

Alternative Models for Moment Inequalities.*

A. Pakes,

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

September 28, 2009.

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 their underlying assumptions.

*This is a revised version of part of my Fisher-Schultz Lecture presented at the World Congress of the Econometric Society in London, August 2005. The paper draws extensively from past interactions with my students and coauthors, and I would like to take this opportunity to express both my intellectual debt and my thanks to them. I like to think they enjoyed the experience as much as I did, though that might have been harder for the students in the group. For help on this paper I owe a particular debt to Robin Lee. I also thank four referees and the editor, Daron Acemoglu, for their helpful comments.

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 that have proven difficult to analyze empirically and have economic importance. Throughout we distinguish between the assumptions needed to estimate the “structural” parameters defined by the underlying 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 role of many of the needed assumptions. The difficulty in analyzing it 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 simplifications available in single agent problems noted). The first approach, the generalized discrete choice approach, 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); a paper which provides 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 conditions

which underlie the two approaches differ and a discussion of when each is likely to be relevant is provided.

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 (who typically then re-market the goods they buy to consumers). At least two problems make empirical research on these markets difficult. First the payments from the buyer to the seller are often proprietary (we do not have access to the “contracts” which determine them), and second there is no generally agreed upon game form to structure the analysis. The importance of these markets stems from the fact that they characterize upstream interactions in many vertical markets. As a result the way payments from the buyer to the seller vary with market structure in these markets determines the relationship between market structure and both consumer prices and investment incentives in these markets.

The section begins by extending the empirical work of Ho(2009) which characterizes the relationship between payments and market characteristics in HMO-hospital networks. We start by showing 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 designed, to the extent feasible, to mimic those used in the empirical analysis. The numerical results from these computations allow us to both investigate the consistency of the empirical results with those obtained from the equilibria to a contracting model, and to engage in a more general examination of the correlates of the markups generated by that model.

Section 4 uses the numerical results from the computed equilibria in a Monte Carlo analysis of the robustness of the two approaches to possible violations of the assumptions underlying them. In doing so it also considers of the computational and the small sample properties of the two approaches. We find that the generalized discrete choice estimator provides tighter confidence intervals than the profit inequality estimator, but the discrete choice estimator has a larger computational burden and is biased as a result of a need to make an assumption on an unknown distribution (though the bias is only 5-10% of the coefficient

value in this example)¹. The biases caused by failure of the other assumptions underlying either of the two estimation approaches seems to be negligible.

Both the numerical results and the empirical results for our examples are encouraging. They indicate that moment inequality estimators can be used to shed empirical light on difficult to analyze, yet economically important, problems.

1 A Single Agent 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 choice models because of the size and complexity of the choice set facing consumers (all possible bundles of goods at all “nearby” stores).

Though large choice sets generate computational problems for discrete choice models, they *facilitate* moment inequality estimators. This because a large choice set implies many possible counterfactual comparisons. The empirical researcher can chose to use those most likely to isolate the effect of interest; in our case 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

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

¹Since 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 econometric research that I do not discuss here; see, in particular, Chiaroscuro, Hong and Tamer (2006), Andrews and Guggenberger (2009), and the papers cited above

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. The empiricists' problem is to chose alternatives that are likely to illuminate the effects of interest. For a particular d_i chose the alternative, say $d'(d_i)$ to be the purchase of

- the *same basket* of goods,
- at a store which is *further away* from the consumer's home than the store the consumer shopped at.

Given the additive separability assumption this choice differences out the impact of the basket of goods chosen on utility; 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 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 we assume that the bundle bought at the chosen store could have been bought at the alternate store, and let $\Delta f(d, d', \cdot) \equiv f(d, \cdot) - f(d', \cdot)$ for any function $f(\cdot)$ then

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

We 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 provided

$$N^{-1} \sum_i \mathcal{E} \left[\Delta \mathbf{e}(\mathbf{b}_i, s_i, s'_i) | \mathcal{J}_i \right] - N^{-1} \sum_i \Delta e(b_i, s_i, s'_i) \rightarrow_P 0, \quad \text{and} \quad (2)$$

$$N^{-1} \sum_i \mathcal{E} \left[\Delta dt(s_i, s'_i, z_i) | \mathcal{J}_i \right] - N^{-1} \sum_i \Delta dt(s_i, s'_i, z_i) \rightarrow_P 0,$$

the fact that $\Delta dt(s_i, s'_i, z_i) < 0$ implies

$$-\frac{\sum_i \Delta e(b_i, s_i, s'_i)}{\sum_i \Delta dt(s_i, s'_i, z_i)} \rightarrow_p \underline{\theta} \leq \theta_0.$$

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

$$-\frac{\sum_i \Delta e(b_i, s_i, s''_i)}{\sum_i \Delta dt(s_i, s''_i, z_i)} \rightarrow_p \bar{\theta} \geq \theta_0,$$

giving 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 $\sum \nu_{2,i} = 0$, so θ_0 is the mean cost of drive time. Then our inequality becomes

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

As a result, if

$$\sum_i \Delta dt(s_i, s'_i, z_i)^{-1} \left[\mathcal{E}[\Delta e(\mathbf{b}_i, s_i, s'_i) | \mathcal{J}_i] - \Delta e(b_i, s_i, s'_i) \right] \rightarrow_P 0, \quad (3)$$

then

$$\frac{1}{N} \sum_i \left(\frac{\Delta e(b_i, s_i, s'_i)}{\Delta dt(s_i, s'_i, z_i)} \right) \rightarrow_P \underline{\theta} \leq \theta_0,$$

and an analogous upper bound to θ_0 from the inequality generated by choosing an alternative where drive time is less than that to the chosen store.

If the agent knows drive times when deciding where to shop then the condition in equation (3) follows from the first line in (2). Equation (3) notes that differences between measured and expected drive times will not interfere with the result provided the difference between actual and measured expenditures is mean independent of the actual drive time (these are obtained from a GIS system).

Discussion. Case 1 uses *a ratio of averages* to bound the parameter of interest while case 2 uses *the average of a ratio*. Some notes on the assumptions underlying these two cases will set the stage for the rest of the paper.

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 $\nu_{2,i}$). Moreover if the unobserved determinant of drive time costs ($\nu_{2,i}$) is correlated with drive time (dt) then Case 1 and Case 2 estimators should be different, if not 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 $\nu_{2,i}$.

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. All we require for equation (2) [or for 3] is that the average of the expectational errors across agents is zero [or a weighted average is zero], and as we shall see below this condition can be weakened further.

Conditions on the Choice Set. All we required of the choice set was one feasible alternative. In particular *we did not need to specify an “outside” choice or chose the form of the choice set* for the inside choices (continuous, discrete, or some mixture). Moreover we only needed to compute returns for two (of the many possible) choices.

Finally note that our concern with 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 Homespun 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. A behavioral discrete choice model of supermarket choice would, for each store; (i) specify the agent's prior distribution of prices, (ii) compute the bundle of goods the agent would buy for each possible realization of the price vector and (iii) form the expected utility of going to the store. It would then base estimation on the distribution of the maxima of the resulting expected utilities. Computing expected utilities for problems with discreteness in the choice set is a demanding task. Moreover single agent discrete choice problems have an appealing reduced form which is much less demanding than the structural model. Regress expected utility on the variables of interest. Provided we are willing to make an assumption on the joint distribution of the disturbances from this regression we can use standard techniques to recover those regression coefficients. We come back to the conditions under which these regression coefficients have a behavioral interpretation below. We note now, however, that such a reduced form *does not exist* for multiple agent problems, at least not without distributional assumptions which are generally considered unacceptable; an alternative estimator is required for that case.

