Moment Inequalities and Their Application

A. Pakes, J. Porter, Kate Ho, and Joy Ishii*

February 28, 2014

Abstract

This paper provides conditions under which the inequality constraints generated by either single agent optimizing behavior, or by the best response condition of multiple agent problems, can be used as a basis for estimation and inference. An application illustrates how the use of these inequality constraints can simplify the analysis of complex behavioral models.

1 Introduction

This paper provides conditions under which the inequality constraints generated by single agent optimizing behavior, or by the best response condition of multiple agent games, can be used as a basis for estimation and inference. The conditions do not restrict choice sets or the equilibrium selection mechanism, and do not require the researcher to specify either a parametric form for the disturbance distributions or the contents of agents' information sets. However, assumptions are required on the difference between the econometrician's measure of returns and the returns the agent is maximizing, and this approach will often result in *partial* (instead of point) identification of the parameters of interest.

The next section begins by assuming that agents maximize their expected returns. This assumption yields a "revealed preference" inequality; the expected returns from the strategy played should be at least as large as that from alternative feasible strategies. We avoid the restrictions that would be imposed by modeling how expectations are formed. Instead we focus on the implications of revealed preference on the difference between

*The authors are from Harvard University, Research, the University of Wisconsin, Columbia University, and Stanford University. We thank four referees, the editor, and numerous seminar participants.

the realized returns from the agent's observed strategy and the returns the agent would have earned had it played a feasible alternative. When there are interacting agents these inequalities are necessary conditions for any (of the possibly many) Nash equilibrium. The advantage of using these inequalities to guide estimation is that we can use them without detailing a model for strategies as functions of the agents' information sets.

We assume that the econometrician constructs an approximation to the returns that depends on agents' actions, other observables, and a finite dimensional parameter vector. The difference between the approximated returns at the observed and an alternative feasible choice is used in estimation. We emphasize two ways in which the econometrician's approximation of profit differences is distinct from the expected profit difference expressed in the best response assumption. One source of (unobserved) error comes from the difference between the actual returns and econometrician's measure of returns. The other source of error comes from observing information on realized, rather than expected, returns. The expectational error and possibly part of the approximation error will be mean independent of the agent's information set (and hence of the choice itself). However there may be a component of the agent's perceived difference in expected profits that the econometrician does not control for and is both a part of the approximation error and a determinant of the choice made. We call this component the "structural" disturbance.

When we can measure profits up to a mean zero measurement error, then there is no structural disturbance. In this case, the proposed identification and estimation algorithm is particularly simple and powerful. When the structural disturbance is nonzero, then a classic selection problem arises. The structural disturbance associated with the observed choice will necessarily come from the possible values that make the observed choice best. As a result, even if the *a priori* mean of the structural disturbances for any fixed choice is zero, the mean of the structural disturbance corresponding to the returns from the observed choice can be non-zero. To deal with this possibility, we propose a condition that can viewed as generalizing an instrumental variables approach to this inequality setting. The formal assumption provides a "high level" sufficient condition for dealing with selection, and we show a number of ways of satisfying this condition in particular examples. The paper closes with an empirical example; estimating the costs of a non-convex (or lumpy) investment choice by banks. The example illustrates ways of circumventing problems posed by the structural error in models with and without boundary conditions. The detailed policy implications of the estimates are discussed in Ishii (2004).

When there is no structural disturbance the proposed framework is a natural extension of the first-order condition estimator for single agent dynamic problems proposed in Hansen and Singleton (1982), and extended to allow for transaction costs, and hence inequalities, by Luttmer (1996). Our generalization allows for arbitrary (including discrete) choice sets and interacting agents. Ciliberto and Tamer (2009) provide alternative methods for estimating models with discrete choice sets and interacting agents that only allows for the structural error. The two approaches are not nested and Pakes (2010) provides both a formal and a Monte Carlo comparison of the two sets of assumptions.[1]

2 A Framework for the Analysis

This section outlines a set of behavioral and statistical assumptions that generate moment inequalities that are informative about payoff or profit functions.[2] We start from a player's best response condition in a simultaneous move game (this reduces to a revealed preference condition for single agent problems). Then we add two assumptions. One assumption allows us to compute counterfactual profits, and the other constrains the relationship between the agent's perceived profits and the profits that we can actually measure.

The Agent's Decision Problem. Suppose there are n decision-making agents indexed $i = 1, \ldots, n$. Let $\mathcal{J}_i \in \mathcal{I}_i$ denote the information set available to agent i when actions are chosen, and $\mathcal{D}_i$ be the set of actions agent i could take. Then the strategy played by agent i is a mapping $s_i : \mathcal{I}_i \to \mathcal{D}_i$. The strategy and information set for each player generate observed decisions $\mathbf{d}_i = s_i(\mathcal{J}_i)$ (so boldface $\mathbf{d}_i$ is a random variable, and d_i will denote the realization of the decision). For notational convenience we assume these are pure strategies.[3] Notice that we place no restrictions on $\mathcal{D}_i$. If $\mathcal{D}_i \subset \mathcal{R}$ it can be either a finite subset ("discrete choice"), countable ("ordered choice"), or uncountable and either bounded (so "corner solutions" are possible) or not. If $\mathbf{d}_i$ is vector-valued, as is typical in

[1] There are also a number of papers that use inequalities to simplify estimation algorithms that, absent computational problems, could be estimated in standard ways; e.g. Bajari, Benkard, and Levin (2007).

[2] A more detailed discussion of these assumptions and examples that use them can be found in the longer version of this paper available on the authors' web sites.

[3] We could obtain a moment inequality of the same form as the inequality derived below from a game in which agents used mixed strategies, provided each pure strategy assigned a positive probability in the mixed strategy has the same expected return. This implies that when using our inequalities there is no need for the econometrician to specify whether the underlying strategies are pure or mixed. However if we did know mixed strategies were being played, and we could distinguish the mixed strategies associated with particular information sets, then more information would be available for use in estimation; see Beresteanu and Molinari (2008).

contracting problems, then $\mathcal{D}_i$ is a subset of the appropriate product space.

The payoff (or profit) to agent i ($\pi(\cdot)$) is determined by its decision, the other agents' decisions ($d_{-i} \in \mathcal{D}_{-i} \equiv \times_{j \neq i} \mathcal{D}_j$), and an additional set of variables $\mathbf{y}_i \in \mathbf{Y}_i$; so $\pi : \mathcal{D}_i \times \mathcal{D}_{-i} \times \mathbf{Y}_i \rightarrow \mathcal{R}$. The functions π and s_i ($i = 1, \ldots, n$), and the joint probability distribution of $(\mathcal{J}_i, \mathbf{Y}_i)_{i=1}^n$ are primitives of the game. So the expectation operator appearing in our assumptions (i.e., $\mathcal{E}(\cdot)$) is with respect to this joint distribution,[4] and observed decisions $\mathbf{d}_i$ are generated by these strategies and information sets.

Assumption 1 (Best Response Condition) *If s_i is the strategy played by agent i*

$$sup_{d \in \mathcal{D}_i} \mathcal{E}[\pi(d, \mathbf{d}_{-i}, \mathbf{y}_i) | \mathcal{J}_i, \mathbf{d}_i = d] \leq \mathcal{E}[\pi(\mathbf{d}_i, \mathbf{d}_{-i}, \mathbf{y}_i) | \mathcal{J}_i, \mathbf{d}_i = s_i(\mathcal{J}_i)], \quad (a.s.\ \mathcal{J}_i),$$

for $i = 1, \ldots n$. ♠

In single agent problems Assumption 1 is a consequence of optimizing behavior. When there are multiple interacting agents, Assumption 1 is a necessary condition for any Bayes-Nash equilibrium. It does not rule out multiple equilibria and does not restrict the selection mechanism used when there are multiple equilibria. Equally important, Assumption 1 does not put any restriction on the contents of $\mathcal{J}_i$ or on the functional form of the mapping from $\mathcal{J}_i$ to $\mathbf{d}_i$ (which depends on the form of the agents' priors); aspects of the problem the analyst typically knows little about.

