Theory and Empirical Work on Imperfectly Competitive Markets.*

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

January 17, 2008.

Abstract

This paper reviews recent developments in the empirical analysis of imperfectly competitive markets highlighting outstanding problems. Some of these problems are econometric. The need for a deeper understanding of the small sample properties of semiparametric estimators and of estimators based on moment inequalities head this list. The modeling problems go back to issues which have been a central part of ongoing research programs in economic theory for some time. We consider ways in which applied work can cope with these problems and, in so doing, also inform theory. The use of estimators based on moment inequalities opens up several possiblities in this regard, and the paper looks in detail at the assumptions that have been used to rationalize these estimators and their likely relevance.

*This is a revised version of my Fisher-Schultz Lecture which was 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 opporunity 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, and to my theory colleagues for helping me to come to terms with parts of their literature. Financial support from the NSF and the Tolousse Network for Information Technologies is greatly appreciated. The word "theory" in the title refers to both economic and econometric theory.

This paper has two goals. It begins by outlining a set of recent developments in the empirical analysis of imperfectly competitive markets. Each development is motivated by a barrier to empirical work. The developments themselves often consist of econometric and computational tools that enabled us to bring to data frameworks that had been used extensively in prior theoretical work.

In describing these developments I will point out several issues empirical work has been less successful in dealing with. Many of these issues are directly related to ongoing research programs in economic theory. The discussion considers ways in which applied work can interact with theory to further our understanding of them.

With these issues in mind, the second part of the paper considers one way of weakening the assumptions needed to obtain our estimators. It details the different sets of assumptions which are implicit in the different ways that have been used to construct estimators based on moment inequalities (including behavioral, informational, and timing assumptions). We then examine how the different assumptions perform in one outstanding problem; the analysis of contracts in buyer-seller networks.

An attempt has been made to make each of the two parts of the paper self-contained, so one can read either of them in isolation. Sections 1 to 3 provide the review of prior work; section 1 provides an overview and a discussion of multiple equilibria, section 2 a review of empirical work using static models, and section 3 a review of dynamic models. Sections 4 to 6 consider moment inequalities and their application; section 4 provides the two sets of assumptions that have been used to generate moment inequalities and considers the estimators that result from them, section 5 provides empirical and numerical results on our buyer-seller network axample, and section 6 provides a Monte Carlo analysis of the impact of specification errors on the alternative estimators.

1 Overview.

The goal is to build up a set of tools that enable us to empirically analyze market outcomes in oligopolistic situations. To keep matters as focused as possible I will begin by assuming; (i) symmetric information and (ii) that the distribution of future states and controls, conditional on the current states and all investment expenditures, does not depend on the current price (or quantity) choice. It is this latter assumption which rationalizes the undergraduate textbook formulation in which price (or quantity) choices can be analyzed in a static framework, and the evolution of the state variables which determine the profits from those static choices in a dynamic one.

As I hope will become evident the modelling choices an empirical I.O. researcher would like to use are largely determined by the the institutional structure of the industry studied.

In this context the independence assumption in (ii) above seems to rule out the analysis of many markets: those with significant learning by doing effects or adjustment costs (then future costs depend on current controls); those in which the experience, durability, and/or the network effects of the products marketed causes future demand to be related to current price or quantity choices; and most models of collusion (then future prices or quantities are partly determinned by the current price or quantity choice). Still one has to start somewhere, and I will mention research on these extensions below.

As for the assumption of symmetric information, most empirical work has not paid alot of attention to the structure of agents' information sets, often chosing whatever structure seems most convenient for the problem at hand. There are many reasons for this, not least of which is that the econometrician usually does not have detailed knowledge of the contents of those information sets. In contrast a great deal of effort has been devoted to the econometric treatment of unobservables that are known to *all of* the agents, just not to the econometrician. Indeed appropriate treatment of these unobservables underlies many of the recent advances in demand and production function estimation reviewed below. The exception here is empirical work on auctions, which has reversed the emphasis and has largely ignored the role of unobservables known to all the agents but not to the econometrician (for a notable exception, see Krasnokutskaya,2006).

I come back to informational assumptions in the second part of the paper. There I show that one can often allow for quite general informational assumptions and still obtain consistent estimates of parameters. Their remains, however, the largely unexplored and important questions of; (i) which informational assumptions are consistent with the data, and (ii) what are the implications of the alternatives that are.

Static Analysis. Static analysis conditions on (i) the goods marketed (or their characteristics) and their cost functions, (ii) consumer's preferences over those goods (or over characteristics tuples), and (iii) "institutional" features like; the type of equilibrium, structure of ownership, and regulatory rules. Many of the recent advances have been directed towards obtaining better estimates of demand and cost systems, and the next section provides a brief review of them, as well as comments on the performance of alternative equilibrium assumptions.

The actual analysis takes these "primitives" as input and then calculates equilibrium prices, quantities, profits and consumer surplus, as a function of the problem's state variables. The latter typically include the characteristics of the goods marketed and the determinants of the cost of producing those goods. The result is a logical framework in which we can consider the implications of policy or environmental changes in the "short-run" (conditional on the state variables prevelant when the analysis is undertaken). This is usually the first step in the analysis of any policy (e.g.s, merger or tariff) or environmental (e.g.s input price or demand) change.

Dynamic Analysis. The goal of the dynamics is to analyze how the state variables that are subject to the firms' controls evolve and the impact of changes in exogenous state variables. The conceptual framework used for most of the applied dynamic work in I.O. has been Markov Perfect equilibrium in investment (broadly defined) stategies. The Markov Perfect notion dates back at least to Starr and Ho (1969) and was used in an influential set of theory papers examining dynamic issues in oligopolistic settings by Maskin and Tirole (1987). It is an equilibrium notion that is particularly well suited to applied work as it allows us to condition on a current state, hopefully a state that we might be able to read off the data, and generate a probability distribution of the next year's state. That distribution can then be used for for either estimation or for numerical analysis.

To use Markov perfection in this way empirical analysis needed a Markov Perfect framework which allowed for the richness of real world data sets; i.e. for firm and industry specific sources of uncertainty (to enable rank reversals in the fortunes of firms as well as profits of competing firms that are positively correlated) and entry and exit. This was initially provided by Ericson and Pakes, 1995, and then extended by a series of authors (see Doraszelski and Pakes,2005, for a recent review). These frameworks are used for the longer run analysis of the likely impacts of policy or environmental changes; for e.g. the impact of mergers on entry or investments and therefore on longer run price changes, or the impact of gas prices on the fuel efficiency of capital.

Multiple Equilibria and Applied Work.

I will largely ignore problems that might arise due to multiple equilibria below, so a short digression on how this possiblity effects applied work and what may be done in response is in order. To an applied person multiplicity is fundamentally a problem of not having a detailed enough theory. There is a set of strategies that are actually selected and used in any given situation, it is just that the equilibrium conditions we are confortable with assuming are not detailed enough to distinguish between the different possibilities. This implies that as we learn more about particular industries we might be able to rule out certain possiblities as not being consistent with observed behavior (or with the institutions governing it). Still in the interim we will have to consider how the possiblity of multiple equilibria effects; (i) estimation, and (ii) the substantive analysis of issues once the primitives have been estimated.

Taking estimation first, it is clear that if there are multiple possible equilibria maximum likelihood estimation can not be used without additional assumptions. To obtain the likelihood of any parameter vector we need to assign a unique outcome to each combination of observables and unobservables determinants of that outcome were that parameter vector correct. When there are multiple equilibria there is no unique outcome. There are a number of likelihood and non-likelihood based ways of circumventing this problem in estimation. The two most frequently used are; estimating off of the necessary conditions for equilibria (e.g. the first order conditions for price setting in Nash equilibria), or making the assumptions that allow one to use the data to pick out the equilibria actually observed (see the discussion of dynamic estimation below). However there remains an important efficiency issue; i.e. we do not know what an efficient estimator is under assumptions that allow for multiple equilibria, and we do not have the heuristics of maximum likelihood to guide us to an answer.

Perhaps more troubling to the applied researcher is the impact of multiple equilibria on our ability to analyze counterfactuals. Much of the detailed work that goes into empirical modelling in I.O. is to enable us to analyze what would happen were a policy or the environment to change. Though under some assumptions we can identify the equilibria actually played in the past (or the selection mechansim actually used in the past), we have no direct evidence on what equilibria would be played (or which selection mechanism would be used) once a change occurs.

There are at least two complementary possible approaches to this problem. First one could attempt to compute all possible equilibria and bound the outcomes of interest. The extent to which this is either practical, or helpful if practical, is likely to vary from problem to problem. An earlier version of this paper illustrated this possiblity with an empirical example from Ishii (2005). Ishii analyzed a two period model of ATM choice; in the first period banks chose the number of their ATMs and in the second they set interest rates and consumers chose between banks based on those rates, the proximity of the banks' branches to their home, and the banks' ATMs. We took the actual estimated demand system for Pittsfield Massachusetts, computed equilibrium interest rates and profits for each possible allocation of fifteen ATM's among Pittsfield's seven banks, and checked which of the possible allocations would lead to a full information Nash equilibrium for different specifications of the cost of ATMs (this excercise is now discussed in a separate note, see Pakes, 2008).

Though there were on the order of 200,000 possible allocations, depending on the cost specification, only one to three of them satisfied the Nash equilibrium conditions. Moreover when there were multiple equilibria for the same cost specification the different equilibria were quite similar to each other. There were no two equilibria for the same cost specification in which one firm differed in its number of ATM's by more than one, and the maximum difference in total number of ATM's across equilibria for a given cost specification was two. Finaly the "comparative static" results on the relationship of the equilibria across cost specifications made economic sense. If an allocation which had been an equilibria was no longer an equilibria when we lowered the cost, that equilibrium was always the equilibrium with the least number of ATM's at the higher cost. If an allocation became an equilibrium allocation when it had not been one at the higher cost, the new equilibrium allocation always has a larger total number of ATM's then the equilibria that are dropped out.

To the extent that these findings are indicative of what might happen in other applied problems, they are good news for the applied researcher. A small number of equilibria makes enummeration possible, and the expected comparative static results (on sets) held when we compared across the equilibria of the different cost specifications. The results do, however, rely, on the fact that the actual profit functions have a substantial amount of heterogeneity built into them. Were we to eliminate the inherited "history" of branch locations and assume that banks chose the number and location of their branches along with the number of their ATMs (or allow them to trade the branches at the current locations), the results would be different. On the other hand empirical work on markets typically does find large assymetries across competitors, and there are costs to change. As a result a a simple procedure for determining whether a given set of assymetries in profit (or value) functions are likely to lead to small numbers of (and/or "well behaved") equilibria would be quite helpful[1].

There is a second approach to analyzing counterfactuals when multiple equilibria are possible. Though we may not feel confortable with assumptions which would select out an equilibrium, we may be willing to model how agents respond to changes in their environment. One possiblity is to employ a model for learning about how a policy or environmental change impacts on the perceived returns from alternative feasible strategies, and assume that the agents choices maximizes their expected perceived returns. If the primitives of the problem and the current equilbrium were known (or estimated), and one were willing to model this learning process, it should be possible to simulate out and attach probabilities to the likely post change equilibria.

We used the ATM example to experiment with this approach also (for more detail see Pakes, 2008). We started at one of the equilibria and assumed there was a change in the cost of operating ATMs. Post regime change firms chose the number of their ATMs before knowing what the costs would be. Cost draws were i.i.d. across time periods and firms. In this environment firms' expected profits depend on their expected costs and on their perceptions on what their competitors will do. We assumed a firm expected its costs to equal the sample mean of the cost shocks it had received since the regime change, and tried two different specifications for a firm's beliefs about its competitors' play; (i) that competitors' play would be the competitors' actual play in the prior period (so the firm

[1]There is some related work in the theory literature; see, for e.g., Mclennan's (2005) analysis of the number of possible Nash equilibria for general payoff functions, and the analysis of the impact of different forms of heterogeneity in global games in Morris and Shin (2002). To date, however, there has been very little applied work that used the results of the theory literature to overcome problems generated by multiplicity (for a notable exception, see Jia,2006).

played its best response to the immediate previous play), and (ii) that the next play of its competitors to be a random draw from the set of tuples of plays observed since the regime change (a "fictitious play" specification). Different specifications for the mean and variance of the cost shock were tried and for each specification we started 1000 runs which used the best reply, and 100 runs which used the fictitious play, dynamic. Each run was allowed to continue until it had stayed at the same location for fifty iterations, and that location was viewed as a "rest point" of the process. All rest points were Nash equilbria of the game where each agent knows its mean costs, and there were no convergence problems.

We found that the variance in the cost shocks can cause a *distribution* of rest points from a given initial condition, and that distribution has a notable dependence on both the cost specification *and* on the learning process. Neither the fact that there were a distribution of rest points nor that they depended on the cost specification is pariuctlarly troubling for applied work, but the dependence of that distribution of on the learning process is. Though there is a recent theory literature which considers different learning models and analyzes their properties (see, in particular, Fudenberg and Levine, 1999, Young, 2004, and the literature they cite), there is little to no *empirical* evidence on when the different models might be appropriate. This is another area where empirical work would be extremely helpful to both theory and applied analysis.

2 Static Analysis.

Static analysis requires; (i) a demand system, (ii) a cost system, and an (iii) equlibrium assumption. I now outline recent work on each of these. The focus is on conceptual issues. I leave the details required to use the techniques to the original articles and their successors.

2.1 Demand Analysis.

Demand and cost systems are the real (in contrast to the strategic) primitives which determine pricing incentives, and through prices and costs, the incentives for product development. Moreover cost data are often proprietary, while (at least) market level data on prices charged, quantities sold, and characteristics of products, are typically not. So costs are often inferred from estimates of a demand system and a behavioral assumption which determines equilbrium play. As a result getting a reasonable approximation to the demand surface is often the most important part of the problem facing an applied researcher.

The last decade has seen at least two major changes in the empirical analysis of demand patterns. First there has been a movement away from representative agent models of demand to models with heterogenous agents, and second there has been a movement away from models set in “product space” to those set in “characteristic space”.

Heterogenous agent models. These have always been used with data that matches individuals to the products they purchased (micro data). However for the most part the I.O. researcher has to get by with market level data on quantities, prices and market characteristics and an ability to obtain distributions of basic consumer attributes from government data sources (e.g. income and family size). Use of heterogeneous agent instead of representative agent models on market level data has two important advantages.

First they enable the researcher to control for differences in the distribution of consumer characteristics across markets. This, in trun, provides an ability both to combine data from different markets in a single estimation algorithm, and to predict demand in markets where existing goods have not yet been introduced. For example a recurrent finding in empirical work on demand is that consumers' price sensitivity depends on their incomes, and income distributions typically varies markedly across local markets (often counties in the U.S.A.). So a representative agent demand system approximation which provides an adequate fit to the demand in one county is unlikely to fit in another, and neither approximation is likely to do very well in predicting demand in a market in which the good (or goods) have not yet been introduced.

The second reason for the use of heterogenous agent models is the fact that they provide an immediate ability to analyze the distributional impacts of policy and environmental changes. This ability is particulary central to the analysis of regulated markets, as regulators are typically either elected officials or are appointed by elected officials, and the distributional impact of their actions is a determinant of voting outcomes.

Research on explicitly aggregating heterogenous agent models to obtain market outcomes dates back at least to Houthakker's (1955) classic paper. However explicit aggregation only lead to tractable forms for the market outcomes that needed to be fit to data if very particular functional forms were chosen for both the primitive micro functions and for the relevant distributions of consumer attributes. Advances in our computation abilities together with the introduction of simulation estimators (this dates to Pakes, 1986) were the enabling factors for the widespread use of heterogenous agent models. Together they allow the researcher to simulate market outcomes for different parameter values from any functional form for demand and any distribution of consumer characteristics that seems appropriate by simply taking random draws from the assumed and/or empirical distribution of consumer attributes, computing the consumption decision for each draw, and adding up the results.

Product vs. Characteristic Space. Models of demand where preferences were defined on products per se rather than on the characteristics of those products have two characteristics which made them particularly problematic for use in empirical I.O. We often have to deal with differentiated product markets with a large number of products (often fifty or more). The first problem is that even a (log) linear product level demand system would then have demand for each good being a function of fifty prices and an income term. As a result to estimate the demand *system* for the fifty goods we would need to estimate on the order of twenty five hundred parameters. We simply do not have data sets that can estimate that many parameters with any precision[2]. Second demand systems set in product space can not provide any indication of what demand would be for a new good before that good is introduced. As a result demand estimates in product space are of no help in analyzing incentives for product development.

The use of characteristic space circumvents (or at least greatly ameliorates) both of these problems. If consumers care only about the characteristics of each product (and not about the product per se), and there are, say, five important characteristics, then a linear model would depend on the joint distribution of preferences over those characteristics. If those preferences were, say, normal, that joint distribution would depend only on twenty parameters (five means and fifteen covariance terms). Knowledge of those twenty parameters would allow us to obtain all twenty five hundred own and cross price elasticities. Similarly if the researcher specifies the characteristics of a new good then the characteristic based models enable the researcher to compute demand for the new good at any set of prices.