A useable reduced form for the supermarket choice problem must also reduce 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 a choice set consisting of ten expenditure bundles at each of the outlets within a given radius of its home. The utility function for a given expenditure bundle is allowed to differ with the amount of shopping trips per week. 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). The utility function is additive in the utility from the bundle purchased, the disutility from expenditures (which is allowed to vary with

household income), the utility from other store characteristics, the disutility from drive time, and “logit” disturbances (for a discussion of alternative ways of building reduced forms for supermarket choice problems and an application; see Beckert et. al.,2009).

There are a number of reasons to doubt the estimates from this model. I will 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 likely 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 this variable, and given that the aversion to drive time is likely to be related to drive time per se, traditional random coefficient assumptions would be suspect.

Results. The specifications used for the models estimated were quite detailed. Those used for the comparison model estimated about forty different parameters for each of three different number of visits per week, while those used for the revealed choice model estimated about fifteen parameters. Both models were estimated with specifications that included outlet characteristics and interactions between expenditure and demographics. Here I suffice with a brief summary of the estimates obtained for the median aversion to drive time (this was also a focus of the original paper).

The multinomial comparison models were estimated using maximum likelihood. The estimates implied that the median aversion to drive time was two hundred and forty dollars per hour. The 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 chain dummies were not ordered intuitively, nearness to the subway stop had a negative coefficient and so on).

The inequalities used in the revealed choice models were obtained from differences be-

tween 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.²

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 (particularly the chain dummies) took on values which accorded with intuition.

2 Conditions For Moment Inequality Estimators.

This section provides two sets of conditions that can be used to justify moment inequality estimators in a more general setting; one that allows for multiple interacting agents (it

²The 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.

considers the single agent problem as a special case) and is explicit about the sources of the disturbances in the specification. For each of the two approaches we consider estimation of both the parameters of the underlying behavioral model and of a particularly easy to interpret “reduced form”. 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. Then the formal requirement is

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

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 Nash equilibrium. As a result, were we to assume equilibrium behavior 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, and though further conditions on beliefs will

be required in order to bring either framework to data, the restrictions required by the two frameworks differ.

Both approaches need 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 a “structural” model is required. To predict what expected profits would be from a choice which was not observed we need to specify what the agent thinks that \mathbf{d}_{-i} and \mathbf{y}_i would be were it to change its own decision. To see this consider the empirical example of the next section which analyzes contracts between a buyer and a seller in a vertical market; a market where a good is sold to a buyer who then re-markets that good to consumers. One component of \mathbf{y}_i in this example is the price the buyer charges to consumers when it resells the seller’s products. That price will typically depend on which sellers contract with which buyers. So when a buyer considers what would have happened if it had rejected a an offer by a seller (an offer it had in fact accepted), the buyer knows that if it had rejected the seller’s offer the equilibrium price at which it would resell the products it does sell to consumers would change. As a result we will need a model for the buyer’s perception of what the price component of \mathbf{y}_i would have been had it rejected the offer. In sequential problems the model must also specify the agent’s beliefs on the impact of a change in its choice on the subsequent choices of its competitors.

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 : $\mathbf{d}_{-i} = d^{-i}(\mathbf{d}_i, \mathbf{z}_i, \theta)$, and $\mathbf{y}_i = y(\mathbf{z}_i, \mathbf{d}_i, \mathbf{d}_{-i}, \theta)$, 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 d_i and \mathbf{z}_i . 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 an exogenous random variable.

The stringency of C2 is likely to vary significantly with the type of 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 often $\mathbf{y}_i = \mathbf{z}_i$. Then C2 is unobjectionable. In multiple

agent *simultaneous move* games $d^{-i}(d', \mathbf{z}_i, \theta) = \mathbf{d}_{-i}$ so there is no need for an explicit model of reactions by competitors, but, as in two stage games, \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 what those agents think are likely to be the responses of the agents who move later.³

Implications of C1 and C2. The implication of C1 and C2 that we take to data is obtained by letting $\Delta\pi(d_i, d', d_{-i}, z_i, \theta_0) \equiv \pi(d_i, d_{-i}, y_i, \theta_0) - \pi(d', d_{-i}(d', z_i), y(z_i, d', d_{-i}), \theta_0)$, where d' is any alternative choice in D_i . Then C1 and C2 together insure

$$\mathcal{E}[\Delta\pi(d_i, d', \mathbf{d}_{-i}, \mathbf{z}_i, \theta_0) | \mathcal{J}_i] \geq 0, \quad \forall d' \in D_i. \quad (4)$$

Equation (4) is the moment inequality delivered by the theory. To move from it to a moment inequality we can use for 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 our two approaches. To proceed we need a measurement framework which nests both their assumptions.

2.2 Measurement Model.

Let $r(d, d_{-i}, z_i^m, \theta)$ be the profit function specified by the econometrician up to an additively separable disturbance (so z_i^m includes observables and any unobservables that do not enter additively), and define $\nu(\cdot)$ to be the difference between the profit function the agent responds to and this specification, so that

$$r(d, d_{-i}, z_i^m, \theta) \equiv \pi(d, d_{-i}, z_i, \theta) + \nu(d, d_{-i}, z_i^m, z_i, \theta). \quad (5)$$

³The fact that we allow for sequential games explains the difference between our C2 and Assumption 2 in Pakes et. al., 2006. The buyer-seller network example in the next section requires this added generality.

Since the agent's decision is based on $\mathcal{E}[\pi(\cdot)|\mathcal{J}_i]$, and $\mathcal{E}[r(\cdot)|\mathcal{J}_i] = \mathcal{E}[\pi(\cdot)|\mathcal{J}_i] + \mathcal{E}[\nu(\cdot)|\mathcal{J}_i]$ we can rewrite the equation for $r(\cdot, \theta_0)$ in a way which is explicit about the sources of difference between our specification for profits ($r(\cdot)$) and the profits the agent bases its decision on ($\mathcal{E}\pi(\cdot)$). In particular

$$r(d, d_{-i}, z_i^m, \theta) \equiv \mathcal{E}[\pi(d, \mathbf{d}_{-i}, \mathbf{z}_i, \theta)|\mathcal{J}_i] + \nu_{2,i,d} + \nu_{1,i,d}, \quad (6)$$

where

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

and

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

Equation (6) shows that the difference between our specification for profits and the $\mathcal{E}[\pi(\cdot)|\mathcal{J}_i]$ that the agent bases its decision on has 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 a result of uncertainty in the exogenous variables (the \mathbf{z}) whose realizations are determinants of returns (in the supermarket example it is the uncertainty in the prices and availability of goods). In multiple agent problems their may also be uncertainty in \mathbf{d}_{-i} (say due to asymmetric information). Note that 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). The multiple agent case is even more difficult in this respect as to compute the distribution of ν_1 in that case we would have to solve for an equilibrium conditional on all possible realizations of \mathbf{d}_{-i} . Not only would this be computationally burdensome it would also typically require an additional assumption that

selects among possible equilibria. The second component of ν_1 , $\nu(d, \cdot) - \mathcal{E}[\nu(d, \cdot)|\mathcal{J}_i]$ is that part of the error in the profit specification that is mean independent of the information the agent bases its decision on (including measurement errors in observables).

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 the goods bought and differences in drive time coefficients (multiplied by drive time). In multiple agent problems d_i might also be a function of $\nu_{2,-i}$.

Selection. We can now explain the selection problem in structural 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}

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

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 $d_i = 1$ if and only if

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

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

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

So as long as x_i is correlated with $\mathcal{E}[\Delta r(d_i = 1, d' = 0, \cdot) | \mathcal{J}_i]$, and any x which is used as an instrument is likely to be correlated with our specification for profit differences, 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 (6) provides the notation needed to clarify the conditions needed to move from the profit inequalities in equation (4) to sample moment inequalities that can be used for inference. 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). Recall that, in addition to the measurement model implicit in the approach, we also need to specify the relationship between the agent's perceptions of expected returns and the returns emanating from the data generation process.