Models that involve maximizing behavior require assumptions on the agent's perceptions of what the outcome would have been had it chosen an alternative action. This generates a need for an assumption on the agent's perceptions of what $\mathbf{d_{-i}}$ and $\mathbf{y}_i$ would have been had the agent taken action d different from the action $s_i(\mathcal{J}_i)$ given by their strategy. In either single agent problems, or multiple agent problems with simultaneous moves, conditional independence of other agents' decisions (of $\mathbf{d}_{-i}$) from $\mathbf{d}_i$ is an assumption of the model.[5] However often the profit function is naturally written as a function of

[4] We could have defined the expectation operator that results from the agents' perceptions, and then put constraints on the relationship between the agents' perceptions and the expectation operator emanating from the data generating process. Though correct perceptions are certainly sufficient for Assumption 1, they are not necessary; see the literature cited in Pakes, 2010.

[5] In non-simultaneous move games, when considering counterfactuals for agents who move early, $\mathbf{d}_{-i}$ typically includes the decisions of those who move later, and the distribution of $\mathbf{d}_{-i}$ conditional on $(\mathcal{J}_i, \mathbf{d}_i = d)$ will typically depend on the value d. One way to construct counterfactuals is to develop a model for the beliefs of the early period agents about the effect of their decisions on the behavior of later period agents. Alternatively one could compute the later agents' responses (to early agent actions) that minimize profits by the early agent. Either strategy could be used to form inequalities based on counterfactuals as is done here (for an empirical example see Crawford and Yurukoglu, 2012).

variables $\mathbf{y}_i$ that would change with different actions d taken by agent i. In our empirical study we analyze the number of ATMs chosen by banks. The profits a bank earns from its ATM investments depend on the equilibrium interest rates in the periods in which those ATMs will be operative. So the profit function is a function of interest rates, the number of ATMs, and other variables. The interest rates, in turn, depend on the number of ATMs installed by the banks. So to compute our counterfactual profits we need a model of what the agent perceives interest rates would have been were it to choose a $d \neq s_i(\mathcal{J}_i)$. Moreover, we require that the model yields interest rates as a function of variables whose distribution is independent of the choice d.

More generally, if $\mathbf{y}_i$ is endogenous in the sense that its distribution depends on the choice d_i, then we require a model which generates $\mathbf{y}_i$ conditional on $\mathbf{d}_{-\mathbf{i}}$ and variables, say $\mathbf{z}_i$, that do not change when $\mathbf{d}_i$ changes.[6] Assumption 2 formalizes this condition. Note that not all of the variables in $\mathbf{z}_i$ need to be observed by the researcher.

Assumption 2 (Counterfactual Condition) $\mathbf{y}_i = y(\mathbf{z}_i, d, \mathbf{d}_{-\mathbf{i}})$ *and the distribution of* $(\mathbf{d}_{-i}, \mathbf{z}_i)$ *conditional on* $\mathcal{J}_i$ *and* $\mathbf{d}_i = d$ *does not depend on* d. ♠

Assumption 2 implies that Assumption 1 can be rewritten without conditioning on different d. That is, if we substitute $y(\mathbf{z}_i, d, \mathbf{d}_{-\mathbf{i}})$ for $\mathbf{y}_i$ in $\pi(d_i, \mathbf{d}_i, \mathbf{y}_i)$, and define

$$\Delta\pi(d, d', d_{-i}, z_i) = \pi(d, d_{-i}, y(z_i, d, d_{-i})) - \pi(d', d_{-i}, y(z_i, d', d_{-i})),$$

then, recalling that $\mathbf{d}_i = s_i(\mathcal{J}_i)$, Assumptions 1 and 2 imply that for any $d' \in \mathcal{D}_i$,

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

Assumptions 1 and 2 are common in econometric models of decision making. The distinction in the current work is that estimation and inference are based directly on the empirical analogues of the profit inequalities in equation (1), rather than on the model's implications for $\mathbf{d}_i$. As a result we require a different set of measurement and informational assumptions. In particular the distribution of $\mathbf{d}_i$ conditional on observables can remain unspecified, but the disturbances in the measures of $\mathcal{E}[\Delta\pi(\mathbf{d}_i, d', \mathbf{d}_{-i}, \mathbf{z}_i)|\mathcal{J}_i]$ will need to be treated. We turn to this task now.

[6]If there is not a one-to-one map between $\mathbf{y}_i$ and $(\mathbf{d}_i, \mathbf{z}_i)$ conditional on $\mathbf{d}_{-i}$ and exogenous variables (or if the researcher is not sure of what that map is), but the researcher can construct a lower bound to the counterfactual profits that the agent could make, the researcher can replace the counterfactual profits in Assumption 2 with that lower bound.

Observables and Disturbances. We assume that the econometrician has a parametric function, say $r(\cdot)$, that approximates $\pi(\cdot)$. $r(\cdot)$ has arguments d_i, d_{-i}, an *observable* vector of the determinants of profits, say $z_i^o \subset z_i$ (so $\mathbf{z}_i^o$ satisfies Assumption 2), and θ. The parameter $\theta \in \Theta$ has an unknown true value of θ_0. Our approximation to $\Delta\pi(d, d', d_-, z)$ is $\Delta r(d, d', d_-, z^o, \theta)$, which is obtained by evaluating $r(\cdot)$ at d and d' and taking the difference, so $\Delta r : \mathcal{D}_i^2 \times \mathcal{D}_{-i} \times Z^o \times \Theta \to \mathcal{R}$. For $i = 1, \ldots, n$ and $(d, d') \in \mathcal{D}_i^2$ define

$$\nu_{2,i,d,d'} = \mathcal{E}[\Delta\pi(d, d', \mathbf{d}_{-i}, \mathbf{z}_i)|\mathcal{J}_i] - \mathcal{E}[\Delta r(d, d', \mathbf{d}_{-i}, \mathbf{z}_i^o, \theta_0)|\mathcal{J}_i], \quad \text{and} \tag{2}$$

$$\nu_{1,i,d,d'} = \nu^{\pi}_{1,i,d,d'} - \nu^{r}_{1,i,d,d'}, \quad \text{where} \tag{3}$$

$$\nu^{\pi}_{1,i,d,d'} = \Delta\pi(d, d', \mathbf{d}_{-i}, \mathbf{z}_i) - \mathcal{E}[\Delta\pi(d, d', \mathbf{d}_{-i}, \mathbf{z}_i)|\mathcal{J}_i], \quad \text{and}$$

$$\nu^{r}_{1,i,d,d'} = \Delta r(d, d', \mathbf{d}_{-i}, \mathbf{z}_i^o, \theta_0) - \mathcal{E}[\Delta r(d, d', \mathbf{d}_{-i}, \mathbf{z}_i^o, \theta_0)|\mathcal{J}_i].$$

We let $\nu_{2,i}$ be the collection of random variables $\nu_{2,i,d,d'}$ over values $(d, d') \in \mathcal{D}_i^2$, and define $\nu_{1,i}$, $\nu^r_{1,i}$, $\nu^\pi_{1,i}$ similarly.

The function $\Delta r(d, d', \mathbf{d}_{-i}, \mathbf{z}_i^o, \theta)$ is the observable measure of the change in profits that would result from a change of $d_i = d$ to $d_i = d'$. The random variables $\nu_{1,i}$ and $\nu_{2,i}$ are the determinants of the true profit difference that are *not observed* by the econometrician. $\nu_{1,i}$ and $\nu_{2,i}$ differ in what agent i knows about them at the time decisons are made. Agent i "knows" its $\nu_{2,i}$ realization *before* it makes its decision ($\nu_{2,i} \in \mathcal{J}_i$), and since $\mathbf{d}_i = s_i(\mathcal{J}_i)$, $\mathbf{d}_i$ can depend on the values of $\nu_{2,i}$. Consequently the expectation of $\nu_{2,i,\mathbf{d}_i,d'}$ (for the $\mathbf{d}_i$ chosen) will depend on $\mathcal{J}_i$ and this can cause a selection problem. In contrast the values of $\nu_{1,i}$ do not change expected profits so $\nu_{1,i,\mathbf{d}_i,d'}$ is mean independent of $\mathcal{J}_i$. The $\nu_{1,i}$ are generated by expectational errors ($\nu^\pi_{1,i}$) and/or approximation errors ($\nu^r_{1,i}$). There are two sources of expectational errors: (i) incomplete information on the environmental variables (on $\mathbf{z}_i$); and (ii) asymmetric information which generates uncertainty in $\mathbf{d}_{-i}$.