Characteristic based models have been used in product placement problems in theoretical I.O. at least since the work of Hotelling (1929). More directly related to our interest was Lancaster's (1955) use of characteristic based models in demand analysis, and McFadden's (1974) incorporating characteristic based models into his analysis of discrete choice econometric models. Though the Lancaster/McFadden framework had great potential for I.O., their use raised two new issues.

First the number of characteristics needed to fully specify consumer products can be very large, too large to include them all in the specification and expect to estimate the parameters of interest with any precision (producer goods tend to be less problematic in this respect). Many of the product characteristics in consumer goods tend to have only small effects on demand patterns, but omitting them entirely caused problems. In particular "high quality" goods tend to contain many small features which, though perhaps individually unimportant, in sum have a noticeable effect on both aggregate demand and price. So ignoring them entirely causes both an "overfitting" and a simulatanaeity problem.

The overfitting problem is that in reasonably sized markets the model will predict a great

[2] A similar "too many parameter problem" motivated much of Gorman's (1959) ingenious work on multi-level budgeting. Gorman was careful to detail both the assumptions required for his approach and the reduction in the number of parameters it implied, and both limit the potential applications in I.O..

deal of precision in the estimates of the aggregate shares. This because the only source of error in the model is the multinomial sampling error, and when predicting aggregate shares this converges to zero with market size. The actual fit at the estimated parameter will not be nearly that good, and statistical tests will let you know there is something wrong with the model's sepecification. The simultanaeity problem is familiar from earlier demand system analysis. An omitted characteristic known to consumers will be known to producers and, in virtually any reasonable equilibrium, be correlated with price. What makes it a more difficult problem in our context is that in characteristic based models that disturbance is buried deep inside an aggregator function which does not have a simple analytic form. As a result the simple solutions used in more standard analysis, like instrumental variables, are not directly applicable.

The second problem with the early characteristic based models is that they made assumptions which forced the model to produce unrealistic own and cross price (and other characteristic) elasticities. The reason could be traced to the independence of irrelvant alternatives (or IIA) problem noted by McFadden(1981) in his work on micro data. The impact of the IIA problem on the models used for market level data was, however, much more dramatic. The early generation of aggregate models assumed that the utility an individual obtains from consuming a given good consisted of a mean utility and an individual specific deviation from that mean which became the disturbance. The disturbances were assumed to distribute independent across choices; indeed they were almost always assumed to be i.i.d. type II extreme value (or logit) deviates. So two individuals who had chosen different goods were assumed to have the same distribution of preferences over any other good (so if price increases to the goods they purchased induced both individuals to chose an alternative good, the probabilities of the different alternatives would be the same for the two individuals). Moreover since two goods with the same share will have the same mean utility, and these means are the only characteristic which differentiates goods, the model implies that goods with the same share *must* have both; (i) the same own price elasticity (and hence the same markup in a price setting model with single product firms), and (ii) the same cross price elasticity with every other good . "Must" here means that if the estimates did not have this property their had to have been an error in the computer program that generated the results. High quality goods with high prices often have similar shares to low quality goods with low prices, but no one believes the two types of goods have similar markups or similar cross price elasticities with other goods.

A paper by Berry Levinsohn and Pakes (1995, hencefoth BLP) provided practical solutions to these two problems. The early distributional assumptions were made primarily to get closed form solutions for the aggregate shares. BLP showed that simulation techniques and modern computers enables the use of much richer distributions of disturbance terms, and this largely eliminated the IIA problem. They also allowed for an unobserved product characteristics and provided a contraction mapping which provided the vector of product specific values for this characteristic as a linear function of the data (conditional on any given value of the parameter vector). Once the unobservable is obtained this way one can control for its impact on price by using any of a number of methods familiar from linear models (including instrumental variables). There have been a number of important papers which have extended these ideas in different ways but they are, for the most part, beyond the scope of this review (though I do return to a subset of them in discussing problems we are having in accounting for equilibrium prices below).

2.2 Cost and Production Functions.

As noted cost data are often propietary, and when they are not proprietary (as is often the case in regulated industries), they are frequently of questionnable quality (partly due to the incentives facing the firm reporting its costs). As a result there has been little recent work on cost functions per se.

In contrast there has been quite a bit of recent work using production functions. This work was largely motivated by two phenomena. First there was a noticeable increase in access to firm (or plant) level panels on production inputs and outputs (often sales and total costs of inputs rather than physical quantities). Much of this data has become available as a result of governmental agencies providing conditions under which researchers could access their data files. Second there has been a desire to analyze the efficiency (or productivity) impacts of a number of major changes in the economic environment. The latter include; the deregulation of important sectors of the economy (telecommunications, electric utility generation, ...), privatization programs (particularly in transition economies), and large changes in tariffs, health care policies, and economic infrastructure.

These two phenomena resulted in a focus on a particular set of substantive and technical issues. From a substantive point of view the availability of micro data provided an ability to distinguish between changes in (i) the efficiency of the output allocation among establishments, and changes in (ii) the productivity of individual establishments (and their correlates). For example, in an early use of the longtidudinal research data files of the US Bureau of the Census, Olley and Pakes (1996) find that the immediate impact of the break-up of A.T. & T. and the consequent partial deregulation of the telecoms sector on the telecommunications equipment industry was an increase in industry productivity due primarily to a reallocation of output to more efficient plants (largely resulting from a reallocation of the industry's capital to those plants). There was no preceptible immediate impact on the productivity of individual establishments.

Productivity is output divided by an index of inputs and the coefficient used to form the input index are usually obtained by estimating the "production function" relating output to the inputs used in producing it. It is important to realize that the output measure is often sales (perhaps divided by some aggregate price index) rather than physical output. Then what we are estimating is a "sales generating" function. One way to obtain its coefficients from a more detailed model is to assume the demand function has a constant elasticity form and incorporating any differences in the scale of demand across firms in the disturbance terms. However then one has to keep in mind that; (i) these coefficients are likely to change if the price elasticity changes (say due to changes in the quantities marketed by competitors), and (ii) before we can derive the welfare implications of any change in productivities we have to separate out price effects from quantity effects, and this will generally require a more detailed modelling effort.

The technical issues surrounding estimation of the coefficients used to form the input index resulted from two characteristics of the micro data; (i) there were large serially correlated differences in "productivity" among plants (no matter how productivity was measured), and (ii) there was lots of entry and exit (see Dunne Roberts and Samuleson, 1982, and Davis and Haltwinger, 1986). This led to worries about simultanaeity biases (or endogeneity of the input choices) on the one hand, and attrition biases (the endogeneity of exit) on the other. Firms whose productivity was positively effected by the changes grew, so input growth was related to productivity growth and the latter was the residual in the production function analysis. Moreover firm's whose productivities were negatively effected by the change floundered and often exited, so the exit decision was not independent of the residual either.

These problems are accentuated by the fact that most research projects were analyzing responses to large structural changes, changes where the relative rankings of firms and their identity changed rather dramatically. Partly as a result bias corrections based on familiar statistical models did not seem rich enough to account for the observed behavior. For example the use of fixed effects to account for the endogeneity of input choices was ruled out because of the changes in the *relative* productivities of different firms. Propensity score corrections for selection were ruled out because they assume a single index model and there clearly was more than one "index" (in I.O. terminology, "state variable") which determines exit behavior (all models included at least productivity and capital as state variables).

The alternative used was to build an economic models of input and exit choices to correct for the endogeneity and attrition biases in the estimates. The intial article of this sort was the Olley and Pakes (1996) article, but work by Levinsohn and Petrin (2003) and Ackerberg Fraser and Caves (2004), provide alternative estimation procedures which have also been extensively used. All these articles emphasize the need for computationally simple economic models that require only minimimal functional form and behavioral assumptions, and rely heavily on the recent econometric literature on semiparametric estimation techniques (see for e.g., Andrews,1994, Newey, 1994, and Chen,forthcoming, and the literature reviewed their). There are ongoing attempts to weaken the assumptions used further (for a review see Ackerberg et. al., 2006), but perhaps the most striking finding in the empirical results thus far is that, for a fixed data set, the estimates from the different estimation algorithms referred to above are most often quite similar.

Partly as a result attention is turning to the use of the estimates to analyze the correlates and timing of productivity growth (see for e.g. De Loecker, forthcoming). The hope is this, in turn, will provide a basis for a deeper empirical analysis of the issues surrounding productivity; a topic of wide interest to I.O., development, growth theory, and trade – and a topic which has been very difficult to analyze empirically in the past.

2.3 Equilibrium Assumptions.

Of the three "primitives" of static analysis, the "equilibrium" assumption (interpreted broadly enough to include issues related to the form of the game played), is the one where there has been least progress. Empirical work has relied heavily on a game in which; sellers set prices or quantities to maximize their current profits, purchasers are price takers and decide how many units of each good to purchase by comparing their current utility from different bundles of goods to the bundles' prices, and the equilibrium is Nash in the seller's price (or quantity) strategy. The reliance on this paradigm is, in large part, a result of the fact that it does so well in tracing out the cross-sectional distribution of prices; particularly in differentiated product consumer goods markets. The fact that this is true even in situations where there is good a priori reasons to think the assumptions of the model are inappropriate reflects the fact that the cross-sectional implications of these simple static models, especially of the discrete choice characteristic based models whose demand functions were described above, are likely to carry over to more complex environments.

The static Nash in prices model has equilibrium price equal to marginal cost plus a particular form for the markup. Marginal costs are typically modeled as a function of characteristics. We know from simple hedonic pricing functions that characteristics by themselves account for a large fraction of the variance in price. For example using the monthly BLS data underlying the consumer price subindex for TV's, Erickson and Pakes (2006) generate adjusted R^2's between 87 and 91% (depending on the month) from regressing log price on just four charactersics (the monthly cross sections average about 240 observations). This is higher than typically seen in I.O. studies (which probably reflects the quality of the BLS data), but R^2's for a regression of log price on even a small set of characteristics are frequently above .6. Moreover the economics of the static markup term from our simple models are both compelling and consistent with what we know about markups. Higher priced goods (which are typically higher quality goods) will be purchased by individuals who are less sensitive to price and as a result will have lower price elasticities and higher markups (at equilibrium prices). This, in turn, is the justification for the investments typically required to develop and produce the higher quality. The model also predicts that the goods located in a crowded part of the characteristic space will have larger price elasticites and lower markups, and that the prices of substitute products owned by the same firm will be higher.

On the other hand there are at least two phenomena which these static equilibrium assumptions seem less well suited for. First the paradigm does less well with shifts in the level (usually of the logs of) prices in a market over time. There is growing evidence that many prices are more "sticky" than our static models would rationalize, a fact which generated a large literature on exchange rate pass through in international trade and is of considerable interest to macroeconomics (see for e.g. Nakamura 2006, and the literature cited their). Also there are markets with clear evidence of phenomena which are inconsistent with a static pricing equilibrium including; "introductory" pricing patterns, price wars, periodic sales in retail outlets, prices noticeably below marginal cost in markets with learning by doing, and so on. All this points to the need for further work on dynamic pricing models, though the particular dynamic model that is relevant is likely to depend on the charactersitics of the market studied.

The easiest cases are those in which it is reasonable to assume that consumers are maximizing static utility functions and prices are dynamic for another reason; say because learning by doing or adjustment costs make today's price or quantity choice effect tomorrow's cost, or because strategic considerations make producer's future price choices depend on current price choices (as in many collusive models). Current price choices then effect both current profits *and* the distribution of future states, so the model requires a notion of consistency between producer's perceptions of the impact of price on future states and the evolution of those states, as well as of producer's perceptions of the prices of its competitors and the actual prices of those competitors. In this case the issues we face in formulating dynamic price setting models that are tractable enough to be used in estimation and policy analysis are similar to the issues faced in chosing investment strategies in the simpler models where price is a static control; see the next section and Benkard (2005) for an early applied example. In models in which consumers as well as producers solve a dynamic problem (as might be required for analyzing markets with durable, storeable, experience, or network goods) the analysis is harder. Then we need to explicitly consider the sense in which consumer's perceptions are consistent with the outcomes of future producer behavior and producer's actions are consistent with their perceptions of future consumer behavior, as well as each group's actions being consistent with the behavior of other members of the same group. Without special structure this increases the size of the state space and makes notions of consistency between perceptions of the evolution of the state variables and their actual evolution more difficult to obtain; for early applied examples which rely on approximations see Melnikov

(2001), Gowrisankaran and Rysman (2007), and Lee (2007)[3]. The complexity of the calculations required also calls into question the ability of consumers (perhaps also producers) to act in accordance with the model's assumptions; a topic we return to below.

Finaly we have *a priori* reason to doubt the static Nash in prices assumption in markets with small numbers of both buyers and sellers, in particular when sellers can sell to several buyers and buyers can purchase from several sellers, as is the case in many vertical markets. These are markets in which the buyer's returns from its actions often has a noticeable dependence on the actions of other buyers. Moreover there is frequently both direct and indirect evidence that the appropriate game form is more complex then that implicit in the Nash in prices assumption; for example the finding that the same goods are sold to different buyers at different prices.

The empirical analysis of how prices are actually formed in these markets is complicated by two facts. First the prices themselves are often implicit in proprietary contracts that researchers do not have access to. This makes it hard to provide reduced form evidence on the determinants of price. Second the theory literature does not seem to have an agreed upon framework for analyzing how these contracts are formed, and this makes the choice of an appropriate structural model difficult. On the other hand the prices and other features of these contracts determine both the split of any net profits from the vertical relationship (and therefore investment incentives) and the efficiency of the output allocation, and so is crucial to an understanding of how these markets work. I come back to this problem as my example of how using inference based on moment inequalities might enable empirical work to characterize the nature of competition in more complex market situations.

3 Dynamics.

Recall that in our simpler models we can solve for profits conditional on the state variables of each active firm (e.g. its capital stocks and/or the characteristics of its products) using the techniques developed to analyze the static equilibrium problem. The dynamic model is designed to analyze how this list of state variables, typically called the market structure, is likely to evolve over time. To do so we construct a dynamic game which mimics, to the extent possible, the situation faced by competitors in the industry of interest, and analyze an equilibrium of that game that is Markov Perfect in investment strategies (defined broadly enough to include entry and exit strategies).

[3]There has also been some work on analyzing dynamic demand without imposing a particular form for the pricing equation (e.g. Hendel and Nevo, forthcoming, and Crawford and Shum, 2006), and on analyzing pricing in models for durable goods where consumers can trade goods without incurring transaction costs see Esteban and Shum (2007).

Incumbents decide whether to exit by comparing continuation values to what they would earn were they to exit. If they continue they chose investment levels. Potential entrants enter if the expected discounted value of net cash flows from entering are greater than the cost of entry. Regularity conditions imply that only a finite set of market structures will ever be observed. So the equilibrium generates a finite state Markov chain in market structures. Given any initial location the current market structure will, in finite time, wander into a recurrent subset of these structures and once within this subset will stay within it forever.

In order to use this framework to analyze an industry's behavior we need: (i) estimators for the parameters needed for the dynamic analysis that do not appear in the static model (the impacts of different types of investments, including the costs of entry and exit), and (ii) an algorithm for computing equilibrium. I begin with a short description of the methods available to accomplish these tasks and then come back to a brief general discussion of the appropriateness of the framework.

3.1 Estimating Dynamic Parameters.

A firm's state is determined by its own state variables and the market structure. Since entry, exit, and investment decisions set likely future states, they are determined by the continuation values from those states. So if we knew the continuation values and the costs of the decisions up to the parameter vector, we would get predictions for these decisions conditional on that parameter vector. Estimates could then be found by finding that value of the parameters that makes predicted decisions "as close as possible" to observed decisions. The problem with this procedure is that the continuation values implied by a given value of the parameter vector are hard to compute. Several recent papers have provided ways of circumventing this computational problem. Their common denominator is the use of nonparametric techniques to estimate the continuation values that are *implicit in the data* from different points in the state space. There are at least two ways to do this.

We can average the continuation values actually earned from a given state. For example consider the case where we know (or have estimated) the parameters of the profit function and have a good idea of the discount rate, so all we need to estimate are the parameters of the entry and/or exit cost distributions. Focusing on entry, assume that there is one potential entrant in each period who draws from an i.i.d. entry cost distribution and enters if its expected discounted value of net cash flows is greater than its draw. Use the profit function parameters, the discount rate, and the costs of investment and exit to compute the average of the realized discounted net returns of firms that had been active at the entry state in the past. If that state has been visited often enough, this will be a consistent estimate of the expected discounted net returns from enterring at that state. So the fraction of times that we observe entry at that state is a consistent estimate of the probability that the costs of entry are less than this value of the discounted net returns. As we vary the discounted net cash flows across market structures these observed fractions will (at least in the limit) trace out the sunk cost distribution we are after (for details see Pakes, Ostrovsky and Berry, 2007). Note that this procedure is implicitly using non-parametric estimates of the transition probabilities from one market structure to the next to obtain its estimates of continuation values.