The multiple agent versions of the generalized discrete choice approach assume

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

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 err⁴.

I rush to note that there are important developing parts of the *single agent* discrete choice literature that *do* allow for uncertainty. These include both the structural discrete choice literature, notably the dynamic single agent discrete choice models that explicitly account for randomness in exogenous variables (for an overview see Keane and Wolpin, 2009,

⁴As 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 at the cost of only notational complexity we could allow for a deterministic relationship between a component of d_{-i} and (d, z_i) .

and the literature cited their), and the literature which uses survey data on expectations in conjunction with choice models to allow for uncertainty (see Manski, 2004, and the literature cited their). However the computational difficulties and cognitive assumptions needed to use Bayesian Nash equilibria in applied work, combined with a lack of data on agents' perceptions on the likely behavior of their competitors, have, at least to date, ruled out using analogous techniques in multiple agent problems.

To appreciate the advance that the generalized discrete choice models enabled for the study of markets, it is useful to recall the reduced form rational for using DC3 in single agent problems. In those problems we can always regress expected returns from alternate decisions onto observed variables and the decision. If we then solve for the optimal d conditional on the observables and the disturbances from the regressions and make an assumption on the joint distribution of the disturbances, we can estimate the reduced form regression coefficients with standard techniques.

It is the fact that an analogous reduced form was not useful in multiple agent settings that led to the development of the generalized discrete choice model. In multiple agent contexts researchers were interested in the relationship between profits and (d_{-i}, z_i) *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 entrant, and models which did not allow for unobserved market characteristics that affected the profitability of all potential entrants often estimated coefficients that implied that a firm's profits increased in the number of competitors (as more profitable markets attracted more entrants). Thus were we to extend the single agent reduced form approach and regress profits onto z_i , d_i and d_{-i} , we would want to allow the disturbance to be correlated with d_{-i} . As we show presently, the generalized discrete choice used novel methods to allow for this correlation in a coherent way.

DC4 provides the restrictions the generalized discrete choice model places on the measurement model in equation (6)

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

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 z_i has both observed (the z_i^o) and unobserved (the $\nu_{2,i,d}$) components, we know the distribution of the unobserved conditional on the observed components up to a parameter vector⁵, 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 they make their decisions, just not to the econometrician. Note that a market is defined by values of $(z_i^o, \{\nu_{2,i,d}\}_d, d_i)$ for all participants, so $F(\cdot; \theta)$ is a distribution *across markets* and must be consistent with the sampling distribution introduced below.

Substituting DC3 and DC4 into the model generated by C1 and C2 (equation 4) and letting $\nu_{2,i} \equiv \{\nu_{2,i,d}\}_d$, with analogous notation for $\nu_{2,-i}$, we obtain

$$\text{Model } D : \quad \forall d' \in D_i, \quad \Delta\pi(d_i, d', d_{-i}, z_i^o, \nu_{2,i}; \theta_0) \geq 0; \quad (\nu_{2,i}, \nu_{2,-i})|_{z_i^o, z_{-i}^o} \sim F(\cdot; \theta_0). \quad (7)$$

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

$$RD^{as} : \quad \forall d \in D_i, \quad \pi(d, d_{-i}, z_i^o, \nu_{2,i}) = \pi^{as}(d, d_{-i}, z_i^o, \theta_0) + \nu_{2,i,d}, \quad (8)$$

and the distribution $\nu_{2,i,d}$ conditional on $\nu_{2,-i}$ has full support $[\forall(i, d)]^6$.

Keep in mind that the additive separability in equation (8) *can not* be obtained definitionally. If we did observe realized profits and regressed it on (d_i, d_{-i}, z_i) we would get a residual, but that residual *is not* the $\nu_{2,i}$ in equation (8) as the regression residuals are mean independent of (d_i, d_{-i}) and $\nu_{2,i}$ are not. The regression would be a reduced form and its coefficients would *not* give us responses which *conditioned on* the $\nu_{2,i,d}$. So $\pi^{as}(\cdot)$ should be derived from the economic primitives of the problem.

⁵There 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.

⁶Allowing 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.

Inequalities for Inference. Index markets by $j = 1, \dots, J$. 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 $Eg(\cdot) = \int g(\cdot) dP(\cdot|x)$. The distribution of $\nu_{2,j}$ conditional on z_j^o which appears in DC4, or $F(\cdot; \theta)$, is constructed from P .

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 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 than 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, can be used as a basis for inference.

More formally define the probability that the model in equation (7) (with a restriction like that in equation 8) is satisfied at a particular d_j given z_j^o for a given θ to be

$$\overline{Pr}\{d_j \mid z_j^o, \theta\} \equiv Pr\{\nu_{2,j} : d_j \text{ satisfy equation (7)} \mid z_j^o, \theta\},$$

the analogous lower bound to be

$$\underline{Pr}\{d_j \mid z_j^o, \theta\} \equiv Pr\{\nu_{2,j} : \text{only } d_j \text{ satisfy equation (7)} \mid z_j^o, \theta\}.$$

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 mechanisms) 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 $Pr\{d_j | z_j^o \theta_0\}$, but we do know that when $\theta = \theta_0$

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

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

$$E\left[J^{-1} \sum_j \left(\overline{Pr}\{d_j | z_j^o, \theta\} - I\{d = d_j\}\right) h(z_j^o)\right] \rightarrow_P J^{-1} \sum_j \left[\left(\overline{Pr}\{d_j | z_j^o, \theta\} - Pr\{d_j | z_j^o, \theta_0\}\right) h(z_j^o)\right] \quad (9)$$

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 satisfied⁷. Since typically neither the upper nor the lower bound are analytic function of θ , simulation techniques are employed 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 (7), 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.

⁷As noted by a referee this routine ignores information. If we can enumerate all possible equilibria, as is assumed in getting the lower bound, we should also be able to use the information in the fact that the equilibrium selection probabilities must sum to one. 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, Hong, and Ryan, (2006). Note, however, that in this case we can not use mixed strategies to circumvent existence problems as that would imply that d_{-i} is not known with certainty, and this contradicts *DC3*.

If we were analyzing markets with N interactive agents each of which had $\#D$ possible choices and we used ns simulation draws on $\{\nu_{2,i}\}_{i=1}^N$ to estimate $(\overline{Pr}\{\cdot|\theta\}, \underline{Pr}(\cdot|\theta))$, 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 structural models of multistage games in which calculation of $\mathbf{y}_i = y(\mathbf{d}_i, \mathbf{d}_{-i}, \mathbf{z}_i^o, \nu_{2,i})$ requires computation of an equilibria to a later stage of that game, as in a two stage entry model where a post entry Nash equilibrium (in either prices or quantities) needs to be calculated for each evaluation⁸. There are a number of ways to reduce the computational burden but, at least to date, the computational burden has been large enough to deter its use many applications.

2.4 Profit Inequalities.

This approach dates to Pakes, Porter, Ho, and Ishii (2006) though I generalize their conditions. 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. PC3 provides the first of these.

PC3: Let $x_i \in J_i$ and $h(\cdot)$ be any positive valued function, then

$$\frac{1}{N} \sum_i \mathcal{E}(\Delta\pi(d_i, d', d_{-i}, z_i)|x_i) \geq 0 \Rightarrow E\left(\frac{1}{N} \sum_i \Delta\pi(d_i, d', d_{-i}, z_i)h(x_i)\right) \geq 0.$$

⁸We could avoid the equilibrium computation if we were willing to settle for the “reduced form” from the regression of $\pi^{as}(z_i, d_i, d_{-i})$ onto its arguments; though this would require both a modification to the algorithm described above and additional assumptions. In particular to construct the outer and inner measures used in the inequalities in equation (9) we would need to make an assumption on the distribution of the regression error conditional on (z_i, d_i, d_{-i}) , draw regression errors *and* ν_2 for each agent, and then check the equilibrium conditions conditional on those draws. The draw on ν_1 would have to be mean independent of (z_i, d_i, d_{-i}) while the draw on ν_2 will, at most, be mean independent of z_i .