The relative importance of $\nu_{1,i}$ and $\nu_{2,i}$ will vary across problems. One advantage of working directly with profit functions, rather than with the implications of those functions on the choice of d, is that some applied problems have access to (usually error prone) profit measures. Then it might suffice to only allow for a ν_1 error. There will also be a need for a ν_2 error if the profit measure omits sources of returns that help determine $d_i = s_i(\mathcal{J}_i)$.

Moment Inequalities. The econometrician only has access to Δr, and equation (1) is expressed in terms of the conditional expectation of $\Delta\pi$. So to use that equation to restrict the moments we obtain from Δr we need restrictions on the distributions of $\nu_{2,i}$ and $\nu_{1,i}$. We now provide conditions under which weighted averages of $\Delta r(\mathbf{d}_i, d', \mathbf{z}_i^o, \theta)$ across values of d' and i have non-negative expectation when $\theta = \theta_0$. The weights can be any nonnegative function of observable components of the agents' information sets.

Assumption 3 *Let $h^i(d'; \mathbf{d}_i, \mathcal{J}_i, \mathbf{x}_{-i}) : \mathcal{D}_i \to \mathbb{R}^+$ be a nonnegative function whose value can depend on the alternative choice considered (on d'), on the information set $\mathcal{J}_i$ (which determines $\mathbf{d_i}$), and on observable components of the other agents' information sets, $\mathbf{x}_{-i} \subset \times_{j\neq i} \mathcal{J}_j$. Assume that*

$$(a) \qquad \mathcal{E}[\sum_{i=1}^{n} \sum_{d' \in \mathcal{D}_i} h^i(d'; \mathbf{d}_i, \mathcal{J}_i, \mathbf{x}_{-i}) \nu_{2,i,\mathbf{d}_i,d'}] \leq 0,$$

and

$$(b) \qquad \mathcal{E}[\sum_{i=1}^{n} \sum_{d' \in \mathcal{D}_i} h^i(d'; \mathbf{d}_i, \mathcal{J}_i, \mathbf{x}_{-i}) \nu^r_{1,i,\mathbf{d}_i,d'}] \geq 0. \quad \spadesuit.$$

Before discussing the economic content of this assumption we show why it suffices. Assumption 3 combined with non-negative expected profit differentials (Assumptions 1 and 2) yields a weighted sum of profit differences that has non-negative expectation. From (2) and (3), $\Delta r(\mathbf{d}_i, d', \mathbf{d}_{-i}, \mathbf{z}_i^o, \theta_0) = \mathcal{E}[\Delta\pi(\mathbf{d}_i, d', \mathbf{d}_{-i}, \mathbf{z}_i)|\mathcal{J}_i] + \nu^r_{1,i,\mathbf{d}_i,d'} - \nu_{2,i,\mathbf{d}_i,d'}$, so

$$\begin{aligned} &\mathcal{E}\left[\sum_{i=1}^{n} \sum_{d' \in \mathcal{D}_i} h^i(d'; \mathbf{d}_i, \mathcal{J}_i, \mathbf{x}_{-i}) \Delta r(\mathbf{d}_i, d', \mathbf{d}_{-i}, \mathbf{z}_i^o, \theta_0)\right] \\ &\quad = \mathcal{E}\left[\sum_{i=1}^{n} \sum_{d' \in \mathcal{D}_i} h^i(d'; \mathbf{d}_i, \mathcal{J}_i, \mathbf{x}_{-i}) \mathcal{E}[\Delta\pi(\mathbf{d}_i, d', \mathbf{d}_{-i}, \mathbf{z}_i)|\mathcal{J}_i]\right] \\ &\qquad + \mathcal{E}\left[\sum_{i=1}^{n} \sum_{d' \in \mathcal{D}_i} h^i(d'; \mathbf{d}_i, \mathcal{J}_i, \mathbf{x}_{-i}) \nu^r_{1,i,\mathbf{d}_i,d'}\right] - \mathcal{E}\left[\sum_{i=1}^{n} \sum_{d' \in \mathcal{D}_i} h^i(d'; \mathbf{d}_i, \mathcal{J}_i, \mathbf{x}_{-i}) \nu_{2,i,\mathbf{d}_i,d'}\right] \end{aligned} \tag{4}$$

Since $\mathcal{E}[\Delta\pi(s_i(\mathcal{J}_i), d'\mathbf{d}_{-i}, \mathbf{z}_i)|\mathcal{J}_i] \geq 0$ by Assumptions 1 and 2, and the weights $(\{h^i(d'; \mathbf{d}_i, \mathcal{J}_i, \mathbf{x}_{-i})\})$ are non-negative, the first term is non-negative. The two conditions of Assumption 3 insure that the last two summands are also nonnegative. So given our three Assumptions

$$\mathcal{E}\left[\sum_{i=1}^{n}\sum_{d'\in\mathcal{D}_i} h^i(d';\mathbf{d}_i,\mathcal{J}_i)\Delta r(\mathbf{d}_i,d',\mathbf{d}_{-i},\mathbf{z}_i^o,\theta_0)\right] \geq 0. \tag{5}$$

Equation (5) depends only on observables and θ_0, so we can form its sample analog and look for values of θ that satisfy the resulting inequality.[7]

We now consider the substantive content of Assumption 3. Notice that if the weight function h^i does not depend on $\mathbf{x}_{-i}$ then Assumption 3(b) is automatically satisfied by the construction of $\nu^r_{1,i}$, since it implies $\mathcal{E}[\nu^r_{1,i,\mathbf{d}_i,d'}|\mathcal{J}_i] = 0$. This implies that Assumption 3(b) is satisfied in single agent problems.[8] We include $\mathbf{x}_{-i}$ in the weight function to provide more flexibility in satisfying Assumption 3(a) in multiple agent problems. Assumption 3(b) will be satisfied with $\mathbf{x}_{-i}$ in the weight function if $\nu^r_{1,i,\mathbf{d}_i,d'}$ is mean independent of $\cup_j\mathcal{J}_j$, since $x_{-i} \in \cup_j\mathcal{J}_j$. This condition holds in symmetric information games. So Assumption 3(b) can only be problematic in games with asymmetric information when the weight function depends on variables that are not a part of i's information set. Our empirical example shows how this can happen when the weights depend nontrivially on $\mathbf{d}_{-i}$.

When $\nu_{2,i,d,d'} \equiv 0$, Assumption 3(a) is automatically satisfied and our measure of the difference in profits (Δr) is an unbiased measure of the actual differenced profits. Assumptions 1 and 2 insure that agents actions are best responses so if Assumption 3(b) is satisfied the weighted average of the expected difference in returns in equation (5) will be nonnegative at $\theta = \theta_0$. This extends Hansen and Singleton (1982)'s first order condition estimator to problems with discrete and/or bounded choice sets and interacting agents.

Now consider the case where $\nu_{2,i,d,d'} \neq 0$ so Assumption 3(a) is not automatically satisfied. Notice that the selected choice, $\mathbf{d}_i = s_i(\mathcal{J}_i)$, must satisfy Assumption 1 which, in turn, implies that larger, positive values of $\nu_{2,i,\mathbf{d}_i,d'}$ are more likely. This selection problem is the chief challenge to weight function selection to satisfy Assumption 3(a), which states that the weighted sum of $\nu_{2,i,\mathbf{d}_i,d'}$ has *nonpositive* expectation. Our empirical example illustrates two weight function choices that satisfy Assumption 3(a). First, we use an ordered choice problem to illustrate that if particular counterfactual choices lead to expressing $\nu_{2,i,\mathbf{d}_i,d'}$ as a linear function of a fixed error distribution, then the selection

[7] Assumptions 1, 2, and 3 are sufficient but not necessary for equation (5) which, in turn, generates our estimates. That is, alternative conditions may also suffice.