Alternatively one could start with nonparametric estimates of all policies (entry, exit, and investment policies) at each state. The estimated policies and the structural model could then be used to simulate continuation values at each state for different values of the parameters of interest. A consistent estimator for the parameter vector could then be formed from the value of that vector that makes the models' implications for the policies as close as possible to the nonparametric estimates of those policies (for details see Bajari, Benkard, and Levin, 2007).

There are related specification and small sample issues associated with these estimators. The specification issue is that currently both estimators assume that, apart from a serially uncorrelated disturbance, the econometrician can observe all the state variables that agents base their decisions on. It is this assumption which enables the econometrician to obtain consistent estimates of the firm's perceived distribution of future states from the observed outcomes from each market structure (at least for the recurrent class of points). This, together with the profit function, allows us to compute continuation values and optimal policies at those market structures[4].

Though the assumption that all serially uncorrelated state variables are observed by the econometrician has been used extensively in the empirical literature on single agent dynamic programming problems (for early influential examples see Rust,1987, and Hotz and Miller,1993), it is an assumption which left empirical researchers uneasy, and it is likely to be even more problematic in the current context. This because the fact that the small sample properties of the nonparametric component of the estimators deteriorate rapidly with the number of state variables, combined with the fact that the cardinality of the state space is now determined by the state variables of several interacting agents, frequently induces researchers to ignore variables which, though perhaps not central to the problem being analyzed, might well jointly have significant impacts on the relevant continuation values (this is similar to the logic that lead us to worry about unobserved characteristics in demand estimation for consumer goods, see section 2.1).

On the other hand we do obtain consistent estimates of some of the objects of interest

[4]Note that under these assumptions this procedure selects out the (generically) unique set of equilibrium actions that are consistent with play on the recurrent class; for details see Pakes, Ostrovsky and Berry, 2007, theorem 1.

regardless of the presence of serially correlated unobserved states, and this implies that we shold be able to incorporate such state variables into the analysis in future research. More precisely the first estimation technique provides estimates of continuation values, and the second provide estimates of policies, which will be consistent estimates of the averages of these objects conditional only on the observed states *regardless* of whether there are state variables the econometrician can not condition on (at least on the recurrent class of points). If there were unobserved state variables and the econometrician conditioned only on an observed state (but one that was visited repeatedly), then the continuation value (or policy) the econometrician estimates would be the averages of the continuation values over the distribution of the unobserved states conditional on the observed states. The consistency of the average value or policy estimates are of interest in themselves. For example an antitrust authority may have some idea of entry costs and therefore could use an estimate of the expected discounted value of an entrant after a merger to evaluate the likelihood of post merger entry. Moreover the relationship between the distribution of continuation values from a given observed state, and the distribution of policies from that state, contains information on the distribution of unobserved states conditional on the oberved state which is not being utilitzed by either of the two estimation techniques.

Coming back to small sample estimation problems, both estimators require nonparametric estimates of objects that take on different values at different states (in one case of policies and in the other of continuation values), and if we discreetize the state variables the cardinality of the state space typically increases geometrically in the maximum number of firms ever active and exponentially in the number of state variables per firm (see below). The precsion of estimates of continuation values from a given state depends also on the number of visits to states that communicate with the given state, and eventually each sample path will wander into a recurrent class of points and stay within that subset. So it is not appropriate to consider the ratio of number of states to the number of observations as an indicator of the extent of the small sample problem. Moreover to some extent one can ameliorate these problems by appropriate choice of estimation algorithms (for a discussion see Pakes, Ostrovsky, and Berry, 2007). However as the number of states in the recurrent class increase (sample size staying constant) the estimates are likely to become unreliable.

So, at least with currently available techniques, applied researchers still face a trade-off between; (i) using a fully dynamic model with only a rough partition of the state space (and there are creative ways to do this, see, Dunne, Klimek, Roberts and Xu, 2006, Collard-Wexler, 2006, and Ryan, 2006), and (ii) estimating the parameters of a far more detailed static equilibrium model and then using the heuristic of a two-period game to gain some understanding of dynamic phenomena. The two period game heurisitic is probably only appropriate in an environment which has been stable for a period of time which is long enough to think that agent's would have changed their decisions if it was profitable to do so, but it is often used as a reduced form summary of the data when truly dynamic models are too difficult to implement (for a dicussion in the context of entry games see Pakes, 2004, and Berry and Reiss, forthcoming.). The terms of this tradeoff will change if we learn how to incorporate serially correlated unobserved state variables into the analysis.

3.2 Computation.

Pakes and McGuire (1994) adapt standard iterative procedures, procedures that are similar to those used in single agent dynamic problems (see for e.g. Rust, 1994, and the literature cited their), to the problem of computing polices and values for Markov Perfect equilibria. Estimates of values (and/or policies) for each incumbent and potential entrant at every point (market structure) in the state space are kept in memory. An iteration cylces through those points and at each point updates the values and policies associated with all incumbents and potential entrants at that point. The computational burden of the technique is proportional to the multiple of, (i) the number of points at which we have to calculate continuation values, and (ii) the number of elements in the summation or expectation which determines those continuation values. Since values and policies are exchangeable in the states of competitors, the number of point at which we have to calculate continuation values will grow geometrically in the maximum number of firms ever active, and without further restrictions, exponentially in the number of state variables per firm.

As a result despite the continual increase in speed and memory of computers, the computational burden of our models will limit the detail we can include when computing equilbria to dynamic problems for some time to come. There have, however, been a number of algorithmic innovations for computing the solution to dynamic games that can provide significant help. I will review two of these that have been used in the context of the models we are discussing. One uses stochastic approximation to approximate the equlibria (Pakes and McGuire,2001; see Bertsekas and Tsikilis, 1995, and the literature cited their for an introduction to stochastic approximation techniques), the other computes equilibria to continuous (rather than a discrete) time models (Doraszelski and Judd, 2006).

The stochastic algorithm uses simulation to (i) approximate the sum over future states that determines continuation values, and (ii) to search for a recurrent class of points generated by the equilibrium. The algorithm is iterative; it stores continuation values and/or policies in memory and updates them every iteration. The iterations are, however, "asynchronous"; instead of cycling through all points at each iteration they select out a single point per iteration. Policies at the selected point are chosen to maximize the expectation of the continuation values in memory at that point. Random draws from the primitives given these policies are then used to both (i) determine the next state visited, and (ii) to update the continuation values in memory. The update of continuation values *acts as if* the draws on competitors' states and on variables which evolve exogenously were simulation draws from the integral determining the true continuation values. I.e. the value in memory associated with the outcome of these draws is computed and averaged with the values of draws taken from the same state at previous iterations to form the new estimate of the state's continuation value. The iterations are stopped periodically to test whether the equilibrium conditions are satisfied by the values in memory, and the algorithm is stopped if they are. I do not know conditions which insure convergence of the algorithm for games which are not zero sum, though convergence has not been a problem for the examples I have dealt with.

Coming back to the issue of computational burden, note that by substituting simulation for summation in computing updates of continuation values, the algorithm changes the burden of computing those values from being exponential in the number of states to being linear in the number of firms. However use of the random draws also introduces "sampling error" in the estimates of the continuation value, and this will only be averaged out if we visit the point (and hence simulate the draw) many times. Fortunately the precision of the estimate for a given number of simulation draws does not depend directly on the number of state variables in the problem, and hence need not increase in the number of those state variables. Note also that the asynchronous steps taken by the algorithm will eventually wander into a recurrent subset of the state space and once in that subset will stay their forever. So after a finite number of iterations the algorithm will only need to update points in this recurrent subset. The size of the recurrent class depends on the economic primitives of the problem at hand and need not grow in any particular way in the number of state variables. In the problems I have worked with the number of points in the recurrent class tends to grow linearly in the number of state variables. Moving from exponential (or geometric) to linear growth in both the number of states and in the sum required for continuation values can generate significant computational savings.

Doraszelski and Judd's (2005) continuous time model focuses on reducing the burden of computing the summation over future states needed for continuation values. They change the timing conventions for the model and assume that there are independent continuous time stochastic processes that generate times when events occur that change the state of the industry. By making the time interval short enough the probability of observing more than one event per interval can be made negligibly small. A limiting argument is used to show that to determine continuation values we need only sum over terms which determine the probability of each possible event occuring times a sum which determines the expectation of the outcome conditional on that event ocurring.

So the continuous time model specifies; (i) hazard rates for each possible event, and (ii) the transition probabilities conditional on the event occurring. In the Ericson Pakes (1995) model with industry specific shocks there are four possible reasons for a change; the investment of one of the incumbents produces a new outcome, the value of the outside alternative increases, an incumbent decides to exit, or a potential entrant decides to enter. Both the hazard rates of these events and/or the transitions conditional on the event occuring can be made a function of investment choices. If there are n incumbents and a single potential entrant there are $2n+2$ possible events that can occur; an investment outcome for an incumbent, a change in an incumbent's exit value, an entrant appearing, and a change in the value of the outside alternative. If, in addition, there are K possible outcomes for each possible event should it occur the summation determining continuation values involves $(2n+2)K$ terms (which again is linear in the number of firms). The continuous time model does not deal with the curse of dimensionality in the number of states, but it could be combined with the asynchronous aspects of the stochastic algorithm to do so.

The different algorithms require somewhat different modeling assumptions and are therefore likely to be more useful in alternative settings. This plus the fact that computatinal burden remains a problem in many examples of interest, particularly in the many extensions to the framework designed to align the framework more closely to empirical work, provides good reason for the computational research which is now gathering momentum (see Doraszelski and Pakes, forthcoming, and the literature cited their).

3.3 Discussion.

Many of the estimation and computational papers that have enabled us to use a dynamic framework which can accomodate the complexities of actual data sets are quite recent and, as noted above, their are still technical limitations to where they can be used. As a result it is only in the last few years that we have begun to see empirical work based on dynamic models and it is probably too early to generalize on how well they do in helping us analyze empirical phenomena (though the early work does indicate that there is at least a rough level of detail at which we do quite well; see for e.g. Benkard,2005). The dynamic framework outlined above has been used to numerically analyze situations that are too complex to admit analytic results (though not as complex as the situations we find in actual data), and this analysis had made it clear that the framework outlined above does provide a useful way of unravelling dynamic incentives.

However the numerical analysis has also made it clear that the dynamic models can often have several equilibria, that the calculations required to compute any one of them are quite complex, and that we often do not have ways of insuring we compute all of them (see Besanko, et. al, 2007). The complexity is especially telling in the extensions of the framework designed to account for consumer (along with producer) dynamics, and those required for games with asymmetric information. It leads to a question of whether agents can actually compute equilbrium strategies, and how to modify our behavioral assumptions to account for any discrepancies. Related issues have lead to a theory literature exploring the relationship between alternative learning mechansims and equilibrium behavior (see Fudenberg and Levine,1993, and the literature cited their). One question, then, is when we might do better by approximating behavior with a learning process than with the implications of an equilibrium notion, and what the differences might be.

To learn from past outcomes there has to be a sense in which the primitives underlying play in the market we are studying have been reasonably stable over some period of time. If that period is long enough we might be willing to assume that the learning process has converged to a limit or rest point; a point at which the perceptions of agents are justified by their objective implications (the perceptions are typically either on the likely outcomes of the agent's actions, or on the likely play of their competitors). The rest points to a learning process generally will satisfy some of the Markov Perfect equilibrium conditions though not all of them; see Fershtman and Pakes (2007) for a discussion of this in the context of reinforcement learning and dynamic games with asymetic information, and Esponda (forthcoming) for a discussion in the context of self-confirming equilibria (Fudenberg and Levine, 1999) in auctions. These papers also consider the closely related question of the testability of equilibrium assmptions; and this should help sort out the implications of equilibrium play we may want to place more weight on. Finaly recall from the discussion of multiple equilibria that another advantage of substituting the responses from a leaning process for equilibrium responses is that they allow us to analyze counterfactuals when multiple equilibira are possible.

It is not as clear how to use learning processes in situations where primitives are changing in fundamental ways as then we would need to specify the perceptions of agents when the new situation unfolds (as well as the learning process). Research on which learning process better approximates behavior in different situations, and on how agents form their initial perceptions, would be helpful.

4 Moment Inequalities.

Estimators based on moment inequalities have a potential for alleviating several of the problems outlined above. This is largely because they can be obtained under weaker assumptions then alternative estimation procedures, and once obtained they can be used to investigate which of the possible more detailed assumptions are, and which are not, at odds with the data. To motivate the discussion consider the following static problem.

Assume that we can use estimates of demand and/or cost primitives to construct profits conditional on at least a subset of the decisions we want to analyze, say (d_i, d_{-i}), and any additional parameters that could not be estimated with the tools outlined in section 2. Now assume that: (i) agents expected the choices they made to lead to returns that were higher than the returns the agents would have made from an alternative feasible choice, (ii) that our model and data are rich enough to provide an "adequate" approximation to both the profits that were earned *and* to the profits that *would have* been been earned had the agent made an alternative feasible choice, and (iii) that agents expectations are not "too much at odds" with what actually happens. Then at the true value of the unknown parameter vector we would expect that, at least on average, the difference between the profits the agents did earn and those that would have been earned had the alternative feasible decision been made should be positive. This is an inequality which can be used as a basis for inference.

To formalize this approach we need more precise definitions of the word "adequate" and the phrase "too much at odds". In particular it would suffice to specify: (i) measures of the profits that resulted from the agent's decision and that would have resulted from a feasible alternative decision, (ii) the relationship between these measured profit differences and the profit differences that underlie the agent's expectations, and (iii) the relationship between the agent's expectations and the sample averages that arise from the actual data generating process. There have been two quite different approaches to filling in the required details, and the goal of this section is to clarify the assumptions underlying each. The next section uses the techniques described here in an empirical and numerical analysis of contracts in one buyer-seller network. Section 6 provides a Monte Carlo analysis of the robustness of the two estimators when applied to our buyer-seller network problem.

The first approach, which dates back to the literature on entry (see Tamer,2003 , Ciliberto and Tamer,2006, and Andrews Berry and Jia,2006), specifies the probability that the observed actions constitute a Nash equilibrium for a particular value of the parameter of interest, say θ. It then computes the difference between; (i) the estimate of the probability that the Nash conditions are satisfied at that θ conditional on the observed determinants of returns, and (ii) the observed frequency of the event conditional on these same variables. Since the Nash conditions are necessary conditions for the assumed equilibrium, any value of θ that makes this difference positive is accepted as an estimator. The second approach, which dates to Pakes, Porter, Ho and Ishii (2006), suggests computing, for each value of θ, the sample average of the difference between the observable part of the actual realized returns and the observable part of returns that would have been earned had an alternative choice been made. They then take, as an estimate of θ, any value that makes that difference non-negative. Note that both approaches require a "structural" model for the returns that would have been earned had an alternative decision been made, and neither requires a model which is detailed enough to single out a unique equilibrium (a fact which is particularly relevant for the more complex market situations that we are having trouble analyzing).

I begin with the assumptions that are common to the two approaches. Throughout I will assume that the relevant model delivers a parametric form for the return function conditional on all decisions, though, at least in principal, non-parametric functions could often be substituted without affecting the basic logic of the discussion.

4.1 Common Assumptions.

The condition that agents expect their choice to lead to higher returns than alternative feasible choices is formalized as follows. 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 what we require is

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

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

Two points about C1 are central to the advantages of both approaches. First there are *no restrictions* on the choice set; for example D_i could be the space of all bilateral contracts (or of all exclusive deals), and d_i may be at a boundary of that choice set. Alternatively d_i could be a vector whose components define the characteristics of a product (say the location and size of a retail outlet). Second, C1 is a necessary condition for a Nash equilibrium (indeed it is necessary for the weaker notions of equilibrium we consider below). As a result were we to assume equilibrium behavior C1 will be satisfied regardless of the equilibrium selection mechanism; indeed if one is careful with the econometric implementation of the approaches outlined below the equilibrium selection mechanism can be allowed to differ across data points. Also keep in mind 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 restrictions will be required, those restrictions will differ between the two estimation algorithms.

Both approaches need a model capable of predicting what expected profits *would be* were the agent to deviate from its observed choice. This, in turn, requires a model of what the agent thinks that $\mathbf{d}_{-i}$ and $\mathbf{y}_i$ would be were it to a change its own decision. For example one component of $\mathbf{y}_i$ in the buyer seller network problem is the price the buyer charges to consumers when it resells the seller's products, and that will typically depend on the which sellers contract with which buyers. So when a buyer considered whether to reject a contract offered by a seller which it in fact had accepted, the buyer knew that if it had rejected the seller's offer the equilbrium 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 preception 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.

