way it is constructed and Assumption 1 to structure the distribution of the $\{\epsilon_{d,j,m}\}$.

I now provide a brief explanation of Ciliberto and Tamer's (2009) GDC algorithm. Let β be a vector which consists of particular values for both the β_d parameters in equation (12) and the parameters of the disturbance distribution. Fix β and let a superscript "o" denote the observed value of a variable. Define $d^o_m = [d^o_{1,m}, \ldots d^o_{J,m}]$, where $d^o_{j,m} = [d^o_{1,j,m}, \ldots d^o_{D_j,j,m}]$, so that d^o_m denotes the choices actually made by participants in market m.

Simulate a draw from the joint distribution of $\{\lambda_{d,m} + \epsilon_{d,j,m}\}$ and check whether the Nash equilibrium conditions are satisfied for that particular market when those conditions are evaluated at β, the draw on the disturbances, and $d_m = d_m^o$. Since this is a full information game, the check entails seeing whether each agent's actual choice is the best choice possible given the actions of the other agents[13]. Simulate many draws from the disturbance distribution, and check whether the Nash conditions are satisfied at d_m^o for each. The fraction of times that the Nash conditions are satisfied gives us an unbiased estimator of the probability that the necessary conditions of a Nash equilibria are satisfied at the observed choice when we evaluate those conditions at that β.

Since the Nash conditions are necessary but not sufficient for a vector of choices to be made, the probability of a particular set of decisions satisfying the Nash conditions conditional on the true value of the parameter vector can be greater than the probability that those decisions represent

---

[13]Notice that the fact that the game admits multiple equilibria implies that there may be more than one set of decisions which will satisfy the Nash conditions. The implicit assumption is that there is a mechanism that selects out the equilibria actually played but this selection mechanism is unknown to the researcher. All the researcher knows is that the actual outcome satisfies one of the set of Nash conditions.

the selected equilibrium. So if we simulate the probability that the Nash conditions are satisfied for a given value of the parameter, the expectation of the simulated probability is an upper bound (or an "outer measure") for the probability of observing the given market structure conditional on that parameter value. The estimation algorithm consists of finding all values of the parameter vector that make the cross market average of the difference between the simulated frequencies of the decisions satisfying the Nash conditions and the observed frequencies of the vector of decisions greater than zero[14].

More formally if $\overline{P}(d_m; \beta)$ is the simulated (outer) probabilities and $X_m = [X_{1,m}, \ldots, X_{J,m}]$, then the model implies that

$$E[\overline{P}(d_m; \beta) - \{d_m^o\}|X_{j,m}] \geq 0, \quad \text{when} \quad \beta = \beta_0,$$

where now the expectation operator integrates out over both the simulated draws which underly the $\{\overline{P}(d_m; \beta)\}$, and the data generating process which underlies the $\{d_m^o\}$. Consequently if $f(X_{j,m})$ is any positive function of $X_{j,m}$

---

[14]One could also use the sufficient conditions for a unique equilibrium to get lower bounds (or an "inner measure") to those probabilities and add an additional set of inequalities to the estimation algorithm, see Pakes (2010).

we can form set estimators of β by finding all those values for β which imply positive values for functions of the form

$$M^{-1} \sum_m \left[\overline{P}(d_m; \beta) - \{d_m^o\} \right] f(X_{m,j}) \geq 0. \tag{14}$$

The GDC estimator requires auxiliary distributional assumptions. Conditional on those assumptions it does account for endogeneity by allowing the disturbance to be correlated with the number and type of competitors. However it does not account for approximation error.

4.3 Monte Carlo Results.

Next we move to the results on the alternative estimators. We begin with the data generated by the model where location is exogenous so the only choice is whether to be active or not. The data generated from this model are used in a Monte Carlo analysis of both the GDC and the Pakes-Porter estimator. We then move on to the data generated by the model where there is a location choice. Our interest here is primarily to see how well the Pakes-Porter estimator behaves when there is a multinomial, in contrast to a binomial, choice. So here we only present the results for the Pakes-Porter

estimator. All confidence intervals presented below are formed using the moment-shifting technique described in the paper by Andrews and Soares (2010).

The Model with Exogenous Location. We are primarily interested in estimators that do not assume that the market size variable is observable, as this reflects the empirical reality that we can not condition on all the determinants of market profitability that we would like to when analyzing the effect of the number and type of competitors on the value of being active. The first row of Table 2 takes the results from the descriptive form that uses market size from Table 1; i.e. it provides the parameter values we are hoping that the estimators uncover. The second row is just a reminder of what one would estimate by OLS when market size is not available and we do not account for endogeneity. The rest of the rows present alternative estimators. We return to them after clarifying some of the details of the estimation algorithms. In particular the GDC estimator requires a parametric assumption on the distribution of disturbances, while the P-P estimator requires a normalization.