[8] As shown in Morales (2011), this can be used to simplify the analysis of dynamic discrete and/or bounded choice problems by eliminating the need for a nested fixed point algorithm.

problem can be avoided. Next, we show that symmetry assumptions, similar to those used by Powell (1986), can be used in conjunction with inequalities to circumvent selection problems when choice sets are bounded.[9]

3 An Example: Semiparametric Ordered Choice.

Our empirical example is based on Ishii (2004) who studies the welfare implications of alternative market designs for ATM networks. Her analysis requires estimates of the cost of installing and operating ATMs. We show that this can be treated as a multiple agent ordered choice estimation problem and that a simple inequality estimator enables us to infer costs without making parametric assumptions on unobserved disturbance distributions. Many "lumpy" investment models can be analyzed similarly.[10] We conclude by noting that the restriction used in this example is a special case of a more general restriction that is appropriate in a wider range of cases.

The framework for Ishii's analysis is a two period model with simultaneous moves in each period. In the first period each bank chooses a number of ATMs to maximize its expected profits given its perceptions on the number of ATMs likely to be chosen by its competitors. In the second period interest rates are set conditional on the ATM network, and consumers choose banks, make deposits, and use ATMs. The second period game is analyzed in Ishii (2004), and her results can be used to generate the profits that would be earned by each bank conditional on any choices for its own and its competitors' ATMs.[11]

We use Ishii's results on the second stage to analyze the game's first stage, the choice of ATMs. Let $d_i \in \mathcal{D} = \mathcal{Z}_+$ denote the number of ATM's chosen by bank i. The second stage provides the bank revenue conditional on d_i, d_{-i}, and observables (z_i^o) which do not change with d. Denote this revenue by $R(d_i, d_{-i}, z_i^o)$. The marginal cost of an ATM for

[9] Ho and Pakes (2012) provide another example where the unobservable driving the selection problem is additively separable and has the same value for two agents (the assumption underlying estimators which use matched observations to analyze continuous dependent variables). In this case, a difference in difference inequality can be constructed to satisfy Assumption 3(a).

[10] For a single agent example without a $\nu_{2,i}$ error see Holmes (2011).

[11] Ishii (2004) estimated a demand system for banking services and an interest rate setting equation. Since we need the equilibrium interest rates that would prevail were alternative possible networks in place we must either assume a unique interest rate setting equilibrium, or common knowledge about which equilibrium is selected. Given the demand system and interest rates, the banks' earnings are calculated as the earnings from the credit instruments funded by the deposits minus the costs of the deposits (including interest costs) plus the fees associated with ATM transactions. The ATM fee revenue is generated when non-customers use a bank's ATMs, and revenue is both generated and paid out when customers use a rival's ATMs.

bank i is $\theta_0 + \eta_i$, where θ_0 is the average (across firms) marginal cost (so $\mathcal{E}\eta_i \equiv 0$). η_i captures bank-specific cost heterogeneity known to the bank when it determines d_i but not observed by the econometrician.

Allowing for mean zero errors (denoted $\nu_{1,i,d}$) in the specification of the banks profits, we have: $\pi(d_i, \mathbf{d_{-i}}, \mathbf{z}_i) = R(d_i, \mathbf{d}_{-i}, \mathbf{z}_i^o) - d_i(\theta_0 + \eta_i) + \nu_{1,i,d_i}$, where we assume $\mathcal{E}(\nu_{1,i,d}|\mathcal{J}_i) = 0$. So if $\Delta\pi$, ΔR, and $\nu_{1,i,d,d'}$ are defined as the corresponding differences, then for any (d, d')

$$\Delta\pi(d, d', \mathbf{d_{-i}}, \mathbf{z}_i) = \Delta R(d, d', \mathbf{d_{-i}}, \mathbf{z}_i^o) + (d' - d)(\theta_0 + \eta_i) + \nu_{1,i,d,d'}.$$

Notice that, in the notation of section 2, the econometrician's approximation to differenced profits is $\Delta r(d, d', \mathbf{d_{-i}}, \mathbf{z}_i^o; \theta) = \Delta R(d, d', \mathbf{d_{-i}}, \mathbf{z}_i^o) + (d' - d)\theta$, so $\nu_{2,i,d,d'} = (d' - d)\eta_i$.[12] Assumptions 1 and 2 insure that if $d' \in \mathcal{D}$, $\mathcal{E}[\Delta\pi(\mathbf{d}_i, d', \mathbf{d_{-i}}, \mathbf{z}_i)|\mathcal{J}_i] = \mathcal{E}[\Delta R(\mathbf{d}_i, d', \mathbf{d_{-i}}, \mathbf{z}_i^o) + (d' - \mathbf{d}_i)(\theta_0 + \eta_i)|\mathcal{J}_i] \geq 0$ regardless of the equilibrium selection rule (and, since this is a network game, there are likely to be many equilibria).

Next we exhibit weights that satisfy Assumption 3(a). Consider a counterfactual choice of $d' = d_i + t$; a fixed, positive number of units (t) away from d_i. Then, $\nu_{2,i,d_i,d_i+t} = t\eta_i$, which does depend on d_i, and

$$\mathcal{E}\nu_{2,i,d_i,d_i+t} = t\mathcal{E}(\eta_i) = 0. \tag{6}$$

Assume that the counterfactual $d' = \mathbf{d}_i + t$ is feasible for all i (the case where it is not is considered below). Taking $h(d'; \mathbf{d}_i, \mathcal{J}_i) = n^{-1}$ if $d' = \mathbf{d}_i + t$, and zero otherwise, we have

$$\mathcal{E}[\sum_{i=1}^{n} \sum_{d' \in \mathcal{D}_i} h(d'; \mathbf{d}_i, \mathcal{J}_i)\nu_{2,i,\mathbf{d}_i,d'}] = \mathcal{E}[n^{-1} \sum_{i=1}^{n} \nu_{2,i,\mathbf{d}_i,\mathbf{d}_i+t}] = n^{-1} \sum_{i=1}^{n} t\mathcal{E}(\eta_i) = 0.$$

Notice that Assumption 3(b) is also satisfied, $\mathcal{E}[n^{-1} \sum_i h(d'; \mathbf{d}_i, \mathcal{J}_i)\nu^r_{1,i,\mathbf{d}_i,d'}]$ $= \mathcal{E}[n^{-1} \sum_i h(d'; \mathbf{d}_i, \mathcal{J}_i)\mathcal{E}(\nu^r_{1,i,\mathbf{d}_i,d'}|\mathcal{J}_i)] = 0$, since the weight function depends only on $\mathcal{J}_i$ (and not on $\mathcal{J}_{-i}$). Additional inequalities that satisfy Assumption 3 can be formed if instrumental variables are available. That is, if $\mathbf{x}_i \in \mathcal{J}_i$ and $\mathcal{E}\eta_i g(\mathbf{x}_i) = 0$ for some positive function $g(\cdot)$, then we can take $h(d'; d_i, \mathcal{J}_i) = g(\mathbf{x}_i)$ if $d' = d_i + t$ and zero otherwise to generate another inequality that satisfies equation (5).

We have just shown that we can derive inequalities from the ordered choice model

[12] Also, $\nu^r_{1,i,d,d'} = \Delta R(d, d', \mathbf{d_{-i}}, \mathbf{z}_i^o) - \mathcal{E}[\Delta R(d, d', \mathbf{d_{-i}}, \mathbf{z}_i^o)|\mathcal{J}_i]$ and $\nu^\pi_{1,i,d,d'} = \nu_{1,i,d,d'} + \Delta R(d, d', \mathbf{d_{-i}}, \mathbf{z}_i^o) - \mathcal{E}[\Delta R(d, d', \mathbf{d_{-i}}, \mathbf{z}_i^o)|\mathcal{J}_i]$ so that $\nu_{1,i,d,d'} = \nu^\pi_{1,i,d,d'} - \nu^r_{1,i,d,d'}$, as in section 2.

that avoid the selection problem. Analogous results are available for other models in which we can find a counterfactual which generates a difference between the actual and counterfactual choices that is a known linear function of the structural error regardless of the observed choice. The vertical discrete choice model (e.g. Bresnahan 1987), and contracting models where the source of the structural error is a component of the transfers among agents (see Pakes 2010), are two examples of other models which can be shown to satisfy this condition.