The model for how the agent thinks $(\mathbf{y}_i, \mathbf{d}_{-i})$ are likely to respond to changes in d_i is likely to depend on other variables, say $\mathbf{z}_i$, which I will assume are "exogenous", i.e. variables that the agent thinks they will not change if the agent changes its decision. Condition 2 formalizes this assumption.

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

Note that if the game is a simultaneous move game then $d^{-i}(d', \mathbf{z}_i) = \mathbf{d}_{-i}$ and there is no need for an explicit model of reactions by competitors (this explains the difference between our C2 and Assumption 2 in Pakes et. al., 2006).

If we let $\Delta\pi(d_i, d', d_{-i}, z_i) \equiv \pi(d_i, d_{-i}, y_i), -\pi(d', d^{-i}(d', z_i), y(z_i, d', d_{-i}))$, where d' is any alternative choice in D_i, then C1 and C2 together insure

$$\mathcal{E}[\Delta\pi(d_i, d', \mathbf{d}_{-i}, \mathbf{z}_i)|\mathcal{J}_i] \geq 0, \quad \forall\, d' \in D_i. \tag{1}$$

Equation (1) 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

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

- the relationship between our constructed profit function and our observable measures of the determinants of those profits on the one hand, and the $\pi(\cdot, \theta)$ and (z_i, d_i, d_{-i}) that appear in the theory on the other.

These are the two aspects of the problem which differ across the approaches I will describe.

4.2 The Full Information, No Errors, Approach.

The first approach to going from equations (1) to inference on θ_0 is a generalization of an approach that emerged from the literature on entry models (see Tamer,2003, and more recently Ciliberto and Tamer,2006, and Andrews, Berry, and Jia,2006). I begin by introducing the assumptions it uses and sketching an estimation algorithm those assumptions imply. Later I come back to discussing the appropriateness of those assumptions for the IO models considered in this essay.

The relationship between the data generating process and the agents' expectations assumed in this literature is that

$$FC3: \quad \forall d \in D_i, \quad \pi(d, d_{-i}, z_i, \theta_0) = \mathcal{E}[\pi(d, d_{-i}, z_i, \theta_0)|\mathcal{J}_i].$$

I.e. it is assumed that all agents know both the decisions of their competitors and the realization of the exogenous variables that will determine profits, when they make their own decision. FC3 *rules out* asymetric and/or incomplete information, and as a consequence, all mixed strategies[5].

To complete the specification we need an assumption on the relationship between the variables we measure and the variables that enter the theoretical model. This approach assumes

FC4. $\pi(\cdot, \theta)$ *is known.* $z_i = (\nu^f_{2,i}, z^o_i)$. $(d_i, d_{-i}, z^o_i, z^o_{-i})$ *are observed, and* $(\nu^f_{2,i}, \nu^f_{2,-i})|_{z^o_i, z^o_{-i}} \sim F(\cdot; \theta)$, *for a known function* $F(\cdot, \theta)$.

FC4 assumes there are no errors in our profit measure; that is were we to know $(d_i, d_{-i}, z_i, z_{-i})$ we could construct an exact measure of profits for each θ. However a (possibly vector valued) component of the determinants of the profits (of the z_i) is not observed by the econometrician (our $\nu^f_{2,i}$). Since FC3 assumes full information, both $\nu^f_{2,i}$ and $\nu^f_{2,-i}$ are assumed to be known to all agents when they make their decisions, just not to the econometrician. FC4 also assumes that there is no error in the observed determinants of profits (in the z^o_i) and that the econometrician knows the distribution of $(\nu^f_{2,i}, \nu^f_{2,-i})$ conditional on (z^o_i, z^o_{-i}) up to a parameter vector to be estimated.

Substituting FC3 and FC4 into equation (1) we obtain

$$Model\ F: \ \forall d' \in D_i, \ \ \Delta\pi(d_i, d', d_{-i}, z^o_i, \nu^f_{2,i}; \theta_0) \ \geq 0; \quad (\nu^f_{2,i}, \nu^f_{2,-i})|_{z^o_i, z^o_{-i}} \sim F(\cdot; \theta_0). \tag{2}$$

To insure that the model assigns positive probability to the condition that

$$\forall d' \in D_i, \ \ \Delta\pi(d_i, d', d_{-i}, z^o_i, \nu^f_{2,i}; \theta) \ \geq 0$$

[5]Unless there was a unique equilibrium, and it was a mixed strategy equilibrium. In this case we could compute the only distribution of $\mathbf{d}_{-i}$ that is consistent with equilibrium play and use it to form expecations. However this would increase the computational burden of the estimator significantly. As stated $C2$ also rules out the analysis of sequential games in which an agent who moves intially 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).

for some θ and all i (as is assumed by the model), we need further conditions on $F(\cdot)$ and/or $\pi(\cdot)$. The additional restrictions typically imposed are that the profit function is additively separable in the unobserved determinants of profits, that is that

$$\textit{Restriction } RF^{as}: \ \ \forall d \in D_i, \ \ \pi(d, d_{-i}, z_i^o, \nu_{2,i}^f) = \pi^{as}(d, d_{-i}, z_i^o, \theta_0) + \nu_{2,i,d}^f, \tag{3}$$

and that the distribution $\nu_{2,i}^f \equiv \{\nu_{2,i,d}^f\}_{d \in D_i}$, conditional on $\nu_{2,-i}^f$, has full support. It is important to realize that the additive separability in equation (3) *can not* be obtained definitionally. We could form the expectation of profits conditional on observables and define an error which is the difference between realized profits and that expectation. However that error would be mean independent of the decision itself. The error in equation (3) is not mean independent of d_i; indeed it is the correlation between $\nu_{2,i}$ and d_i that is the major reason for using the inequality estimator instead of a more standard estimation procedure.

Though this model does specify a parametric distribution for the $(\nu_{2,i}^f, \nu_{2,-i}^f)$ conditional on the observables, it is not detailed enough to deliver a likelihood. This because the conditions required by the model can be satisfied by multiple tuples of (d_i, d_{-i}) 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 variable on the other.

However we can check whether the conditions of the model are satisfied at the observed (d_i, d_{-i}) for any $(\nu_{2,i}^f, \nu_{2,-i}^f)$ and θ, and this, together with $F(\cdot, \theta)$, enable us to calculate the probability of those conditions being satisfied at any θ. Since these are necessary conditions for the choice to be made, when $\theta = \theta_0$ the probability of satisfying them must be greater then the probability of actually observing (d_i, d_{-i}). Indeed we could compute the probability that a subset of those conditions were satisfied, and this would have to be greater than actually observing (d_i, d_{-i}). If for that θ we could, in addition, check whether (d_i, d_{-i}) are the only values of the decision variables to satisfy the necessary conditions for all $(\nu_{2,i}^f, \nu_{2,-i}^f)$, we could provide a lower bound to the probability of actually observing (d_i, d_{-i}) given θ. 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 (2) (with a restriction like that in equation 3) is satisfied at a particular (d_i, d_{-i}) for a given θ to be

$$\overline{P}\{(d_i, d_{-i}) \,|\theta\} \equiv Pr\{(\nu_{2,i}^f, \nu_{2,-i}^f) : (d_i, d_{-i}) \text{ satisfy equation (2)} | z_i^o, z_{-i}^o, \theta\},$$

the analogous lower bound to be

$$\underline{P}\{(d_i, d_{-i}) \,|\theta\} \equiv Pr\{(\nu_{2,i}^f, \nu_{2,-i}^f) : \text{only } (d_i, d_{-i}) \text{ satisfy equation (2)} | z_i^o, z_{-i}^o, \theta\},$$

and the actual likelihood of (d_i, d_{-i}) for a given θ to be

$$P\{(d_i, d_{-i})|\theta\} \equiv Pr\{(d_i, d_{-i}) \;|z_i^o, z_{-i}^o, \theta\}.$$

As noted the likelihood can not be calculated without a mechanism which selects among multiple equilibria but we do know that for the true selection mechansim when $\theta = \theta_0$

$$\overline{P}\{(d_i, d_{-i}) \;|\theta_0\} \quad \geq P\{(d_i, d_{-i})|\theta_0\} \quad \geq \underline{P}\{(d_i, d_{-i}) \;|\theta_0\}.$$

Let $\{\cdot\}$ be the indicator function which takes the value one if the condition inside the brackets is satisfied and zero elsewhere, $h(\cdot)$ be a function which only takes on positive values, and $E(\cdot)$ provide expectations conditional on the process actually generating the data (including the equilibrium selection process). Then the model's assumptions imply that

$$E\Big((\overline{P}\{(d_i, d_{-i}) \;|\theta\} - \{d = d_i, d^{-i} = d_{-i}\})h(z_i^o, z_{-i}^o)\Big) \tag{4}$$
$$= (\overline{P}\{(d_i, d_{-i}) \;|\theta\} \;-\; P\{(d_i, d_{-i})|\theta_0\})h(z_i^o, z_{-i}^o) \;\geq 0 \;\; at \;\; \theta = \theta_0.$$

In cases where we can construct an estimate of $\underline{P}\{\cdot\}$, an analgous moment condition can be constructed from $P\{(d_i, d_{-i})|\theta_0\} - \underline{P}\{(d_i, d_{-i}) \;|\theta_0\}$.

The estimation routine constructs unbiased estimates of $(\underline{P}(\cdot|\theta), \overline{P}(\cdot|\theta))$, substitutes them for the true values of the probability bounds into these moments, and then accepts values of θ for which the moment inequalities are satisfied. 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 $FC4$, and for each random draw check the necessary conditions for an equilibrium, i.e. the conditions in equation (2), at the observed (d_i, d_{-i}). Estimate $\overline{P}(d_i, d_{-i}|\theta)$ 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 $\underline{P}(d_i, d_{-i}|\theta)$ by the fraction of the draws for which (d_i, d_{-i}) is the only such value.

If we were analyzing markets with N interactive agents each of which had δ possible choices and we used ns simulation draws on $\{\nu_{2,i}^f\}_{i=1}^N$ to estimate $(\overline{P}\{\cdot \;|\theta\}, \underline{P}(\cdot|\theta))$, then for each market and each θ evaluated in the estimation routine we need to evalute up to $ns \times \delta \times N$ inequalities if we only obtained estimates of $\overline{P}\{\cdot \;|\theta\}$, and we need to evaluate up to $ns \times \delta^N$ inequalities if we also evaluated $\underline{P}(\cdot|\theta)$. This can be computationally expensive, particularly when the functions determining $\mathbf{d}_{-i} = d^{-i}(\mathbf{d}_i, \mathbf{z}_i)$, and $\mathbf{y}_i = y(\mathbf{z}_i, \mathbf{d}_i, \mathbf{d}_{-i})$ are difficult to calculate. In particular inequality estimators are often used to characterize the first stage decision in a multistage game. In this case to obtain $\mathbf{y}_i = y(\mathbf{z}_i, \mathbf{d}_i, \mathbf{d}_{-i})$ (and/or

$d^{-i}(\mathbf{d}_i, \mathbf{z}_i)$) from a structural model we typically work backwards from the computed equilibrium conditions for the last stage of that game, and this can be computationally prohibitive (as is the case with the empirical example below)[6].

4.2.1 Discussion: Uses of the Algorithm and Its Assumptions.

If there is a component of d_i which is discrete, which is the leading case for applications of this approach, there is a logical problem with this framework, as then there is no guarantee that the model in equation (2) has an equilibrium in pure strategies for any given θ and all points in the support of $F(\cdot)$. [7] In what follows I am simply going to assume that this existence condition is satisfied.

There are at least two ways in which to rationalize the estimation algorithm outlined in this subsection, and I want to comment on the assumptions implicit in both of them. Empirical applications have predominantly (indeed perhaps even exclusively) used this algorithm to provide a reduced form summary of the relationship between (d_{-i}, z_i^o) and the profitability of agent i. The reason for allowing for $\nu_{2,i}^f$ in this context is that the researcher is interested in the relationship between $\pi_i(\cdot)$ and (d_{-i}, z_i^o) *conditional on* unobservable determinants of profits, particularly those that are corrleated with d_{-i}. For example in the entry models the usual focus is on the relationship of profitability to the number of entrants, and the researcher wants to understand this relationship conditional on unobserved, as well as observed, market characterisitics.

What additional assumptions are needed to obtain reduced form parameters which can be interpreted in this way? Say that the model that described how agent's made their decisons were the model in equation (2), with a profit equation built up from a rich set of demand and cost primitives (for e.g., one of those described in section 2), and for simplicity assume that the restriction in equation (3) is appropriate. Now consider regressing the $\pi^{as}(z_i^o, d_i, d_{-i}, \theta_0)$

[6] There are a number of ways to reduce the computational burden. We noted that we could check a fraction of the inequalities in equation (2) for each agent; though this is likely to increase the size of the identified set. Use of variance reduction techniques should increase the precision of the estimates of $(\underline{P}(\cdot|\theta), \overline{P}(\cdot|\theta))$ for a given ns. Alternatively one might be able to formulate the estimation problem as a minimization problem subject to a set of constraints, as in Judd (2007), and this might reduce the computational burden. At least to date, however, the computational burden of this technique has been large enough to deter its use in a number of applications.

[7] I.e. for any given $(z_i^o, z_{-i}^o, \nu_{2,i}^f, \nu_{2,-i}^f)$, at the true θ_0 there may not be a $(d_i, d_{-i}) \in D_i \times D_{-i}$ which satisfies the necessary conditions in equation (2). Since we actually observe play that is assumed to satisfy these conditions, this implies that the support of $(\nu_{2,i}^f, \nu_{2,-i}^f)$ is restricted in a complicated parameter depedent way. We can insure existence in discrete games by allowing for mixed strategies (see Bajari,Hong and Ryan, 2006, for more on this), but the use of mixed strategies implies that d_{-i} is not known with certainty, and so contradicts the assumptions of the model.

onto a polynomial "reduced form" function of (z_i^o, d_i, d_{-i}), say $r^f(\cdot)$, so that

$$\pi(d_i, d_{-i}, z_i, \theta_0) = \pi^{as}(d_i, d_{-i}, z_i^o, \theta_0) + \nu_{2,i,d}^f \equiv r^f(d_i, d_{-i}, z_i^o, \theta_0^*) + \nu_{2,i,d}^f + \nu_{1,i,d}^f, \tag{5}$$

where the $\nu_{1,i,d}^f$ is the error from that regression and hence is, by contruction, mean independent of the terms in the polynomial, and θ_0^* are the reduced form polynomial coefficients of interest. The difference between this model and the model in equation (3) is that this model contains an unobservable which is mean independent of the observable determinants of profits (including the d_i), whereas the model in (3) does not. Since the $\{\nu_{1,i,d}^f\}$ determine profits, we have to account for them when checking the Nash conditions in equation 2; (else we will not obtain an inconsistent estimate of θ^*). Accounting for them will be difficult since they have distributions which should be derived from the true unknown profit functions, and are likely to exhibit both correlation across choices and dependence on the observed z_i^o.

The second way of rationalizing the estimation algorithm described in this subsection is more in tune with the rest of this essay. Instead of relying on a reduced form approximation to profits, specify demand and cost primitives directly and construct the implied profit function conditional on (d_i, d_{-i}) and any parameters one is unable to estimate from more direct estimation procedures. We then have to consider the appropriateness of $FC3$ and $FC4$. Though the full information assumption in $FC3$ is not likely to be particularly appealing as a general approximation to how decisions are actually made, it might be appropriate for characterizating the "rest point" to an environment which is sufficiently stable. I.e. there might be good reasons to think FC3 holds in a market where there are no serially correlated changes in the important determinants of profits and the actual decisions of the participants have not changed over a reasonable period of time. As noted above this is often the environment that rationalizes the use of two stage games in empirical work, and hence is of some importance.

This bring us to $FC4$, and there are two reasons why one might worry about it. First there is the question of the robustness of the results to the assumption on the distribution of $(\nu_{2,i}^f, \nu_{2,-i}^f)$ conditional on (z_i^o, z_{-i}^o). As explained below the $\nu_{2,i}^f$ generate a selection problem and, as a result, even if we assume the additively separable model in equation 3 and are reasonably sure about the specification for its conditional mean, the estimator will not be consistent unless more detailed properties of the distribution, in particular expectations conditonal on d_i, are correct. We provide some Monte Carlo results on the impact of the choice of the distribution function for the buyer-seller network problem below.

