Alternative models for moment inequalities

A. Pakes

Alternative Models for Moment Inequalities.*

A. Pakes,

(Harvard University and the National Bureau of Economic Research.)

July 6, 2010.

Abstract

Behavioral choice models generate inequalities which, when combined with additional assumptions, can be used as a basis for estimation. This paper considers two sets of such assumptions and uses them in two empirical examples. The second example examines the structure of payments resulting from the upstream interactions in a vertical market. We then mimic the empirical setting for this example in a numerical analysis which computes actual equilibria, examines how their characteristics vary with the market setting, and compares them to the empirical results. The final section uses the numerical results in a Monte Carlo analysis of the robustness of the two approaches to estimation to their underlying assumptions.

1

Behavioral choice models generate inequalities which, when combined with additional assumptions, can be used as a basis for estimation. This paper considers two sets of assumptions which suffice and uses them in two examples which have been difficult to analyze empirically. In doing so we distinguish between the assumptions needed to estimate the "structural" parameters defined by the primitives of the choice problem and the "reduced form" coefficients obtained from regressing profits on variables of interest.

I begin with a single agent discrete choice problem; a consumer's decision of which supermarket to shop at. This provides a transparent setting to illustrate the assumptions underlying alternative estimators and motivates the more formal discussion in the rest of the paper. The difficulty in analyzing this example arises from the size of its choice set; all possible bundles of goods at "nearby" locations. Its importance stems from the need to analyze similar problems to understand the implications of alternative local policies (zoning laws, public transportation alternatives and the like).

Section 2 of the paper formalizes two sets of assumptions that take one from the choice model to an estimation algorithm. This is done in a multiple agent setting (with the single agent simplifications noted). The first approach is labeled the generalized discrete choice approach as it generalizes familiar discrete choice theory to allow for multiple interacting agents. The ideas behind this approach date to Tamer (2003), and are developed in more detail in papers by Ciliberto and Tamer (2007) and Andrews, Berry and Jia (2007). It was first considered in the context of analyzing two stage entry games, but is easily adapted to other multiple agent problems.

The second approach is based upon the inequalities generated by the difference between the expected profits from the choice made and those from an alternative feasible choice; so we refer to it as the "profit inequality" approach. It is preceded by the first order (or Euler) condition estimators for single agent dynamic models provided in Hansen and Singleton (1984) and extended to incorporate transaction costs, and hence inequalities, by Luttmer (1996). The approach considered here is a direct extension of the work in Pakes, Porter, Ho and Ishii (2007) who provide assumptions that enable us to take "revealed preference" inequalities to data (for related work on revealed preference in demand analysis, see Varian, 1982, and in the analysis of auctions, see Haile and Tamer, 1996). The two approaches are

not nested and a comparison of their assumptions closes this section.

Section 3 applies the frameworks developed in Section 2 to the analysis of markets in which a small number of sellers interact with a small number of buyers. This is typical of upstream interactions in many vertical markets; markets where sellers re-market the goods they buy to consumers. A difficulty in analyzing them is that the contracts that establish the buyers' costs are typically proprietary, and these costs determine both the prices the buyer charges to consumers and the sellers' investment incentives. Costs are also often proprietary in consumer goods markets. However because there are many consumers in those markets we typically assume a Nash in price (or quantity) equilibrium in them. Then the first order conditions from that equilibrium can be used to back out marginal cost. The analogous procedure for vertical markets leads us to moment inequalities: we observe who contracts with whom and ask what features must the buyer's cost functions have for each agent to be doing better under the observed set of contracts than what they could have expected from changing their contracting behavior.

The section begins by extending the empirical work of Ho(2009) which characterizes cost functions in HMO-hospital networks. It shows that her approach can be extended to allow for disturbances that are known to the agents when they make their decisions but not to the econometrician. It then compares the empirical results she obtains to those obtained once we allow for these disturbances. Next we compute equilibria for markets similar to those used in the empirical analysis. The numerical results allow us to both investigate the consistency of the empirical results with those obtained from an equilibrium computation, and to engage in a more general examination of the correlates of equilibrium markups.

Section 4 uses the data underlying the numerical results of Section 3 in a Monte Carlo analysis of the two approaches to estimation introduced in Section 2. It focuses on the behavior of the two estimators when one or more of the assumptions needed to derive their properties is violated. At least in our example the estimators were robust to all assumptions but that on form of the disturbance distribution required only for the generalized discrete choice model[1].

---

[1]Since the theoretical restrictions brought to data are moment inequalities, they typically lead to set valued estimators. Methods of inference for set valued estimators are an active and important area of

1 A Motivating Example.

I begin with a single agent example taken from an unpublished thesis by Michael Katz (2007; I thank him for permission to use it). Katz's goal was to estimate the costs shoppers assign to driving to supermarkets. Transportation costs are central to understanding store location decisions and hence to the analysis of the impact of regulations (e.g. zoning laws) and policy changes (e.g. public transportation projects) on retail trade. They have been difficult to analyze empirically with traditional discrete choice models because of the size and complexity of the choice set facing consumers (all possible bundles of goods at all "nearby" stores). In contrast large choice sets *facilitate* moment inequality estimators, as they give the empirical researcher a greater ability to chose a counterfactual that is likely to isolate the effect of interest (here the cost of travel time).

Assume that the agents' utility functions are additively separable functions of the utility from the basket of goods the agent buys, expenditure on that basket, and drive time to the supermarket. The agent's decision, say d_i, consists of buying a basket of goods, say b_i, at a particular store, say s_i, so $d_i = (b_i, s_i)$. If z_i represents individual characteristics, $U(b_i, z_i)$ and $dt(s_i, z_i)$ provides individual i's utility from b_i and drive time to s_i, respectively, and $e(b_i, s_i)$ is the expenditure required to buy b_i at s_i, then the agent's utility from (b_i, s_i) is

$$\pi(d_i, z_i, \theta) = U(b_i, z_i) - e(b_i, s_i) - \theta_i dt(s_i, z_i), \tag{1}$$

where I have normalized the coefficient on expenditure to one so θ_i, the dis-utility of a unit of drive time, is in dollars.

To proceed using moment inequalities we need to compare the utility from the choice the individual made to the utility from a choice the individual could have made but chose not to. This is a *sample design* question as it determines what variance is used to estimate the parameter. For a particular d_i we chose the alternative, say $d'(d_i)$, to be the purchase of; (i) the *same basket* of goods, (ii) at a store which is *further away* from the consumer's home then the store the consumer shopped at. Note that, given the additive separability assumption, this choice differences out the impact of the basket of goods chosen on utility;

econometric research that I do not discuss here; see, in particular, Chernozhukov, Hong and Tamer (2006), Andrews and Soares (2010), and the papers cited above.

i.e. it allows us to hold fixed the dimension of the choice that is not of direct interest and investigate the impact of travel time in isolation.

Assume the agent makes its choice of store by maximizing its expected utility (equation 1) conditional on the information at its disposal when it chooses the store to shop at. We denote the agent's expectation operator by $\mathcal{E}(\cdot)$ and its information set by $\mathcal{J}_i$. Note that when the agent makes this choice the goods that will be bought at the store are a random variable, say $\mathbf{b}$, as are their prices and hence total expenditure $\mathbf{e}(\cdot)$. If the bundle bought at the chosen store could have been bought at the alternate store then (s', b'_i), where b'_i is what would have been bought had the agent gone to s', is preferred over (s'_i, b_i). Since s_i was preferred over s'_i, transitivity of preferences insures that if, for any function $f(\cdot)$, $\Delta f(d, d', \cdot) \equiv f(d, \cdot) - f(d', \cdot)$

$$\mathcal{E}\Big[\Delta\pi(\mathbf{b}, s_i, s'_i, z_i)|\mathcal{J}_i\Big] = \mathcal{E}\Big[-\Delta\mathbf{e}(\mathbf{b}, s_i, s'_i) - \theta_i\Delta dt(s_i, s'_i, z_i))|\mathcal{J}_i\Big] \geq 0.$$

We let $\nu_{1,i,s,s'}$ be the difference between realized and expected utility, i.e.

$$\nu_{1,i,s,s'} \equiv \Delta\pi(b, s_i, s'_i, z_i) - \mathcal{E}\Big[\Delta\pi(\mathbf{b}, s_i, s'_i, z_i)|\mathcal{J}_i\Big]$$

and consider two different assumptions on the distribution of the θ_i.

Case 1. Assume $\theta_i = \theta_0$, or more generally, that all determinants of the costs of drive time are observed and incorporated in the econometrician's specification. Letting $\rightarrow_P$ denote convergence in probability, then since $\Delta dt(s_i, s'_i, z_i) < 0$

$$N^{-1}\sum_i \nu_{1,i,s,s'} \rightarrow_P 0, \text{ implies } -\frac{\sum_i \Delta e(b_i, s_i, s'_i)}{\sum_i \Delta dt(s_i, s'_i, z_i)} \rightarrow_p \underline{\theta} \leq \theta_0. \qquad (2)$$

To obtain an upper bound for θ_0 consider an alternative store (s''_i) which was closer to the individual. An analogous argument shows

$$N^{-1}\sum_i \nu_{1,i,s,s''} \rightarrow_P 0, \text{ implies } -\frac{\sum_i \Delta e(b_i, s_i, s''_i))}{\sum_i \Delta dt(s_i, s''_i, z_i)} \rightarrow_p \overline{\theta} \geq \theta_0. \qquad (3)$$

I.e. provided the average of the expectational errors converge to zero, equations (2) and (3) give us asymptotic bounds for θ_0.

Case 2. Now assume that there is a determinant of the cost of drive time that the agent knows but is not observed by the econometrician, i.e. $\theta_i = (\theta_0 + \nu_{2,i})$, where $\sum \nu_{2,i} \equiv 0$, so θ_0 is the mean cost of drive time. Then our inequality becomes

$$\mathcal{E}\Big[\Delta \mathbf{e}(\mathbf{b}_i, s_i, s_i') - (\theta_0 + \nu_{2,i})\Delta dt(s_i, s_i', z_i)|\mathcal{J}_i\Big] \geq 0.$$

Now assume the agent knows drive time in deciding where to shop, or $\Delta dt(\cdot) \in \mathcal{J}_i$. Then

$$\frac{1}{N}\sum_i \Delta dt(s_i, s_i', z_i)^{-1}\nu_{1,i,s,s'} \rightarrow_P 0, \textbf{ which implies } \frac{1}{N}\sum_i \left(\frac{\Delta e(b_i, s_i, s_i')}{\Delta dt(s_i, s_i', z_i)}\right) \rightarrow_P \underline{\theta} \leq \theta_0.$$

$$(4)$$

An analogous upper bound to θ_0 is generated by choosing an alternative whose drive time is less then that of the chosen store.

Discussion. Case 1 uses *a ratio of averages* to bound the parameter of interest while case 2 uses *the average of a ratio*. The following points should be kept in mind.

Case 1 vs. Case 2. Case 2 allows for unobserved heterogeneity in the coefficient of interest and does not need to specify what the distribution of that unobservable is. In particular the unobservable can be *freely correlated* with the right hand side variable. "Drive time" is a choice variable, so we might expect it to be correlated with the perceived costs of that time (with $\nu_{2,i}$). If it is then Case 1 and Case 2 estimators should be different, otherwise they should be the same. So there is a test for whether any unobserved differences in preferences are correlated with the "independent" variable, and that test *does not* require us to specify a conditional distribution for $\nu_{2,i}$. Similar issues arise in analyzing the choice of a durable good when there is an intensity of use decision that follows the choice of the durable (for a classic example see Dubin and McFadden,1984).

Behavioral Conditions. This is a "two-stage" model with uncertainty; an initial choice of where to shop is made before knowing what prices are, and a choice of what to buy is made after arriving at the store. Note, however, that *we did not have to specify* either the *information on prices* the agent had at its disposal when it made its initial decision, *or the form of the agent's prior* price distribution conditional on that information. These are objects econometricians seldom have access to.

6

Conditions on the Choice Set. All we required of the choice set was one feasible alternative. In particular *we did not* need to specify and compute returns for the many possible "inside" choices and we did not need to specify an "outside" alternative.

Finally note that our focus on the drive time coefficient lead us to chose an alternative that differenced out any heterogeneity in preferences over bundles of goods. If instead we were interested in the utility of a particular good we would compare baskets with and without that good at the same store. If we had multiple observations on the same individual there are many more (largely unexplored) possibilities.

1.1 Estimates from the Inequality and A Comparison Model.

Katz (2007) estimates his model using the Nielsen Homescan Panel, 2004, for household expenditures and data from Retail Site Database of TradeDimensions for the characteristics of stores. He uses the shopping trips of about 1,300 families in Massachusetts and surrounding counties, and compares the results that use inequalities to the results that he obtains from estimating a discrete choice comparison model.

The Comparison Model. To obtain the econometric implications of a behavioral model of supermarket choice we would (i) specify the agent's prior distribution of prices at each store, (ii) compute the bundle of goods the agent would buy for each possible realization of the price vector at the store and (iii) form the expected utility of going to the store. These are demanding tasks. Similar considerations lead most (though not all, see below) analysis of single agent discrete choice problems to reduced forms. The reduced form can be given an appealing interpretation by constructing it from the regressions of expected returns from each choice on variables of interest. If we then make a sufficiently powerful assumptions on the joint distribution of the regressions function disturbances, the functions themselves can be estimated. It is the fact that an analogous reduced form *is not useful* in multiple agent problems that lead to the generalized discrete choice model considered below.

Unfortunately this reduced form can not be used in the supermarket choice problem without first reducing the dimension of the choice set. Katz assumes the number of weekly visits made to supermarkets is distributed as a Poisson random variable. At each visit the

consumer chooses between ten expenditure bundles at each of the outlets within a given radius of its home. The utility function for a given expenditure bundle and store is allowed to differ with the number of shopping trips per week, but for a fixed number of trips is given by equation (1) augmented with an additive "logit" error. The expenditure bundles are constructed from typical purchase patterns for a given amount of expenditure which are then priced at each outlet (giving us the expenditure level for each choice)[2].