Since PC3 allows for uncertainty, a word on what it assumes on information sets and prior distributions is in order. 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, condition C1 insures that PC3 is satisfied. These assumptions are, however, stronger than are needed for PC3. Several authors have noted that agents' expectations on profits can be correct without them having such detailed information⁹. More strikingly though correct expectations about profit differences are sufficient for PC3, they are *not necessary*. A weaker sufficient condition is

$$\frac{1}{N} \sum_i \left(\mathcal{E}[\Delta\pi(d_i, d', d_{-i}, z_i)|x_i] - E[\Delta\pi(d_i, d', d_{-i}, z_i)|x_i] \right) h(x_i) \geq 0.$$

This implies that agents can have *incorrect* expectations on $\Delta\pi_i(\cdot, \theta_0)$ provided their expectational error is not systematically related to (i.e. are *mean independent* of) x_i . Indeed PC3 would be satisfied even if agents were *incorrect on average*, provided they were *overly optimistic* about the incremental profits emanating from their decisions.

The final condition used in this estimation strategy is designed to deal with the selection problem induced by the possible presence of the $\{\nu_{2,i,d}\}$. The conditions thus far insure that if the average of individual's expectations, $N^{-1} \sum_i \mathcal{E}(\Delta\pi(\cdot)|x_i) \geq 0$, then the sample average, $E(N^{-1} \sum_i \Delta\pi(\cdot)h(x_i)) \geq 0$. However we do not measure $N^{-1} \sum_i \Delta\pi(\cdot)h(x_i)$, what

⁹The agent's expectation of $\Delta\pi_i(\cdot, \theta_0)$ will be correct if the agent's perception of the joint distribution of (d_{-i}, \mathbf{z}_i) conditional on x_i was correct. This does not require the agent to know either its competitors' strategies or their information sets (for more detail see see Auman and Brandenburger, 1995). Moreover if the decisions being analyzed are decisions that have been made before, the conditional distribution of (d_{-i}, \mathbf{z}_i) is an object which the agent might learn about directly from past play. There are also cases where we do not require the agent to know the full conditional distributions of (d_{-i}, \mathbf{z}_i) . For example in an auction the agents' expectations on $\Delta\pi_i(\cdot)$ would be correct if the agents beliefs about the conditional joint distribution of their own and the highest bid were correct (see Dekel, Fudenberg, and Levine (1999) both for this example, and for more on the relationship between correctness in the sense used here and the closely related notion of self-confirming equilibrium), while if the profit function were quadratic in $(\mathbf{z}_i, \mathbf{d}_{-i})$ all that would be required is that the agents' conditional means and variances of $(\mathbf{z}_i, \mathbf{d}_{-i})$ were correct.

we measure is $N^{-1} \sum_i \Delta r(\cdot) h(x_i)$. From equation (6) the covariance of $\Delta r(\cdot)$ with $h(x)$ will converge to that of $\Delta \pi(\cdot) + \Delta \nu_2$, with $h(x)$. Moreover subsection (2.2) noted that even if all the $\{\nu_{2,d}\}$ are unconditionally mean independent of x , once we condition on the decision actually made, $E[\Delta \nu_{2,d_i,d'} | d_i, x] \neq 0$. However if we can insure that the population average of $\Delta \nu_{2,d_i,d'} h(x)$ is negative, then $E[N^{-1} \sum_i \Delta r(\cdot) h(x_i)] \geq 0$ implies $E[N^{-1} \sum_i \Delta r(\cdot) h(x_i)] \geq 0$. Condition PC4 nest various restrictions on the measurement model and the form of the inequalities chosen that insure the population average of $\Delta \nu_{2,d_i,d'} h(x)$ is negative.

PC4 : For observed $(x_{1,i}, x_{2,i}) \in \mathcal{J}_i$ and a positive function $c(\cdot)$

$$E\left(\frac{1}{N} \sum_i \left(\sum_{j \in D_i} I\{d_i = j\} c(j, d'(j), x_{1,i}) (\nu_{2,i,j} - \nu_{2,i,d'(j)}) \right) h(x_{2,i})\right) \leq 0,$$

where $I\{\cdot\}$ and $h(\cdot)$ are the indicator and positive valued functions introduced above.

To see what underlies *PC4* assume temporarily that $c(j, d'(j), x_{1,i}) \equiv 1$. Then it suffices that an average over agents of

$$\sum_{j \in D_i} E[(\nu_{2,i,j} - \nu_{2,i,d'(j)}) | d_i = j, x_i] \Pr\{d_i = j | x_i\}$$

is not positive. This is an *unconditional* average, it sums over each possible choice for each individual (the choice actually made enters in the same way as any other possibility). For each choice the researcher is free to chose any alternative provided the average of the *differences* between the ν_2 associated with the choice and its alternatives is not positive. The function $c(j, d'(j), x_{1,i})$ allows the researcher to weight the different differences differently. Consider the supermarket example in which the cost of drive time differed across shoppers. We first chose an alternative which was buying the same bundle at different stores thus differencing out the impact of heterogeneity in preferences over bundles, then weighted the resultant utility differences across individuals by the inverse of the difference in drive time thus isolating the drive time coefficients, and only then averaged over individuals to obtain the average cost of drive time.

Assumptions which imply PC4. A sufficient condition for PC4 is that $\nu_{2,i,d} = \nu_{2,i}, \forall d$; i.e. that the component of profitability that the agent knows but the econometrician can not condition on is a “fixed effect” in the sense that it does not vary across choices¹⁰. A related case occurs when a variable that does differ across choices is unobserved at the micro level (say because of confidentiality or a lack of reporting requirements), but is observed, or at least can be computed as a function of observables and a parameter to be estimated, at a higher level of aggregation (say from the census or other official data sources; see De Loecker, et. al. in process, for a firm location choice example).

Pakes et. al (2006) provide a number of examples where PC4 is satisfied when $\nu_{2,d}$ does vary across d . They are constructed as combinations of two underlying properties of the model. In one a feasible alternative choice with the same $\nu_{2,d}$ value as the actual choice is available. Then that alternative is used to form a moment condition which differences out the $\nu_{2,d}$ (in the supermarket example the alternative chosen had the same bundle of goods as that actually bought). A generalization of this occurs when the same $\nu_{2,d}$ appears in the utility of two or more agents. Then to obtain a moment that differences out the $\nu_{2,d}$ the researcher often needs to add inequalities from two or more agents together. Consider a model in which firms are considering entering a market which has a location specific unobservable which enters the value of entry additively and is known to the firms but not the econometrician. Now add the increment in value an entering firm earned from entering to the savings in profits a firm that did not enter saved from not entering. Both differences have positive expectations and their sum differences out the unobservable.¹¹

The second property of a model that allows us to account for $\{\nu_{2,d}\}$ when they vary across decisions is an ability to find inequalities which are additive in the source of error no matter

¹⁰This latter assumption is also the basis of Hansen and Singleton’s (1982) classic article which estimates from the first order conditions generated by Euler’s perturbations to dynamic models. The use of inequalities simply allows us to provide conditions which enable us to extend their analysis to richer choice sets, choices which are on boundaries of those sets, and multiple interacting agents.