The second reason for worrying about $FC4$ is that it is unlikely that we can specify primitives accurately enough to be able to ignore all sources of error in our measures of profits, particularly for the profits from the counterfactuals, and it will be difficult to allow for such errors. Assume, for example, that there was an error in our model for $r^f(\cdot)$, so that the construct we observe, say $r^o(\cdot) = r^f(\cdot) + \nu_1^m$. Now to proceed in a manner analagous to what is done without the ν_m^1 error, for any given draw on ν_2^f we would have to determine whether

$$\Big(r^o(d_i, d_{-i}, z_i^o, \theta_0) - \nu_{1,i,d}^m\Big) + \nu_{2,i,d}^f,$$

satisfies the equilibrium conditions in model (5). Even if we are willing to assume a parameteric distribution for $\nu_{1,i}^f$, this will be difficult to do. This because though an assumption on the distribution of the measurement error would enable us to construct the distribution of the observed profits conditional on the structural profits, what we need to check the equilibrium condtions is the distribution of structural profits given observed profits. To find that conditional distribution we would, at least in general, need to solve an integral equation.

There is one restriction on either the $\{\nu_1^f\}$ errors in the reduced form model of equation 5), or on the $\{\nu_1^m\}$ in the structural model above, which does rationalize the full information no error algorithm. It would suffice that

$$\textit{Restriction } RF^r: \quad \forall d \in D_i, \quad \nu_{1,i,d} = \nu_{1,i}. \tag{6}$$

I.e. it would suffice that the values of these disturbances do not differ across the various decisions the agent might take. This may or may not be an attractive assumption, but it is the only assumption currently available which allows for differences between the profit variable the agent acts on and the profits we measure in this algorithm.

4.3 Profit Inequalities.

This approach is due to Pakes, Porter, Ho and Ishii (2006), though I will present it in a slightly different way. It will be helpful to begin by being explicit about the relationship between the profit construct that enters the theoretical model and the observed measure of profits. Let $r^I(d, d_{-i}, z_i^o, \theta_0)$ be our *observable* approximation to $\pi(\cdot)$ evaluated at the true $\theta = \theta_0$, and define $\nu(\cdot)$ as the difference between the profits that actually accrue to the agent and this approximation, that is

$$\nu(d, d_{-i}, z_i^o, z_i, \theta_0) \equiv r^I(d, d_{-i}, z_i^o, \theta_0) - \pi(d, d_{-i}, z_i). \tag{7}$$

The measurement model in equation (7) allows us to be clear about the sources of the disturbances in our models and their relationship to different estimation strategies. Rewrite $r(\cdot, \theta_0)$ as a sum of three components

$$r^I(d, d_{-i}, z_i^o, \theta_0) \equiv \mathcal{E}[\pi(d, \mathbf{d}_{-i}, \mathbf{z}_i)|\mathcal{J}_i] + \nu_{2,i,d}^I + \nu_{1,i,d}^I, \tag{8}$$

where

$$\nu^I_{2,i,d} \equiv \mathcal{E}[\nu(d, \mathbf{d}_{-i}, \mathbf{z}^o_i, \mathbf{z}_i, \theta_0)|\mathcal{J}_i],$$

and

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

Note that $\forall d \in D_i$, $\mathcal{E}[\nu^I_{1,i,d}|\mathcal{J}_i] = 0$, by construction, a property shared by the $\{\nu^f_{1,i,d}\}$ in the full information model of equation (5). On the other hand $\mathcal{E}[\nu^I_{2,i,d}|\mathcal{J}_i] \neq 0$, a property shared with the $\{\nu^f_{2,i,d}\}$ of the model in equation (5). The fact that these two disturbances differ in their conditional expectations implies that they have different impacts on the desirability of different estimators. We begin by considering the sources of the two disturbances, as this should provide guidance on when we need to worry about them.

ν^I_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 observe. So though it is not known to the econometrician, $\nu^I_{2,i} \in \mathcal{J}_i$, and since $d_i = d(\mathcal{J}_i)$, d_i will generally be a function of $\nu^I_{2,i}$ (and depending on the information structure of the game, perhaps also of $\nu^I_{2,-i}$). In contrast ν^I_1 is not known to *either* the econometrician or to the agent when it makes its decision, and so d_i will not be a function of it.

$\nu_{1,i}$ 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. It is a result of uncertainty in the exogenous variables that will eventually help determine profits (in our $\mathbf{z}_i$) and/or assymetric information (which causes uncertainty in $\mathbf{d}_{-i}$). It follows that 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 the possible play of its competitors and to realizations of $\mathbf{z}_i$, and then repeatedly solve for an equilibrium (a process which typically would require us to select among equilibria). This is likely to require both more information (e.g. knowledge of what agent knows about the other agent) and more computational power than the econometrician has available. The second component of ν^I_1, $\nu(d,\cdot) - \mathcal{E}[\nu(d,\cdot)|\mathcal{J}_i]$ is that part of the error in our measure of profits that is mean independent of the information the agent basis its decision on. One source of this component that is likely to be particularly relevant for the models considered here (since they require estimates of profits from counterfactuals) is measurement and modelling error in our profit measures.

Selection. If we temporarily ignore any difference between the agent's expectations (our $\mathcal{E}(\cdot)$ operator), and the expectations generated by the true data generating process (our $E(\cdot)$ operator), the definitions above provide a straightforward explanation of the selection problem in structural models. Assume that x is an "instrument" in the sense that $\mathcal{E}[\nu^I_2|x] = 0$, and, in addition, that $x \in \mathcal{J}$. Then

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

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^I is mean independent of any function of $\mathcal{J}$, $\mathcal{E}[\nu_1^I|x,d] = 0$. In contrast the model implies a relationship between $\mathcal{E}[r(d_i,\cdot;\theta_0)|\mathcal{J}_i]$ and $\nu_{2,i}^I$. That is if d_i was chosen and the observable part of the expected returns to d_i were less than those to d', then the unobservable part of expected returns to d_i must have been higher than that to d'. Formally if $\Delta(\cdot)$ designates the difference operator, so that $\Delta r(d_i, d', \cdot) = r(d_i, \cdot) - r(d', \cdot)$, then $\mathcal{E}[\Delta r(d_i, d', \cdot;\theta_0)|\mathcal{J}_i, d(\mathcal{J}_i)] \leq 0$ implies $\nu_{2,i,d_i}^I - \nu_{2,i,d'}^I > 0$. So even if we knew that x was an "instrument" in the sense that $\forall d \in D_i$, $\mathcal{E}[\nu_{2,i,d}^f|x] = 0$, as long as $\mathcal{E}[\Delta r(\cdot)|\mathcal{J}]$ is correlated with x, $\mathcal{E}[\nu_{2,i,d_i}^I|x_i, d_i] \neq 0$.

As a result any estimation algorithm based on accepting any value for θ which makes the sample average of our observable proxy for the difference in profits (of $\Delta r(\cdot, \theta)$), or its covariance with a positive valued instrument, positive *should not*, in general, be expected to lead to an estimated set which includes θ_0 (even asymptotically). To see this recall that equation (1) implies that $\mathcal{E}[\Delta\pi(\cdot)|x_i, d_i] \geq 0$, while equation (8) and our definitions imply that

$$\mathcal{E}[\Delta\pi(d_i, d', d_{-i}, z_i)|x_i, d_i] = \mathcal{E}[\Delta r(d_i, d', d_{-i}, z_i^o, \theta_0)|x_i, d_i] + \mathcal{E}[\nu_{2,i,d_i}^I - \nu_{2,i,d'}^I|x_i, d_i].$$

Thus $\mathcal{E}[\Delta\pi(\cdot)|x_i, d_i] \geq 0$ only implies that $\mathcal{E}[\Delta r(\cdot)|x_i, d_i] \geq 0$ if $\mathcal{E}[\nu_{2,i,d_i}^I - \nu_{2,i,d'}^I|x_i, d_i] \leq 0$, and the fact that x is an instrument, that is that $\mathcal{E}[\nu_{2,i,d_i}^I - \nu_{2,i,d'}^I|x_i] = 0$ *does not* insure the latter inequality.

For example if this was a single agent binary choice problem, i.e. if $d_i \in \{0,1\}$ with $d_i = 1$ if and only if

$$\mathcal{E}[\Delta\pi(d_i = 1, d' = 0, z_i)|\mathcal{J}_i] = \mathcal{E}[\Delta r(d_i = 1, d' = 0, z_i^o, \theta_0)|\mathcal{J}_i] + \nu_{2,i}^I \geq 0,$$

and $\nu_{2,i}^I$ was centerred at zero, then

$$\mathcal{E}[\nu_{2,i}^I|d_i = 1] = \mathcal{E}\Big(\nu_{2,i}^I|\nu_{2,i}^I \geq -\mathcal{E}[\Delta r(d_i = 1, d' = 0, z_i^o)|\mathcal{J}_i]\Big) \geq 0,$$

which violates our condition.

4.3.1 Conditions for the Profit Inequality Approach.

We begin with the assumption on the relationship between the expectation operator underlying agents' decisions (our $\mathcal{E}(\cdot)$), and the expectation conditional on the process actually generating the data (our $E(\cdot)$). This will help clarify the sense in which the behavioral model can be misspecified without invalidating the properties of the estimator.

We assume that we observe a subset of the variables which are contained in $\mathcal{J}_i$, say x_i, that satisfy the condition that if $h(\cdot)$ is a positive valued function, then

$$IC3: \;\; \frac{1}{N}\sum_i \mathcal{E}\Big(\Delta\pi(d_i, d', d_{-i}, z_i)|x_i\Big) \geq 0 \;\; \Rightarrow \;\; \frac{1}{N}\sum_i E\Big(\Delta\pi(d_i, d', d_{-i}, z_i)h(x_i)\Big) \geq 0 \; .$$

Clearly if the 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, the assumption that the agents' choices constitute a Nash equilibrium (condition C1) insures that $IC3$ is satisfied.

These assumptions are, however, stronger then the assumptions needed for $IC3$. One sufficient condition for $IC3$ is that agents' expectations of profit difference are *correct*; i.e. they equal the expectation of the $\Delta\pi_i(\cdot, \theta_0)$ conditional on x_i resulting from the data generating process. The 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 (a point made by Auman and Brandenburger, 1995) or their information sets, and 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. Further correctness of the agents' conditional expectations of $\Delta\pi(\cdot)$ does not require correctness of the conditional distributions for $(d_{-i}, \mathbf{z}_i)$. For example if the profit function were quadratic in $(\mathbf{z}_i, \mathbf{d}_{-i})$ the agents conditional expectations about $\Delta\pi_i(\cdot, \theta_0)$ would be correct if their conditional means and variances of $(\mathbf{z}_i, \mathbf{d}_{-i})$ were correct. 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 (1995) 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.

The discussion thus far justifies IC3 by conditions which insure that agents have correct expectations about profit differences. Though this is sufficient, it is by no means necessary. For $IC3$ it suffices that

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

I.e. agents can have *incorrect* expectations provided the difference between their expectations on $\Delta\pi_i(\cdot, \theta_0)$ and the expectation emanating from the data generating process is uncorrelated with the choice of x. Indeed IC3 would be correct even if agent's were incorrect *on average*, provided they were overly optimistic about the incremental profits emanating from their decisions.

The final requirement of this estimation strategy is that there be an $x \in \mathcal{J}_i$ that is observed by the econometrician and a function $c(\cdot) : D_i \times D_i \to \mathcal{R}^+$, such that

$$IC4: \quad \frac{1}{N}E\sum_i \Big(\sum_{j\in D_i} \chi\{d_i = j\}c\big(j, d'(j)\big)\big(\nu^I_{2,i,j} - \nu^I_{2,i,d'(j)}\big)\Big)h(x_i) \ \leq 0$$

where $h(\cdot)$ is a positive valued function, and $\chi\{d_i = j\}$ is the indicator function which takes the value of one if $d_i = j$. To understand what is required for $IC4$ to hold it is helpful to rewrite it as the iterated expectation

$$\frac{1}{N}\sum_i \Bigg(\Big(\sum_{j\in D_i} c\big(j, d'(j)\big)E[\big(\nu^I_{2,i,j} - \nu^I_{2,i,d'(j)}\big)|d_i = j, x_i]\Pr\{d_i = j|x_i\}\Big)h(x_i)\Bigg) \ \leq 0.$$

This shows that what we require is that an *unconditional* average, an average that does not condition on d_i, of the *differences* betweem the ν_2 associated with the decision and the alternative for that decision be less than or equal to zero. Note that we are free to vary both; (i) the weights assigned to the possible differences (the $\{c\big(j, d'(j)\big)\}$), and (ii) the alternative we compare to should the decision be $d = j$ (i.e. the $\{d'(j)\}$). As we illustrate below this enables us to use an assortment of primitive conditions to insure $IC4$.

To see how the assumptions of this subsection generate moment inequalities which can form the basis of an estimation algorithm note that $C1$ and $C2$ imply that

$$0 \ \leq \frac{1}{N}\sum_{i=1}^{N}\mathcal{E}\Big[\big(\sum_j \chi\{d_i = j\}\Delta\pi(j, d'(j), \cdot)\big)h(x_i)\Big] \tag{9}$$

which from $IC3$

$$\leq \ \frac{1}{N}\sum_i E\Big[\sum_j \chi\{d_i = j\}\Big(\Delta\pi(j, d'(j), \cdot)\Big)h(x_i)\Big],$$

which from $IC4$ and the definitions in equation (8)

$$\leq \frac{1}{N}\sum_i E\Big[\sum_j \chi\{d_i = j\}\Delta r(j, d'(j), \cdot, \theta_0) h(x_i)\Big].$$

Since this last inequality is in terms of *observable* moments it can be used as a basis for estimation.

Assumptions which imply $IC4$. Recall that the restriction we used to insure the full information algorithm of the last subsection was appropriate when there was either modelling or measurement errors was that those errors, i.e. the $\{\nu^f_{1,i,d}\}$, did not vary across d (see restriction 6). Analogously $IC4$ will hold if the unobservable known to the agent when it makes its choice but not observed by the econometrician, i.e. the $\{\nu^I_{2,i,d}\}$, is constant across choices[8]. The first restriction is on the disturbances that the agent does not take account of in making its decision while the second is on the disturbances that the agent does take account of. As a result they will be appropriate for different problems.

There are cases in which one can form moments which satisfy $IC4$ even though the $\{\nu^I_{2,i,d}\}$ do vary across d. One subset of the cases occurs when the effect of the $\{\nu^I_2\}$ can be eliminated by appropriate choice of $d'(j)$ and $c(j, d'(j))$. Pakes et. al. (2006) note that when the $\nu_{2,i,d}$ vary across decisions but the same value of $\nu^I_{2,i,d}$ appears in more than one of them (so there are "group effects"), one can form inequalities which "difference out" the ν^I_2. Examples include; entry models in which $\nu^I_{2,i,d}$ is a location specific fixed effect, social interaction models where the interaction effects are group specific, panel data discrete choice models in which the $\nu^I_{2,i,d}$ are choice specific fixed effects, and cross sectional discrete choice models where the same $\nu^I_{2,i,d}$ appear in more than one choice. Alternatively when a variable is unobserved at the micro level, but is observed at a higher level of aggregation (say from census data), then a summation of inequalities will do away with the $\nu_{2,i}$[9].

A different subset of cases which satisfy $IC4$ even though the $\{\nu^I_{2,i,d}\}$ do vary across d are cases in which inequalites can be formed which are a linear function of the same $\nu^I_{2,i}$ regardless of the realization of d_i. Then as long as we have a variable that is in the agent's information set when the agent makes its decision that the $\nu_{2,i}$ is mean independent of (an "instrument"), we can form the uncenterred sample covariance of the error in the inequality and any positive function of that variable, and its expectation will be zero. Pakes et. al.

[8]As noted in Pakes et. al, this latter assumption is also used in Hansen and Singleton's (1982) classic article. 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.

[9]A similar procedure applies if variables are measured with error at the micro level but that error averages out at a higher level of aggregation.

(2006) show that the ordered choice models (defined broadly enough to include the vertically differentiated demand model used in I.O.) is one example of this case. Another occurs in the analysis of contracts in buyer seller networks when the terms of contracts are known to the agents but not to the econometrician. These procedures are often helpful in the familiar case in which there is a determinant of costs that is unknown to the econometrician but known to the agents. For example if a firm is buying a discrete number of units, so that $d_i \in \mathcal{Z}_+$ and the orderred choice model is relevant, and we do not observe a determinant of unit cost, then taking $d'(j) = j + 1$ provides us with a linear function of the unobserved cost variable no matter the d chosen. In the buyer seller contracting model the unobserved component of the payments emanating from the contract are a cost to the buyer. Then if the inequality used when the contract is established is the difference between the seller's profit with and without the contract, and the inequality used when the contract is rejected is the difference between the buyer's actual profits and what its profits would have been were the contract established, then our inequality will include the unobserved component of the transfer regardless of whether the contract is established (see below).