There are a number of reasons to doubt the estimates from this model. I focus on those directly related to the price and drive time variables. First the prices for the expenditure class need not reflect the prices of the goods the individual actually is interested in. So there is an error in the price variable and if the individual shops at stores where the goods they are interested in are less costly, that error is negatively correlated with the price itself. Second the model does not allow for expectational errors. So agents are assumed to know all relevant prices when store choice decisions were made (and there are a lot of them). Finally the model does not allow for unobserved heterogeneity in the aversion to drive time. One could allow for a random coefficient on drive time and integrate it out but this would require an assumption on the conditional distribution for that variable, and given that the aversion to drive time is likely to be related to drive time per se, a traditional random coefficient assumption would be suspect.

Results. The specifications used for the models estimated were quite detailed; the comparison model estimated about forty different parameters for each of three different number of visits per week, while the revealed choice model estimated about fifteen. Both models were estimated with specifications that included outlet characteristics and interactions between expenditure and demographics; so the aversion to drive time varied with observed individual characteristics. Like the original paper I focus on the estimates of the median (of the mean) aversion to drive time.

The multinomial comparison models were estimated using maximum likelihood. The median aversion to drive time was estimated at two hundred and forty dollars per hour. The

---

[2]For a discussion of alternative ways of building reduced forms for supermarket discrete choice problems and an application; see Beckert et. al.,2009.

median wage in the region was seventeen dollars an hour, so this estimate is implausibly high. Also several of the other coefficients had the "wrong" sign.

The inequality estimators were obtained from differences between the chosen store and four different counterfactual store choices (chosen to reflect price and distance differences with the chosen store). Each comparison was interacted with positive functions of twenty six "instruments" (variables that were assumed to be mean independent of the expectational errors), producing over a hundred moment inequalities. As is not unusual for problems with many more inequalities than bounds to estimate, the inequality estimation routine generated point (rather than interval) estimates for the coefficients of interest (there was no value of the parameter vector that satisfied all of the moment inequalities). However tests indicated that one could accept the null that this result was due to sampling error.[3]

The inequality estimators that corresponded to case 1 above, i.e. those that did not allow for unobserved heterogeneity in the drive time coefficient, produced median aversions to drive time of about four dollars per hour. The estimators that corresponded to case 2 above, the case that did allow for heterogeneity in the drive time coefficient, generated estimates of the median aversion to drive time that varied between sixteen to eighteen dollars per hour, depending on the specification. The difference between the two estimators is consistent with their being unobserved heterogeneity in the drive time coefficient that is negatively correlated with drive time itself; a result one would expect from a model where drive time itself was a choice variable. Moreover in the model which allowed for heterogeneity the other coefficients took on values which accorded with intuition.

---

[3]The finding that there is no value of the parameter vector that satisfies all the inequalities is not unusual in moment inequality problems with many inequalities. Consider the one parameter case. When there are many moment inequalities there are many upper and lower bounds for that parameter. The estimation routine forms an interval estimate from the least upper and the greatest lower bound. The approximate normality of finite sample means implies that in finite samples the least upper bound will have a negative bias and the greatest lower bound will have a positive bias. So the two can easily cross. The test is a test of whether such crossings could have been a result of sampling error.

2 Conditions For Moment Inequality Estimators.

This section provides two sets of conditions that can be used to justify moment inequality estimators in more general (both multiple and single agent) settings. For each of the two approaches we consider estimation of both the parameters of the underlying behavioral model and a "reduced form" constructed by regressing expected profits on variables of interest. Each approach is defined by four assumptions, two of which are common across approaches and two differ. I begin with the two common assumptions.

2.1 Common Assumptions.

The first condition is that agents expect their choice to lead to higher returns than alternative feasible choices. Let $\pi(\cdot)$ be the profit function, d_i and d_{-i} be the agent's and its competitors' choices, D_i be the choice set, $\mathcal{J}_i$ be the agent's information set, and $\mathcal{E}$ be the expectation operator used by the agent to evaluate the implications of its actions. We assume

$$C1: \quad sup_{d \in D_i} \mathcal{E}[\pi(d, \mathbf{d}_{-i}, \mathbf{y}_i, \theta_0)|\mathcal{J}_i] \leq \mathcal{E}[\pi(d_i = d(\mathcal{J}_i), \mathbf{d}_{-i}, \mathbf{y}_i, \theta_0)|\mathcal{J}_i],$$

where $\mathbf{y}_i$ is any variable (other than the decision variables) which affects the agent's profits, and the expectation is calculated using the agent's beliefs on the likely values of $(\mathbf{d}_{-i}, \mathbf{y}_i)$. Throughout variables that the decision maker views as random will be boldface while realizations of those random variables will be represented by standard typeface.

Three points about C1 are central to what follows. First, there are *no restrictions* on either the choice set or the objective function. In particular the objective function need not be concave in d, D could be discrete (e.g.'s; a choice among bilateral contracts, ordered choice, ...) or continuous (e.g., the choice of the location and size of a retail outlet), and when continuous d_i can be at a corner of the choice set. Second, C1 is a necessary condition for a best response. As a result, were we to assume a Nash equilibrium C1 will be satisfied regardless of the equilibrium selection mechanism. Finally note that C1 is meant to be a rationality assumption in the sense of Savage (1954); i.e. the agent's choice is optimal with respect to the agent's beliefs. In itself it does not place any restrictions on the relationship

of those beliefs to the data generating process. We will need to restrict beliefs to evaluate estimators, but the restrictions used differ between the two approaches to estimation.

Both approaches require a model capable of predicting what expected profits would be were the agent to deviate from its observed choice. This is the sense in which both require a "structural" model. To predict what expected profits would be from a counterfactual choice we need to model what the agent thinks that y_i and d_{-i} would be were it to change its own decision. For example consider a two period model to determine the number of outlets (d_i) a retailer builds. Returns from the choice of d_i will be a function of post entry prices (which are in y_i) and the number of competing outlets (d_{-i}). If either of these are likely to change when d_i is changed to d' we need a model of that change in order to construct the firm's profits from its counterfactual choice. $C2$ formalizes this requirement.

We say $\mathbf{y}_i$ and/or $\mathbf{d}_{-i}$ are *endogenous* if they change in response to a change in d_i. $\mathbf{z}_i$ will represent a set of *exogenous* variables, i.e. variables whose distributions do not change in response to changes in d_i. Then our second condition is

$C2$: $\mathbf{y}_i = y(\mathbf{z}_i, d, \mathbf{d}_{-i}, \theta)$, *and* $\mathbf{d}_{-i} = d^{-i}(d, \mathbf{z}_i, \theta)$, *and the distribution of* $\mathbf{z}_i$ *conditional on* $(J_i, d_i = d)$ *does not depend on* d.

In words C2 states that if either $\mathbf{y}_i$ or $\mathbf{d}_{-i}$ is endogenous we need a model for its response to changes in d_i, and the model must produce a value for the endogenous variable which depends only on decisions and exogenous variables. The condition that the distribution of $\mathbf{z}_i$ does not depend on the agent's choice is what we mean when we say that $\mathbf{z}_i$ is exogenous.

The restrictiveness of C2 will vary with the problem. In *single agent* problems profits do not depend on d_{-i} and the agent's decision is typically not thought to effect environmental conditions so $\mathbf{y}_i$ is exogenous. Then C2 is unobjectionable. In multiple agent *simultaneous move* games $d^{-i}(d \neq d_i, \mathbf{z}_i, \theta) = \mathbf{d}_{-i}$, so there is no need for an explicit model of reactions by competitors, but $\mathbf{y}_i$ often contains price and/or quantity variables which are endogenous. If, in a *sequential move* game, we want to consider counterfactuals for agents who move early, we need a model for the responses of the agents that move later (for $\mathbf{d}_{-i}$)[4].

---

[4]Often it is natural to write that model recursively, so that each agent's decision depends on the decisions of the agents who move prior to it. The fact that we allow for sequential games explains the difference

Implications of C1 and C2. Let $d' \in D_i$ and define

$$\Delta\pi(d_i, d', d_{-i}, z_i, \theta_0) \equiv \pi(d_i, d_{-i}, y_i, \theta_0) - \pi(d', d^{-i}(d', z), y(z_i, d', d_{-i}), \theta_0).$$

Then together C1 and C2 imply the inequality

$$\mathcal{E}[\Delta\pi(d_i, d', \mathbf{d}_{-i}, \mathbf{z}_i, \theta_0)|\mathcal{J}_i] \geq 0, \quad \forall\, d' \in D_i. \tag{5}$$

To move from (5) to a moment inequality we can use in estimation we need to specify

- a measurement model which determines the relationship between the $\pi(\cdot, \theta)$ and (z_i, d_i, d_{-i}) that appear in the theory and the measures of them we use in estimation, and

- the relationship between the expectation operator underlying the agents decisions (our $\mathcal{E}(\cdot)$) and the sample moments that the data generating process provides.

These are the two aspects of the problem which differ across the two approaches. We begin with a measurement model which nests both their assumptions.

2.2 Measurement Model.

Let $r(d, d_{-i}, z^o, \theta)$ be the profit function specified by the econometrician up to an additively separable disturbance (so the z^o are observed), and define $\nu(\cdot)$ to be the difference between the profit function the agent responds to and this specification, so that

$$r(d, d_{-i}, z_i^o, \theta) \equiv \pi(d, d_{-i}, z_i, \theta) + \nu(d, d_{-i}, z_i^o, z_i, \theta). \tag{6}$$

The agent's decision is based on $\mathcal{E}[\pi(\cdot)|\mathcal{J}_i]$. We observe $r(\cdot)$ and have constructed ν so that $\mathcal{E}[r(\cdot)|\mathcal{J}_i] = \mathcal{E}[\pi(\cdot)|\mathcal{J}_i] + \mathcal{E}[\nu(\cdot)|\mathcal{J}_i]$. It follows that

$$r(d, d_{-i}, z_i^o, \theta) \equiv \mathcal{E}[\pi(d, \mathbf{d}_{-i}, \mathbf{z}_i, \theta)|\mathcal{J}_i] + \nu_{2,i,d} + \nu_{1,i,d}, \tag{7}$$

where

$$\nu_{2,i,d} \equiv \mathcal{E}[\nu(d, \mathbf{d}_{-i}, z_i^o, \mathbf{z}_i, \theta)|\mathcal{J}_i],$$

---

between our C2 and Assumption 2 in Pakes et. al., 2006. The buyer-seller network example in the next section is sequential and illustrates the types of assumptions needed to model $\mathbf{d}_{-i}$.

and

$$\nu_{1,i,d} \equiv \Big(\pi(d,\cdot) - \mathcal{E}[\pi(d,\cdot)|\mathcal{J}_i] \Big) + \Big(\nu(d,\cdot) - \mathcal{E}[\nu(d,\cdot)|\mathcal{J}_i] \Big).$$

Equation (7) expresses the difference between the researcher's specification of profits (i.e. $r(\cdot)$) and the function the agent bases its decision on (i.e. $\mathcal{E}[\pi(\cdot)|J_i]$) as a sum of three components; two of which I have grouped together into ν_1. The grouping was done because, when evaluated at $\theta = \theta_0$, they both are "mean independent" of $\mathcal{J}_i$ under the agent's expectation operator (under $\mathcal{E}$) by construction. ν_2 does not share this property and it is this distinction which forces us to keep track of two separate disturbances below. The relative importance of the two disturbances will differ with the application.

Sources of ν_1. ν_1 is a sum of two terms. $\pi(d,\cdot) - \mathcal{E}[\pi(d,\cdot)|\mathcal{J}_i]$ provides the difference between the agent's expectation of profits at the time the agent makes its decision and the realization of profits. In single agent problems it is solely a result of uncertainty in the exogenous variables whose realizations help determine returns (in the supermarket example, the uncertainty in the prices). In multiple agent problems their may also be uncertainty in $\mathbf{d}_{-i}$. In either case to compute the distribution of $\pi(d,\cdot) - \mathcal{E}[\pi(d,\cdot)|\mathcal{J}_i]$ we would have to specify the probabilities each agent assigns to different outcomes conditional on their information sets (objects we often know little about). In order to compute the distribution of ν_1 in the multiple agent case we would also have to solve for an equilibrium conditional on all possible realizations of $\mathbf{d}_{-i}$. This would both be computationally burdensome and require an assumption that selects among possible equilibria. The second component of ν_1, $\nu(d,\cdot) - \mathcal{E}[\nu(d,\cdot)|\mathcal{J}_i]$ results from either measurement error in observables or from a specification error in $r(\cdot)$ that is mean independent of $\mathcal{J}_i$. Note that such a "specification" error occurs when $r(\cdot)$ is formed by regressing the true profit function onto variables of interest to obtain a "reduced form" whose coefficients become the focus of investigation.

Sources of ν_2. ν_2 is defined to equal that part of profits that the agent can condition on when it makes its decisions but the econometrician does not include in the specification. So though it is not known to the econometrician, $\nu_{2,i} \in \mathcal{J}_i$, and since $d_i = d(\mathcal{J}_i)$, d_i will generally be a function of $\nu_{2,i}$. In the supermarket example $\nu_{2,i}$ has two components; the utility from

13

the goods bought (the $U(b_i, z_i)$ in equation 1) and, in case 2, the differences between the individual's and the average drive time coefficients $(\theta_i - \theta_0)$. In multiple agent problems d_i might also be a function of $\nu_{2,-i}$.

Selection. We can now explain the selection problem in behavioral models. Assume that x is an "instrument" in the sense that $\mathcal{E}[\nu_2|x] = 0$, and, in addition, that $x \in \mathcal{J}$. Then

$$\mathcal{E}[\nu_1|x] = \mathcal{E}[\nu_2|x] = 0.$$

These expectations do not, however, condition on the decision actually made (our d_i), and any moment which depends on the selected choice requires properties of the disturbance conditional on the d_i the agent selected. Since d_i is measurable $\sigma(\mathcal{J}_i)$, and ν_1 is mean independent of any function of $\mathcal{J}$

$$\mathcal{E}[\nu_{1,i,d}|x, d] = 0, \quad however \quad \mathcal{E}[\nu_{2,i,d}|x, d] \neq 0.$$

As a result the sample average of x and the residuals will typically not be zero when $\theta = \theta_0$, the condition we generally require of an "instrument".