Boundaries. To construct the (unconditional) moment used to estimate the parameter of the ordered choice model, the weight function h placed positive weight only on counterfactuals $d' = d_i + t$ for fixed (positive) t. More generally, we could consider counterfactuals $d' = d_i + t_i$ where t_i depends on i, if the t_i are fixed (i.e. nonstochastic) and if the t_i have the same sign for all i. In this case, weight proportional to $1/|t_i|$ satisfy Assumption 3. If the choice set is *bounded* this kind of weighting is not always possible. In the ATM choice model, different weight functions can be used to generate different inequalities for identification and estimation. Typically, we want at least one inequality based on weighting positive t_i counterfactuals and one inequality based on weighting negative t_i counterfactuals in order to get both upper and lower bounds for θ_0. For any agents with $d_i = 0$, there are no feasible counterfactuals with $d' = d_i + t$ for any $t < 0$. Dropping the observations with $d_i = 0$ before forming the inequalities generates a standard truncation problem. A similar problem will occur when controls are continuous but bounded from one side (as in a tobit model, or in an auction model where there is a cost to formulating the bid which causes some agents not to bid).

Recall that $\nu_{2,i,d_i,d_i+t} = t\eta_i$ in the ATM model. By definition of the parameter θ_0, $\mathcal{E}\eta_i = 0$. To deal with the boundary problem, we make an additional assumption that the η_i are i.i.d. with a distribution that is symmetric (about zero). Extending the argument of Powell (1986), the symmetry assumption allows for the use of the information from the untruncated direction (e.g. ν_{2,i,d_i,d_i+t} with positive t) to obtain a bound in the truncated direction (e.g. ν_{2,i,d_i,d_i-t}). We use the truncation of choices $d_i \geq 0$ in the ATM model to illustrate, but the idea extends to other one-sided boundary models.

Let $L = \{i : \mathbf{d}_i > 0\}$ denote the set of firms that install a positive number of machines and so are *not* on the boundary, and let n_L be the number of firms in L. It will be helpful to use standard order statistic notation, i.e. $\eta_{(1)} \leq \eta_{(2)} \leq \cdots \leq \eta_{(n)}$. Let $L_\eta = \{i : \eta_i \leq \eta_{(n_L)}\}$ and $U_\eta = \{i : \eta_i \geq \eta_{(n_L+1)}\}$. Similarly, let $\Delta R_i^+ = \Delta R(\mathbf{d}_i, \mathbf{d}_i + 1, \mathbf{d}_{-i}, \mathbf{z}_i)$

and $\Delta R^+_{(1)} \leq \Delta R^+_{(2)} \leq \cdots \leq \Delta R^+_{(n)}$. Let $U_R = \{i : \Delta R^+_i \geq \Delta R^+_{(n_L+1)}\}$. Sets L and U_R are observable to the econometrician, but sets L_η and U_η are not. Consider the following choice of weight function

$$h^i(d'; \mathbf{d}_i, \mathcal{J}_i) = n^{-1}\Big[\ \mathbf{1}\{d' = \mathbf{d}_i - 1\}\mathbf{1}\{i \in L\} +\ \mathbf{1}\{d' = \mathbf{d}_i + 1\}\mathbf{1}\{i \in U_R\}\Big],$$

and form

$$\begin{aligned}
&\sum_i \sum_{d' \in \mathcal{D}_i} h^i(d'; \mathbf{d}_i, \mathcal{J}_i, \mathbf{x}_{-i}) \Delta r(\mathbf{d}_i, d', \mathbf{d}_{-i}, \mathbf{z}^o_i, \theta_0) \\
&= \frac{1}{n}\sum_{i \in L} \Delta r(\mathbf{d}_i, \mathbf{d}_i - 1, \mathbf{d}_{-i}, \mathbf{z}^o_i, \theta_0) + \frac{1}{n}\sum_{i \in U_R} \Delta r(\mathbf{d}_i, \mathbf{d}_i + 1, \mathbf{d}_{-i}, \mathbf{z}^o_i, \theta_0) \\
&\geq \frac{1}{n}\sum_{i \in L} \Delta r(\mathbf{d}_i, \mathbf{d}_i - 1, \mathbf{d}_{-i}, \mathbf{z}^o_i, \theta_0) + \frac{1}{n}\sum_{i \in U_\eta} \Delta r(\mathbf{d}_i, \mathbf{d}_i + 1, \mathbf{d}_{-i}, \mathbf{z}^o_i, \theta_0) \\
&= \frac{1}{n}\sum_{i \in L} \left\{\mathcal{E}[\Delta\pi(\mathbf{d}_i, \mathbf{d}_i - 1, \mathbf{d}_{-i}, \mathbf{z}_i)|\mathcal{J}_i] - \nu_{2,i,\mathbf{d}_i,\mathbf{d}_i-1} + \nu^r_{1,i,\mathbf{d}_i,\mathbf{d}_i-1}\right\} \\
&\quad + \frac{1}{n}\sum_{i \in U_\eta} \left\{\mathcal{E}[\Delta\pi(\mathbf{d}_i, \mathbf{d}_i + 1, \mathbf{d}_{-i}, \mathbf{z}_i)|\mathcal{J}_i] - \nu_{2,i,\mathbf{d}_i,\mathbf{d}_i+1} + \nu^r_{1,i,\mathbf{d}_i,\mathbf{d}_i+1}\right\} \qquad (7) \\
&\geq -\frac{1}{n}\left\{\sum_{i \in L} \nu_{2,i,\mathbf{d}_i,\mathbf{d}_i-1} + \sum_{i \in U_\eta} \nu_{2,i,\mathbf{d}_i,\mathbf{d}_i+1}\right\} + \frac{1}{n}\left\{\sum_{i \in L} \nu^r_{1,i,\mathbf{d}_i,\mathbf{d}_i-1} + \sum_{i \in U_\eta} \nu^r_{1,i,\mathbf{d}_i,\mathbf{d}_i+1}\right\}.
\end{aligned}$$

The first inequality holds by the definition of U_R and noting $\Delta r(\mathbf{d}_i, \mathbf{d}_i + 1, \mathbf{d}_{-i}, \mathbf{z}^o_i, \theta_0) = \Delta R^+_i + \theta_0$. The second follows from Assumptions 1 and 2.

We first focus on the first term in the last expression in (7) to show Assumption 3(a) is satisfied,

$$\begin{aligned}
-\frac{1}{n}\left\{\sum_{i \in L} \nu_{2,i,\mathbf{d}_i,\mathbf{d}_i-1} + \sum_{i \in U_\eta} \nu_{2,i,\mathbf{d}_i,\mathbf{d}_i+1}\right\} &= \frac{1}{n}\left\{\sum_{i \in L} \eta_i - \sum_{i \in U_\eta} \eta_i\right\} \\
\geq \frac{1}{n}\left\{\sum_{i \in L_\eta} \eta_i - \sum_{i \in U_\eta} \eta_i\right\} &= \frac{1}{n}\left\{\sum_{i=1}^{n_L} \eta_{(i)} - \sum_{i=n_L+1}^{n} \eta_{(i)}\right\}.
\end{aligned}$$

Under the assumption that η_i are i.i.d. and symmetrically distributed about zero, the last term above has mean zero. So, $\mathcal{E}\left[-n^{-1}\sum_{i \in L} \nu_{2,i,\mathbf{d}_i,\mathbf{d}_i-1} - n^{-1}\sum_{i \in U_\eta} \nu_{2,i,\mathbf{d}_i,\mathbf{d}_i+1}\right] \geq 0$.