¹¹Other examples with a similar structure include social interaction models where the interaction effects are group specific, panel data discrete choice models in which the $\nu_{2,d}$ are choice specific fixed effects (here we need individuals who make a particular choice in some periods and not in others), and cross sectional discrete choice models where the same $\nu_{2,i,d}$ appear in more than one choice.

the choice made (in the supermarket example dividing by drive time made the individual's drive time coefficient enter the inequality additively). One example of a class of models where this is possible that is often used in Industrial Organization is *ordered* choice models. This includes the vertically differentiated demand model and most models designed to analyze the firm's decision of how many units of a given input to purchase (machines, discrete size classes for outlets,...;). In the latter case the cost of the machine often differs across agents (due to local search, installation, or complementary costs) and are known to the agent but not the econometrician (as cost data are usually proprietary). If d is the number of machines bought, to form an inequality that satisfies *PC4* set: $d'(j) = j + 1$, $c(j, j + 1) = 1$ and all other $c(\cdot) = 0$ in that condition. This gives us the difference between what profits would have been had the agent chosen one more machine than it did chose and the actual profits. Given our other conditions the average of this difference in incremental profits across firms will be less than the average cost of a machine. Note that the alternative we compared the choice to differed for different firms, but for each firm we can form an inequality which is linear in ν_2 , so there is *no selection* when we take averages.

I introduce another case which satisfies *PC4* even though $\nu_{2,d}$ varies across d in the next section. It involves the analysis of contracts between buyers and sellers when the terms of contracts are known to the agents but not to the econometrician. In this case $d = \{1, 0\}$ depending on whether the seller's contract offer is accepted, and the $\nu_{2,d}$ represent unobserved components of the payments from the buyer to the seller defined by the contract offer. Note that ν_2 is a cost to the buyer and a revenue to the seller. For each possible buyer-seller combination take; (i) the difference between the seller's profit with and without the contract when the contract is established (when $d = 1$) and (ii) the difference between the buyer's actual profits and what its profits would have been were the contract established when the contract is rejected (when $d = 0$). Since for each possible outcome we have an inequality which is linear in ν_2 , when we average we do not incur a selection problem.

Inequalities for Inference. $C1$ and $C2$ imply that

$$0 \leq \frac{1}{N} \sum_{i=1}^N \mathcal{E} \left[\sum_j \left(I\{d_i = j\} c(j, d'(j), x_{1,i}) \Delta \pi(j, d'(j), \cdot) \right) h(x_i) \right],$$

which from *PC3* implies

$$0 \leq E\left(\frac{1}{N} \sum_i \left[\sum_j \left(I\{d_i = j\} c(j, d'(j), x_{1,i}) \Delta\pi(j, d'(j), \cdot) \right) h(x_i) \right] \right),$$

which from *PC4* and the definitions in equation (6) implies

$$0 \leq E\left(\frac{1}{N} \sum_i \left[\sum_j \left(I\{d_i = j\} c(j, d'(j), x_{1,i}) \Delta r(j, d'(j), \cdot, \theta_0) \right) h(x_i) \right] \right). \quad (10)$$

The terms in equation (10) are *observable* up to θ_0 . So it provides an inequality that can be used as a basis for inference.

Computation of this inequality does not require simulation. However, as in the generalized discrete choice model, when we analyze multistage structural models computation of the counterfactuals for an early stage of the game will generally require $\mathbf{y}_i = y(\mathbf{d}, \mathbf{d}_{-i}, \mathbf{z}_i, \theta)$ for $d \neq d_i$, and this typically requires computing equilibria to a later stage of that game, a task that can be computationally burdensome. On the other hand if we were able to satisfy the assumptions of the profit inequality model but found that calculating the structural form of the profit function was difficult, we could regress profits onto (z_i, d_i, d_{-i}) to obtain reduced form coefficients with a transparent interpretation, and then form inequalities that are designed for inference on these coefficients. We can do this because the regression errors that are added to equation (10) by this process have averages that converge to zero with sample size and hence do not effect the limiting properties of the estimators. It is this reduced form which we analyze in our next example.

2.5 Comparing the Approaches.

The two approaches differ in their informational and measurement assumptions. The informational assumption from the profit inequality approach (*PC3*) nests that of the generalized discrete choice model (*DC3*). In particular *PC3* allows for uncertainty, and it does so without having to specify either the agents' information sets (and hence what each agent knows about either nature or the other agents) or the form of their subjective probability distribution conditional on those information sets; objects we typically know little about. *PC3* does

require those distributions to satisfy conditions, but it suffices that the expectational errors they generate are uncorrelated with the variables used as instruments. On the other hand though the full information restrictions in *DC3* are not likely to be particularly appealing as a general approximation to how decisions are actually made, they might be appropriate for characterizing “rest points” to environments which are sufficiently stable over time. This is often the environment that rationalizes the use of two stage games in cross sectional empirical work in Industrial Organization, and hence is of some importance.¹²

The two measurement assumptions (*PC4* and *DC4*) are not nested, so the different assumptions are likely to be more relevant to different applications. Indeed the simplest cases of the two approaches embody almost diametrically opposite assumptions; the profit inequality approach can always be used when $\nu_{2,i,d} = \nu_{2,i}$ for all d , and, provided one is willing to make an assumption on the joint distribution of the $\nu_{2,i,d}$, the generalized discrete choice model can always be used $\nu_{1,i,d} = \nu_{1,i}$ for all d .

We emphasized the major concern with *PC4* above; the possibility of inadequate controls for selection and the biases that might cause when $\nu_{2,i,d} \neq \nu_{2,i}, \forall d$. The major concerns with *DC4* are; (i) the possibility of a misspecification in the distributional assumption and (ii) the fact that there may be a specification or measurement error which causes $\nu_{1,i,d} \neq \nu_{1,i}, \forall d$. The distributional assumption is a concern because the generalized discrete choice model undertakes an explicit correction for selection and that correction will depend on the properties of the tail of the assumed distribution, properties we typically know little about. Specification error is a concern because it is hard to integrate into the analysis. If we allow for errors in our profit measure, say a $\nu_{1,i,d}$ that represented measurement error in a z_i^o whose impact varied across d , we would have $r(d_i, d_{-i}, z_i^o, \theta_0) = \pi^{as}(d_i, d_{-i}, z_i^o, \theta_0) + \nu_{2,i,d} + \nu_{1,i,d}$. The agent bases its decision only on $\pi^{as}(\cdot) + \nu_2$. So to check the Nash conditions for any given draw on ν_2 we would have to determine whether $\left(r(d_i, d_{-i}, z_i^o, \theta_0) - \nu_{1,i,d}\right) + \nu_{2,i,d}$ satisfied them. For this we would need to take draws on ν_1 conditional on $r(\cdot)$. Even if we were willing

¹²Most two stage models used to date rely on reduced form approximations to the second stage value functions. As noted above to use the generalized discrete choice model to analyze the reduced form coefficients obtained in this way we need assumptions on the approximation error per se and an estimation algorithm which accounts for them.

to specify a parametric distribution of ν_1 that was independent of $\pi^{as}(\cdot)$ and ν_2 , to obtain the conditional distribution of ν_1 given $r(\cdot)$ we would need to solve an integral equation.

There is the question of whether the specification errors that are likely to occur in either approach will have serious consequences on the estimators derived from them. We come back to this question in section 5 where we provide a Monte Carlo analysis of the impact of those errors for the buyer-seller contracting example we move to now.

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

Empirical analysis of upstream interactions in vertical markets has proven to be difficult. There are often a small number of both sellers and buyers (who resell the products they buy to consumers) with most buyers buying from more than one seller and most sellers selling to more than one buyer. The payments made by the buyer are negotiated and vary with underlying market conditions. The structure of these payments determines the welfare benefits from different market structures as they determine both the costs sellers must factor into the prices they set for consumers and investment incentives (whether a seller will invest in cost savings will depend on whether the resulting savings are passed on to the buyer or not). The structure of payments has, however, been difficult to unravel for at least two reasons. First the payments themselves are typically proprietary; a seller who bargains with many buyers often does not want one buyer to know the terms of its contracts with other buyers. Second there is no generally agreed upon theory for how such contracts are set, and as a result, there is no obvious starting point for a structural model of the negotiating process.¹³ My second example explores whether moment inequalities can help.