4.3.2 Discussion; Assumptions and Uses of the Approaches.

$IC3$ is as general a behavioral assumption as has been used in structural models ($FC3$, the full information assumption, is a special case of it). Moreover the generality in $IC3$ is likely to be helpful both because; (i) it enables us to proceed without having to specify either the agents' information sets (and hence what each agent knows about the other agents) or the form of the probability distribution the agents use to form expectations; and these are objects we typically know little about, and because (ii) it allows for expectational, measurement, and modelling errors (as long as those errors are not correlated with the instruments used). As a result when $IC4$ is appropriate, the assumptions underlying the profit inequality approach are attractive. The profit inequality approach is also attractive from a computational point of view as to generate a moment inequality from it we need only compute the actual profits and compare them to the profits that would have been earned from a single feasible alternative policy.

Of course assumption $IC4$ need not be appropriate. Indeed neither the profit inequality approach nor the full information approach is likely to fit any particular application exactly, and the choice between approaches in a particular application is likely to depend upon: (i) the relative importance of differences in the two sources of error, ν_1 and ν_2, across decisions, (ii) a comparison of our ability to control for the important sources of ν_2 nonparametrically in the profit inequality approach with the ability of the parametric assumptions used in the full information approach to control for ν_2, (iii) and the extent to which any error in our assumptions is likely to affect the empirical results of interest. Below we provide a Monte

Carlo comparison of the two approaches in the context of analyzing contracts in buyer-seller networks, that looks into these issues in one case of interest.

To conclude this subsection I want to illustrate the flexibility built into the profit inequality approach through the fact that it allows the researcher to chose different counterfactuals for different choices (the illustration is taken from Katz, 2007). The problem is to analyze the costs shoppers assign to driving to a supermarket. These costs are of considerable importance to the choice of supermarket locations and, as a result, to the analysis of the impact of zoning regulations. Moreover they have proven difficult to analyze empirically with standard choice models because of the complexity of the choice set facing consumers (all possible bundles of goods at all possible supermarkets).

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. Since utilities are only defined up to a montone transformation, there is a free normalization for each individual, and we normalize the coefficient on expenditure for each individual to equal one. We want to allow for heterogeneity in the cost of drive time that is known to the agents when they make their decision but unobserved by the econometrician, so this will be one component of $\nu_{2,i}$. The counterfactuals possible are the purchase of any bundle of goods at any store.

For a particular d_i chose $d'(d_i)$ to be the difference in utility between the choice actually made and the utility that would have been obtained from purchasing; (i) the *same basket* of goods, (ii) at a store which is *further away* from the consumer's home then the store the consumer shopped at. This choice of alternative (of $d'(d_i)$) will allow us to difference out the impact of the basket of goods chosen on utility. I.e. if $e(d)$ and $dt(d)$ provide the expenditure and the drive time for store choice d, and $(\theta + \nu_{2,i})$ is agent i's cost of drive time (in units of expenditure), this choice of alternative gives us the inequality

$$\mathcal{E}\left[\sum_j \chi\{d_i = j\}\Delta\pi(j, d'(j), z_i)|\mathcal{J}_i\right] =$$

$$\mathcal{E}\left[\sum_j \chi\{d_i = j\}\Big(e(j) - e(d'(j)) + (\theta + \nu_{2,i})(dt(j) - dt(d'(j)))\Big)|\mathcal{J}_i\right] \geq 0, \quad at \;\; \theta = \theta_0.$$

Assuming, as seems reasonable, that $(dt(d_i), dt(d'(d_i))) \subset \mathcal{J}_i$, this together with the fact that $dt(d'(j)) - dt(j) > 0$ by choice of alternative, implies that

$$\mathcal{E}\left[\sum_j \chi\{d_i = j\}\left(\frac{e(j) - e(d'(j))}{dt(d'(j)) - dt(j))} - (\theta_0 + \nu_{2,i})\right)|\mathcal{J}_i\right] \geq 0.$$

Let θ be the average of the cost of drive time across consumers, so $\sum_i \nu_{2,i} = 0$ by construction, and assume $IC3$. Then

$$E\left[\frac{1}{N}\sum_i \Big(\frac{e(d_i) - e(d'(d_i))}{dt(d'(d_i)) - dt(d_i))}\Big)\right] - \theta \geq 0, \quad at \;\; \theta = \theta_0.$$

This provides an upper bound to θ. Were we to also consider a second alternative in which the bundle of goods purchased was the same as in the actual choice but the counterfactual store required *less drive time*, we would also get a lower bound to θ.

Note that to obtain these inequalities we chose an alternative which allowed us to difference out the the impact of the bundle of goods chosen on utility (differencing out our "group" effect), and then rearranged these differences to form a moment which was linear in the remaining source of ν_2 variance no matter d_i (the source being differences in the costs of travel time) . Were we interested in the impact of a particular good purchased on utility, we would have considered baskets of goods which differed only in that good and goods which had cross partials with that good in the utility function, at the *same* supermarket (thus differencing out the effects of travel time and other components of utility). Alot more options would present themselves were we to have data on multiple shopping trips for each household.

4.4 A Note on Inference.

Since the theoretical restrictions we bring to data are moment inequalities, our estimators will typically be set valued. Methods of inference for set valued estimators are an active and important area of econometric research. There are a number of papers which prove "consistency" for set valued estimators; i.e. they give conditions which insure that the set of parameter values that satisfy the sample moment inequalities (the "estimated set") converges to the set of values of θ that satisfy the population moment inequalities (the "identified set") in the Hausdorff metric. In the sections that follow I will require some results on the distribution of estimates, so I now provide a brief explanation of the issues that arise in obtaining them.

Assume the identified set is compact and convex (as it is in our examples which have linear moment inequalities), so that the set of values of any component of θ that are contained in this set is a bounded interval. If the first component of θ is $\theta_{0,1}$, with bounds $[\underline{\theta}_{0,1}, \overline{\theta}_{0,1}]$ which are obtained at $\theta = \underline{\theta}_1$ and $\theta = \overline{\theta}_1$ respectively, we can obtain conservative α level confidence intervals for either $\theta_{0,1}$ or for $[\underline{\theta}_{0,1}, \overline{\theta}_{0,1}]$ if we have an $(\hat{a}, \hat{b})$ such that

$$\Pr\left\{\theta_{0,1} \in [\hat{a}, \hat{b}]\right\} \;\geq\; \Pr\left\{[\underline{\theta}_{0,1}, \overline{\theta}_{0,1}] \subset [\hat{a}, \hat{b}]\right\} \;\geq\; 1 \,-\, \Pr\left\{\hat{a} > \underline{\theta}_{0,1}\right\} \,-\, \Pr\left\{\hat{b} < \overline{\theta}_{0,1}\right\} = 1-\alpha. \tag{10}$$

If we find estimators for $(\underline{\theta}_{0,1}, \overline{\theta}_{0,1})$ whose distributions can be approximated with satisfactory precision, we can use the $(1-\alpha/2)$ quantile of the estimator for $\underline{\theta}_{0,1}$ as $\hat{a}$, and the $\alpha/2$ quantile of the estimator for $\overline{\theta}_{0,1}$ as $\hat{b}$.

Consistent estimators for $(\underline{\theta}_{0,1}, \overline{\theta}_{0,1})$ are the lowest and highest value of θ_1 which satisfy all the sample moment inequalities. The standard asymptotic approximation to the distribution of those estimators is obtained by analyzing the impact of the variance in the moments that define these parameters in the *given sample* on the estimates of the parameters of interest. If there are K parameters, there will be K moments that hold with equality at the estimate of $\underline{\theta}_1$, and it is the impact of the variance in these moments on the estimate of the parameters which will determine this estimate of the variance. This calculation, however, ignores the fact that in samples of the size we use in economics, the sampling variance in the moments will often cause *different* moments to bind in different samples. As a result though the standard approximation may provide an adequate approximation to the sampling variance of the intervals that satisfy the inequalities associated with the moments which bind in the sample, it does not provide an adequate approximation to the distribution of the intervals that satisfy all the inequalities.

There are a number of approaches to obtaining confidence intervals which provide a closer approximation to the true finite sample distributions of the estimated intervals currently being investigated. I will use the bootstrap methodology introduced in Pakes, Porter, Ho and Ishii (2006). I do so not because I have any reason to believe it is more accurate than the alternatives currently available, but rather because; (i) it is easy to use, and (ii) I have done Monte Carlo experiments on precisely the problem we focus on in the analysis of specification errors below and found that for the sample sizes used their this bootstrap provides almost exact coverage for the estimated intervals[10].

There are a number of other estimation issues which have a different structure when doing inference based on moment inequalities (in contrast to on moment equalities). Perhaps most important, though it has received little attention, is the question of the choice of moment

[10]These results were deleted from this version of the paper for space considerations, but are available from the author on request. I note that the bootstrap does seem to "undercover" for smaller samples than those used in our specification analysis. In particular for samples comparable in size to those used in the empirical example coverage for the entire interval in experiments designed to provide a 95% confidence interval for that interval was just over 90%. However coverage for the parameter itself was 100% and the magnitude of the difference between the estimated bounds and the actual bounds in cases where the estimated bounds did not cover the actual bounds was quite small.

inequalities. In the inequality context that choice determines the boundaries of the identified set, as well as the variance of the estimator of boundary points. Testing is an issue which has received some attention. Though typically there will be more than one value of θ that satify the population moment inequalities, there may well not be any value which satisfies the sample moment inequalities. Then the set estimate will reduce to a point. This does not necessarily imply that there is no value of θ that satisfies all the population inequalities. The fact that in finite samples the sample moments distribute approximately normally means that point estimates can occur even if there are values of θ at which all the inequalities are satisfied by the population moments. This is clearly more likely to occur the larger the number of inequalities used in estimation, and the tests are designed to distinguish whether the fact that we find a point estimate is due to sampling error, or due to model misspecifications. Developments in the econometric literature are likely to accompany applied work using inequality estimators, and we can expect the two literatures to reinforce each other.

5 An Example: Buyer-Seller Networks.

We now come back to the empirical analysis of the nature of contracts in markets in which there are a smaller number of both buyers and sellers with buyers able to buy from more than one seller and sellers able to sell to more than one buyer; a market structure I will refer to as a "buyer-seller network". This section uses an empirical example to develop and illustrate alternative moment inequality estimators for such markets, and then computes equilibria to a buyer-seller network game designed to be similar, to the extent possible, to the empirical example. The next section uses the computed equilibria in a Monte Carlo excercise designed to compare the robustness and computational burdens of the full information to the profit inequality approach to inequality esitmators.

The example is taken from Ho (2007) who analyzes the contracts established between HMOs and hospitals in forty-two markets. Ho's estimates are based on an assumption that there are no structural disturbance (or $\nu_2 \equiv 0$) in her data. I begin by showing that her model can be used to generate moment inequalities that allow for both stuctural and non-structural disturbances. The inequalities that this generates differ from those used in Ho's article and we compare the results from estimators that allow for structural errors to those that do not.

Next I compute contracting equilibria for markets which are constructed to be similar to the markets analyzed in Ho, and then compare the characteristics of the computed equilibrium contracts to the characteristics obtained from the two different estimators. To compute the equilibria we need a more complete set of assumptions then are needed for the estimation algorithm, and we consider two candidates from the theory literature. The computed equilibria allow us to engage in a broader investigation of features of the environment that are correlated with the markups implicit in equilibrium contracts then is possible in the empirical work. The results highlight two facts. First though the moment inequality estimators' characterization is reduced form in a sense that we will make clear, the empirical results from Ho's data do pick up important features of the contracts generated by our computed equilibria. Second the contracts that emanate from the complete structural model have features that are familiar to IO economists who have studied pricing models in more traditional oligopolistic markets (both in markets where a unilateral pricing assumption, and in markets where a bargaining assumption, seems relevant).

5.1 Empirical Analysis.

Recall that the terms of the contracts that govern behavior in buyer seller networks are often proprietary and not available to the econometrician. The reason that moment inequalities might help sort out the likely correlates of the transfers implicit in different contracting environments is that though we may not observe either the details of the contracts or the actual transfers from the buyers to the sellers, we do typically observe who contracts with whom in these markets. Provided we assume a game form which details an alternative network that a contracting party could have chosen, and we are able to obtain a sufficiently good approximation to the profits that would be earned had that alternative network been formed, the information on who contracts with whom should enable us to bound the transfers implicit in the contracts that were formed.

For example assume that we have a structural model which can approximate what buyers earn from reselling the products they buy from the sellers. Then were we able to specificy the networks that the buyer thought would have been established had it rejected a contract offer that it accepted (or visa versa), we could use the moment inequality framework to investigate the properties contract offer must have had in order to have supported the networks we actually observed. Of course the specification for what the agent believes would have happened requires assumptions, and the way we make those assumptions might change the inequalities we bring to data and hence the estimated parameters (a possiblitiy we consier below). On the other hand at least in principal one can test whether one or more of the possiblities are in fact consistent with the data and the insitutional information available.

Moreover there are a number of reasons to think that the flexibility of the moment inequality framework will be helpful in uncovering the characteristics of contracts which support buyer-seller networks in different settings. First in many cases of interest the information at the disposal of the analyst is not rich enough to associate a unique equilibrium with the incentives emanating from a given set of institutions, so allowing for multiple equilibria is likely to be essential. Second, our ability to use moment inequalities does not depend on the form of the choice set, so we can make the feasible set of contracts as complicated as reality demands. Finaly the fact that moment inequality estimators can allow for an assortment of errors is likely to be helpful in enabling us to get a start on problems as complex as this one.

5.1.1 The Analytic Framework.

It is assumed there are two periods. In the first period contracts between HMO's 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 HMO's engage in a premium setting game which we assume has a unique Nash equilibria.

The second period equilibrium generate revenues for *each* HMO conditional on *any* configuration of hospital networks, and the number of patients each HMO sends to each hospital. 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 HMO's, these revenues and quantities will be denoted by $R_{\mathtt{m}}(H_{\mathtt{m}}, H_{-\mathtt{m}}, z)$, and $q_{\mathtt{m},h}(H_{\mathtt{m}}, H_{-\mathtt{m}}, z)$, repsectively. The parameters needed for these calculations are obtained and the calculation is done using the techniques described in section 2 of this paper (for detail see Ho, 2005, 2006).

The profits of the HMO are the revenues from the second period game minus the payments the HMO makes to the hospital in its networks, say $T_{\mathtt{m},h}$ or

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

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

$$\pi^H_h(M_h, M_{-h}, z) = \sum_{\mathtt{m} \in M_h} T_{\mathtt{m},h}(H_{\mathtt{m}}, H_{-\mathtt{m}}, z) - c_h \sum_{\mathtt{m} \in M_h} q_{\mathtt{m},h}(H_{\mathtt{m}}, H_{-\mathtt{m}}, z).$$

Throughout we shall assume that the HMO revenues and hospital costs obtained in this way are correct up to an error which is mean independent of the variables we use as instruments.

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

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

Note that the error from this projection is known to both agents when they make their decsions (i.e. which is why it is labelled as a ν_2 error). Substituting it into the profit equations and defining $r_{\mathtt{m}}(H_{\mathtt{m}}, H_{-\mathtt{m}}, z; \theta)$ to be the observed portion of profits, we have

$$\pi^M_{\mathtt{m}}(H_{\mathtt{m}}, H_{-\mathtt{m}}, z) = r^M_{\mathtt{m}}(H_{\mathtt{m}}, H_{-\mathtt{m}}, z; \theta) - \sum_{h \in H_{\mathtt{m}}} \nu_{2,\mathtt{m},h} \tag{11}$$

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

$$\pi^H_h(M_h, M_{-h}, z) = r^H_h(M_h, M_{-h}, z; \theta) + \sum_{\mathtt{m} \in M_h} \nu_{2,m,h} \tag{12}$$

We use the inequalities generated by the first stage game to obtain information on θ.

Inequalities Used in the Empirical Work. To obtain those inequalities we need assumptions on the first period interactions between sellers (the hospitals sell services) and the buyers (the HMOs). It is assumed that there is an alternating move game with simultaneous moves on each side and no contracts established until a final period in which the seller makes offers to the buyer. As in Hart and Tirole (1991), contract offers are assumed to be proprietary: each HMO knows the offers made to it but not to its competitors, and each hospital knows the offers it makes but not those of its competitors.

We observe which HMO's contracted with which hospital. In order to obtain profit inequalities we need to be able to compute what would happen were either the HMO or the hospital to change its behavior. A major reason for making the assumptions of the last paragraph is that they make it relatively easy to compute the profits for any change in final period HMO behavior. The assumptions imply that the HMO could reject any offer it accepted or accept any offer it rejected without changing the behavior of any other agent. So we obtain one set of inequalites by reversing the HMO's acceptance/rejection decision with each of the hospitals in the market, and leaving all other contracts unchanged.

To calculate the returns the hospital would make from a different set of offers we need another assumption. We need to know what the hospital thinks the HMO would infer about the offers made to other HMO's if that HMO were to be offered a different contract. For the empirical work we make a "passive beliefs" assumption; 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 $d^{-i}(d', z_i) = d_{-i}$. Here we also assume that the hospital could always offer a null contract which is never accepted and that the HMO response to this contract offer leaves the hospital worse off then it would be were the HMO to respond by leaving its other contracting decisions unchanged[11]. These assumptions are computationally convenient as they insure that the hospital expects its profits from offers that are accepted to be larger than they would be if an accepted contract were rejected and no other contract changes.