To see why $\mathcal{E}[\nu_{2,i}|x, d] \neq 0$ consider a single agent binary choice problem ($d_i \in \{0, 1\}$). Then $d_i = 1$ implies

$$\mathcal{E}[\Delta\pi(d_i = 1, d' = 0, \cdot)|\mathcal{J}_i] = \mathcal{E}[\Delta r(d_i = 1, d' = 0, \cdot)|\mathcal{J}_i] + \Delta\nu_{2,i} \geq 0,$$

where $\Delta\nu_{2,i} = \nu_{2,i,d=1} - \nu_{2,i,d=0}$. So for every agent with $d_i = 1$

$$\Delta\nu_{2,i} > -\mathcal{E}[\Delta r(d_i = 1, d' = 0, \cdot)|\mathcal{J}_i].$$

If x_i is correlated with $\mathcal{E}[\Delta r(d_i = 1, d' = 0, \cdot)|\mathcal{J}_i]$, and if x is used as an instrument it is likely to be correlated $\Delta\pi(\cdot)$, then the expectation of $\Delta\nu_{2,i}$ given x_i and $d_i = 1$ will be not be zero, regardless of whether $\mathcal{E}[\nu_{2,i,d=1}|x_i] = \mathcal{E}[\nu_{2,i,d=0}|x_i] = 0$. In words if d_i was selected then the difference in the unobservable part of the incremental expected returns to d_i must have been greater than the (negative of the) difference in the observable part of the incremental returns, and the latter will typically be correlated with our instruments.

2.3   The Generalized Discrete Choice Approach.

The measurement model in equation (7) provides the notation needed to clarify the conditions needed to move from the profit inequalities in equation (5) to sample moments for our two approaches. Recall that we also need to clarify the relationship between the agent's perceptions of expected returns and the returns emanating from the data generation process embedded in those approaches. We begin with the generalized discrete choice approach (the approach originally developed to handle entry games in the papers by Tamer,2003, Ciliberto and Tamer, 2007, and Andrews Berry and Jia , 2007).

The multiple agent versions of the generalized discrete choice approach assume

$$DC3: \quad \forall d \in D_i, \quad \pi(d, d_{-i}, z_i, \theta_0) = \mathcal{E}[\pi(d, \mathbf{d}_{-i}, \mathbf{z}_i, \theta_0)|\mathcal{J}_i],$$

or that there is no uncertainty in either the exogenous variables (in $\mathbf{z}_i$) or in the actions of the firm's competitors (in $\mathbf{d}_{-i}$). Together C1 and DC3 imply that agents never err[5].

It is important to note that there are parts of the *single agent* discrete choice literature that *do* allow for uncertainty. These include both the dynamic single agent discrete choice models that explicitly account for randomness in exogenous variables, and the literature which uses survey data on expectations in conjunction with choice models to allow for uncertainty (see Keane and Wolpin, 2009, and Manski, 2004, respectively, and the literature they cite). However computational difficulties and a lack of information on agents' perceptions on the likely behavior of their competitors have made it difficult to use analogous techniques in multiple agent problems.

DC4 provides the restrictions the generalized discrete choice model places on the measurement model in equation (7).

$DC4.$ $\forall d \in \mathcal{D}_i,$ $r(d, d_{-i}, z_i^o, \theta) = \pi(d, d_{-i}, z_i, \theta) + \nu_{1,i},$ *for a known* $\pi(\cdot, \theta),$ *and*
$z_i = (\{\nu_{2,i,d}\}_d, z_i^o)$, *with* $\{(\nu_{2,i,d}, \nu_{2,-i,d})\}_d|_{z^o, z_{-i}^o} \sim F(\cdot; \theta),$ *for a known* $F(\cdot, \theta).$

---

[5]As stated $DC3$ also rules out the analysis of sequential games in which an agent who moves initially believes that the decisions of an agent who moves thereafter depends on its initial decision. However with only notational costs we could allow for a deterministic relationship between a component of d_{-i} and (d, z).

The first line in DC4 states that there are no decision specific errors in our profit measure ($\nu_{1,i}$ does not have a d subscript). So if we knew $(d_i, d_{-i}, z_i, z_{-i})$ we could construct an *exact* measure of profit *differences* for each θ. The second line states that z_i has both observed (the z_i^o) and unobserved (the $\nu_{2,i,d}$) components and provides their properties. The distribution of the unobserved conditional on the observed components are known up to a parameter vector[6], and there is no measurement error in the observed components. Since DC3 assumes full information all the $\nu_{2,i,d}$ are known to *all* agents when decisions are made, just not to the econometrician. Note that given $DC3$ the distribution $F(\cdot|\cdot)$ appearing in $DC4$ is a distribution of realized values, and hence must be consistent with the data generating process (more on this below).

Though the assumptions used in $DC3$ and $DC4$ may seem restrictive, they clearly advanced the study of discrete choice models in multiple agent settings. Recall that single agent discrete choice models can always be given an intuitive reduced form interpretation. Simply regress expected returns from alternate decisions onto observed variables and the decision itself, then solve for the optimal d conditional on the observables and the disturbances from the regressions, and finally make an assumption on the joint distribution of those disturbances that enables identification. The analogous reduced form for multiple agent problems proved not to be useful. In multiple agent contexts researchers were interested in the relationship between profits and (d_{-i}, z) *conditional on* unobservable determinants of profits, particularly those that were correlated with d_{-i}. For example in the entry models that stimulated this literature there was a focus on the relationship of profitability to the number of entrants. Models which did not allow for unobserved market characteristics that affected the profitability of all potential entrants in a market often estimated coefficients that implied that a firm's profits increased in the number of competitors (since more profitable markets attracted more entrants). So to provide a "reduced form" of interest for the relationship between profits and (z_i, d_i, d_{-i}) we needed to allow for a disturbance that was correlated with d_{-i}. This is the problem the generalized discrete choice approach sought to solve.

---

[6]There are papers in the single agent discrete choice literature which have allowed for classification errors in d; see for example Hausman and Scott Morton, 1998. At least in principal such errors could be added to any of the models considered here.

Substituting DC3 and DC4 into the model generated by C1 and C2 (equation 5) and letting $\nu_{2,i} \equiv \{\nu_{2,i,d}\}_d$, with analogous notation for $\nu_{2,-i}$, we obtain

$$Model\ D: \quad \forall d' \in D_i, \quad \Delta\pi(d_i, d', d_{-i}, z_i^o, \nu_{2,i}; \theta_0) \geq 0; \quad (\nu_{2,i}, \nu_{2,-i})|_{z_i^o, z_{-i}^o} \sim F(\cdot; \theta_0). \quad (8)$$

To insure that there exists a θ for which the event $\Delta\pi(d_i, d', d_{-i}, z_i^o, \nu_{2,i}; \theta) \geq 0$, has positive probability $\forall d \in D_i$ and all agents in each market, we need further conditions on $F(\cdot)$ and/or $\pi(\cdot)$. The additional restriction typically imposed is that the profit function is additively separable in the unobserved determinants of profits, that is

$$RD^{as}: \quad \forall d \in D_i, \quad \pi(d, d_{-i}, z_i^o, \nu_{2,i}) = \pi^{as}(d, d_{-i}, z_i^o, \theta_0) + \nu_{2,i,d}, \quad (9)$$

and the distribution $\nu_{2,i,d}$ conditional on $\nu_{2,-i}$ has full support $[\forall(i,d)]^7$.

Keep in mind that the additive separability in equation (9) *can not* be obtained definitionally. If we did observe realized profits and regressed it on (d_i, d_{-i}, z_i^o) we would get a residual, but that residual *is not* the $\nu_{2,i}$ in equation (9). The regression residuals are mean independent of (d_i, d_{-i}), while $\nu_{2,i}$ is not. So for the specification in equation (9) to be correct $\pi^{as}(\cdot)$ and $\nu_{2,i}$ have to be derived from the primitives of the problem.

Inequalities for Inference. Index markets by $j = 1, \ldots, J$. For inference I need the distribution of $\{(\nu_{2,i,j}, z_{i,j}^o, d_{i,j})\}_{i=1}^{n_j} \equiv (\nu_{2,j}, z_j^o, d_j)$ across the population of markets, which we denote by $P(\cdot)$. The data consist of random draws on $(\nu_{2,j}, z_j^o, d_j)$ from $P(\cdot)$. The expectations of any function $g(\cdot)$ of a draw conditional on x will be given by $E[g(\cdot)|x] = \int g(\cdot)dP(\cdot|x)$, so $E(\cdot)$ is defined by the data generating process (the DGP). The distribution of $\nu_{2,j}$ conditional on z_j^o in DC4, or $F(\cdot; \theta)$, is assumed to be consistent with this DGP.

The model's conditions can be satisfied by multiple vectors of d_j for any value of θ (i.e., there can be multiple equilibria). As a result there is not a one to one map between observables unobservables and parameters on the one hand, and outcomes for the decision variables on the other; so the model is not detailed enough to deliver a likelihood. However Cliberto and Tamer (2006) and Andrews Berry and Jia (2004) note that we can check

---

[7]Allowing for additional unobservables, for example unobservables random coefficients on the z_i^o, would increase the notational burden but would not change our ability to obtain any of the results below.

whether the model's conditions are satisfied at the observed d_j for any ν_j and θ, and this, together with $F(\cdot;\theta)$, enable us to calculate conditional probabilities of satisfying those conditions. Since these are necessary conditions for observing the choices made when $\theta = \theta_0$ the probability of satisfying them must be greater then the probability of actually observing d_j. In addition if we checked whether the d_j are the only values of the decision variables to satisfy the necessary conditions for each $\nu_{2,j}$ at that θ we could construct the probability that d_j is the unique equilibrium. That probability must be lower than the true probability of observing d_j at $\theta = \theta_0$. These are inequalities that not all values of θ will satisfy, and as a result they can be used as a basis for inference.

More formally define the probability that the model in equation (8) (with a restriction like that in equation 9) is satisfied at a particular d_j given z_j^o for a given θ to be

$$\overline{Pr}\{d_j \mid z_j^o, \theta\} \equiv Pr\{\nu_{2,j} : d_j \text{ satisfy equation (8)} | z_j^o, \theta\},$$

and the analogous lower bound to be

$$\underline{Pr}\{d_j \mid z_j^o, \theta\} \equiv Pr\{\nu_{2,j} : \text{only } d_j \text{ satisfy equation (8)} | z_j^o, \theta\}.$$

Letting $I\{\cdot\}$ be the indicator function which takes the value one if the condition inside the brackets is satisfied and zero elsewhere, the true probability (determined in part by the equilibrium selection mechanism) is

$$Pr\{d_j \mid z_j^o, \theta_0\} \equiv E[I\{d = d_j\} \mid z_j^o].$$

Since do not know the selection mechanism we do not know $Pr\{d_j \mid z_j^o, \theta_0\}$, but we do know that when $\theta = \theta_0$

$$\overline{Pr}\{d_j \mid z_j^o, \theta_0\} \geq Pr\{d_j \mid z_j^o, \theta_0\} \geq \underline{Pr}\{d_j \mid z_j^o, \theta_0\}.$$

Let $h(\cdot)$ be a function which only takes on positive values and $\rightarrow_P$ denote convergence in probability. Since $E(\cdot)$ provides cross markets averages we have

$$E\Big[J^{-1}\sum_j \Big(\overline{Pr}\{d_j|z_j^o,\theta\} - I\{d = d_j\}\Big)h(z_j^o)\Big] \rightarrow_P J^{-1}\sum_j \Big[\Big(\overline{Pr}\{d_j|z_j^o,\theta\} - Pr\{d_j|z_j^o,\theta_0\}\Big)h(z_j^o)\Big]$$

(10)

18

which is non-negative at $\theta = \theta_0$. An analogous moment condition can be constructed from $Pr\{d_j | z_j^o, \theta_0\} - \underline{Pr}\{d_j | z_j^o, \theta\}$. The estimation routine constructs unbiased estimates of $(\underline{Pr}(\cdot|\theta), \overline{Pr}(\cdot|\theta))$, substitutes them for the true values of the probability bounds into these moments, and then accepts values of θ for which the moment inequalities are satisfied[8].

Since typically neither the upper nor the lower bound are analytic function of θ simulation is used to obtain unbiased estimates of them. The simulation procedure is straightforward, though often computationally burdensome. Take pseudo random draws from a standardized version of $F(\cdot)$ as defined in $DC4$, and for each random draw check the necessary conditions for an equilibrium, i.e. the conditions in equation (8), at the observed (d_i, d_{-i}). Estimate $\overline{Pr}(d_i, d_{-i}|\theta)$ by the fraction of random draws that satisfy those conditions at that θ. Next check if there is another value of $(d, d_{-i}) \in D_i \times D_{-i}$ that satisfy the equilibrium conditions at that θ and estimate $\underline{Pr}(d_i, d_{-i}|\theta)$ by the fraction of the draws for which (d_i, d_{-i}) is the only such value. If we were analyzing markets with N interactive agents each of which had $\#D$ possible choices and used ns simulation draws on $\{\nu_{2,i}\}_{i=1}^N$, then for each market and each θ evaluated in the estimation routine we need to evaluate up to $ns \times \#D \times N$ inequalities to obtain estimates of $\overline{Pr}\{\cdot | \theta\}$, and we need to evaluate up to $ns \times (\#D)^N$ inequalities if we also estimated $\underline{Pr}(\cdot|\theta)$. This can be computationally expensive, particularly in multistage games solved by backward recursions as then to solve for each $\pi(d_i, d_{-i}, \cdot)$ we need to compute equilibria to later stages of the game.

2.4 Profit Inequalities.

An earlier version of this approach appears in Pakes, Porter, Ho, and Ishii (2006). Recall that what we need is an assumption on the relationship between; (i) the agents' perceptions of expected returns and the returns emanating from the data generating process and (ii) the profit measure the agent uses and the one the econometrician specifies.

---

[8]As noted by a referee this routine ignores information. If we can enumerate all possible equilibria, as is assumed if we use the lower bound, we could use the fact that the equilibrium selection probabilities must sum to one (for more detail see Beresteanu and Molinari, 2008). Also, I have implicitly assumed that there is an equilibrium in pure strategies for each point evaluated. This need not be the case; for a discussion of the implications of existence problems for econometric work on discrete games see Bajari,et.al.,(2006).

From equation (5), $\mathcal{E}[\Delta\pi(\cdot,\theta_0)|\mathcal{J}_i] \geq 0$ which implies $\mathcal{E}[\Delta\pi(\cdot,\theta_0)|x_i] \geq 0$ if $x_i \in J_i$. PC3 relates these expectations to averages from the data generating process (our $E(\cdot)$ operator).