To generate the GDC estimator we assumed that the firm specific cost shock was exponential (as it was in the simulated equilibrium), and that the market specific shock ($\{u_m\}$) was normally distributed. We estimated the variance of the market specific effect, and will report it, but when we calculate confidence intervals for the other parameters estimated by the GDC estimator we assume that this variance is known and fixed at its estimated value. This because then we estimate the same number of parameters with the P-P estimator as we do with the GDC estimator (otherwise we would be estimating one more parameter for the GDC estimator).

The normalization we used for the P-P estimator was to set the dummy variable for the "west" location, or $l_w = 1$. There is a normalization implicit in the GDC estimator through the choice for the distribution of idiosyncratic disturbance (or ϵ). Still were we not to impose the $l_w = 1$ constraint with the GDC estimator it would be estimating more parameters than the Pakes-Porter estimator. So we present one set of results for the GDC estimator where l_w was set equal to one, and one where this constraint was not imposed. When the constraint is imposed the GDC estimator

estimates the same number of parameters as we does the P-P estimator; otherwise it estimates one additional parameter.

Computational Burden   Before proceeding to the results a note on computational burden is in order. All calculations are done on a Dell T7500 Workstation.[15] The compute time for one thousand function evaluations of the Pakes-Porter algorithm was 1.93 seconds, while for the GDC algorithm it was 25.3 seconds. We evaluated 21 grid points in each of our four dimensions (i.e. 21^4 grid points). The total computational time was about 6 minutes for Pakes Porter, and 82 minutes for GDC. As can be seen from the description of the algorithms, the GDC algorithm is also significantly harder to code up and de-bug than the Pakes-Porter algorithm. So one advantage of the Pakes Porter algorithm is that it requires considerably less compute and programming time.

Results   For all rows of estimators in Table 2 we had enough data to condition on the number of potentially active firms at each location-quality node (this gave us about seventy-five inequalities). The potential location

---

[15] This had 72GB of RAM, Windows 7, with Intel Xeon X5690 CPU, 3.47GHz with 6 cores.

is the location that an incumbent could exit from, and for a potential entrant it was the location the entrant could enter into. When we assume the market size dummy variable is available we condition on the number of potentially active firms at each location and the market size. So our instruments for the estimators in equations (13) and (14) were indicator functions for all possible vectors of the number of firms who could be active at each quality-location node[16].

We present estimators that assume that the market size variable is observable in rows 3 and 4 of the table, and the estimators which do not use the market size variable in rows 5 to 7. Since the P-P estimator uses differences in the probability of being active in a given market across agents, the algorithm differences out the effect of market size. Hence the only difference in the estimation algorithm underlying row (3) (when we do have access to the markets size variable) and row (5) (when we do not have access to that variable) is that row (3) uses more instruments. In actual

---

[16]For the GDC estimator this resulted in 735 moments for the estimates in the column that used market size dummies and 435 moments for the estimates that did not use them. For the P-P estimator there were 470 moments when we used market size dummies and 278 moments when we did not.

applications we would often have a variable that is correlated with, but not an exact measure of market profitability, so we might expect estimates somewhere between rows (3) and (5) for the Pakes-Porter estimator.

Table 2 Monte Carlo Results (Exogenous Location):

Confidence Intervals for the GDC and P-P Estimators.

Estimator	$n_{l,q}$	n_{-l}	n_{-q}	q
Descriptive Form				
1. OLS with Mkt Size	-0.73 (.01)	-0.58 (.01)	-0.07 (.01)	1.59 (.01)
2. OLS without Mkt Size	-0.48 (.01)	-0.17 (.02)	0.18 (.01)	1.69 (.02)
With Market Size				
3. P-P Inequalities	[-1.02,-0.68]	[-0.69,-0.25]	[-0.22,-0.02]	[1.22,1.88]
4. GDC Inequalities	[-1.10,-0.63]	[-0.91, -0.35]	[-0.13,0.17]	[1.09,1.53]
Without Market Size				
5. P-P Inequalities	[-1.14,-0.65]	[-0.81,-0.44]	[-0.27,-0.03]	[0.91,1.27]
6. GDC, $l_w = 1$	[-1.50,-0.11]	[-1.05,-0.61]	[-0.55,0.05]	[0.81, 1.97]
7. GDC, $l_w \neq 1$	[-0.60,-0.49]	[-0.22,-0.05]	[0.08,0.17]	[0.39,0.96]

[1] For rows 6 and 7, the GDC estimator also estimated the variance of a random market effect. In all cases we got a point estimate and set it at its estimated value when calculating the confidence intervals for the remaining coefficients. The estimate of the variance of the market effect for row 6 was .45 and its confidence interval was [.44, .48], and for row 7 it was .48 with a confidence interval of [.41,.51].