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. If $\nu_2 \neq 0$ these inequalities are invalid because they involve the ν_2 *and* condition on the decision made. Were the ν_2 generated by a structural model one might have reason to believe that they obey certain constraints; perhaps the most familiar structure being that of HMO and/or hospital fixed effects. Appendix 1 derives the inequalities that this generates. Their is no particular reason to assume this structure for our problem and when we allowed for fixed effects 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 significant at the 5% level. In the model with fixed effects, about a third of the inequalites were negative and 10% were significant at the 5% level.

Moreover there are two other inequalities available which do not require the special fixed effects structure for the $\{\nu_2\}$ so we focus on them. One of these does not involve the ν_2 while the other does but does not need to condition on the decision made. For the first inequality 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 $\nu_{2,m,h}$). Formally if we 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 $\Delta r_h^H(M_h, M_h/h, M_{-h}, z; \theta)$ be defined accordingly, while $\Delta\pi_m^M(H_m, H_m \cup h, H_{-m}, z)$ be the difference between the HMO's profit when the network of the HMO excludes hospital h

[11] We can relax the assumption that the HMO response to this contract offer leaves the hospital worse off then it would be were the HMO to respond by leaving its other contracting decisions unchanged. What we require is a lower bound to the profits the hospital could make as a result of actions the HMO might take if it received a null contract from the given hospital, leaving all the other HMOs' contracts unchanged. This bound can be obtained by assuming the HMO choses the set of contracts which minimizes the given hospital's profits. On the other hand if the HMO does change its other contracts it is likely to add a hospital and increase its number of patients, none of which will be sent to the given hospital. Indeed when we did a partial relaxation of this assumption, allowing the HMO to chose another hospital in response to the loss of the given hospital, we obtained results which were similar to those reported below.

and when it includes it and $\Delta r_m^M(H_m, H_m \cup h, H_{-m}, z; \theta)$ be defined accordingly, we have

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

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

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, $h(\cdot)$

$$E\Big[\chi_{m,h}\Delta r_h^H(M_h, M_h/h, \cdot; \theta_0) + (1-\chi_{m,h})\Delta r_m^M(H_m, H_m \cup h, \cdot : \theta_0)\Big]h(x) \;\geq 0. \qquad (13)$$

For the second inequality note that the sum 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 contains information on θ (since if the contract is not established there is a change in transfers to other agents), and must have positive xpectation if a contract is established (at least if contract offers are proprietary, see Hart and Tirole, 1991). So provided $x \in \mathcal{J}_m \cap \mathcal{J}_h$

$$E\Big[\chi_{m,h}\Big(\Delta r_h^H(M_h, M_h/h, \cdot; \theta_0) + \Delta r_m^M(H_m, H_m/h, \cdot; \theta_0)\Big)\Big]h(x) \;\geq 0. \qquad (14)$$

Empirical Results. 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 on the table are the coefficients of the markup implicit in the per patient payment.

Table 1: Determinants of Hospital/HMO Contracts.

Data	Real Data				Simulated Data			
Estimator	**Inequality Estimators**				**OLS Regresssion**			
	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.
Per Patient Markup (Units = $ thousand/patient)								
Const.	9.5	15.4/4.8	8.2	15.2/3.3	8.9	.09	3.7	.24
CapCon.	3.5	8.6/1.4	13.5	16.1/2.3	1.2	.10	.48	.11
Cost/Adm.	-.95	-1.5/-.57	-.58	-.2/-1.1	-.39	.01	–	–
Ave.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 HMO's, 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-Average cost" is the cost per admission of the hospital minus the averge of that over the hopsitals 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 HMOmargin is the HMO's average premium minus its average cost.

The estimate of Θ_0 from every specification was a singleton, so there was no parameter vector that satisfied all the inequality constraints. On the other hand none of the test statistics we computed were significant at the 5% level, though since all specifications had eighty-eight or more inequality constraints and some of them had large variance this might not be considered surprising. There is some indication that the model that allows for ν_2 is more consistent with the data, as only six of its inequalities were negative at the estimated parameter value (the ν_1-only model had eleven), and none of them were significant at the 5% level (in contrast to one for the ν_1 only model).

Ho (2006) reports a series of robustness checks on the ν_1-only estimates. 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 columns (1) above. These are of the same sign, but quite different in magnitude, then those in columns (3).

Both sets of point estimates imply an equilibrium configuration where the majority of cost savings from low cost hospitals are captured by the HMO's who do business with those hospitals, and in which markups increase sharply when a hospital is capacity constrained. These are troubling findings, for if they were interpreted as causal they would imply significantly lower incentives for hospitals to invest in either cost savings or in capacity expansion than would occur in a price-taking equilibrium. The difference between the ν_1-only estimates and those that allow for ν_2 is that the former imply that almost all the cost savings from low cost hospitals go to the HMO's, while the latter imply that only just over 50% of those savings do and that a larger fraction of profits go to capacity constrained hospitals (as one might expect low cost hospitals tend to be more capacity constrained then high cost hospitals, so the two variables are negatively correlated).

Might we expect contracts with these characteristics to emanate from a Nash equilibrium? If so which of our two specifications is likely to lead to coefficients which better approximate the characteristics of equilibrium contracts and should we interpret those coefficients to mean that an increase in the right hand side variable would, ceterus parabus, generate the markup response we estimate? To shed some light on these issues we computed equilibria in markets with charactersitic 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 HMO's in each market (this made it possible to compute equilibria for many different markets in a reasonable amount of time)[12].

5.1.2 Numerical Analysis.

We compute a full information Nash equilbirum to a game in which hospitals make take it or leave it offers to HMO's[13]. 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 simultaenously 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

[12]As in Ho(2005) demand was calculated using a discrete choice model. The market characteristic distributions for the demand model and the costs were both random draws from distributions which mimiced those in Ho's data.

[13]The closest excercise I know of in the literature is in a paper by Gal-Or (1997). By judicious choice of primitives she is able to provide 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.

equilibrium to a premium setting game among the HMOs). The HMOs then simultaneously accept or reject the offers. At equilibrium each hospital is making the best offers it can given the offers of the other hospital and the responses of the HMOs, and each HMO is doing the best it can do given the actions of its competitor and the offers made by the hospitals. An iterative process with an initial condition in which both hospitals contract with both HMOs choses among the equilibria when there are multiple equilibria[14].

Note that the game form used here is different then that which we assumed to generate the data analyzed in the empirical work. The full information Nash assumptions used here insures that the equilibrium is renegotiation proof. Of particular importance to us is the fact that the different informational assumptions change the necessary conditions for an equilibrium, and this in turn changes the inequalities that we can take to data. For example, in the full information game HMO knows that if it had rejected an offer which it had in fact accepted, another HMO which had rejected the given hospital's offer might decide to accept, and the original HMO takes that change into account when computing what would happen were it to reject the hospital's offer.

The related questions of which equilibrium notion is likely to be more appropriate for a given institutional setting, and whether the estimation results are likely to be sensitive to the notion assumed, are two of the many questions that research on buyer seller networks will have to sort out. What is clear is that though often the contents of contracts are proprietary, at any given point in time who contracts with whom is usually known to all participants. So if we were trying to model a set of relationships which have been stable over for 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. A problem with using an equilibrium notion with this property is that when we compute counterfactuals we will have to consider the reaction of competitors to changes in the contracting outcome of two participants, and this will be computationally burdensome. Of course the market we are studying may be constantly changing and negotiations might

[14]It starts at the lowest set of contract offers. The algorithm 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 repete 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 repete this process until we find a Nash equilibrium to the offers. For about 3% of the random draws of characteristics we could not find an equilibria, and those markets were dropped from the analysis.

be costly. Then we might not expect the data to abide by this renegotiation proof criteria, at least not one with costless renegotiations.

We have raised these issues in a very simple contracting environment. In reality the contracting process may be more complex than our simple take it or leave it markup game, and the environment facing the agent changes over time. What we need in order to use the observed networks to make inferences on the properties of contracts is a way of obtaining a lower bound to the *expected* profits from a counterfactual choice. This is an area where theoretical insights into when different counterfactuals might be appropriate and their implications for the inequalities we can take to data would be of great help. The stronger the conditions we have the tighter our bounds are likely to be, but we could start with weaker conditions and then investigate when stronger ones might be appropriate[15].

Column (5) through (8) of Table 1 present the results from projecting the computed full information Nash equilibirum markups onto variables of interest. The first two colums 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 row adds variables and reruns the projection. The original three variables maintain their signs and remain significant, but there are noticeable changes in their magnitudes, and those magnitudes seem to depend on other features of the contracting environment.

The coefficients of the additional variables illuminate different aspects of the 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%. 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, i.e. in markets where the ratio of population to the number of hospital beds is lower. Moreover once we account for this effect the effect of capacity constraints is greatly reduced (though not eliminated). HMO's 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 HMO's they are dealing with can charge their members higher markups. Their is alot of economic intuition underlying the signs (though not the magnitudes) of these results, and one might think that it would carry over to more complex environments.

[15]For example we might start by allowing for all the full information Nash equilibria, and then test the additional restrictions implied if we thought that no two firms could simultaneously deviate and increase their profits. Note that the weaker conditions could also include mutually exclusive sets of possiblities. We would then obtain our inequalities from the difference between the profits of the chosen alternative and the minimum profits from counterfactuals in any of those sets.

One final point. After adding the extra variables 20% of the variance in markups, which works out to 8% of the variance in transfers, is not accounted for by our observables. Since the contracts are known to both agents when they make their decisions, this variance would be classified as variance in the structural error (or in ν_2) in our prior discussion. Even if our behavioral, informational, and functional form assumptions were perfect, in an actual empirical project there is likely to be substantial measurement error in hospital costs, and this would constitute a ν_1 error. So there is a question of how well our alternative inequality estimators, those based on the ν_1-only and the ν_2-only inequalities as well as the robust inequalities which allow for both types of error, would do with reasonable amounts of variance in both types of disturbances. We now turn to an analysis of the performance of the alternative estimators both when they are, and when they are not, based on data which satisfies their assumptions.

6 Specification Errors and Alternative Estimators.

This section presents a comparison of both the computational requirements and the performance of different moment inequality estimators in the context of the empirical analysis of contracts in buyer-seller networks. The estimators we consider differ in what they assume about the disturbances defined in equation (7). The first only allows for a ν_1 disturbance, the second only allows for a ν_2 disturbance, and the third allows for both types of disturbances. Recall that the ν_1-only estimator is always available for the profit inequality approach to building moment inequality estimators, the ν_2-only estimator is always available for the full information no error approach, and estimators which allow for both ν_1 and ν_2 are available for some models based on profit inequalities. Most of the statistical results will be Monte Carlo results using the data from the full information equilibrium computed in the last section, and we consider results for the ν_1-only and the ν_e-only estimators both when their assumptions are the assumptions generating the data, and when they are not. Where possible, we will also present results from Ho's data.

Details of the Monte Carlo Analysis. The counterfactuals used to generate inequalities in the Monte Carlo analysis 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 and the empirical work is based on a model with partial information, these counterfactuals generate different inequalities for the Monte Carlo sample than those used in the empirical work. In particular when the hospital offers a null contract to an HMO which had accepted its equilibrium offer, now both HMO's reoptimize. If $o_{M,H}$ is the contract offered by hospital H to HMO M in equilibrium, ϕ designates the null contract, and H1's offer was accepted by HMO1, then the counterfactual requires the profits of H1 from the HMO equilibrium responses to the tuple $(\phi, o_{1,2}, o_{2,1}, o_{2,2})$[16].

We estimate a model with only one parameter; the average markup. To obtain the true value of that parameter for the simulated data sets we took the transfers implicit in the equilibrium offers and projected them onto the number of patients and the variables we used as instruments[17]. 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. When all we require is a ν_1 error we also treat the residual from the projection as known, and then add pseudo random draws on measurement error to the costs of each hospital (and sometimes also to the population, and hence the patient flows in the market, see below). When we require a ν_2 error we let the residual from our parametric transfer function be unknown. Note that this insures that ν_2 has zero covariance with our instruments *before* we condition on the outcome (as is required of our "instruments"). On the other hand the distribution of ν_2 conditional on x that results from this procedure may well depend on x.

We used the algorithm described in the last subsection to compute equlbria for about twenty thousand markets with two HMOs and two hospitals in each. Monte Carlo data sets were obtained by taking random draws of one thousand three hundred and eighty five markets (without replacement) from these simulated markets. This gives us datas set that have about the same number of contracts as in Ho's data, but *many fewer* inequalities per market (Ho's data has about eight hospitals and ten HMO's per market, but only forty markets). The small number of inequalities implies that the identified set from the Monte Carlo data can be quite large.

We did the Monte Carlos for the profit inequality approach in two ways. In the first we simply assigned ν_1 errors and held them fixed. We then drew two hundred Monte Carlo data sts, obtained estimators from each, and tabulated the results. 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. The latter case provides

[16]Note that the hospital does not necessarily make the lowest offer that is consistent with the HMO accepting, because different offers change the HMO costs per patient, and hence change the outcome of the premium setting game that the HMO's engage in. Note also that we are not using all the necessary conditions for equilibrium. At the cost of increasing the computational burden of the estimator we could have used the inequalitiess obtained from simultaneously switching each HMO's (each hospital's) behavior with respect to both hospitals (both HMOs), and if more details of the actual contracts were available to the researcher, yet more inequalities would be available.

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

confidence intervals that condition on the observables (averaged over these observables), while the former does not. The difference in results was too small to be of much interest, so we only report the first case.

To compute the inequalities from the full information no error algorithm we take two hundred draws on a ν_2 vector of length equal to the number of possible contracts from the ν_2 distribution specified below for each market, and hold them fixed for the entire algorithm. Then for each θ evaluated in the estimation algorithm we compute the simulated probabilities that our three Nash conditions are satisfied at the observed market structure for each market. This gives us the upper bound to the probabilities. To get a lower bound we would have to check if the observed structure was the only structure which satisfied all the Nash inequalities (not just our three conditions) for all possible market structures. This was too computationally demanding even for our two by two problem. Computational concerns also limited the Monte Carlo to one hundred and twenty data sets for the full information no error approach.

Results. Table 5, which presents the results, is split into panels. Panel A provides estimates obtained from using the ν_1-only inequalities, Panel B from using the ν_2-only inequalities (the inequalities from the full information no error model), and Panel C uses the inequalites that allow for both ν_1 and ν_2 disturbances. The true value of θ_0 is 18.77, and the interval we obtain from the "population" moment inequalities when there are no errors of any kind is [15.43,20.62].

The first three rows of panel A provide results from the Monte Carlo data that only has ν_1 errors; so the estimators in these rows are consistent estimators of the "identified" set. Row 1 adds measurement error in costs equal to 25% of the true measured variance in cost. The estimated lower bound is slightly lower than the true lower bound, while the estimated upper bound is almost exactly equal to the truth. The 95% confidence interval for the interval covers the true interval and is not too different from the estimated interval per se, indicating that our bound estimates are quite precise (and this sample is not large by modern IO standards). When we add an expectational error to the population, and hence to the patient flows from the HMO's to the hospitals, the estimated interval gets substantially larger and it is estimated with less precision. This is a little unfair to the inequality estimator since, though there may be some uncertainty in the relevant population size and patient flows variables when contracts are signed, we would generally expect to be able to construct good instruments for them from current population size and predicted flows, and we did not use those instruments here. Keeping this case, however, 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 four and five use the data set with both ν_1 and ν_2 errors, but the inequalities from the ν_1-only model. The ratio of the variance in ν_2 to the variance in the dependent variable is 12.7%. Now the estimated bounds are inconsistent; in particular 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. The problem is that they can overshoot, leaving us either with an interval which does not cover the true θ_0 or a point estimate. 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 where the only measurement error is in costs, the specification error introduced by adding ν_2 increases the estimated lower bound by about 25% but the upper bound hardly changes at all, producing a confidence interval that is almost exactly equal to the true identified interval. The effect of ν_2 on the lower bound is much smaller when there is also measurement error in population, but now the upper bound, which was estimated very imprecisely when we did not have a ν_2 error, falls by about 15%. However, the estimated interval still covers the true interval. Apparently estimates from the ν_1-only inequalities do not change dramatically when there is a reasonable amount of ν_2 error (at least in our buyer-seller network example).