$PC3:$ There is a positive valued $h(\cdot)$ and an $x_i \in J_i$ for which

$$\frac{1}{N}\sum_i \mathcal{E}\Big(\Delta\pi(d_i, d', \mathbf{d}_{-i}, \mathbf{z}_i, \theta_0)|x_i\Big) \geq 0 \;\; \Rightarrow \;\; E\Big(\frac{1}{N}\sum_i \Delta\pi(d_i, d', \mathbf{d}_{-i}, \mathbf{z}_i, \theta_0)h(x_i)\Big) \geq 0 \;.$$

PC3 nests DC3 as it allows for uncertainty, and it does so without requiring us to fully specify how the agent forms its expectations. If agents know (i) the other agents' strategies, i.e. $\mathbf{d}_{-i}(\mathcal{J}_{-i})$, and (ii) the joint distribution of other agents' information sets and the primitives sources of uncertainty (i.e. of $(\mathcal{J}_{-i}, \mathbf{z}_i)$ conditional on $\mathcal{J}_i$), then, provided all expectations exist, our optimality condition (C1) insures that $PC3$ is satisfied. These assumptions are, however, stronger than are needed for $PC3$. Several authors have noted that agents' expectations can satisfy C1 without them having such detailed information (see for e.g. Dekel, Fudenberg, and Levine,1999). Further, though correct expectations about profit differences are sufficient for PC3, they are *not necessary*. A weaker sufficient condition is

$$\frac{1}{N}\sum_i \mathcal{E}\Delta\pi(d_i, d', \mathbf{d}_{-i}, \mathbf{z}_i, \theta_0)h(x_i) - E\Big(\frac{1}{N}\sum_i \Delta\pi(d_i, d', \mathbf{d}_{-i}, \mathbf{z}_i, \theta_0)h(x_i)\Big) \geq 0, \qquad (11)$$

where again $E(\cdot)$ is defined by the DGP. If agents have *incorrect* expectations on $\Delta\pi(\cdot, \theta_0)$ but their expectational error is not systematically related to x_i (i.e. are *mean independent* of x_i), then (11) is satisfied with equality. Indeed PC3 is satisfied even if agents are *incorrect on average*, provided they are overly optimistic about the relative profitability of their choices.

The final condition used in this estimation strategy is designed to deal with the selection problem caused by the $\{\nu_{2,i,j,d}\}$ for the $i = 1, \ldots, N_j$ agents in market j. $C1$, $C2$, and the definitions in equation (7) imply that our model generates the restriction that

$$\mathcal{E}[\Delta\pi(\cdot; \theta_0)|\mathcal{J}] = \mathcal{E}[\Delta r(\cdot; \theta_0)|\mathcal{J}] - \Delta\nu_2 \geq 0.$$

$PC3$ insures that this implies that if $x \in \mathcal{J}$ *sample* averages of $\Delta\pi(\cdot; \theta_0)h(x) = \Delta r(\cdot; \theta_0)h(x) - \Delta\nu_2 h(x)$ have positive expectation. For consistency we require that the sample average of

the *observable* $\Delta r(\cdot; \theta)h(x)$ have positive expectation at $\theta = \theta_0$. This will be the case if the expectation of the average of $\nu_2 h(x)$ is non-negative. Above we stressed that even if x is an instrument in the sense that $\{\nu_{2,d}\}$ is mean independent of x in the population at large, the mean of $\nu_{2,d}$ conditional on x for those who made particular decisions *will* typically depend on x. Pakes et al (2006) present a general condition which insures that the selection resulting from conditioning on agents choices does not impact the consistency of our estimators. Here I consider three ways of forming moments that satisfy that condition that I have seen used, often in combination, in applied work. They are based on the researcher finding weighted averages of differences between actual and counterfactual choices that either; (i) difference out the effect of the ν_2, (ii) insure that we average over the ν_2 of every agent (so that there is no selection), or (iii) sum to an observable which controls for a weighted average of the ν_2.

PC4a: Differencing. Let there be G groups of observations indexed by g, counterfactuals $d'_{i,g} \in \mathcal{D}_{i,g}$, and positive weights $w_{i,g} \in \mathcal{J}_{i,g}$, such that $\sum_{i \in g} w_{i,g} \Delta \nu_{2,i,g,d_{i,g},d'_{i,g}} = 0$; i.e. a within-group weighted average of profit differences eliminates the ν_2 errors. Then

$$G^{-1} \sum_g \sum_{i \in g} w_{i,g} \Big(\Delta r(d_{i,g}, d'_{i,g}, \cdot; \theta_0) - \mathcal{E}[\Delta \pi(d_{i,g}, d'_{i,g}, \cdot; \theta_0) | \mathcal{J}_{i,g}] \Big) \rightarrow_P 0,$$

provided $G^{-1} \sum_g \sum_i w_{i,g} \Delta r(d_{i,g}, d'_{i,g}, \cdot; \theta_0)$ obeys a law of large numbers.

Our case 1 supermarket example is a special case of $PC4a$ with $n_g = w_{i,g} = 1$. There $d_i = (b_i, s_i)$, $\pi(\cdot) = U(b_i, z_i) - e(b_i, s_i) - \theta_0 dt(s_i, z_i)$ and $\nu_{2,i,d} \equiv U(b_i, z_i)$. If we measure expenditures up to a $\nu_{1,i,d}$ error, $r(\cdot) = -e(b_i, s_i) - \theta_0 dt(s_i, z_i) + \nu_{2,i,d} + \nu_{1,i,d}$. We chose a counterfactual with $b'_i = b_i$, so $\Delta r(\cdot) = \Delta \pi(\cdot) + \Delta \nu_{1,\cdot}$, and the utility from the bundle of good bought is differenced out. "Matching estimators", i.e. estimators based on differences in outcomes of matched observations, implicitly assume $PC4a$ (no differences in unobservable determinants of the choices made by matched observations). For other single agent examples see Pakes et. al.(2006).

For a multiple agent example consider two period entry games with common unobservable determinants of market profitability; the problem that stimulated the literature on using inequality estimators in multiple agent settings. For specificity consider two retailers, say $i = \{W, T\}$, deciding whether to enter different markets, so $d^i_j = \{1, 0\}$ and $d^i_j = 1$ indicates

the firm i enters market $j = [1, \ldots, J]$. If there are market specific unobservables known to the agents but not to the econometrician then $r_j(d_{i,j}, d_{-i,j}, z_i, \theta) = \mathcal{E}[\pi_j(d_{i,j}, \mathbf{d}_{-i,j}, \mathbf{z}_i, \theta)|J_i] + \nu_{2,j} + \nu_{1,i,j}$. The z_i include sources of cost differences (like warehouse and central office locations). Let $w_{i,j} = w_j \in \{0, J^{-1}\}$ with $w_j = J^{-1} \Leftrightarrow d_j^T = [1 - d_j^W]$; i.e. $\sum_j \sum_i w_{i,j} \Delta \pi_{i,j}(\cdot)$ only puts weight on markets where the two agents make opposite decisions.

The only possible counterfactual is $d_j^{i,\prime} = [1 - d_j^i]$. So if $w_j = 1$ and $d_j^W = 1$; $d_j^T = 0$, $\Delta r_j^W(\cdot) = \mathcal{E}[\Delta \pi_j^W(\cdot)|\mathcal{J}_W] + \nu_{2,j} + \nu_{1,W,j}$, and $\Delta r_j^T(\cdot) = \mathcal{E}[\Delta \pi_j^T(\cdot)|\mathcal{J}_T] - \nu_{2,j} + \nu_{1,T,j}$. Since $\nu_{2,j}$ enters the two inequalities with opposite signs, it cancels when we sum over $i = \{W, T\}$ and

$$\sum_j w_j \sum_{i=W,T} \Delta r_j^i(d^i, d^{i,\prime}, \mathbf{d}_{-i}, \mathbf{z}_i, \theta) = \sum_j w_j \sum_{i=W,T} \left(\mathcal{E}[\Delta \pi_j^i(d^i, d^{i,\prime}, \mathbf{d}_{-i}, \mathbf{z}_i, \theta)|\mathcal{J}_i] + \nu_{1,i,j} \right)$$

which will be non-negative at $\theta = \theta_0$.

Notice that since we assume $w_{i,j} \in \mathcal{J}_{i,j}$, these weights imply $d_{-i,j} \in \mathcal{J}_{i,j}$, as in the generalized discrete choice model, but we do not require assumptions on the distribution of $\mathbf{z}_i$ or of $\nu_{i,\cdot}$. Also more moment inequalities can be generated from appropriate instruments and the model can be enriched to explicitly allow for differences in the firm's responses to the market specific shock (replace $\nu_{2,j}$ by $\theta z_j^i \nu_{2,j}$, where $z_j^i \geq 0$ and $z_j^i \in J_i$, and then divide each difference by z_j^i). Similar structures appear in a number of other familiar problems (e.g. social interaction models where the interaction effects are additive and group specific).

In simultaneous move games where the market allocation mechanism is known one can often construct counterfactuals which difference out *individual* specific (rather than group specific) ν_2 effects. For e.g. in electricity auctions with known allocation mechanisms we can compute the difference between the revenues and quantities actually allocated to the agent and those the agent would have obtained had the agent submitted a different bid (holding the realizations of environmental variables and competitors' bids constant). Profits are revenues minus fixed and variable costs. The fact that the expectation of the difference in profits from the two bids should be positive allows us to bound the variable cost function without restricting agent-specific fixed costs in any way (as they are differenced out).

PC4b: Unconditional Averages and Instrumental Variables. Assume that $\forall d \in D_i$, there is a $d' \in D_i$ and a $w_i \in \mathcal{J}_i$ such that

$$w_i \Delta r(d_i, d'_i, \cdot; \theta) = w_i \mathcal{E}[\Delta \pi(d_i, d'_i, \cdot; \theta)|\mathcal{J}_i] + \nu_{2,i} + \Delta \nu_{1,i,\cdot},$$

Then if $x_i \in \mathcal{J}_i$, $E[\nu_{2,i}|x_i] = 0$, and $h(\cdot) > 0$

$$N^{-1} \sum_i w_i \Delta r(d_i, d'_i, \cdot; \theta_0) h(x_i) \to_P N^{-1} \sum_i w_i \mathcal{E}[\Delta \pi(d_i, d'_i, \cdot; \theta_0)|\mathcal{J}_i] h(x_i) \geq 0,$$

provided $N^{-1} \sum \nu_{1,i,\cdot} h(x_i)$ and $N^{-1} \sum \nu_{2,i} h(x_i)$ obey laws of large numbers. PC4b assumes there is a counterfactual which gives us an inequality that is additive in ν_2 *no matter the decision the agent made.* Then we can form averages *which do not condition on d*, and hence do not have a selection problems. This form of *PC4b* suffices for the examples in this section but the next section requires a more general form given in the footnote below.[9]

Case 2 of our supermarket example had two ν_2 components; a decision specific utility from the goods bought, $\nu_{2,i,d} = U(b_i, z_i)$ (like in case 1), and an agent specific aversion to drive time, $\theta_i = \theta_0 + \nu_{2,i}$. As in case 1, taking $d' = (b_i, s'_i)$ differenced out the $U(b_i, z_i)$. Then $\Delta r(\cdot) = -\Delta e(\cdot, s_i, s'_i) - (\theta_0 + \nu_{2,i})\Delta dt(s_i, s'_i, z_i) + \Delta \nu_{1,\cdot}$. Divide by $\Delta dt(s_i, s'_i, z_i) \leq 0$. Then $C1$ and $C2$ imply that $\mathcal{E}[\Delta e(s_i, s'_i, b_i)/\Delta dt(s_i, s'_i, z_i)|\mathcal{J}_i] - (\theta_0 + \nu_{2,i}) \leq 0$. This inequality is; (i) linear in $\nu_{2,i}$, and (ii) is available for every agent. So if $E[\nu_2] = 0$, $PC3$ and a law of large numbers insures $N^{-1} \sum_i \nu_{2,i} \to_P 0$, and $\sum_i \Delta e(s_i, s'_i, b_i)/\Delta dt(s_i, s'_i, z_i) \to_P \underline{\theta}_0 \leq \theta_0$; while if $E[\nu_2|x] = 0$ we can use x to form instruments. Notice that $\nu_{2,i}$ can be correlated with $dt(z_i, s_i)$ so this procedure enables us to analyze discrete choice models when a random coefficient affecting tastes for a characteristic is correlated with the characteristics chosen.

For a multiple agent example of *PC4b* we look at within market expansion decisions. Agents chose a number of outlets, a $d_i \in \mathcal{Z}_+$ (the integers) to maximize expected profits. Formally the model is a multiple agent two period *ordered* choice model; a model with many

---

[9]The more general version is as follows. Assume G groups of observations with n_g members in group g, and that for each $\{d_{i,g} \in D_{i,g}\}_i$ there is a counterfactual $\{d'_{i,g} \in D_{i,g}\}_i$ and positive weights $w_{i,g} \in \mathcal{J}_{i,g}$ such that $\sum_{i \in g} w_{i,g} \Delta r(d_{i,g}, d'_{i,g}, \cdot, \theta_0) = \sum_{i \in g} w_{i,g} \left(\mathcal{E}[\Delta \pi(d_{i,g}, d'_{i,g}, \cdot, \theta_0)|\mathcal{J}_{i,g}] + \nu_{2,g} + \nu_{1,i,g} \right)$, Then if $x_{i,g} \in \mathcal{J}_{i,g}$, $E[\nu_{2,g}|x_{i,g}] = 0$ and $h(\cdot) > 0$, $G^{-1} \sum_g \sum_{i \in g} h(x_{i,g})(w_{i,g} \Delta r(d_{i,g}, d'_{i,g}, \cdot; \theta_0) \to_P G^{-1} \sum_g \sum_{i \in g} h(x_{i,g})(w_{i,g} \mathcal{E}\left[\Delta \pi(d_{i,g}, d'_{i,g}, \cdot; \theta_0)|\mathcal{J}_{i,g}\right] \geq 0$, provided laws of large numbers hold.