Rows (5) to (7) of the table make it clear that either correction for endogeneity improves on the OLS estimates that does not correct for endogeneity considerably (compare these to the estimates in row 2 above). Recall that both estimators require different assumptions than those that generated the data. The P-P estimator requires that the distribution of the disturbances for the different agents in a given market be identical, while the GDC estimator goes further and imposes a parametric assumption on the distribution of the disturbance and incorrectly assumes that the agent specific component of the assumed distribution (the $\{\epsilon_{j,d,m}\}$) are correlated with the right hand side variables. So perhaps the most important lesson from the table is that some correction for endogeneity, even if that correction is based on assumptions which are likely to be false, is quite helpful. I am going to focus on the row (6) results for the GDC estimator, as they are more directly comparable to the results from the P-P estimator (then both estimators estimate the same number of parameters).[17]

---

[17]The interpretation of the parameters is then relative to the l_w coefficient. The argument for presenting the results in row (7) is that a researcher analyzing this problem would probably rely solely on the normalization implicit in the assumed distribution functions. Row (7) tells us what would happen if that were the case.

What is striking about the estimates in rows 5 and 6 is that they indicate that the confidence intervals for each parameter produced by both the P-P estimator and, at least the $l_w = 1$, GDC estimator cover their true values. With this much data these confidence intervals are close to the boundaries of the identified set obtained from the various estimators. The lengths of the the confidence intervals for the two estimators do differ. The P-P estimators confidence intervals are shorter for all variables, and sometimes the difference is quite noticeable. This despite the fact that the GDC estimator uses a parametric distribution for the disturbances, while the P-P estimator does not. Apparently the GDC-estimator's miss-specification in the distribution function is reflected in the tightness of the bounds it produces on, particularly, the n_{-l} and q coefficients [18].

---

[18]Interestingly, the only other time I have done a Monte Carlo comparison of this sort was to estimate a buyer-seller network problem in Pakes (2010). Their the GDC estimator generated point estimates that were quite close to the true values but confidence intervals that were extremely tight; so tight that they did not cover the true values (quite the opposite of what we find here). Perhaps not surprisingly the impact of the miss-specification of the GDC estimator seems to depend on the problem at hand.

The Model with Endogenous Location Choices. This subsection uses the P-P estimator to evaluate the model with an endogenous location choice. Recall that in that model potential entrants are presented with a trinomial choice (enter east, enter west, or do not enter), while the incumbents face the binomial choice of whether to exit or remain active. Also there are now observable differences between markets that can be used as instruments (the fixed costs of the three qualities differ by market).

As a result this model provides an opportunity to examine several different aspects of the P-P estimator. First it enables us to evaluate the difference between a P-P estimator based on only the indicator function differences for the first ranked index function difference between agents and an estimator based on two differences in indicator functions (that in the first, and the indicator for the combined first and the second, ranked differences in index functions across agents). Second we can distinguish between models that only compare agents with the same choice sets (as is required by the Theorem) and between those that lump potential entrants and incumbents together (this is what is done in most applications of entry

models). Finally the additional observable variables provide us instruments that are not used as right hand side variables in the estimation, and most prior work on entry models makes intensive use of observable instruments of this sort.

The estimates in the first two rows of Table 3 are the descriptive form and the OLS estimate of it that omits the market size variable for the model with endogenous location choices (they are the last four columns of Table 1). The rest of the rows in Table 3 present results from estimates based on alternative versions of the P-P theorem. All use inequalities constructed by interacting differences in indicator functions for different choices between two agents with the instruments used in the last section *plus* those instruments interacted with the fixed cost variables (which essentially doubles the amount of instruments). When using this many instruments we obtain point estimates (rather than sets). However the models could always accept the null that the fact that no parameter vector satisfied all the inequality constraints was a result of sampling error.

Table 3. Monte Carlo Results (Endogenous Location):

P-P Estimators.