It extends the work in Ho (2009) on the HMO/hospital market. The goal is to obtain a reduced form characterization of the relationship between payments and observable measures of market structure and costs in this market; information that might both provide some

¹³This contrasts rather sharply with models of markets for consumer goods where, because there are a large number of consumers, we typically assume all power to set prices (or quantities) lies with the seller and use a Nash in prices (or quantities) assumption as a starting point. In these markets costs are usually proprietary, and we often use the equilibrium assumption to back them out.

insights into the likely determinants of the payment structure and help theory chose between alternative modeling frameworks. The analysis works off the fact that though we might not observe the contracts per se, we frequently do observe which sellers establish contracts with which buyers. A model which specifies the inequalities that must be satisfied for the observed relationships to be incentive compatible can be used as a basis for estimation. Note, however, that to use moment inequalities we will have to specify counterfactuals; what would have happened if either the buyer or the seller opted to change its contracting behavior. Counterfactuals require assumptions, though the assumptions need not generate a unique equilibria nor specify the form of the set of feasible contracts.

Ho provides empirical results from forty markets. She uses a particular game form and assumes that there are no structural disturbance 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 used we can develop an estimation algorithm for her model that allows for both ν_2 and ν_1 errors that vary over d . I then compare the results from estimators that do, and those that do not, allow for ν_2 errors. Next actual equilibria are computed for a buyer-seller network game that has primitives which are similar to those in the empirical example. The results from these computations are used in two ways. First we examine the reduced form from the computed equilibria to both; provide an indication of the robustness of the empirical results, and to provide insights into variables that might be important in determining the structure of HMO-hospital contracts. Second we use the computed equilibria in a Monte Carlo examination of the robustness and computational burdens of our two approaches to moment inequality estimation.

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. These determine both the network of hospitals the HMO's members can access and the transfers from the HMO to a hospital for each patient hospitalized. In the second period the HMOs engage in a premium setting game which is assumed to have a unique Nash equilibria.

The second period equilibrium generate revenues for *each* HMO conditional on *any* configuration of hospital networks. Letting 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 H_{-m} specify the networks of the competing HMOs, these revenues and quantities will be denoted by $R_m(H_m, H_{-m}, z)$, and $q_{m,h}(H_m, H_{-m}, z)$, respectively. The parameters needed for these calculations are obtained from a careful study of hospital demand and costs (for details see Ho, 2006, 2009).

The profits of the HMO are the revenues from the second period game minus the payments (the “transfers”) the HMO makes to each hospital in its networks, say $T_{m,h}$ or

$$\pi_m^M(H_m, H_{-m}, z) = R_m(H_m, H_{-m}, z) - \sum_{h \in H_m} T_{m,h}(H_m, H_{-m}, z).$$

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

$$\pi_h^H(M_h, M_{-h}, z) = \sum_{m \in M_h} T_{m,h}(H_m, H_{-m}, z) - c_h \sum_{m \in M_h} q_{m,h}(H_m, H_{-m}, z).$$

We are after a reduced form characterization of the per patient transfers from the HMOs to the hospitals (our $T_{m,h}$). So the equation of interest is a projection of $T_{m,h}$ onto a set of interactions of $q_{m,h}(H_m, H_{-m}, z)$ with a vector of hospital, HMO, and market characteristics, say $x_{m,h}$, which we write as

$$T_{m,h}(H_m, H_{-m}, z) = x_{m,h}(H_m, H_{-m}, z) \theta + \nu_{2,m,h}.$$

Note that since the agents know the details of the contracts they sign (though the econometrician who does not) the disturbance generated by this projection is known to both agents when they make their decisions; i.e. it is a ν_2 error in both agents' profit functions. So if we define $r_m^M(\cdot, \theta) \equiv R_m^M(\cdot) - \sum_{h \in H_m} x_{m,h}(\cdot) \theta$, and $r_h^H(\cdot, \theta) \equiv \sum_{m \in M_h} x_{m,h} \theta - c_h \sum_{m \in M_h} q_{m,h}(\cdot)$, to be the observable parts of hospital and HMO profits we have

$$\begin{aligned} \pi_m^M(H_m, H_{-m}, z) &= r_m^M(H_m, H_{-m}, z; \theta) - \sum_{h \in H_m} \nu_{2,m,h}, \quad \text{and} \\ \pi_h^H(M_h, M_{-h}, z) &= r_h^H(M_h, M_{-h}, z; \theta) + \sum_{m \in M_h} \nu_{2,m,h}. \end{aligned} \tag{11}$$

Equation (11) provides the profits agents obtain from any given buyer-seller network. To obtain our moment inequalities we have to consider the profits, and hence the network, a contracting agent thought might obtain had it changed its contracting decision.

Counterfactuals. To obtain the counterfactuals needed for the moment inequalities we need assumptions on the first stage contracting game. Ho’s assumption are both familiar and computationally convenient (we consider alternatives below).

Assume that sellers simultaneously make take it or leave it offers to buyers, who then simultaneously accept or reject. As in Hart and Tirole (1991) also assume that contract offers are 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 the returns to any network. So to form our moment inequalities all we need is the network that would be established were either an HMO or a hospital to change its behavior.¹⁴

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 need to specify; (i) an alternative feasible offer the hospital could make, and (ii) what the hospital thinks the particular HMO would infer about the offers the hospital made to other HMOs were the HMO to be offered that alternative. We assume; (i) that the hospital could always offer a null contract (a contract which is never accepted), and, at least for the empirical work, (ii) “passive beliefs”, i.e. the hospital believes that the HMO will not change its perceptions about the likelihood of different offers being made to its competitors were it to receive a different offer, or in terms of our earlier notation, that $d^{-i}(d', z_i) = d_{-i}$. Still if an HMO that had accepted a hospital’s

¹⁴Since 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.

offer received the null contract instead, it might change its contracts with other hospitals. So to compute the hospital profit inequality that results from the offer of the null contract we either have to specify how the hospital thinks the HMO would have changed its contracting decisions with other hospitals if it had received the null contract, or we have to compute a lower bound to the profits the hospital could make as a result of the actions the HMO might take. In the empirical results presented here we assume that the hospital thinks the HMO does not change its behavior with other hospitals were it to receive the null contract, but we have also done the analysis with a counterfactual which allowed the HMO to add a different hospital with little difference in empirical results.

Profit Inequalities Used. We begin with the assumption that $\nu_2 \equiv 0$. Then we can obtain our inequalities by interacting the differences in profits from the three counterfactuals listed above with positive valued functions of variables that are known to the decision maker when it makes its decision (our “instruments”). We call these inequalities the ν_1 only inequalities, since, if $\nu_2 \neq 0$, these inequalities are invalid.

We considered alternative ways of allowing for ν_2 , starting with the familiar specification in which the $\nu_{2,m,h} = \nu_{2,m}, \forall(m, h)$; i.e. there are HMO specific fixed effect in forming contracts. With multiple agents on each side of the market this generates the quite detailed set of inequalities provided in Appendix 1¹⁵. There is no particular reason to assume a fixed effects structure for our problem and when we allowed for them the empirical results seemed to accentuate the problems with the ν_1 -only model (though the parameter estimates from the two specifications were not too different from one another). 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. So we look for inequalities which allow for ν_2 but do not require them to be fixed effects.

¹⁵A 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 the Appendix. 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.