IO applications (e.g. Ishii, forthcoming). In the first period the agents chose a number of outlets and in the second they obtain the variable profits from sales at those outlets. So $\pi_i(\cdot) = vp(d_i, d_{-i}, \cdot) - (c_0 + \nu_{2,i})d_i$ where $vp(\cdot)$ are variable profits and $(c_0 + \nu_{2,i})$ represent the costs of building and maintaining the outlets. These costs differ across firms in ways known to the firms but not to the econometrician, and c_0 is defined to be their average (so $\sum \nu_{2,i} \equiv 0$). We measure variable profits up to a $\nu_1(\cdot)$ measurement error, so define $r_i(\cdot) = \pi_i(\cdot) + \nu_{1,d_i} = vp_i(\cdot) - c_o d_i + \nu_{2,i} d_i + \nu_{1,d_i}$, where $vp_i(\cdot)$ and d_i are observed.

C1 and C2 imply that the incremental profits from choosing one more machine than was actually chosen (a $d_i' = d_i + 1$) are expected to be less than its cost, or $\mathcal{E}[\Delta \pi(d_i, d_i + 1, \cdot)|\mathcal{J}_i] \leq 0$. But $\Delta r(d_i, d_i + 1, \cdot) = \mathcal{E}[\Delta \pi(d_i, d_i + 1, \cdot)|\mathcal{J}_i] + \nu_{2,i} + \Delta \nu_{1,i,\cdot}$. So since $\sum \nu_{2,i} \equiv 0$

$$N^{-1} \sum_i \Delta r(d_i, d_i + 1, \cdot) \to_P N^{-1} \sum_i \Delta \pi(d_i, d_i + 1, \cdot) = N^{-1} \sum_i [\Delta vp(d_i, d_i + 1, \cdot) - c_0] \leq 0,$$

where we have assumed PC3 and a law of large numbers. I.e. $N^{-1} \sum_i [\Delta r(d_i, d_i + 1, \cdot) \to_P \underline{c}_0 \leq c_0$. An upper bound for c_0 can be obtained by choosing $d_i' \leq d_i$ (see Pakes et.al, 2006, for the case where some observations are at $d_i = 0$, in which case the counterfactual $d_i' < d_i$ is infeasible). Additional moments can be obtained by forming covariances with $h(x_i)$ that are (unconditionally) uncorrelated with $\nu_{2,i}$. Notice that $E[v_2 h(x)] = 0$ is our only assumption; in particular we do not require $d_{-i} \in \mathcal{J}_i$ as in our previous multiple agent example.

PC4c: Control Functions. Often variables that are not available at a disaggregated level are available as aggregates, and this can be used to develop a control function for ν_2. Two familiar examples are; (i) firm level exports to different countries are typically unavailable but aggregate trade flows by product and country of destination are recorded, and (ii) product level input, cost, and sales data of multiproduct firms are not typically available, but both firm level aggregates over products, and product level aggregates over firms, often are. We illustrate with a firm location application adapted from de Loecker et. al (*in process*) that uses the trade data. An alternative illustration would have been to use the available data on multiproduct firms in conjunction with the product level aggregates over firms, to analyze multiproduct cost functions.

The output of each firm at each location is known, but where that output is sold is not.

The fixed costs of firm i in location d are, $f(z_i, d, \theta)$, while its marginal costs $m(z_i, d; \theta)$ (both vary with location of production and firm characteristics). Let $q_{i,d}$ be the quantity firm i produces, and $q_{i,e} = q_e + \nu_{2,i}^e$ be the quantity it exports to market e. Here q_e is average exports to e so $\sum_i \nu_{2,i}^e \equiv 0$ $(\forall e)$. $c(d, e)$ is the transportation costs (which vary with the location of production, d, and consumption,e). We observe $q_{i,d}$ and z_i and measure the firm's total cost up to a ν_1 error, or $r(\cdot)$ where

$$r(z_i, d; \theta) = f(z_i, d; \theta) + m(z_i, d; \theta)q_{i,d} + \sum_e c(d, e, \theta)q_e + \sum_e c(d, e, \theta)\nu_{2,i}^e + \nu_{1,i,d}.$$

Each firm producing in d chose its location to minimize expected costs and each could have produced in counterfactual location d' without changing the countries it sold to. Summing the expected cost difference in moving all firms from d to d', letting Q^e be (the observed) total exports to e, and using $\sum_i \sum_e c(d, e)\nu_{2,i}^e = 0$ we have

$$N^{-1} \sum_i \Delta r(d, d', \cdot, \theta) \to_P N^{-1} \sum_i \left(\Delta f(d, d', \cdot, \theta) + \Delta m(d, d', \cdot, \theta)q_{i,d} \right) + \sum_e \Delta c(d, d', e, \theta)Q^e \leq 0$$

at $\theta = \theta_0$, which allows us to combine the micro data on firms costs and aggregate data on exports to bound the parameters of interest. Notice that we do so without having to either estimate demand functions or make a pricing assumption in each country.

2.5   Comparing the Two Approaches.

The two approaches differ in their informational ($DC3$ vs $PC3$), and their measurement, ($DC4$ vs $PC4$) assumptions. They also differ in their computational properties but these are discussed in the context of the Monte Carlo example in section 4.

$PC3$ nests $DC3$. $PC3$ allows for uncertainty and does so *without having to specify* either the agents' information sets or their subjective probability distributions conditional on those information sets; objects we typically know little about. It also allows for expectational errors provided they are mean independent of the instruments. $DC3$ assumes agents know the returns from every choice and correctly optimize. Partly as a result $DC3$'s primary use in multiple agent settings has been as a characterization of "rest points" of environments which are "stable" over time; a setting often invoked to justify the use of two period games

to structure cross sectional empirical work. Notice that even in this setting the the combination of $DC3$ with $DC4$ only leads to consistent bounds if the profit (or value) function is constructed from correctly specified primitives. If instead a reduced form from regressing profits on variables of interest is used, the model contain a regression error. This violates the assumptions of the generalized discrete choice model and will lead to inconsistent estimates of the bounds on the reduced form parameters. We show how to adapt the discrete choice model to accommodate reduced forms below.

The two measurement assumptions ($PC4$ and $DC4$) are not nested. Indeed the simplest of the measurement assumptions used to justify the two models are distinctly different. Sufficient conditions for $PC4$ are that the $\{\nu_{2,i,d}\}$ do not vary over d and that there are instruments which are uncorrelated with the $\{\nu_{1,i,d}\}$. $DC4$ requires that the $\{\nu_{1,i,d}\}$ do not vary over d and that there be a known joint distribution for the $\{\nu_{2,i,d}\}$. When $\{\nu_{2,i,d}\}$ does vary over d then to use the profit inequalities approach we need to control for a selection problem. Our ability to do so typically depends on the richness of the set of feasible counterfactual choices, and the appropriate form(s) for the heterogeneity. When $\{\nu_{1,i,d}\}$ does vary over d then the generalized discrete choice model will not deliver consistent bounds. $\{\nu_{1,i,d}\}$ will vary over d if there is uncertainty about outcomes, measurement error, or specification error. For some of these cases we can modify the generalized discrete choice model and its estimation algorithm to deliver consistent bounds.

For a familiar example consider the case where we are analyzing the profits from entering markets in which the true profit from entering is additively separable in ν_2, or $\pi(d_{i,j} = 1, d_{-i,j}, z^o_{i,j}, \nu_{2,i,j}) = \pi^{as}(d_{i,j} = 1, d_{-i,j}, z^o_{i,j}, \theta_0) + \nu_{2,i,j}$, as in equation (9), and the profits from not entering are normalized to zero. As in most entry models, we are interested in the reduced form obtained by regressing $\pi^{as}(\cdot)$ onto $z^o_{i,j}$ and the number of competitors, say $\sum_i \{d_{i,j} = 1\}$. So

$$r(d_{i,j} = 1, d_{-i,j}, z^o_{i,j}, \theta) = z^o_{i,j}\theta_z + \sum_i \{d_{i,j} = 1\}\theta_d + \nu_{2,i,j} + \nu_{1,i,j}.$$

where $\nu_{1,i,j}$ is the regression error, so $E[\nu_{1,i,j}|z^o_{i,j}, \sum_i\{d_i = 1\}, \nu_{2,i,j}] = 0$ by construction. Recall that to obtain the inequalities used in estimation in this model we have to check whether the Nash conditions are satisfied at a given θ; i.e. to check whether $d_{i,j}$ maximizes

$z_{i,j}^o \theta_z + \sum \{d_{i,j} = 1\}\theta_d + \nu_{2,i,j} + \nu_{1,i,j}$ conditional on $d_{-i,j}$ for each i. To do this we will need an assumption on the joint distribution of the $\nu_{1,i,j}$, as well as for the $\{\nu_{2,i,j}\}$; and the $\nu_{1,i,j}$ must be mean independent of the $\{\nu_{2,i,j}\}$, $d_{i,j}$ and $z_{i,j}$, while the $\{\nu_{2,i,j}\}$ are determinants of $d_{i,j}$. Given the additional distributional assumption the estimation algorithm is analogous to that in section 2.3, though a more complex simulator must be used.

It is not as easy to accommodate measurement error in the generalized discrete choice framework. If $z_{i,j}^o = z_{i,j}^* + \nu_{1,i,j}$ and the agent responds to z^* (not z^o), estimation requires checking whether $z_{i,j}^* \theta_z + \sum_i \{d_{i,j} = 1\}\theta_d + \nu_{2,i,j,d}$ satisfies the Nash conditions. To obtain a simulator that would allow us to do so we would need to draw from the distribution of z^* conditional on z^0. This is typically not known nor is it easy to estimate. In contrast classical measurement error does not affect the consistency of the profit inequality estimator.

Finally $DC4$ requires explicit distributional assumptions on the $\{\nu_{2,i,j}\}$, while the profit inequality model relies only on mean independence assumptions. The need for an explicit distributional assumption is a concern, as the generalized discrete choice corrects for selection by finding the probability of ν_2-values which induce the agent to make particular choices. Those probabilities depend on the properties of the tail of the ν_2 distributions; properties we often know little about. The Monte Carlo example in section 4 is designed to investigate the robustness of the two models to violations of their assumptions and indicates that it is the distributional assumption of the generalized discrete choice model that is most problematic.

3 A Multiple Agent Example (Buyer-Seller Networks).

Vertical markets typically contain a small number of both sellers and buyers (who resell the products they buy to consumers). Most buyers buy from more than one seller while most sellers sell to more than one buyer. The terms of the payments the buyer makes to the seller are negotiated and vary with underlying market conditions. These terms determine both the costs buyers factor into the prices they set when they re-market the goods they sell to consumers, and the split of the profits between the sellers and the buyers and hence the sellers' incentives to invest in cost reductions (or product improvements). Unfortunately those terms are often proprietary; a seller bargaining with many buyers may not want one

buyer to know the terms of its other contracts.

Costs are also often proprietary in consumer goods markets. However since these are markets with many purchasers we typically assume sellers have the power to set prices (or quantities) in them. Then the first order conditions from a Nash equilibrium can be used to back out costs; i.e. we can find the marginal costs that insure that no firm has an incentive to deviate from the observed prices. This section uses moment inequalities to unravel features of the payment structure in vertical markets in an analogous way. We observe which sellers establish contracts with which buyers and, were we to know the buyer's cost function, could compute approximations to both the buyers and the sellers profits from; (i) the existing arrangement, and from (ii) a counterfactual in which one of the observed relationships is changed. So we proceed by parameterizing the buyer's cost function and look for values of the parameter vector that, on average, make the profits from the observed contracts larger than those from possible counterfactuals; i.e. values of θ that make the observed relationship in the interests of both agents.

We analyze an HMO/hospital example. To see how market characteristics effect payments in this example consider two different situations. In one a hospital with excess capacity in a neighborhood with several other similar non-capacity constrained hospitals is bargaining with an HMO. The HMO has already contracted with other neighborhood hospitals. Since there are similar options for consumers who require a hospital, the HMO's attractiveness to consumers is relatively insensitive to the inclusion of the given hospital in its network. As a result were the HMO to include that hospital it would not, in equilibrium, increase the premium it charges to consumers. So for the hospital's contract offer to be accepted by the HMO the contract would have to set hospital prices low enough for the HMO to prefer sending patients to that particular hospital rather than to its neighbors. On the other hand if the hospital was the only hospital in the neighborhood the HMO would be unlikely to attract any customers from that neighborhood without having the hospital in its network. Then, provided it is in the HMO's interest to operate in the neighborhood, the hospital should be able to extract nearly all the (hospital related) premiums that would be generated in that neighborhood.

To use this logic in estimation we will have to specify what the buyer (seller) would have

expected to happen if it had made a counterfactual choice. This requires assumptions, though the assumptions need not specify the form of the contracts. We suffice with a reduced form for the payments generated by the contracts obtained by regressing the HMO's per patient payments on hospital characteristics. In these respects this paper follows the assumptions used in Ho's (2009) work on HMO/hospital markets. Ho's analysis assumes that there are no structural disturbances in her data (in our notation, $\nu_{2,i,d} = \nu_{2,i}, \forall d$). I begin by showing that, by changing the moment inequalities taken to data, we can develop an estimation algorithm for her model that allows for both ν_2 and ν_1 errors. I then compare the results from estimators that do, and those that do not, allow for ν_2 errors. Next full information equilibria are computed from a structural buyer-seller network game with primitives similar to those in the empirical example. The reduced form implied by the computed equilibria is calculated, compared to the estimates obtained from the inequality estimators, and explored for possible additional correlates of contract characteristics.

3.1   Empirical Analysis.

Ho (2009) uses a two period game to structure the analysis; in the first period contracts between HMOs and hospitals are established and in the second period the HMOs engage in a premium setting game which conditions on those contracts (and is assumed to have a unique Nash equilibria). Once the premiums are set consumers chose HMO's and, if the need arises, chose a hospital in their HMO's network.

The premium setting game generates revenues for *each* HMO conditional on *any* configuration of hospital networks. Let H_m be a vector of dimension equal to the number of hospitals whose components are either zero or one, a one indicating the hospital is in HMO m's network and a zero indicating it is not. H_{-m} specifies the networks of the competing HMOs. The revenues the HMO receives from the premium setting game, say $R_{\mathtt{m}}(H_{\mathtt{m}}, H_{-\mathtt{m}}, z)$, and the number of patients HMO m sends to hospital h, say $q_{\mathtt{m},h}(H_{\mathtt{m}}, H_{-\mathtt{m}}, z)$, depend on these networks and exogenous variables (our z).