Estimator	$n_{l,q}$	n_{-l}	n_{-q}	q
Descriptive Form				
1. OLS with Mkt Size	-1.04 (.03)	-0.36 (.04)	-0.15 (.01)	1.43 (.06)
2. OLS without Mkt Size	-0.51 (.01)	-0.13 (.01)	0.62 (.05)	1.33 (.06)
Only Use Indicators for Largest Difference in Indexes; All Choice Sets.				
3. Point Estimate	-0.76	-0.58	-0.03	1.06
4. Confidence Interval	[-1.49,-0.18]	[-1.49,-0.15]	[-0.66,0.04]	[0.90,1.94]
Use Indicators for All Available Differences in Indexes; All Choice Sets.				
5. Point Estimates	-0.85	-0.37	-0.19	1.39
6. Confidence Interval	[-1.11,-0.66]	[-0.49,-0.25]	[-0.24,-0.13]	[1.13,1.58]
Use Indicators for All Available Differences in Indexes; Distinguish Choice Sets				
7. Point Estimates	-0.88	-0.36	-0.19	1.45
8. Confidence Interval	[-1.14 -0.60]	[-0.56,-0.30]	[-0.32,-0.07]	[1.19,1.63]

Rows (3) and (4) provide the results from interacting only the difference

in indicator functions for the first ranked difference in choices with the instruments. It does this for every couple of agents that could be active in the market (so it compares incumbents to potential entrants). This despite the fact that the choice sets for the potential entrants and incumbents are different (so formally the assumptions that underlie the P-P estimator are not satisfied). All confidence intervals cover the true parameter estimates, but they are rather wide. Indeed for $n_{l,q}$ and n_{-l} they also cover the estimates obtained from the OLS regression coefficients which do not control for market size.

Rows (5) and (6) interact both the difference in indicator functions for the first ranked differences in index functions, and the difference in indicator function for the combined first or second ranked differences in index functions (when the latter was available) with our instruments. It still compares all couples, so sometimes one of the pair was a potential entrant and the other was an incumbent. Again all confidence intervals cover the true value of the coefficients, but now the length of the confidence intervals are a third to a fifth of those in rows (3) and (4) (and as a consequence,

we can now reject each of the coefficients estimates in row (2)). The addition of the extra difference in indicators for the difference in index function ranks, that is comparing indicator functions for the combined first or second ranked difference in index functions, contained a lot of information.

Rows (7) and (8) again uses both the difference in indicators for first ranked difference in index functions, and the difference in indicator functions for the combined first or second ranked difference in index functions (when that is available). However now we assume we can distinguish between potential entrants and incumbents, and only compare incumbents to incumbents and potential entrants to potential entrants. The results are very similar to those from rows (5) and (6). This is encouraging for cross sectional analysis of entry models using P-P estimators as many of those data sets do not distinguish between new entrants and incumbents. On the other hand we did use a model where the unobservable (market size) had the same value for both entry locations.

5   Conclusion.

When approaching a new choice problem (or a new choice data set) we often begin with descriptive results. Their role is to summarize relationships in the data that an appropriate behavioral model needs to rationalize. Typically the focus is on covariates which might be determinants of the choice. The purpose of this paper is to suggest a formalization of this endeavor that; has an interpretation in terms of fit, can be compared across data sets, and has a clear relationship to the behavioral form of the model. We call this formalization the "descriptive form" of the model.

The descriptive form of discrete choice models is the regression function of the value the agent assigns to each choice onto a set of variables known to the agent when the agent's decision was made. We began by deriving the descriptive form from the behavioral form of the model. This necessarily generates an approximation error which is mean independent to the included right hand side variables by construction. We then showed that at least in some cases the relationship between the descriptive and the behavioral forms of the model can be used to test for a weak form of ra-

tionality of agents' expectations; in particular we can test whether agents' expectations are correct on average. We can do this test conditional on any set of observables the agent knew when decisions were made, and when agents expectations are biased we can quantify the bias as a function of those observables.

We then considered estimation algorithms for the descriptive form, focusing on the case where the dependent variable is discrete. Often the descriptive form we want to use for one or more of the choices conditions on an unobserved factor with a particular structure. In our entry example the unobservable with particular structure was a market specific fixed effect. As a result it is useful to have an estimation algorithm that allows for group and choice specific fixed effects. We show that a rather simple semiparametric multinomial choice estimator introduced by Pakes and Porter (2013) can accommodate these fixed effects. The estimator requires only relatively weak assumptions on the disturbances (the approximation errors), and can allow for choice specific fixed effects whose cardinality can grow with sample size.