There are at least two such inequalities available. One does involve the ν_2 but does not need to condition on the decision made, while the other does not involve ν_2 . Let $\chi_{m,h} = \{1, 0\}$ index the two possible contracting outcomes, with $\chi_{m,h} = 1$ if HMO m accepts hospital h 's offer and zero otherwise. If $\chi_{m,h} = 1$ the increment in the hospital's profit from offering the given contract instead of the null contract is expected to be positive and contains the transfer (including $\nu_{2,m,h}$), while if $\chi_{m,h} = 0$ the increment in the HMO's profit from rejecting (instead of accepting) the contract saves the transfer (including $\nu_{2,m,h}$). This gives us an inequality which does not condition on $\chi_{m,h}$.

More formally let $\Delta\pi_h^H(M_h, M_h/h, M_{-h}, z)$ be the difference between the hospital's profit when the network of the hospital includes HMO m and when it does not, and $\Delta\pi_m^M(H_m, H_m \cup h, H_{-m}, z)$ be the difference between the HMO's profit when the network of the HMO excludes hospital h and when it includes it. Then from equation (11)

$$\begin{aligned} & \chi_{m,h}\Delta\pi_h^H(M_h, M_h/h, M_{-h}, z) + (1 - \chi_{m,h})\Delta\pi_m^M(H_m, H_m \cup h, H_{-m}, z) = \\ & \chi_{m,h}\Delta r_h^H(M_h, M_h/h, M_{-h}, z; \theta) + (1 - \chi_{m,h})\Delta r_m^M(H_m, H_m \cup h, H_{-m}, z; \theta) + \nu_{2,m,h}, \end{aligned}$$

is expected to be positive and is linear in ν_2 *no matter the outcome*. So provided we have an $x \in \mathcal{J}_m \cap \mathcal{J}_h$ that is an instrument in the sense that $E[\nu_2|x] = 0$, then for any positive function, $h(\cdot)$

$$E\left[\chi_{m,h}\Delta r_h^H(M_h, M_h/h, \cdot; \theta_0) + (1 - \chi_{m,h})\Delta r_m^M(H_m, H_m \cup h, \cdot; \theta_0)\right]h(x) \geq 0. \quad (12)$$

For the second inequality note that the sum of the increments in profits to the HMO and the hospital does not contain the transfers between them (and hence not $\nu_{2,m,h}$), 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 if a contract is established (at least if contract offers are proprietary). So provided $x \in \mathcal{J}_m \cap \mathcal{J}_h$

$$E\left[\chi_{m,h}\left(\Delta r_h^H(M_h, M_h/h, \cdot; \theta_0) + \Delta r_m^M(H_m, H_m/h, \cdot; \theta_0)\right)\right]h(x) \geq 0 \quad (13)$$

where $\Delta r_m^M(H_m, H_m \cup h, H_{-m}, z; \theta)$ is the change in HMO m 's profits were it to refuse a contract it in fact accepted.

Neither the ν_1 -only nor the model which allowed for ν_2 could be rejected by our formal tests which is not surprising given the sample size. However the results did seem to favor the model that allows for ν_2 as only six of the inequalities were negative at the 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).

Parameter Estimates. The first four columns of Table 1 present the empirical results. We subtracted costs per patient 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 (and all our specifications had eighty-eight or more inequality constraints). The algorithm then generates a point estimate for θ_0 which is that value which minimizes a squared metric in the negative part of the sample moments. These are the values reported in the table.¹⁶

Though the small sample size implied that we could not do a very detailed investigation of the correlates of markups the estimates we do get are eye-opening, indicating just how important a deeper understanding of the determinants of these contracts might be. This is because both sets of point estimates imply an equilibrium configuration where the majority of cost savings from low cost hospitals are captured by the HMOs who do business with those hospitals, and in which markups increase sharply when a hospital is capacity constrained. If these findings were interpreted as causal they would imply significantly lower incentives

¹⁶Though we have not formally introduced them, all our estimates allow for errors in our measures of HMO revenues and hospital costs provided they are mean independent of the variables we use as instruments. There were about 450 plans and 630 hospitals in our markets, and plan and market characteristics were used as instruments. 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). Finally Ho (2009) reports a series of robustness checks on the ν_1 -only estimates of a model which is only slightly different from the model presented here. Though more robust specifications, particularly those which add right hand side variables, can increase the confidence intervals quite a bit, the parameter estimates do not change much from those in column (1) of the table.

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 over 50% of those savings do and a larger fraction of profits go to capacity constrained hospitals. As one might expect low cost hospitals tend to be more capacity constrained than high cost hospitals, so the two variables are negatively correlated.

3.2 Numerical Analysis.

Might we expect contracts with these characteristics to emanate from a Nash 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 in markets with characteristic distributions 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 different markets in a reasonable amount of time).¹⁷

We compute a full information Nash equilibrium to a game in which hospitals make take it or leave it offers to HMOs. More specifically 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

¹⁷We 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.

best it can do given the actions of its competitor and the offers made by the hospitals.¹⁸

Note that these assumptions differ from those assumed to generate the data in the empirical analysis. In particular in the 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) which equilibrium notion is likely to be more appropriate for a given institutional setting, and (ii) whether the estimation results are likely to be sensitive to the profit inequalities they generate, are two of the many 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 (long enough so that each agent could have responded to the situation if it wanted to), we might only want to consider equilibria which are renegotiation proof in the sense that 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

¹⁸An 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 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 vary 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 offer is 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.

need for the empirical work is a way of obtaining a *lower bound* to the expected profits from a counterfactual choice, one possibility is to 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 of counterfactuals the less tight our bounds are likely to be and the larger the computational burden; for alternative game forms see Hart and Tirole ,1990, McAfee and Schwartz, 1994, Kranton and Minehart,2001, Rasmussen and Ramseyer,1991, and the literature they cite).

Numerical Results. Column (5) through (8) of Table 1 present the results from projecting the computed full information Nash equilibrium markups onto different sets of variables. 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 (this translates into over 85% of the variance in transfers). The second two columns add variables of interest. The original three variables maintain their signs and remain significant but there are noticeable changes in their magnitudes. So though the empirical results do seem to pick up important features of the contracts generated by our computed equilibria, the numerical results suggest that they are, not surprisingly, only a starting point for a more detailed causal analysis.

The coefficients of the additional variables illuminate the likely relationships embodied in equilibrium outcomes. 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 (when the ratio of population to the number of hospital beds is lower) and once we account for this effect 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 can charge their members higher markups. Apparently the contracts that emanate from the complete structural model are complex, but have many familiar features.

One final point. After adding the extra variables 20% of the variance in markups, or 8% of the variance in transfers, is not accounted for by our observables. As noted this variance would be classified as variance in the structural error (or in ν_2) in our prior discussion. On the other hand even if our behavioral, informational, and functional form assumptions were perfect we would expect substantial measurement error in hospital costs in any data set and this would constitute a ν_1 error. So there is reason to believe that both types of errors are present. This is likely to be more generally true in applications of moment inequality estimators and there is a question of how the different errors impact alternative estimation algorithms. We turn to this question now.

4 Specification Errors and Alternative Estimators.

I now apply both the generalized discrete choice and the profit inequality approach to the analysis of contracts in our HMO-hospital example. Most of the results will be Monte Carlo results based on the data obtained from computing equilibria to about one hundred thousand of the markets described in the last section. Estimates are presented *both* when the models' assumptions are the assumptions generating the data, and when they are not. This enables us to examine the implications of specification errors on the parameter estimates from the two models. Where possible we will also present results from Ho's data.

Details of the Monte Carlo Analysis. 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.¹⁹ The function obtained from this projection is treated as the parametric transfer function. The coefficients of the instruments in this function 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

¹⁹For 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.

treat these ν_2 as known and add pseudo random draws on a normal measurement error to the costs of each hospital (and sometimes also to the population size, and hence the patient flows, in the market).