Next we consider the second term in the last expression in (7) to verify Assumption

3(b). The mean of the first sum is:

$$\mathcal{E}\left[\frac{1}{n}\sum_{i\in L}\nu^r_{1,i,\mathbf{d}_i,\mathbf{d}_i-1}\right] = \frac{1}{n}\sum_{i=1}^{n}\mathcal{E}[\mathbf{1}\{\mathbf{d}_i>0\}\nu^r_{1,i,\mathbf{d}_i,\mathbf{d}_i-1}] = \frac{1}{n}\sum_{i=1}^{n}\mathcal{E}[\mathbf{1}\{\mathbf{d}_i>0\}\underbrace{\mathcal{E}(\nu^r_{1,i,\mathbf{d}_i,\mathbf{d}_i-1}|\mathcal{J}_i)}_{=0}] = 0$$

since $\mathbf{d}_i \subset \mathcal{J}_i$. The second sum is $\frac{1}{n}\sum_i \mathbf{1}\{i\in U_\eta\}\nu^r_{1,i,\mathbf{d}_i,\mathbf{d}_i+1}$, and since the event, $\{i\in U_\eta\}$, depends on η_{-i}, the expectation of this sum will depend on the information structure of the game. If $\eta_{-i}\in\mathcal{J}_i$, as would be the case in a symmetric information game, then the fact that $\mathcal{E}[\nu^r_{1,i,\mathbf{d}_i,\mathbf{d}_i+1}|\mathcal{J}_i]=0$ would insure that $\mathcal{E}[\frac{1}{n}\sum_i \mathbf{1}\{i\in U_\eta\}\nu^r_{1,i,\mathbf{d}_i,\mathbf{d}_i+1}]=0$. If there was asymmetric information then signing the mean would require assumptions on the relationship between the unexpected part of agent i's profit measure and η_{-i}.

So taking expectations of the weighted sum in the first term of (7) and using the derived inequalities with symmetric information, we find that Assumptions 3(a) and 3(b) hold, leading to:

$$\mathcal{E}\left[\frac{1}{n}\sum_{i\in L}\Delta r(\mathbf{d}_i,\mathbf{d}_i-1,\mathbf{d}_{-i},\mathbf{z}^o_i,\theta_0)+\frac{1}{n}\sum_{i\in U_R}\Delta r(\mathbf{d}_i,\mathbf{d}_i+1,\mathbf{d}_{-i},\mathbf{z}^o_i,\theta_0)\right]\geq 0.$$

We have provided a set of assumptions which generates a lower bound for the parameter of interest despite the fact that the choice set is bounded from below. The Appendix shows that we can use instruments along with a symmetry assumption to generate more moment inequalities for the lower bound.

Moments. For now, assume that there are no observations on a boundary. Then two necessary conditions for Assumption 1 are that the expected increment to returns from the last ATM the bank installed ($t=-1$) were greater than its cost, while the expected increment to returns from adding one ATM more than the number actually installed ($t=+1$) was less than the corresponding cost.[13] Let the vector of differences be $\mathbf{\Delta r}_i(\theta)' \equiv [\Delta r(\mathbf{d}_i,\mathbf{d}_i-1,\mathbf{d}_{-i},\mathbf{z}^o_i,\theta),\,\Delta r(\mathbf{d}_i,\mathbf{d}_i+1,\mathbf{d}_{-i},\mathbf{z}^o_i,\theta)]$. Also suppose there are "instruments" $\mathbf{x}_i\in\mathcal{J}_i$ with $\mathcal{E}[\eta_i|\mathbf{x}_i]=0$. If $\mathbf{g}(\cdot)$ is a vector of nonnegative functions while $\otimes$ denotes the

[13] These conditions will also be sufficient if the expectation of π is concave with respect to the discrete values of d_i for all values of d_{-i}. We can not check this condition without specifying information sets, but the realizations of profits evaluated at the estimated value of θ were concave in d_i for almost all banks.

Kronecker product, sample analog moments formed as

$$m(\mathbf{w}, \theta) = \frac{1}{n} \sum_i \mathbf{\Delta r_i}(\theta) \otimes \mathbf{g}(\mathbf{x}_i)$$

have non-negative expectations at $\theta = \theta_0$. It is useful to consider inference based directly on these sample moments. In the empirical work we add a market index (j) and sum $m(\mathbf{w}^j, \theta)$ over markets.

To start, consider using only the moment conditions generated by $g(x_i) \equiv 1$, i.e. $m(\mathbf{w}^j, \theta) = n^{-1} \sum_i \mathbf{\Delta r_i^j}(\theta)$. Then the moment condition that results from decreasing the value of d_i, or the change "to the left", is $(n^j)^{-1} \sum_i [\Delta R(d_i^j, d_i^j - 1, d_{-i}^j, z_i^{oj}) - \theta] \equiv \Delta R_L^j - \theta$, while the moment condition from that results from increasing the value of d_i^j, or the change to the right, is $(n^j)^{-1} \sum_i [\Delta R(d_i^j, d_i^j + 1, d_{-i}^j, z_i^{oj}) + \theta] \equiv \Delta R_R^j + \theta$. Averaging across markets yields $\Delta \overline{R}_L \equiv J^{-1} \sum_j \Delta R_L^j$ which should be positive and provide an estimated upper bound on the average cost of an ATM, and $\Delta \overline{R}_R = J^{-1} \Delta R_R^j$ which should be negative with $-\Delta \overline{R}_R$ yielding an estimated lower bound on the average ATM cost. So our estimate of an interval that covers θ_0 is simply

$$\hat{\Theta}_J = \{\theta : -\Delta \overline{R}_R \le \theta \le \Delta \overline{R}_L\}.$$

Now suppose we add instruments (or x_k with $\mathcal{E}(\eta_i | \mathbf{x}_{k,i}) = 0$) indexed by k. Each instrument generates a pair of inequalities $0 \le \mathcal{E}[(\Delta R(\mathbf{d}_i, \mathbf{d}_i - 1, \mathbf{d}_{-i}, \mathbf{z}_i^o) - \theta_0) g(\mathbf{x}_{k,i}) | \mathcal{J}_i]$, and $0 \le \mathcal{E}[(\Delta R(\mathbf{d}_i, \mathbf{d}_i + 1, \mathbf{d}_{-i}, \mathbf{z}_i^o) + \theta_0) g(\mathbf{x}_{k,i}) | \mathcal{J}_i]$.[14] Averaging over i and j to form sample versions of these moments, yields additional sample upper and lower bounds: $\Delta \overline{R}_{k,L} \equiv \frac{\frac{1}{J} \sum_j \frac{1}{n^j} \sum_i (\Delta R(d_i^j, d_i^j - 1, d_{-i}^j, z_i^{oj})) g(x_{k,i}^j)}{\frac{1}{J} \sum_j \frac{1}{n} \sum_i g(x_{k,i}^j)}$ and $-\Delta \overline{R}_{k,R} \equiv - \frac{\frac{1}{J} \sum_j \frac{1}{n^j} \sum_i (\Delta R(d_i^j, d_i^j + 1, d_{-i}^j, z_i^{oj})) g(x_{k,i}^j)}{\frac{1}{J} \sum_j \frac{1}{n^j} \sum_i g(x_{k,i}^j)}$. So provided $max_k\{-\Delta \overline{R}_{k,R}\} \le \ min_k\{\Delta \overline{R}_{k,L}\}$ an interval estimate for θ_0 is

$$\hat{\Theta}_J \ = \ [max_k\{-\Delta \overline{R}_{k,R}\}, \ min_k\{\Delta \overline{R}_{k,L}\}],$$

based on taking the greatest lower bound and the least upper bound over all the sample bounds for each instrument.

Empirical Results. The data set consists of a cross-section of all banks and thrifts in Massachusetts metropolitan statistical areas in 2002. A market is defined as a primary

[14] Andrews and Shi (2012) provides a method of inferences based on *conditional* moment inequalities, as given here.

metropolitan statistical area, and the sample is small: it has 291 banks in 10 markets.[15] The number of banks varies considerably across markets (from 8 to 148 in Boston), as does the number of distinct ATM locations per bank (which averages 10.1 and has a standard deviation of 40.1). Just under 5.5% of the banks have no ATMs ($d_i = 0$).

Table 1 contains the inequality estimators of the cost parameter.[16] The first two rows provide the results when only a constant term is used as an instrument; row (1) uses only observations with $d \geq 1$ when calculating the upper bound while row (2) keeps all the observations and uses the symmetry assumption to correct for truncation (banks with $d_i = 0$ may have had higher than average ATM costs). The lower bounds provided in the two rows are based on the same estimator and hence identical, but we expect the upper bound from row (2) to be higher due to correcting the truncation bias in row (1). Indeed, we do find that the corrected truncation bias leads to a larger upper bound, though the difference is small (25,283 vs 26,644), about equal in percentage terms to the fraction of observations with $d_i = 0$. Even after the correction, the estimate of the identified interval is quite short [24,452, 26,444], with a confidence interval of [20,472, 30,402].