The profits of the HMO are the revenues from the second period game minus the transfers

the HMO makes to the hospitals in its network in payment for their services, say $T_{\mathtt{m},h}$ or

$$\pi_{\mathtt{m}}^{M}(H_{\mathtt{m}}, H_{-\mathtt{m}}, z) = R_{\mathtt{m}}(H_{\mathtt{m}}, H_{-\mathtt{m}}, z) - \sum_{h \in H_{\mathtt{m}}} T_{\mathtt{m},h}(H_{\mathtt{m}}, H_{-\mathtt{m}}, z).$$

Analogously if c_h is the per patient costs of hospital h and M_h is the hospital's network of HMOs, the hospital's profits are

$$\pi_{h}^{H}(M_{h}, M_{-h}, z) = \sum_{\mathtt{m} \in M_{h}} T_{\mathtt{m},h}(H_{\mathtt{m}}, H_{-\mathtt{m}}, z) - c_{h} \sum_{\mathtt{m} \in M_{h}} q_{\mathtt{m},h}(H_{\mathtt{m}}, H_{-\mathtt{m}}, z).$$

We project $T_{m,h}$ onto a set of interactions of $q_{\mathtt{m},h}(\cdot)$ with a vector of hospital characteristics, say z_h, and look for bounds on the resulting reduced form parameters; i.e. if $x_{\mathtt{m},h}(\cdot) = q_{m,h}(\cdot) \times z_h$ are the interactions, we estimate the θ in $T_{m,h}(H_{\mathtt{m}}, H_{-\mathtt{m}}, z) = x_{m,h}(H_{\mathtt{m}}, H_{-\mathtt{m}}, z)\,\theta + \nu_{m,h}$, where $\nu_{m,h}$ are uncorrelated with $x_{m,h}$ by construction. Note that if agents know more about the details of the contracts they sign than is captured by $x_{\mathtt{m},h}(\cdot)$, $\nu_{m,h}$ has a component which is known to both agents when they make their decisions (a "ν_2" component). Substituting this form of $T_{\mathtt{m},h}(\cdot)$ into the two profit functions we obtain

$$\pi_{\mathtt{m}}^{M}(\cdot, \theta) \equiv R_{m}^{M}(H_{\mathtt{m}}, H_{-\mathtt{m}}, z) - \sum_{h \in H_{m}} x_{m,h}(H_{\mathtt{m}}, H_{-\mathtt{m}}, z)\theta - \sum_{h \in H_{m}} \nu_{m,h}, \quad \text{and} \tag{12}$$

$$\pi_{h}^{H}(\cdot, \theta) \equiv \sum_{m \in M_{h}} x_{m,h}(H_{\mathtt{m}}, H_{-\mathtt{m}}, z)\theta - c_{h} \sum_{m \in M_{h}} q_{m,h}(H_{\mathtt{m}}, H_{-\mathtt{m}}, z) + \sum_{m \in M_{h}} \nu_{m,h}.$$

These equations determine actual (in contrast to measured) realized profits. Our measured variables, $(R_{\mathtt{m}}^{o}(\cdot), x_{m,h}^{o}(\cdot), q_{\mathtt{m},h}^{o}(\cdot), c_{h}^{o})$ are either obtained directly from data or from a careful study of hospital demand and the formation of HMO premiums described in Ho (2009). We assume that they are correct up to a mean zero measurement error. That is our measure of profits for HMO m and hospital h given a value for θ are

$$r_{\mathtt{m}}^{M}(\cdot, \theta) \equiv R_{m}^{o}(\cdot) - \sum_{h \in H_{m}} x_{m,h}^{o}(\cdot)\,\theta, \quad \text{and} \quad r_{h}^{H}(\cdot, \theta) \equiv \sum_{m \in M_{h}} x_{m,h}^{o}\theta - c_{h}^{o} \sum_{m \in M_{h}} q_{m,h}^{o}(\cdot), \tag{13}$$

and our assumptions imply

$$r_{\mathtt{m}}^{M}(\cdot, \theta) \equiv \mathcal{E}[\pi_{m}^{M}(\cdot, \theta)|\mathcal{J}_{m}] + \mathcal{E}[\sum_{h \in H_{m}} \nu_{m,h}|\mathcal{J}_{m}] + \nu_{1,m,M_{h},M_{-h}}, \quad \text{and}$$

$$r_{h}^{H}(\cdot, \theta) \equiv \mathcal{E}[\pi_{h}^{H}(\cdot, \theta)|\mathcal{J}_{h}] - \mathcal{E}[\sum_{m \in M_{h}} \nu_{m,h}|\mathcal{J}_{h}] + \nu_{1,h,H_{m},H_{-m}}.$$

Counterfactuals. Equation (12) provides the profits agents obtain from the observed network. To obtain our moment inequalities we have to consider the profits, and hence the network, the agents thought would have obtained from counterfactual behavior, and this requires assumptions on the contracting game. Ho's assumption are both familiar and computationally convenient (we consider a less convenient alternative below). She assumes that sellers simultaneously make take it or leave it offers to buyers, who then simultaneously accept or reject. As in Hart and Tirole (1991) the contract offers are assumed to be proprietary: each HMO knows the offers made to it but not to its competitors, and each hospital knows the offers it makes but not those of its competitors. We observe which HMOs contracted with which hospitals and can compute our measures of returns from any network. To form our moment inequalities we need to know the network that would be established were either an HMO or a hospital to change its behavior[10].

The HMO's act last. So our assumptions imply that the HMO could reverse any of its decision without changing the behavior of any other agent. Accordingly our HMO counterfactuals are obtained by reversing the HMO's acceptance/rejection decision with each of the hospitals in the market, leaving all other contracts unchanged, and computing the difference in the HMO's profits between the actual and the counterfactual networks.

To obtain a profit inequality for the hospital we have to; (i) specify an alternative offer the hospital could make, and (ii) either specify what the hospital thinks the particular HMO would do were it offered the alternative contract, or compute a lower bound to the profits the hospital could make as a result of the actions the HMO might take in response to the alternative contract. We assume that the hospital could always offer a null contract (a contract which is never accepted). What the hospital thinks the HMO would do if offered this contract depends on how the hospital thinks receiving the alternative contract would affect the HMO's beliefs about the contracts offered to other hospitals, and given those beliefs, whether the hospital thinks the HMO would change its replies to the contracts offered by

---

[10]Since we assume that the premium setting game is a full information game, our assumptions are what McAfee and Schwartz (1994) refer to as "ex poste observability"; the HMO's do not know each other's offer in the first period, but the costs in each accepted contract are revealed before the second stage of the game. This assumption could be relaxed at a cost of increasing the computational burden of estimation.

other hospitals. We assume "passive beliefs", i.e. the hospital believes that the HMO will not change its beliefs about the offers the hospital makes to other HMOs were it to receive the counterfactual offer, and present results which assume that the hospital thinks the HMO would not change its behavior with other hospitals were it to receive the null contract. However we have also done the analysis assuming the hospital thinks the HMO might add a different hospital with little difference in empirical findings.

Inequalities Used. We began with Ho's assumption that $\mathcal{E}[\nu_{m,h}|\mathcal{J}_i] = 0$ for $i = (m,h)$. Then the only disturbances in equations (13) are ν_1 disturbances, so we can form our inequalities by interacting positive functions of variables that were known to the decision maker when the decision was made with the difference between our models' estimates of the profits actually earned and those that would have been generated by our counterfactuals. Recall that these are the HMO's profits from reversing its decision with each hospital, and hospital's profits from offering a null contract to an HMO which had accepted its offer.

Next we considered alternative ways of allowing $\nu_{m,h}$ to have a ν_2 component, i.e. we allowed $\mathcal{E}[\nu_{m,h}|\mathcal{J}_i] \equiv \nu_{2,m,h} \neq 0$ and the same value for $i = m, h$. We first tried $\nu_{2,m,h} = \nu_{2,m}$, $\forall(m,h)$; i.e. that the ν_2 are HMO specific fixed effects. As shown in Appendix 1, we can then use $PC4a$ to generate a quite detailed set of inequalities. There is no *a priori* reason to assume a fixed effects structure here and when we did it accentuated the problems with the ν_1-only model[11]. So we used the generalized version of $PC4b$ in footnote 9 to develop an estimator for the buyer-seller network problem that allows for a $\nu_{2,m,h}$ of a general form.

Recall that the $\nu_{m,h}$ are a component of transfers, so the same $\nu_{m,h}$ value that goes into a hospital's revenues is a component of an HMO's costs. Let $\chi_{m,h}$ be the indicator function for whether a contract is established between m and h with $\chi_{m,h} = 1$ if it is established

---

[11]In the ν_1 only model about 12% of the inequalities were negative but under 2% were individually significant at the 5% level. In the model with fixed effects, about a third of the inequalities were negative and 10% were significant at the 5% level. A more complete analysis of effects models in buyer-seller networks would allow for both buyer and seller effects. This is a straightforward, though somewhat tedious, extension of the results in Appendix 1. We examine the HMO effects case in detail because all the contract correlates we use in our analysis are hospital specific, and we wanted to make sure that the absence of HMO characteristics did not bias the analysis of the impacts of these hospital specific variables.

and zero if not. These are the only two outcomes possible. So to satisfy $PC4b$ we need an inequality which is additively separable in $\nu_{m,h}$ regardless of whether $\chi_{m,h} = 0$ or 1.

Let $\Delta\pi_h^H(M_h, M_h/m, M_{-h}, z)$ be the difference between the hospital's profit when the network of the hospital includes HMO m and when it does not. If $\chi_{m,h} = 1$ this contains $\nu_{m,h}$. Let $\Delta\pi_m^M(H_m, H_m \cup h, H_{-m}, z)$ be the difference between the HMO's profit were it to reject hospital h's contract and were it to accept it. If $\chi_{m,h} = 0$ this includes the savings in $\nu_{m,h}$ from rejecting the contract. Note that if $\chi_{m,h} = 1$, $\mathcal{E}[\Delta\pi_h^H(M_h, M_h/m, M_{-h}, z)|\mathcal{J}_h] \geq 0$, while if $\chi_{m,h} = 0$, $\mathcal{E}[\Delta\pi_m^M(H_m, H_m \cup h, H_{-m}, z)|\mathcal{J}_m] \geq 0$. Using analogous notation for $\Delta r(\cdot)$, equation (13) implies

$$\chi_{m,h}\Delta r_h^H(M_h, M_h/m, M_{-h}, z; \theta) \;+\; (1 - \chi_{m,h})\Delta r_m^M(H_m, H_m \cup h, H_{-m}, z; \theta) \;=$$

$$\chi_{m,h}\Big(\mathcal{E}[\Delta\pi_h^H(M_h, M_h/m, \cdot)|\mathcal{J}_h] + \nu_{2,m,h} + \Delta\nu_{1,m,\cdot}\Big) + (1 - \chi_{m,h})\Big(\mathcal{E}[\Delta\pi_m^M(H_m, H_m \cup h, \cdot)|\mathcal{J}_m] + \nu_{2,m,h} + \Delta\nu_{1,h,\cdot}\Big)$$

$$= \;\chi_{m,h}\Big(\mathcal{E}[\Delta\pi_h^H(M_h, M_h/m, \cdot)|\mathcal{J}_h] + \Delta\nu_{1,m,\cdot}\Big) + (1 - \chi_{m,h})\Big(\mathcal{E}[\Delta\pi_m^M(H_m, H_m \cup h, \cdot)|\mathcal{J}_m] + \Delta\nu_{1,h,\cdot}\Big) + \nu_{2,m,h},$$

which is additive in $\nu_{2,m,h}$ *regardless* of whether $\chi_{m,h}$ is one or zero. So there is *no selection* and provided $x \in \mathcal{J}_m \cap \mathcal{J}_h$, $E[\nu_{2,m,h}|x] = 0$, and $h(\cdot) \geq 0$, $PC3$ insures that at $\theta = \theta_0$

$$E\Big(\chi_{m,h}\Delta r_h^H(M_h, M_h/m, M_{-h}, z; \theta) \;+\; (1 - \chi_{m,h})\Delta r_m^M(H_m, H_m \cup h, H_{-m}, z; \theta)\Big)h(x) \geq 0. \tag{14}$$

The model also delivers an inequality that does not depend on the $\nu_{m,h}$ (as in $PC4a$). The sum of the increments in profits to the HMO and the hospital when a contract is established does not contain the transfers between them (and hence $\nu_{m,h}$), does contain information on θ (since if the contract is not established there is a change in transfers to other agents), and must have positive expectation (at least if contract offers are proprietary). So if H_m/h is the observed network of hospital m minus hospital h, our conditions on x also insure

$$E\Big[\chi_{m,h}\Big(\Delta r_m^M(H_m, H_m/h, \cdot; \theta_0) + r_h^H(M_h, M_h/m, \cdot; \theta_0)\Big)h(x)\Big] \geq 0. \tag{15}$$

Estimates. Neither the ν_1-only nor the model which allowed for ν_2 could be rejected by our formal tests. This is not surprising given the sample size[12]. However the results did

---

[12]There were 40 markets containing about 450 plans and 630 hospitals. The market characteristics used as instruments were indicators for; the quartile of the market's population size, high (greater than mean) share

seem to favor the model that allows for ν_2 as only six of the inequalities were negative at its estimated parameter value (the ν_1-only model had eleven), and none of them were individually significant at the 5% level (in contrast to one for the ν_1 only model).

The first four columns of Table 1 present the empirical results. We subtracted our estimate of hospital costs from the revenues in all specifications, so the coefficients appearing in the table are the coefficients of the markup implicit in the per patient payment. Despite the fact that none of the test statistics we computed were significant at the 5% level, there was no value of θ which satisfied all the inequality constraints in any specification, a finding that is not unusual when there are many inequalities (all our specifications had eighty-eight or more of them). The algorithm then generates a point estimate equal to the θ value that minimizes a squared metric in the negative part of the sample moments.