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; random draws from the empirical distribution of the actual ν_2, and a normal distribution. The first option would not be available to empirical researchers, but might be closer to the true population distribution. However when using the empirical distribution we are only using an asymptotically correct approximation to the true distribution if the ν_2 were truly independent of our instruments (rather than just mean independent), and this is not likely to be ture (so there is also a misspecification in this model). Note also that the identified set for the ν_2-only estimator depends on the unknown true distribution of ν_2. There will be incorrect "identified sets" associated with the distributions we use but they are difficult to characterize, so we will not be able to say whether the estimated confidence interval covers that set.

Both estimators, that based on the empirical distribution and the normal distribution for the ν_2, generate point estimates, and have estimated confidence intervals that are unusally short. The confidence interval for the estimator which uses the bootstrap distribution does not cover the true θ_0, but the one that uses the normal distribution does. The confidence intervals from the empirical distribution and the normal distribution do not overlap, indicating that the choice of functional form for the ν_2 distribution has a significant impact on the estimators. On the other both estimators appear to be close enough to the truth for most applied issues. The problem is only in the shortness of the confidence intervals, giving what appears to be a misleading impression of the precision of the estimates. This is probably caused by the (likely incorrect) assumption on the distribution of ν_2, a problem which would be hard to avoid in applied work using the ν_2-only estimator.

Just as we added ν_2 variance to the algorithm which uses the ν_1-only inequalities, rows 9 and 10 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 this is what one might expect an empirical researcher would do (though use of the bootstrap distribution produced similar estimates). We still obtain point estimates. The estimated confidence intervals from the data that had ν_1 errors in both costs and population did not cover the true θ_0, but that from the data that only had errors in cost did. The point estimates themselves seem reasonably close the true value of the parameter (though a little further away with the ν_1 errors in both population and costs), but the confidence intervals are still unusually short. That is, at least in this example, the primary problem with the nu_2-only estimator appears to be in providing a misleading impression of the precision of the estimates.

We tried to compute estimates from both the ν_1 and the ν_2-only algorithms using Ho's actual data set. The ν_1-only algorithm generated a point estimate, but with a reasonably large confidence interval (column 6 panel A). The ν_2-only estimator *could not* be computed on the real data set; its computational burden is just too large. 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 to come. In contrast none of the other estimates that used Ho's data took more than an hour of computer time.

Panel C provides the estimates from the algorithm which uses the robust inequalities introduced in the last subsection. The fact that we have so few inequalities from our simulated data implies that the robust inequalities do not deliver an upper bound with these data. The lower bound is lower than the bound obtained when we used the ν_1 only inequalities, but it is not that much lower. Using Ho's data 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.

We conclude that the ν_1-only estimator is easy to use and, at least in the buyer-seller network problem, seems reasonably robust to a moderate amount of ν_2 variance. The ν_2-only estimator suffers from two problems. First it requires a distributional assumption which is hard to verify and does have an impact on the estimated confidence interval, and second it can have a very large computational burden. In our problem the distributional assumption used seemed to have only small effects on the actual estimates; its major effect was to give a misleading impression of the precision of those estimates. Of course the estimators which use the robust inequalities are least subject to consistency and misleading precision problems, but they can lead to rather large identified sets.

Table 1: **Inequality Estimators: Simulated and Real Data.**

Disturbances	Not In IV	Average LB	Average UB	95% CI of θ LB	95% CI of θ UB
Simulated data: True θ=18.77; true interval with no errors (15.43, 20.62).					
A: Using ν_1 inequalities.					
Only ν_1 disturbances; Simulated Data.					
1. 25% Cost	Cost	14.25	20.77	13.59	21.12
2. 25% Cost,5% pop	Cost, $N_{j,k}$,Pop	13.36	39.85	12.89	47.50
3. 25% Cost,5% pop	Cost	14.24	20.68	13.54	21.13
ν_1 & ν_2 disturbances; Simulated Data.					
4. ν_2, costs, pop	Cost, $N_{j,k}$,Pop	14.26	34.94	13.88	41.06
5. ν_2, costs, pop	Cost	16.22	21.02	15.48	21.27
Actual disturbances; Real Data.					
6. actual disturbances	Cost	8.2	8.2	2.3	16.4
B: Using ν_2 inequalities.					
Only ν_2 disturbances; Simulated Data.					
7. ν_2 (bootstrap dist)		18.59	18.59	18.50	18.75
8. ν_2 (normal dist)		18.80	18.80	18.75	19.00
ν_1 & ν_2 disturbances; Simulated Data.					
9. $\nu_2 \sim \mathcal{N}$, Costs, Pop	Costs Pop	18.33	18.33	18.00	18.50
10 $\nu_2 \sim \mathcal{N}$, Costs	Costs	18.79	18.79	18.75	19.00
Actual Disturbances; Real Data.					
11. Assume ν_2 normal	Could Not Compute.				
C: Using Robust inequalities.					
Robust Inequalities; Simulated Data.					
12. ν_2, costs	Cost	14.52	n.b.	14.15	n.b.
13. ν_2, costs, pop	Cost,$N_{j,k}$,Pop	14.25	n.b.	13.88	n.b.
Actual Disturbances; Real Data.					
14. Actual Disturbances	Cost	11.7	11.7	3.6	17.9

Notes. Instruments for ν_1 and robust inequalities case (unless indicated as omitted); constant, $N_{j,k}$, hospital cost characteristic and capacity measures, market cost capacity and population measures, HMO charactersitics, and some interactions among above. Instruments for ν_2 inequalities are market averages and sums of above variables. The model used to estimate on Ho's data allowed also for a cost coefficient; without that coefficient the estimate of the average markup was negative.

7 Concluding Remarks.

It is important to note that there is a sense in which this review is seriously incomplete. The most notable immediate benefits from the advances in the analysis of imperfectly competitive markets over the last decade or two have come from the applied work which uses the techniques outlined here to provide a deeper understanding of the causes and effects of historical events and/or the likely impacts of possible policy changes. Every review must chose among topics, and I have focused on a set of methodological questions which seem to be important to enabling applied work to go further. However there clearly is also a need for a review of applications.

The methodological issues that arose had a lot in common. On the one hand they were questions which have been part of economic theory's research program for some time; examples include the literature on equilibrium conditions in markets with a small number of agents on each side, the literature on learning, and the related material on equilibrium selection. On the other there has been little interaction between those research programs and empirical work on real data sets[18].

There are at least three reasons to be hopeful that a more integrated approach to these problems is likely to be beneficial. One is the simple fact that the emprical researcher now has a better idea of the functional forms that are likely to be relevant in different settings, and this should give the theorist a more concrete base to work from. Second computers, computerized data sets, and some of the tools reviewed here have made it much easier to "test" theorists conjectures, and consequently to go back and forth between theory and empirical work. Finaly, at least in Industrial Organization, the current generation of empirical people tend to be much more theory literate than their predecessors.

References.

Ackerberg, D., K. Caves, and G. Fraser, 2005, "Structural Identification of Production Functions," *mimeo*, UCLA.

Ackerberg, D., Berry, S., Benkard, L., and Pakes, A., (forthcoming);"Econometric Tools for Analyzing Market Outcomes", in J. Heckman and E. Leaner (ed.s) *The Handbook of Econometrics*, Volume 6.

Andrews, D., 1994, "Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity", *Econometrica* 62, 1994, pp.43-72.

[18]The empirical literature has also faced econometric and computational challenges, but here there seems to have been more direct interaction with theory (good examples are the literatures on semiparametrics and on inequality estimators).

Andrews, D., Berry, S., and P. Jia (2004), “Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location,” *mimeo* Yale University.

Bajari, P., L. Benkard, and J. Levin (2006), “Estimating Dynamic Models of Imperfect Competition,” *Econometrica*, forthcoming.

Benkard, L . (2004), “A Dynamic Analysis of the Market for Wide-Bodied Commercial Aircraft,” *Review of Economic Studies*, vol. 71, no. 3, pp. 581-612.

Berry S., and Reiss P., (forthcoming), “Empirical Models of Entry and Market Structure”, *The Handbook of Industrial Organization*, M. Armstrong and R. Porter editors.

Berry, S., J. Levinsohn, and A. Pakes, 1995, ”Automobile Prices in Market Equilibrium,” *Econometrica*, vol. 63, no. 4, pp. 841-890.

Berry, S., J. Levinsohn, and A. Pakes, 2004, ”Estimating Differentiated Product Demand Systems from a Combination of Micro and Macro Data: The New Car Model,” *Journal of Political Economy*, vol. 112, no. 1, pp. 68-105.

Bertsekas, D., J. Tsitsiklis (1996), *Neuro-Dynamic Programming*, Athena Scientific Publications.

Besanko, D., Doraszelski, U., Kryukov, Y., and M. Satterthwaite (2007); “Learning-by-Doing, Organizational Forgetting, and Industry Dynamics”, *mimeo* Northwestern University.

Collard Wexler, A., 2006 “Demand Fluctuations and Plant Turnover in the Ready-Mix Concrete Industry”, *mimeo*, New York University.

Chen, X., forthcoming, “Large Sample Sieve Estimation of Semi-nonparametric Models” in J. Heckman and E. Leaner (ed.s) *The Handbook of Econometrics*, Volume 6.

Cliberto, and E. Tamer, 2006, “Market Structure and Multiple Equilibria in the Airline Markets,” manuscript.

Crawford, G. and M. Shum, 2006,

De Loecker J., forthcoming, “Do Exports Generate Higher Productivity? Evidence from Solovenia.”, *Journal of International Economics*.

Davis, S. and J. Haltwanger, 1992, ”Gross Job Creation, Gross Job Destruction, and Employment Reallocation”” *Quarterly Journal of Economics*, 819-64.

Doraszelski, U. and K. Judd (2004), “Solution Methods for Markov Perfect Nash Equilibria of Continuous Time, Finite State Stochastic Games,”*mimeo*, Hoover Institution.

Doraszelski, U., and A. Pakes (forthcoming); “A Framework for Dynamic Analysis in Applied I.O; Computation.”, *The Handbook of Industrial Organization*, M. Armstrong and R. Porter (ed.s)

Dunne,T., S. Klimek, M.Roberts and Y. Xu, 2005, “Entry and Exit in Geographic Markets”, *mimeo*, Pennsylvania State Univeristy.

Dunne, T., M. Roberts, and L. Samuelson, 1988, ”Patterns of Firm Entry and Exit in U.S. Manufacturing Industries,” *Rand Journal of Economics*, vol. 19, no. 4, pp. 495-515.

Erickson, T., and A. Pakes, 2007, “An Experimental Component Index for the CPI: From Annual Computer Data to Monthly Data on Other Goods”, *mimeo* Harvard University.

Ericson, R., and A. Pakes, 1995, “Markov-Perfect Industry Dynamics: A Framework for Empirical Work,” *Review of Economic Studies*, vol. 62, no. 1, pp. 53-83.

Esponda, I, forthcoming, “Information Feedback in First Price Auctions”, the *Rand Journal of Economics*.

Esteban, S., and M. Shum, 2007, “Durable Goods Oligopoly iwth Secondary Markets: The Case of Automobiles” the *Rand Journal of Economics*.

Fershtman, C. and A. Pakes, 2005, “Finite State Dynamic Games with Asymmetric Information: A Computational Framework,” mimeo, Harvard University.

Fudenberg, D. and D. Levine, 1993, “Self-Confirming Equilibrium”, *Econometrica*, Vol. 61, pp. 523-45. Cambridge: MIT Press.

Fudenberg, D. and D. Levine, 1999, *Learning in Games*, Cambridge: MIT Press.

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

Gorman, T., 1959, “Separable Utility and Aggregation”, *Econometrica* Vol. 27, pp. 469-81.

Gowrisnakaran G., and M. Rysman (2007)

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.

Hendel, I. and A. Nevo, forthcoming,

Ho, K.,2006, "Insurer-Provider Networks in the Medical Care Market " *mimeo* Columbia University.

Ho, K.,2006, "The Welfare Effects of Restricted Hospital Choice in the US Medical Care Market" *Journal of Applied Econometrics.*

Hotz, J. and R. Miller

Hotelling, H., 1929, "Stability in Competition," *Economic Journal*, vol. 39, pp. 41-57.

Houthakker, H.S., 1955, "The Pareto Distribution and the Cobb-Douglas Production Function," *Review of Economic Studies*, vol. 23, no. 1, pp. 27-31.

Ishii, J. ,2006, "Interconnection Pricing, Compatibility, and Investment in Network Industries: ATM Networks in the Banking Industry", *mimeo* Stanford University.

Jia, P.,2006, "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the DIscount Retail Industry",*mimeo* MIT.

Judd K.,2007,

Katz, M., in process, "Supermarkets and Zoning Laws", *mimeo* Harvard University.

Krasnokutskaya, E, 2006, "Identification and Estimation in Highway Procurement Auctions under Unobserved Auction Heterogeneity," *mimeo* University of Pennsylvania.

Lancaster, K., 1971, Consumer Demand: A New Approach, Columbia University Press, New York.

Lee, R. ,2007,

Levinsohn, J., and A. Petrin, 2003, "Estimating Production Functions Using Inputs to Control for Unobservables," *The Review of Economic Studies*, vol. 70, no. 2, pp. 317-341.

Maskin, E. and J. Tirole (1988) "A Theory of Dynamic Oligopoly: I and II" *Econometrica*, Vol.56, pp.549-99.

McFadden, D., 1974 "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka eds. Frontiers of Econometrics, Academic Press, New York.

McFadden, D., 1981, "Econometric Models of Probabilistic Choice," in C. Manski and D. McFadden, eds. Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, MA.

Mclennan, A., 2005 "The Expected Number of Nash Equilibria of a Normal Form Game" *Econometrica*, pp. 141-74.

Melnikov, O., 2001,.

Morris, S. and H. Shin, 2002, “Heterogeneity and Uniqueness in Interaction Games”, *mimeo* Yale University.

Nakamura, E., 2007, “Accounting for Incomplete Pass-Through” *mimeo*, Harvard University.

Newey, W., 1994, “ The Asymptotic Variance of Semiparametric Estimators”, *Econometrica* 62, 1994, pp.1249-82.

Olley, S., and A. Pakes (1996), “The Dynamics of Productivity in the Telecommunications Equipment Industry,” *Econometrica*, vol. 64, no. 6, pp. 1263-1298.

Pakes, A., 1986, ”Patents as Options: Some Estimates of the Value of Holding European Patent Stocks,” *Econometrica*, vol. 54, pp. 755-784.

Pakes, A., 2004, ”Common Sense and Simplicity in Industrial Organization” the *Review of Industrial Organization*, vol. 23, pp. 193-215.

Pakes, A., 2008, “A Note on the Analysis of Multiple Equilibria in an Applied Setting”, *mimeo*, Harvard University.

Pakes, A. and P. McGuire (1994), “Computing Markov-Perfect Nash Equilibria: Numerical Implications of a Dynamic Differentiated Product Model,” *Rand Journal of Economics*, vol. 25, no. 4, pp. 555-89.

Pakes, A., and P. McGuire (2001), “Stochastic Approximation for Dynamic Models: Markov Perfect Equilibrium and the ‘Curse’ of Dimensionality,” *Econometrica*, vol. 69, no. 5, pp. 1261-81.

Pakes, A., M. Ostrovsky, and S. Berry (2006), “Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry/Exit Examples),” *RAND Journal of Economics*, forthcoming.

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

Rust, J. 1987

Rust, J. (1994), “Structural Estimation of Markov Decision Processes,” in R. F. Engle and D. L. McFadden (eds.), *Handbook of Econometrics*, Volume IV, Elsevier Science, Amsterdam.

Ryan, S.

Starr, A.W. and Y.C. Ho (1969); “Nonzero-Sum Differential Games”, *Journal of Optimization Theory and Applications*, pp.1984-208.

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

Young, H.P.,2004, *Strategic Learning and Its Limits*, Oxford UK, Oxford University Press.

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[19]; i.e. that $\forall(h,m),\ \ \nu_{2,\mathtt{m},h} = \nu_{2,\mathtt{m}}$. 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 $\tilde{h} \notin H_{\mathtt{m}}$ and $h \in H_{\mathtt{m}}$ we have

$$\Delta\pi^M_{\mathtt{m}}(H_{\mathtt{m}}, H_{\mathtt{m}} \cup \tilde{h}, \cdot) + \Delta\pi^M_{\mathtt{m}}(H_{\mathtt{m}}, H_{\mathtt{m}} \backslash h, \cdot) = \Delta r^M_{\mathtt{m}}(H_{\mathtt{m}}, H_{\mathtt{m}} \cup \tilde{h}, \cdot) + \Delta r^M_{\mathtt{m}}(H_{\mathtt{m}}, H_{\mathtt{m}} \backslash h, \cdot),$$

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

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

For the second inequality note that if $\nu_{2,m,h} = \nu_{2,m}$ we can use the logic leading to equation (13) in the text to show that for any positive valued function, $h(\cdot)$

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

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

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

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

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

[19]A more complete analysis of effects models in buyer seller networks would allow for both buyer and seller effects. This is a straightforward, though somewhat tedious, extension of the results below. 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.