Rows 3 and 4 repeat the exercise in Rows 1 and 2 using the market population, the number of banks in the market, and the number of branches of the bank (its mean is 6 and standard deviation is 15), as well as the constant term, as instruments. $\hat{\Theta}_J$ reduces to a singleton, reflecting the fact that there was no value of the parameters that satisfied all of the inequalities, and the estimated value closest to satisfying all the inequalities is less than the lower bounds in rows 1 and 2. The fact that we obtained a point estimate could be due to sampling error[17] or a misspecification; say a violation of Assumption 3. A test for misspecification rejected the null, casting doubt on the estimates in rows 3 and 4. We also tried adding profit differences based on additional counterfactuals $d' = d_i \pm 2$, i.e. adding the rows $[\Delta r(\mathbf{d}_i, \mathbf{d}_i - 2, \mathbf{d}_{-i}, \mathbf{z}_i^o, \theta),\ \Delta r(\mathbf{d}_i, \mathbf{d}_i + 2, \mathbf{d}_{-i}, \mathbf{z}_i^o, \theta)]'$ to the vector $\mathbf{\Delta r}_i(\theta)$. The fraction of banks with $d_i < 2$ is much larger than the fraction with $d_i < 1$ (25.5% *vs* 5.5%),

[15]The data set is described in Ishii (2004), and is carefully put together from a variety of sources including the Summary of Deposits, the Call and Thrift Financial Reports, the 2000 Census, the Massachusetts Division of Banks, and various industry publications.

[16]Confidence intervals are calculated using the technique described in the longer version of this paper.

[17]The greatest lower bound is the maximum of a finite number of moments each of which will, in finite samples, distribute approximately normally. So in small samples we should expect a positive bias in the estimated lower bound which is (weakly) increasing in the number of inequalities. Analogously we should expect a negative bias in the upper bound. That is, the simple maximum and minimum estimators tend to be "inward biased" even if the model is correctly specified, and the inward bias tends to generate the a singleton for $\hat{\Theta}_J$. Hirano and Porter (2009) discuss implications of the bias from taking the minimum and maximum as endpoint estimates.

Table 1: **Inequality Method, ATM Costs***

	θ_J	95% CI for θ LB	UB
1. $h(x) \equiv 1, d \geq 1$ for u.b. $\hat{\theta}$	[24,452, 25,283]	20,544	29,006
2. $h(x) \equiv 1, d \geq 0$	[24,452, 26,444]	20,472	30,402
$h(x)$ =(1,pop, # Banks in Mkt, # Branches of Bank)			
3. $d \geq 1$ for u.b. $\hat{\theta}$	19,264	16,130	23,283
4. $d \geq 0$	20,273	17,349	24,535
$\{d : \lvert d - d_i\rvert = 1, 2\},\ h(x) = 1$			
5.$\{d : \lvert d - d_i\rvert = 1, 2\};\ d \geq 1$ for u.b. $\hat{\theta}$	[24,452, 25,283]	20,691	28,738
6.$\{d : \lvert d - d_i\rvert = 1, 2\};\ d \geq 0$	[24,452, 26,644]	20,736	29,897
First Order Conditions (analogue of Hansen and Singleton,1982)			
7. h(x)=1	28,528	23,929	33,126
8. h(x)=(1,pop, # Banks in Mkt, # Branches of Bank)	16,039	11,105	20,262

* There are 291 banks in 10 markets. The first order condition estimator requires derivatives with respect to interest rate movements induced by the increment in the number of ATMs. We used two-sided numerical derivatives of the first order conditions for a Nash equilibria for interest rates.

so the truncation issue corresponding to the counterfactual $d' = d_i - 2$ is potentially worse. Rows 5 and 6 present the results from interacting all four profit differences with a constant term. The interval estimates from using only the data on $d_i \geq 1$ for the upper bound are unchanged by this addition, as would be expected if the profit function was concave in discretely valued d_i. Moreover, as expected, when we use all the data and correct for truncation, the estimated upper bound increases, but only by a small amount.

Alternative Estimators. At least two alternative estimation techniques have been used in similar problems: ordered probit (or logit), and estimators based on first order conditions. In our notation, ordered probit sets $\nu_{1,i} \equiv 0$, assumes independent normal distributions for each level of marginal cost heterogeneity η_i conditional on the other determinants of profits, and forms the likelihood of the observed d_i. To apply this to an interacting agent model we need to assume a unique equilibrium and that $\eta_{-i} \notin \mathcal{J}_i$. The first order condition estimator ignores the discrete nature of the control and goes to the opposite extreme: it assumes that $\nu_{2,i} \equiv 0$ but does not restrict the $\nu_{1,i}$ distribution nor does it require a unique equilibrium.

The probit model can not be estimated with our data. The issue is that for some of our observations the revenue difference from the left ($\Delta R_{L,i}$) is less than the revenue difference to the right ($\Delta R_{R,i}$). In this case there is no value of $\theta + \eta_i$ that rationalizes d_i so the likelihood is not well defined for these observations (the model when combined with the data say that if it was profitable to purchase the last ATM, it must have been profitable to purchase the next ATM). In a world with either uncertainty or approximation errors in revenues we would not be surprised to find at least one agent with $\Delta R_{L,i} < \Delta R_{R,i}$. However, adding either approximation error or uncertainty would require additional assumptions and complicate the ordered probit model significantly.

If we act as if the control is continuous and assume that agents maximize expected profits, the first order condition for agents with $d_i > 0$ must have an expectation of zero conditional on their information sets. We approximated the derivative of the profit function by averaging the incremental profits obtained from changing d_i to $d_i + 1$ and to $d_i - 1$ and then obtained the estimate of θ by minimizing $\| \frac{1}{n} \sum_i \{d_i > 0\} (\frac{\partial R(d, d_{-i}, z_i^o)|_{d=d_i}}{\partial d} - \theta) \times g(x_i) \|$. So, operationally, row 7 in Table 1 is obtained by adding the two moment inequalities used in row 1 together for observations with $d \geq 1$, dividing by two, and searching for a θ that makes the average as close as possible to zero. Row 8 is similar with the added instruments. Again the estimates that use the added instruments are lower than the bounds in rows 2 and 4. The row 7 estimate is outside of the interval estimate obtained from the inequality estimators (by 15 to 20%), but about equal to the upper bound of the confidence intervals in those rows. Interestingly the confidence interval from this (point) estimator has about the same length as that from the moment inequalities.

4 Conclusion

This paper provides conditions which ensure that the inequality constraints generated by either single agent optimizing behavior, or by the best response functions of problems with interacting agents, can be used for inference. The conditions do not place any restrictions on the choice sets of the agents, the contents of the agents' information sets, or the equilibrium selection mechanism.

We work directly with a model for expected profits rather than modeling the strategy for making choices. As a result, we focus on the difference between our approximation for profits and the agent's expectations for profits. This difference can generate two types of disturbances. The first which we label $\nu_{1,i}$, has zero expectation conditional on the agent's information set, and can be caused by either realizations of variables that were unknown to the agent when it made its decision or approximation error. The distribution of $\nu_{1,i}$ can be quite complex as it depends on the information sets of agents and, in multiple agent problems, on the details of the equilibria selected by the market participants. However, the fact that $\nu_{1,i}$ has zero conditional expectations allows the formation of moment inequalities which account for this complexity without either needing to specify these details or computing an equilibria.

The other possible disturbance arises when there is a source of profit differences that the econometrician does not observe and is a determinant of the choice. Formally it is the difference between the agent's conditional expectation of the profit variable and the conditional expectation that is implicit in the researchers' model for realized profits, a difference which we label $\nu_{2,i}$. We provide conditions which suffice to obtain inequality constraints for the parameters of interest when both types of disturbances are present. The conditions do impose restrictions on the structure of these errors, but they do not require a parametric specification for the form of the joint distribution of $\nu_{1,i}$ and $\nu_{2,i}$ while allowing for endogenous regressors and general choice sets.