Sample size limited the right hand side variables we could use in the investigation. Still the estimates we do get, though reduced form, are eye-opening. They imply an equilibrium configuration in which; the majority of cost savings from low cost hospitals are captured by the HMOs, and markups increase sharply when a hospital is capacity constrained (CapCon measures whether the hospital would be capacity constrained if all hospitals contracted with all HMOs). Though these are not structural estimates they do lead us to worry about the possibility of significantly lower incentives for hospitals to invest in either cost savings or in capacity expansion than would occur in a price-taking equilibrium. The difference between the ν_1- only estimates and those that allow for ν_2 is that the former imply that almost all the cost savings from low cost hospitals go to the HMOs, while the latter imply that just

---

of population aged 55-64, and hospitals integrated into systems. The plan characteristics were indicators for; whether the plan was local, its quartile of the breast screening distribution, the quality of its mental health services, and an interaction between the last two variables. The hospital cost measure was not used as an instrument because we were worried about measurement error in that variable. The results reported here weighted the market averages of the moment inequalities by the square root of the number of plans in the market, as this produced slightly smaller confidence intervals (interestingly weighting by the variance of the moment inequalities did not improve those intervals). Confidence intervals for each dimension are computing using the techniques in Pakes et. al (2010; a Monte Carlo study of their properties is available from the author). Finally Ho (2009) reports a series of robustness checks on the ν_1-only estimates of a model which is similar to the model presented here. Though specifications which add right hand side variables sometimes increase the confidence intervals quite a bit, the qualitative results in our column (1) are never reversed.

over 50% do and a larger fraction of profits go to capacity constrained hospitals. Low cost hospitals tend to be more capacity constrained so the two variables are negatively correlated.

3.2 Numerical Analysis.

Might we expect contracts with these characteristics to emanate from a contracting equilibrium and should we interpret those coefficients to mean that an increase in the right-hand side variable would, *ceteris paribus*, generate the markup response we estimate? To shed some light on these issues we computed equilibria to a structural contracting model in markets with characteristics similar to those in Ho's data, but with population scaled down to a size where we would expect to have two hospitals and two HMOs in each market (this made it possible to compute equilibria for many markets in a reasonable amount of time).[13]

We compute a full information Nash equilibrium to a game in which hospitals make take it or leave it offers to HMOs. The algorithm assumes that both hospitals chose among a finite set of couples of markups, one for each HMO, and that these markups are offered simultaneously to the HMOs. The offers are public information, as are the HMO premiums that would result from any set of contracts (these are obtained as the Nash equilibrium to a premium setting game among the HMOs). The HMOs then simultaneously accept or reject the offers. At equilibrium each hospital is making the best offers it can given the offers of the other hospital and the responses of the HMOs, and each HMO is doing the best it can do given the actions of its competitor and the offers made by the hospitals[14].

---

[13]We used a discrete choice model of demand and market characteristics determined by random draws from demand and cost characteristic distributions that mimicked those in Ho's data. The closest exercise I know of is in a paper by Gal-Or (1997). By judicious choice of primitives she is able to provide analytic results from a full information Nash bargaining game between two HMO's and two hospitals. She focuses on when her assumptions would generate exclusive dealing and its effects on consumers.

[14]An iterative process with an initial condition in which both hospitals contract with both HMOs chooses among the equilibria when there are multiple equilibria. The choice set included fifty possible markups for each of the two hospitals. The algorithm starts with the lowest ones. It then determines whether HMO1 wants to reject one (or both) of the contracts conditional on HMO2 being contracted to both hospitals. This requires solving for equilibrium premiums and profits for HMO1 given each possible choice it can make and the fact that HMO2 is contracted to both hospitals. HMO2 then computes its optimal responses to

Note that these assumptions *differ* from those used in the empirical analysis. In this full information game the necessary conditions for an equilibrium guarantee an outcome which is renegotiation proof while the necessary conditions for the asymmetric information game we took to data do not. The related questions of; (i) when the different equilibrium notions are appropriate, and (ii) whether the estimation results are sensitive to this choice, are questions that research on buyer-seller networks will have to sort out. Though the contents of contracts are often proprietary, typically who contracts with whom is not. So if we were trying to model a set of relationships which have been stable over some time we might only consider equilibria in which no two agents would find it profitable to recontract given the information on who is contracting with whom. Of course the market we are studying may be constantly changing and negotiations might be costly. Then we might not expect the data to abide by a renegotiation proof criteria, at least not one with costless renegotiation. Since all we need for estimation is a way of obtaining a *lower bound* to the expected profits from a counterfactual choice, we could, at least in principal, obtain our inequalities from the difference between the actual profits and the minimum of the profits from a group of counterfactuals chosen to reflect different possible game forms (though the larger the group, the less tight our bounds and the larger the computational burden).

Numerical Results. Column (5) through (8) of Table 1 present OLS estimates from regressing the computed markups onto variables of interest. The first two columns show that the three variables that the empirical study focused on have the appropriate signs, are significant, and account for a large fraction, about 70%, of the variation in markups (or about

---

HMO1's decisions in the same way. This process is repeated until we find a Nash equilibrium for the HMOs' responses. No matter the offers, we always found an equilibrium to this subgame. We then optimize over the first hospital's (say H1) offers, holding H2's offers fixed. For each offer we repeat the process above until we find a Nash equilibrium for the HMOs' responses. This gives us H1's optimal offers given the initial offers by H2. Next H1's offers are held fixed and H2 optimize against that. We repeat this process until we find a Nash equilibrium in offers. For 3% of the random draws of characteristics we could not find an equilibria, and those markets were dropped from the analysis. Note that when a hospital contracts with an HMO in equilibrium it does not necessarily contract at the lowest offer that is consistent with the HMO accepting. Different offers change the HMO costs per patient. This changes the outcome of the premium setting game that the HMOs engage in and feeds back into hospital profits.

85% of the variance in transfers). Columns 7 and 8 add variables. The original three variables maintain their signs and remain significant but have noticeably different magnitudes; i.e. though the empirical results do pick up important correlates of the equilibrium payments, the reduced form parameter estimates should not be thought of as causal responses.

This coefficients of the additional variables in column (7) are instructive. They imply that when the average hospital cost in the market goes up by 1% the markups of the hospitals in the market go down by .23%, but if the *difference* between a hospital's cost and the average hospital cost goes up by 1%, the hospitals markup goes down by .56%. So a hospital's markup over costs depends on the costs of the other hospitals it is competing with. Hospitals earn higher markups in "tighter" markets (markets with lower ratios of population to the number of hospital beds) and once we account for this the effect of capacity constraints is greatly reduced (though not eliminated). HMOs seem to get a small quantity discount (the markups they pay are lower when they send more patients to the hospital), and hospitals earn higher markups when the HMOs they are dealing with charge their members higher markups.

Finally note that 20% of the variance in markups, or 8% of the variance in transfers, is not accounted for by our observables. Given the full information assumptions, this is ν_2 variance. Even in a world where our equilibrium and functional form assumptions are correct, measurement error in hospital costs would cause ν_1 error. So in this, and we suspect in most, empirical examples both types of errors are likely to be present.

4 Specification Errors and Alternative Estimators.

The generalized discrete choice model ignores ν_1 errors and requires an *a priori* specification of the ν_2 distribution; both assumptions which, if incorrect, can generate an inconsistency in its estimators. The profit inequality model which pays inadequate attention to possible sources of ν_2 error will generate selection biases. This section asks what the impacts of these specification errors are likely to be in the context of our buyer-seller network example. It presents Monte Carlo results from using each of the two model's estimators *both*; (i) when that models' assumptions are the assumptions generating the data, *and* (ii) when they are not. Where possible we will also present results from Ho's data.

Details of the Monte Carlo Analysis. The Monte Carlo results are based on a population of 100,000 markets whose equilibria were computed using the algorithm described in the last section. We estimate one parameter; the average per patient markup. To obtain the true value of that parameter we took the transfers implicit in the equilibrium offers and projected them onto the number of patients and the variables we used as instruments.[15] The function obtained from this projection is treated as the parametric transfer function. The coefficients of the instruments are treated as known and the coefficient of the patient variable is the coefficient to be estimated. The residual from this projection is the ν_2 error. This insures that ν_2 has zero covariance with our instruments *before* we condition on the outcome. When all we require is a ν_1 error we treat these ν_2 as known and add pseudo random draws on a normal measurement error to hospital costs and/or population size.

The Monte Carlo results are based on four hundred data sets each obtained as independent draws from our "population" of markets. The sample size was set so that the number of contracts in each sample' matched the number of contracts in Ho's (2009) data set. Since each of the computed equilibria had only two hospitals and two HMO's this gave us a larger number of markets (1,385 markets per sample), but *many fewer* contracts per market, than in Ho (2009). The ν_1 and ν_2 draws are taken independently across samples.[16]

The inequalities used to estimate the profit inequality model are the same as those used in the empirical work; each HMO reverses its equilibrium decision with each hospital, and each hospital replaces its equilibrium contract offer to each HMO with a null contract. However since the Monte Carlo data is generated from a full information Nash equilibrium, when the hospital offers a null contract to an HMO that hospital considers the profits that would accrue to it were both HMOs to reoptimize[17]. For the generalized discrete choice approach

---

[15]For accepted offers these were the actual transfers, for the offers that were rejected these are the transfers that would have resulted if the last offer had been accepted.

[16]Actually we did the analysis in two ways. In the second we drew a Monte Carlo data set, took two hundred draws on vectors of ν_1 errors for that data set, tabulated the results for each data set, and then averaged over data sets. This provides confidence intervals that condition on the observables, while the results reported in the text do not. The results from the two procedures were virtually identical.

[17]To obtain the $\pi(\cdot)$ resulting from the null contract offer, let $o_{m,h}$ be the contract offered by hospital h to HMO m in equilibrium, and ϕ be the null contract. If $h = 1$ contracted with $m = 1$, its profits from offering ϕ are obtained from the HMO equilibrium responses to the tuple $(\phi, o_{1,2}, o_{2,1}, o_{2,2})$.

we used the inequalities generated by the necessary conditions for equilibrium.[18]

Results. Table 2, which presents the results, is split into panels. Panel A provides estimates obtained using the ν_1-only inequalities, Panel B from using the ν_2-only inequalities (the inequalities from the generalized discrete choice model), and Panel C uses the inequalities that allow for both ν_1 and ν_2 disturbances. The "true" value of θ_0 from the simulated data was 16.76. The "identified set", that is the θ interval that satisfies the population moment conditions, differs across panels.

Since the instruments are orthogonal to the disturbance by construction, the identified set for Panel A is the θ interval which generates positive population profit inequalities when we set all disturbances to zero; $[13.47, 18.59]$. The true identified set for the generalized discrete choice model depends on the true joint distribution of the ν_2's conditional on the market's instruments. This is not known and is too complex to estimate non-parameterically; a problem which is likely to recur in empirical work. To get a sense of the identified set generated by this approach we set all the disturbances to zero and for each possible network structures found the set of θ which lead to positive values for; (i) the averages of the differences between the indicator functions for satisfying the Nash conditions and the "observed" equilibrium outcome, and (ii) did the same after interacting the difference in indicator functions for each network structure with the variables used as instruments. The interval for (i) was $[16.7, 16.95]$, while for (ii) it was $[16.7, 16.85]$; both rather amazingly short.

The first two rows of Panel A provide results from the ν_1-only profit inequality model when there are only ν_1 errors (so its estimators are consistent). Row 1 adds measurement error in costs equal to 25% of the true cost variance. The average of the estimated lower bounds is 8% lower than the true lower bound ($\underline{\theta}_0$), while that of the upper bound is 2.5% higher than $\overline{\theta}_0$. Moreover the bounds are precisely estimated, less than 2.5% of the lower

---

[18]Note that the estimating equations used do not exhaust the information in the data in either approach. At the cost of increasing the computational burden we could have; (i) used the inequalities obtained from simultaneously switching each HMO's behavior with respect to both hospitals in the profit inequality approach (and if more details on the contracts were available yet other inequalities would become available), and (ii) for the generalized discrete choice approach we could have computed the probability that the observed equilibrium was unique.

(upper) bounds were more than 10% different from their true values. When we add an expectational error to the population, and hence to the patient flows from the HMOs to the hospitals, the estimated interval gets substantially larger. This is unfair to the model since, though there may be uncertainty in the relevant population size and patient flows variables when contracts are signed, we should be able to construct good instruments for them from current population size and flows, and we did not do that here. We keep this case because it allows us to examine the impact of specification errors in one setting where the bounds define a short interval and one where they do not.

Rows 3 and 4 use a simulated data set that contains both ν_1 and ν_2 errors but the inequalities from the ν_1-only model. The ratio of the variance in ν_2 to the variance in the dependent variable is 12.7%. Now the estimated bounds are inconsistent; the lower bound will, in the limit, be too large, while the upper bound will be too low. This makes the bounds move towards θ_0, but they may overshoot, leaving us with an estimator which does not cover the true θ_0. Adding ν_2 also adds variance to the estimators, so in any finite sample the estimated bounds may be smaller or larger with ν_2 errors than without them. In the case with only measurement error in costs, the case in which the interval was tightly estimated, adding specification error in the form of the ν_2 has little effect on any of the estimates. When there is also measurement error in population and the estimated intervals are larger, the effect of the specification error is to lessen the loosely estimated upper bound, but only by 5%. At least in this example estimates from the ν_1-only inequalities do not change much when we allowed for ν_2 error. Apparently when we add ν_2 variance its biasing effects on the estimates are largely offset by the effect of increased data variance on those estimates.

The last row of panel A provides the estimates when we use Ho's (2009) data with this specification. This generates a point estimate about a third lower than the lower bound estimate from the simulated data and a confidence interval of length between that of the model with errors in the population and that without those errors.

Panel B provides the results when we use the ν_2-only inequalities. To use the ν_2-only algorithm we need an assumption on the joint distribution of the ν_2. We tried two assumptions; (i) random draws from the empirical distribution of the actual ν_2, and (ii) a normal distribution. The first option would not be available to empirical researchers but might be

closer to the true population distribution (it would be if the ν_2 were truly independent, not just mean independent, of the instruments and had no within market correlation).

Regardless of whether we use just the constant term (rows 6 & 7) or all of our instruments (rows 8 & 9) and regardless of the choice of ν_2-distribution, the ν_2-only model generates point estimates whose values are larger than the true θ_0. The distribution of estimates had little variance, so the interval formed from 95% of the point estimates does not cover the true θ_0 either. Apparently the lack of information on the ν_2-distribution leads to an inconsistency. Though this is disturbing the asymptotic bias is not large; the lower bound of the (normal) confidence interval is about 2.6% larger, and the point estimate 6.9% larger, than θ_0.