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.²⁰

The counterfactuals used to generate inequalities in the profit inequality approach 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 which had accepted its equilibrium offer that hospital considers the profits that would accrue to it were both HMOs to reoptimize²¹. For the generalized discrete choice approach we used the inequalities generated by the necessary (Nash) conditions for equilibrium.²²

Results. Table 5, 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

²⁰Actually 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.

²¹If $o_{m,h}$ is the contract offered by hospital h to HMO m in equilibrium, ϕ designates the null contract, and $h=1$ ’s offer was accepted by $m=1$, then the counterfactual requires the profits of $h=1$ from the HMO equilibrium responses to the tuple $(\phi, o_{1,2}, o_{2,1}, o_{2,2})$.

²²This does not exhaust the possible inequalities for 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 actual contracts were available to the researcher yet other inequalities would become available), and (ii) computed the probability that the observed equilibrium was unique for the generalized discrete choice approach.

(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 joint distribution of the ν_2 's conditional on the market's instruments. This is not known and too complex to estimate accurately. To get a sense of the narrowness of the interval generated by the generalized discrete choice approach we set all the disturbances to zero and for each possible network structures we; (i) found the set of θ which lead to positive differences between the averages of the indicator functions for the Nash conditions being satisfied and the indicator functions for the “observed” equilibrium, 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 simulated data adding 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 bound (θ_0), while that of the upper bound is almost exactly equal to $\bar{\theta}_0$. Moreover the bounds are precisely estimated, less than 2.5% of the lower (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 inequality estimator 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 the simulated data set after adding both ν_1 and ν_2 errors and 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 very 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 the ν_2 error.

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 what we are estimating 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 a distributional assumption for the ν_2 disturbances. 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 (we would expect it to be if the ν_2 were truly independent of our instruments, rather than just mean independent, but this is unlikely).

Regardless of whether we use just the constant term as an instrument or all of our instruments both the algorithms based on the empirical distribution of the ν_2 and that based on the normal distribution always generate point estimates whose point estimates are larger than the true value of the parameter and whose distribution falls in a narrow range about the point estimates. In all of the cases the interval formed from 95% of the point estimates does not intersect with our approximation to the identified set, and does not cover the true value of θ_0 . Clearly the choice of distribution for ν_2 does matter, and the obvious choices lead to biased estimates. Though this is disturbing the magnitude of the resulting

bias is small; the point estimate is never more than 7%, and the lower bound of the confidence interval is never more than 5%, larger than the true 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 , as an empirical researcher would not have access to the bootstrap distribution. Adding ν_1 errors does increase the parameter estimates further, but by a surprisingly small amount. We tried but could not compute estimates the ν_2 -only algorithms using Ho's actual data set; the number of HMOs and hospitals in Ho's data imply that there are on the order of 100,000 outcomes for which premium setting equilibria and profits must be calculated for each ν_2 draw and each θ evaluated. This 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 (12) and (13). 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 that much lower. 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 encouraging. The ν_1 -only estimator is easy to use and seems quite robust to ν_2 variance. The ν_2 -only estimator does seem sensitive to distributional assumptions, but the bias resulting from using a standard normal assumption was small, and the bias caused by the presence of ν_1 error was even smaller. Perhaps the biggest problem with using the ν_2 only estimator in our two stage game example is its computational burden. The estimator which use the robust inequalities are least subject to biases, but they do lead to larger identified sets.

5 Summary.

Behavioral choice models generate inequalities. To use these inequalities in estimation we require a set of auxiliary assumptions. This paper formulates and compares two sets of assumptions which suffice. Their relevance differs with; (i) whether the researcher is interested in estimating the parameters of the primitives underlying the behavioral model or in estimating the reduced form coefficients obtained from regressing the profit function on its determinants, and (ii) with whether the problem involves interactions among multiple agents or is a single agent problem. The two sets of assumptions generate different estimators with different computational burdens. We illustrate their use in two problems of importance to Industrial Organization and consider their robustness to specification errors in one of them.

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 P. Guggenberger ,2009; "Validity of Subsampling and Plug-in Asymptotic Inference for Parameters Defined by Moment Inequalities" Cowles Foundation Discussion Paper No. 1620, forthcoming *Econometric Theory*
- Aumann, R., and A. Brandenburger ,1995; "Epistemic Conditions for Nash Equilibrium", *Econometrica*, 63, 1161-80.
- 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.
- 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.
- 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.
- Ho, K.,2006, "The Welfare Effects of Restricted Hospital Choice in the US Medical Care Market" *Journal of Applied Econometrics*, 21(7), pp 487-545.
- Kranton, R. and Minehart, D. 2001, "A Theory of Buyer-Seller Networks", *American Economic Review*, Vol. 91 (3), pp 485-508.
- 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 "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.
- Rasmusen, E., and Ramseyer, M.,(1991); "Naked Exclusion" *American Economic Review*, Vol. 81 (5), pp 1137-45.
- 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 $\forall(h, m), \nu_{2,m,h} = \nu_{2,m}$. 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 $\tilde{h} \notin H_m$ and $h \in H_m$ we have

$$\Delta\pi_m^M(H_m, H_m \cup \tilde{h}, \cdot) + \Delta\pi_m^M(H_m, H_m \setminus h, \cdot) = \Delta r_m^M(H_m, H_m \cup \tilde{h}, \cdot) + \Delta r_m^M(H_m, H_m \setminus h, \cdot),$$

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

$$E\left[\Delta r_m^M(H_m, H_m \cup \tilde{h}, \cdot; \theta_0) + \Delta r_m^M(H_m, H_m \setminus h, \cdot; \theta_0)\right]h(x) \geq 0.$$

For the second inequality note that if $\nu_{2,m,h} = \nu_{2,m}$ we can use the logic leading to equation (12) in the text to show that for any positive valued function, $h(\cdot)$

$$\begin{aligned} 0 &\leq E\left[\frac{1}{\#H} \sum_h \left(\chi_{m,h} \Delta\pi_h^H(M_h, M_h/h, \cdot) + (1 - \chi_{m,h}) \Delta\pi_m^M(H_m, H_m \cup h, \cdot)\right)\right]h(x) = \\ E\left[\frac{1}{\#H} \sum_h \left(\chi_{m,h} \Delta r_h^H(M_h, M_h/h, \cdot; \theta) + (1 - \chi_{m,h}) \Delta r_m^M(H_m, H_m \cup h, \cdot; \theta)\right) + \nu_{2,m}\right]h(x) \\ &\equiv E\left[\bar{S}^r(\mathbf{m}, \cdot; \theta) + \nu_{2,m}\right]h(x). \end{aligned}$$

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

$$E\left[\Delta r_m^M(H_m, H_m \setminus h, \cdot; \theta_0) + \bar{S}^r(\mathbf{m}, \cdot; \theta_0)\right]h(x) \geq 0.$$

Table 1: Determinants of Hospital/HMO Contracts.

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)
Variable	θ	95% CI UB/LB	θ	95% CI UB/LB	θ	s.e.	θ	s.e.
<i>Per Patient Markup (Units = \$ thousand/patient)</i>								
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 UB	95% of θ in LB UB
A: Using ν_1 inequalities.			
Simulated data: $\theta_0=16.76$, $[\underline{\theta}_0, \bar{\theta}_0]=[13.47, 18.59]$.			
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: $\theta_0=16.76$, $''[\underline{\theta}_0, \bar{\theta}_0]''=[16.7, 16.95]$.			
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: $\theta_0=16.76$, $''[\underline{\theta}_0, \bar{\theta}_0]''=[16.7, 16.85]$.			
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.