Our main example shows how the framework proposed here can enable us to obtain information on parameters of interest in environments where estimation has proven difficult in the past, and which are of significant applied interest.

References

Andrews, D., and X. Shi (2013) Inference Based on Conditional Moment Inequality Models, *Econometrica*, 81(2).

Bajari, P., L. Benkard, and J. Levin (2007): "Estimating Dynamic Models of Imperfect Competition" *Econometrica,* 1331-1370-814..

Beresteanu, A., and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models" *Econometrica,* 76(4), 763-814..

Bresnahan T. (1987) "Competition and Collusion in the American Automobile Industry: The 1955 Price War," *Journal of Industrial Economics,* vol. 35, no. 4, pp. 457-482.

CILIBERTO, F. AND E. TAMER (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77(6), 1791-1828.

CRAWFORD, G. AND A. YURUKOGLU (2012): "The Welfare Effects of Bundling in Multichannel Television" *American Economic Review*, 102(2), 643-85.

HANSEN, L., AND K. SINGLETON (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models", *Econometrica*, 50, 1269-86.

HIRANO, K., AND J. PORTER (2009): "Asymptotics for Statistical Treatment Rules," *Econometrica*, 77(5), 1683-1701.

HO, K., AND A. PAKES (2011): "Physician Responses to Financial Incentives: Evidence from Hospital Discharge Data," manuscript, Columbia University.

HOLMES, T. (2011): "The Diffusion of Wal-Mart and Economies of Density," *Econometrica*, 79(1), 253-302.

ISHII, J. (2004): "Compatibility, Competition, and Investment in Network Industries: ATM Networks in the Banking Industry," *Unpublished Ph.D. Thesis*, Harvard University.

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

MORALES, E. (2011): "Gravity and Extended Gravity: Estimating a Structural Model of Export Entry," *Unpublished Ph.D Thesis* Harvard University.

PAKES, A. (2010): "Alternative Models for Moment Inequalities," *Econometrica*, Vol. 78, No. 6, pp 1783-1822.

POWELL, J. (1986): "Symmetrically Trimmed Least Squares Estimation of Tobit Models", *Econometrica*, 54, 1435-1460.

Appendix: Boundaries Generalization

In this Appendix, the derivation of moment inequalities in the Boundaries section is generalized in two ways. First, we allow for instruments $\mathbf{x}_i \in \mathcal{J}_i$ in the formation of our moment inequalities. Second, a more general marginal cost parametrization that can depend on observed random variables will be allowed. For example, the suppose the marginal cost function for firm i is $c(\mathbf{z}_i^o, \theta_0) + \eta_i$, where η_i represents firm heterogeneity, as before, and

$c(\mathbf{z}_i^o, \theta_0)$ is the average marginal cost function across firms. Then, the profit function takes the form: $\pi(d_i, \mathbf{d_{-i}}, \mathbf{z}_i) = R(d_i, \mathbf{d}_{-i}, \mathbf{z}_i^o) - d_i(c(\mathbf{z}_i^o, \theta_0) + \eta_i) + \nu_{1,i,d_i}$. The econometrician's differenced profit function is then $\Delta r(d, d', \mathbf{d}_{-i}, \mathbf{z}_i^o; \theta) = \Delta R(d, d', \mathbf{d}_{-i}, \mathbf{z}_i^o) + (d' - d)c(\mathbf{z}_i^o, \theta)$.

We assume that $(\mathbf{x}_i, \eta_i)$ is i.i.d. over $i = 1, \ldots, n$. As before, assume that η_i is distributed symmetrically about zero. In addition, assume η_i is independent of $\mathbf{x}_i$.

Let $g\eta_i$ denote $g(x_i)\eta_i$, and let gr_i^+ denote $g(x_i)\Delta r(\mathbf{d}_i, \mathbf{d}_i + 1, \mathbf{d}_{-i}, \mathbf{z}_i^o, \theta_0)$. Then, using order statistic notation $g\eta_{(1)} \leq g\eta_{(2)} \leq \cdots \leq g\eta_{(n)}$, and $gr_{(1)}^+ \leq gr_{(2)}^+ \leq \cdots \leq gr_{(n)}^+$. Define $L = \{i : \mathbf{d}_i > 0\}$ and n_L, as before. Also, define $L_{g\eta} = \{i : g\eta_i \leq g\eta_{(n_L)}\}$, $U_{g\eta} = \{i : g\eta_i \geq g\eta_{(n_L+1)}\}$, and $U_{gr} = \{i : gr_i^+ \geq gr_{(n_L+1)}^+\}$. Finally, define the weight function to include a positive function g of the instrument: $h^i(d'; \mathbf{d}_i, \mathcal{J}_i, gr_{-i}^+) = n^{-1}\Big[g(\mathbf{x}_i)\ \mathbf{1}\{d' = \mathbf{d}_i - 1\}\mathbf{1}\{i \in L\} + g(\mathbf{x}_i)\ \mathbf{1}\{d' = \mathbf{d}_i + 1\}\mathbf{1}\{i \in U_{gr}\}\Big]$.

Now the derivation follows the steps in (7) with $U_{g\eta}$ replacing U_η, and U_{gr} replacing U_R. So, we have

$$\sum_i \sum_{d' \in \mathcal{D}_i} h^i(d'; \mathbf{d}_i, \mathcal{J}_i, gr_{-i}^+)\Delta r(\mathbf{d}_i, d', \mathbf{d}_{-i}, \mathbf{z}_i^o, \theta_0) \geq -\frac{1}{n}\left\{\sum_{i \in L} g(\mathbf{x}_i)\nu_{2,i,\mathbf{d}_i,\mathbf{d}_i-1} + \sum_{i \in U_{g\eta}} g(\mathbf{x}_i)\nu_{2,i,\mathbf{d}_i,\mathbf{d}_i+1}\right\}$$
$$+\frac{1}{n}\left\{\sum_{i \in L} g(\mathbf{x}_i)\nu^r_{1,i,\mathbf{d}_i,\mathbf{d}_i-1} + \sum_{i \in U_{g\eta}} g(\mathbf{x}_i)\nu^r_{1,i,\mathbf{d}_i,\mathbf{d}_i+1}\right\}$$

As in the Boundaries section, it can be argued that the second term in the last expression has mean zero in symmetric information games.

Now focus on the first term in the last expression. As before,

$$-\frac{1}{n}\left\{\sum_{i \in L} g(\mathbf{x}_i)\nu_{2,i,\mathbf{d}_i,\mathbf{d}_i-1} + \sum_{i \in U_{g\eta}} g(\mathbf{x}_i)\nu_{2,i,\mathbf{d}_i,\mathbf{d}_i+1}\right\} \geq \frac{1}{n}\left\{\sum_{i=1}^{n_L} g\eta_{(i)} - \sum_{i=n_L+1}^{n} g\eta_{(i)}\right\}$$

Note that since η_i is symmetrically distributed about zero and η_i is independent of $\mathbf{x}_i$, $g\eta_i$ is also distributed symmetrically about zero. Hence, as argued before, the last term above has expectation zero, and $\mathcal{E}\left[-\frac{1}{n}\left\{\sum_{i \in L} g(\mathbf{x}_i)\nu_{2,i,\mathbf{d}_i,\mathbf{d}_i-1} + \sum_{i \in U_{g\eta}} g(\mathbf{x}_i)\nu_{2,i,\mathbf{d}_i,\mathbf{d}_i+1}\right\}\right] \geq 0$.

So, we can conclude that

$$\mathcal{E}\left[\frac{1}{n}\sum_{i \in L} g(\mathbf{x}_i)\Delta r(\mathbf{d}_i, \mathbf{d}_i - 1, \mathbf{d}_{-i}, \mathbf{z}_i^o, \theta_0) + \frac{1}{n}\sum_{i \in U_{gr}} g(\mathbf{x}_i)\Delta r(\mathbf{d}_i, \mathbf{d}_i + 1, \mathbf{d}_{-i}, \mathbf{z}_i^o, \theta_0)\right] \geq 0.$$