Just as we added ν_2 variance to the algorithm which uses the ν_1-only inequalities, rows 10 and 11 add ν_1 variance to the algorithm which uses the ν_2-only inequalities. The estimates presented in these rows use the normal distribution of the ν_2 (an empirical researcher would not have access to the bootstrap distribution). Adding ν_1 errors does tend to increase the parameter estimates further, but by a surprisingly small amount. We could not use the ν_2-only algorithms on Ho's actual data. To do so we would have had to compute about 100,000 premium setting equilibria and their implied profits for each ν_2 draw and each θ evaluated in estimation; a task that will be beyond our computational abilities for some time.

Panel C provides the estimates obtained when we used the inequalities that allow for both ν_1 and ν_2 disturbances; the "robust" inequalities in equations (14) and (15). The fact that there are only two agents on each side of the simulated markets implies that the robust inequalities do not deliver an upper bound. The lower bound is lower than the bound obtained when we used the ν_1 only inequalities on this data, but not by much. When we move to Ho's (2009) data and use the robust inequalities we get an estimate which is larger than the estimate which allows for only ν_1 errors but a confidence interval of similar length, and both confidence intervals cover both estimates. Interestingly once we allow for ν_2 errors the estimates from Ho's data is closer to, and the confidence interval covers, the value of the parameter obtained from the numerical analysis.

The results from the Monte Carlo are quite encouraging. It seems that the most salient problem is the requirement of an assumption on the distribution of ν_2 in the generalized discrete choice model. In multistage games that estimator also carries a large computational

burden. However the worry that a moderate amount of ν_1 variance in the generalized discrete choice model, or a moderate amount of ν_2 variance in the profit inequality model, would severely bias the estimates is, at least in this example, not warranted. The addition of the unaccounted for error adds variance, as well as bias, to the estimates. This variance tends to move the bounds in the opposite direction as does the bias and in our example the net effect was small. The estimator which uses the robust inequalities is least subject to bias but does generate larger identified sets.

5   Summary.

This paper formulates two set of assumptions that enable one to bring behavioral models, both their structural and their reduced forms, to data and applies them to two empirical problems. Our first example illustrates that the assumptions underlying traditional discrete choice estimators are not always the most sensible choice for discrete choice problems. This motivates an enumeration of assumptions that justify alternative estimators in both multiple, and single, agent settings. An empirical example illustrated how the multiple agent estimator can be used to analyze a problem which is central to the determinants of prices and investment incentives in vertical markets; the correlates of the profit split between buyers and sellers in those markets. Though the results were reduced form and had to make do with both limited data and the auxiliary assumptions required to obtain counterfactual profits, they were broadly consistent with results obtained from a numerical analysis of equilibrium contracts in markets which were similar to those used in the empirical analysis. A Monte Carlo analysis indicated that the estimators from both models were surprisingly robust to all likely sources of problems but one; the need to assume a distribution for the generalized discrete choice model. It seems that moment inequalities open up possibilities for empirically analyzing market interactions in relatively unexplored, yet important, settings.

References.

Andrews, D., Berry, S., and P. Jia, 2006; "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location," *mimeo* Yale University.

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

Beckert W., Griffith, R., and Nesheim, L., 2009; "Disaggregate Demand Elasticities at the Extensive and Intensive Margins: An Empirical Analysis of Store and Shopping Basket Choices", *mimeo*, Institute for Fiscal Studies, London.

Beresteanu, A., and F. Molinari, 2008; "Asymptotic Properties for a Class of Partially Identified Models", *Econometrica*,Vol.76, pp763-814.

Chernozhukov, V., H. Hong and E Tamer, 2007;" Estimation and Confidence Regions for Parameter Sets in Econometric Models" *Econometrica*, Vol. 75 (5), pp. 1243-1284.

Cliberto, F. and E. Tamer, 2006, "Market Structure and Multiple Equilibria in the Airline Markets," *mimeo*, Northwestern University.

Dekel, E., Fudenberg, D. and D. Levine, 1993, "Payoff Information and Self-Confirming Equilibrium", *Journal of Economic Theory*, 89, pp. 165-85.

Dubin, J. and D. McFadden, 1984, "An Econometric Analysis of Residential Electric Appliance Holding and Consumption", *Econometrica*, 62, pp.345-62.

Gal-Or, Esther, 1997, "Exclusionary Equilibria in Health-Care Markets" *Journal of Economics and Management Strategy*", Vol. 6, No. 1, pp.5-43.

Haile, P. and E. Tamer, 2003, "Inference with an Incomplete Model of English Auctions", *Journal of Political Economy*, Vol.111, No. 1, pp 1-5.

Hansen, Lars Peter, and Kenneth J. Singleton, 1982, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models", *Econometrica*, vol. 50, no. 5, pp. 1269-86.

Hart O., and J. Tirole, 1990, "Vertical Integration and Market Foreclosure" *Brookings Papers on Economic Activity; Microeconomics*, pp.205-76.

Hausman, J., J. Abrevaya, and F. Scott-Morton, " "Misclassification of the Dependent Variable In a Discrete Choice Setting", *Journal of Econometrics*, vol. 87(2) pp 239-269.

Ho, K.,2009, "Insurer-Provider Networks in the Medical Care Market " *American Economic Review*, 99(1) pp 393-430.

Katz, M., 2007, "Supermarkets and Zoning Laws", *Unpublished Ph.D. dissertation* Harvard University.

Keane, M., and K. Wolpin, 2009, " Empirical Applications of Discrete Choice Dynamic Programming Models", *Review of Economic Dynamics*, Vol. 12, pp 1-22.

Luttmer, E. , 1996, "Asset Pricing in Economies with Frictions", *Econometrica* Vol.64 (6), pp 1439-1467.

Manski, C. 2004, "Measuring Expectations", *Econometrica* vol.72 No.5, pp 1329-76.

McAfee, R.. and M. Schwartz, 1992, "Opportunism in Multilateral Vertical Contracting: Nondiscrimination, Exclusivity, and Uniformity" *American Economic Review*, Vol.84 (1) pp.210-230.

Pakes, A., J. Porter, K. Ho, and J. Ishii, 2006, "Moment Inequalities and Their Application," mimeo, Harvard University.

Savage L., (1954); "The Foundations of Statistics", John Wiley and Sons.

Tamer, E.,2003, "Incomplete Simultaneous Discrete Response with Multiple Equilibria", *Review of Economic Studies*, pp.147-65.

Varian, H.,1982, " The Nonparametric Approach to Demand Analysis", *Econometrica*, Vol.50 (4) pp. 945-65.

Appendix: Inequalities for Buyer Seller Network With Fixed Effects.

We use the notation introduced for the hospital HMO problem in subsection 4.1.1, and consider the case in which the $\{\nu_{2,m,h}\}$ are HMO fixed effects; i.e. that $\forall (h,m)$, $\nu_{2,\mathfrak{m},h} = \nu_{2,\mathfrak{m}}$. These restrictions generate two sets of inequalities.

The first is a *difference in difference* inequality. If an HMO accepts at least one hospital's contract and rejects the contract of another, then the sum of the increment in profits from accepting the contract accepted and rejecting the contract rejected; (i) differences out the HMO effect and (ii) has a positive expectation. More formally for every $\tilde{h} \notin H_\mathfrak{m}$ and $h \in H_\mathfrak{m}$ we have

$$\Delta \pi_\mathfrak{m}^M(H_\mathfrak{m}, H_\mathfrak{m} \cup \tilde{h}, \cdot) + \Delta \pi_\mathfrak{m}^M(H_\mathfrak{m}, H_\mathfrak{m} \backslash h, \cdot) = \Delta r_\mathfrak{m}^M(H_\mathfrak{m}, H_\mathfrak{m} \cup \tilde{h}, \cdot) + \Delta r_\mathfrak{m}^M(H_\mathfrak{m}, H_\mathfrak{m} \backslash h, \cdot),$$

which implies that provided $x \in \mathcal{J}_m \cap \mathcal{J}_h$ and $h(\cdot)$ is a positive valued function

$$E\Big[\Delta r_\mathfrak{m}^M(H_\mathfrak{m}, H_\mathfrak{m} \cup \tilde{h}, \cdot; \theta_0) + \Delta r_\mathfrak{m}^M(H_\mathfrak{m}, H_\mathfrak{m} \backslash h, \cdot; \theta_0)\Big] h(x) \geq 0.$$

For the second inequality note that if $\nu_{2,m,h} = \nu_{2,m}$ we can use the logic leading to equation (14) in the text to show that for any positive valued function, $h(\cdot)$

$$0 \leq E\left[\frac{1}{\#H} \sum_h \Big(\chi_{m,h} \Delta \pi_h^H(M_h, M_h/h, \cdot) + (1 - \chi_{m,h}) \Delta \pi_m^M(H_m, H_m \cup h, \cdot)\Big)\right] h(x) =$$

$$E\left[\frac{1}{\#H} \sum_h \Big(\chi_{m,h} \Delta r_h^H(M_h, M_h/h, \cdot; \theta) + (1 - \chi_{m,h}) \Delta r_m^M(H_m, H_m \cup h, \cdot; \theta)\Big) + \nu_{2,m}\right] h(x)$$

$$\equiv E\Big[\overline{S}^r(\mathfrak{m}, \cdot; \theta) + \nu_{2,m}\Big] h(x).$$

This implies that $E\overline{S}^r(\mathfrak{m}, \cdot; \theta_0) h(x) \geq -E\nu_{2,m} h(x)$, and consequently that for any $x \in \mathcal{J}_m \cap \mathcal{J}_h$

$$E\Big[\Delta r_\mathfrak{m}^M(H_\mathfrak{m}, H_\mathfrak{m} \backslash h, \cdot; \theta_0) + \overline{S}^r(\mathfrak{m}, \cdot; \theta_0)\Big] h(x) \geq 0.$$

Data	Real Data				Simulated Data			
Estimator	Inequality Estimators				OLS Regression			
	ν_1 only		ν_1 & ν_2		Actual Markups			
column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	θ	95% CI	θ	95% CI	θ	s.e.	θ	s.e.
Variable		UB/LB		UB/LB				
Per Patient Markup (Units = $ thousand/patient)								
Const.	9.5	15.4/4.8	8.2	15.2/3.3	8.9	.09	3.7	.24
CapCon.	3.5	8.6/1.4	13.5	16.1/2.3	1.2	.10	.48	.11
Cost/Adm.	-.95	-1.5/-.57	-.58	-.2/-1.1	-.39	.01	–	–
Av.Cost	–	–	–	–	–	–	-.23	.01
Cost-AC	–	–	–	–	–	–	-.56	.01
Pop/bed	–	–	–	–	–	–	.11	.01
# patient	–	–	–	–	–	–	-.09	.01
HMOmarg	–	–	–	–	–	–	1.4	.10
R^2	–	–	–	–	.71		.80	

Notes: Real Data. There are 40 markets. CapCon measures whether the hospital would be capacity constrained if all hospitals contracted with all HMOs, Cost/Adm = hospital cost per admission. Costs and admissions $\notin$ IV.

Simulated Data. These are least squares regressions coefficients from projecting computed markups onto the included variables. See below for the calculation of equilibrium markups. There are 1385 markets with 2 HMOs and 2 Hospitals in each. This generates approximately the same number of buyer-seller pairings as in the data set used in the empirical analysis. Additional variables are defined as follows; "Cost-AC" is the cost per admission of the hospital minus the average of that over the hospitals in the market, Pop/bed is population over total number of hospital beds in the market, # patients is number of patients the HMO sends to the hospital, and HMO margin is the HMO's average premium minus its average cost.

Table 2: Inequality Estimators: Simulated and Actual Data.

Disturbances	Not In IV	Average LB	Average UB	95% of θ in LB	95% of θ in UB
A: Using ν_1 inequalities.					
Simulated data: θ_0=16.76, $[\underline{\theta}_0,\bar{\theta}_0]$=[13.47,18.59].					
Only ν_1 disturbances					
1. 25% Cost	Cost	12.39	18.72	12.12	19.05
2. 25% Cost,5% pop	Cost, $N_{j,k}$,Pop	11.43	37.34	11.30	45.88
ν_1 & ν_2 disturbances.					
3. ν_2, costs	Cost	12.25	18.42	12.01	18.86
4. ν_2, costs, pop	Cost, $N_{j,k}$,Pop	11.69	35.91	11.55	43.97
Ho's (2009) data with ν_1 inequalities.					
5. actual disturbances	Cost	8.2	8.2	2.3	16.4
B: Using ν_2 inequalities.					
Simulated data, Only Constant as IV: θ_0=16.76, $"[\underline{\theta}_0,\bar{\theta}_0]"$=[16.7,16.95].					
6. ν_2 (bootstrap dist)		17.75	17.75	17.25	18.1
7. ν_2 (normal dist)		17.92	17.92	17.2	18.45
Simulated data, All IV: θ_0=16.76, $"[\underline{\theta}_0,\bar{\theta}_0]"$=[16.7,16.85].					
8. ν_2 (bootstrap dist)		17.84	17.84	17.40	18.25
9. ν_2 (normal dist)		18.02	18.02	17.65	18.5
Simulated data, ν_1 (in costs) & ν_2 disturbances.					
10. $\nu_2 \sim \mathcal{N}$, IV=Only Constant		18.02	18.02	17.35	18.5
11. $\nu_2 \sim \mathcal{N}$, All IV	Costs	18.11	18.11	17.64	18.65
Ho's (2009) data with ν_2 inequalities.					
12. Assume ν_2 normal	Could Not Compute.				
C: Using Robust inequalities.					
Simulated data.					
13. ν_2, costs	Cost	11.86	n.b.	11.72	n.b.
14. ν_2, costs, pop	Cost,$N_{j,k}$,Pop	11.69	n.b.	11.55	n.b.
Ho's (2009) data.					
15. Actual Disturbances	Cost	11.7	11.7	3.6	17.9

Notes. Instruments for panels A & C (unless omitted); constant, $N_{j,k}$, hospital cost and capacity measures, market cost capacity and population measures, HMO characteristics, and interactions among above. Instruments for ν_2 inequalities are market averages of above variables. The model for line 15 allowed also for a cost coefficient; without it the average markup was negative.