This estimator allows for unobservables which are, and which are not, related to the included right hand side variables. The unobservables that are related to the right hand side variable, or the "endogenous" unobservables, must be group specific and can be choice specific. This is a constraint, but when using a descriptive model it is a natural constraint as it makes clear what the researcher wants to hold constant in analyzing the determinants of a choice. The unobservables that are not related to the independent variables result naturally from the approximation error obtained in deriving the descriptive form from a behavioral model.

We illustrate with a Monte Carlo analysis. We begin by noting that a two period entry game can be viewed as a descriptive form of a more complicated Markov Perfect dynamic model of industry evolution. The endogeneity of the error term, which motivated the empirical literature on two period entry models, becomes a market specific unobservable that we want to condition on when investigating the impact of the number and type of competitors on the value of being active. We first show that allowing for the market specific unobservable is essential if we want to investigate

the importance of the number and type of competitors on the value of being active. Following the literature we chose a descriptive form for that value which was linear in the number and types of competitors. We tried two estimators that allowed for endogeneity of the disturbance term; the GDC estimator of Ciliberto and Tamer (2009) and the semiparametric multinomial choice estimator that allows for choice specific fixed effects in Pakes and Porter (2013). Both these estimators are set estimators. The GDC estimator requires parametric assumptions and does not allow for an approximation error which is orthogonal to the included right hand side variables. Our Monte Carlo analysis indicated that both set estimators cover the true parameter vector. Somewhat surprisingly the Pakes-Porter estimator, which is easier to both program and compute and does not require explicit distributional assumptions, produced narrower confidence bands than the parametric GDC estimator.

Affiliations: Harvard University and NBER.

References.

- Andrews, D., and G. Soares, "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection Procedures," *Econometrica* 78 (2010), 119-157.

- Benkard, L., "A Dynamic Analysis of the Market for Wide-Bodied Commercial Aircraft," *The Review of Economic Studies* 71 (2004), 581-611.

- Berry S., and E. Tamer, "Identification in Models of Oligopoly Entry," in R. Blundell, W.K. Newey and T. Persson, eds., Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, vol. 2, (Cambridge University Press, 2007), 46-85.

- Bresnahan, T., and P. Reiss "Entry and Competition in Concentrated Markets," *Journal of Political Economy* 99 (1991), 977-1009.

- Browning, M, P. Chiappori and Y. Weiss, Family Economics (Cambridge University Press, forthcoming).

- Ciliberto F., and E. Tamer, "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica* 77 (2009), 1791-1828.

- Crawford, G. and A. Yurukoglu, "The Welfare Effects of Bundling in Multichannel Television," *American Economic Review"* 102 (2012), 643-685.

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

- Goettler, R. and B Gordon, "Does AMD Spur Intel to Innovate More?" *Journal of Political Economy* 119 (2011), 1141-1200.

- Grennan M., "Price Discrimination and Bargaining: Empirical Evidence from Medical Devices," *American Economic Review* 103 (2013), 145-177.

- Heckman, J., "Dummy Endogenous Variables in Simultaneous Equation Systems," *Econometrica* 46 (1978), 931-960.

- Heckman, J., and E. Vytlacil, "Local Instrumental Variables and La-

tent Variable Models for Identifying and Bounding Treatment Effects, Proceedings of the National Academy of Sciences, 96 (1999), 4730-4734.

- Hendren, N., "Private Information and Insurance Rejections," *Econometrica* 81 (2013), 1713-1762.

- Ho, K., "Insurer-Provider Networks in the Medical Care Market," *American Economic Review* 99 (2009), 393-430.

- Ho, K. and A. Pakes, "Hospital Choices, Hospital Prices, and Financial Incentives to Physicians," *American Economic Review* (forthcoming).

- Imbens, G., and J. Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica* 62 (1994), 467-475.

- Manski, C., "Measuring Expectations," *Econometrica* 72 (2004), 1329-1376.

- Mazzeo, M., "Product choice and oligopoly market structure," *RAND Journal of Economics* 33 (2002), 221-242.

- Pakes, A., "Alternative Models for Moment Inequalities," *Economet-*

rica 78 (2010), 1783-1822.

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

- Pakes, A. and J. Porter, "Moment Inequalities for Semiparameteric Multinomial Choice with Fixed Effects", *mimeo* Harvard University, 2013.

- Savage, L.J., <u>The Foundations of Statistics</u> (New York: Wiley, 1954).

- Seim, K., "An empirical model of firm entry with endogenous product-type choices," *The RAND Journal of Economics* 37 (2006), 619-640.

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