AEA Continuing Education Program: Industrial Organization.

Third Topic; Moment Inequalities and Their Application in Industrial Organization.

Lecturer: Ariel Pakes, Harvard University.

I am going to try and cover three topics. I will omit parts of each topic from the lectures due to time constraints. However I will overview the material not covered in lectures and point you to the part of the notes that cover them for those of you who want to access the relevant material.

Topic 1.

I will begin with the behavioral model that leads to moment inequalities (including examples). This is the analogue of revealed preference in the analysis of utility, but to bring it to data we will need to allow for the disturbances that arise in applications. I will then move to a more detailed discussion of product repositioning. Finally, I will conclude with a note on analyzing counterfactuals in situations where multiple equilibrium are likely.

Topic 2.
The topic here is the econometric of inequality estimators. It begins by explaining the econometric issues that arise in moment inequality estimators that do not arise on estimators based on moment equalities. It then moves on to techniques available to derive confidence sets for the partially identified models generated by moment inequalities. Emphasis is given to practical issues which arise in getting confidence intervals for parameters.

Topic 3.
The use of inequalities in choice theory. Again this is based on revealed preference. We focus on discrete choice problems that have been difficult to analyze with traditional discrete choice methods. These include models with; (i) errors in the right hand side variables, (ii) and models with choice specific fixed effects.

Profit Inequalities: The Behavioral Model.

- Econometrican observes a set of choices made by various agents.

- Assume agents expected the choices they made to lead to returns that were higher than the returns the agents would have earned had they made an alternative *feasible* choice.

- Assume a parametric return function and for each value of θ compute the difference between the observable part of the actual realized returns and the observable part of returns that would have been earned had the alternative choice been made.

- Estimator: accept any value of θ that, on average, makes the observed decisions better than the alternative.

- Question: When do such (possibly set valued) estimators enable us to make valid inferences on the parameters of interest?

Pakes (2010) provides two (non-nested) sets of conditions where they do, and develops the actual estimators. The ideas behind these estimators date, respectively to

- Tamer (2003),

- Pakes, Porter, Ho, and Ishii (2015).

I start with a simple example, designed, I hope, to get your interest. Later I come back to multiple agent problems.

Static Example: In M. Katz (2007) unpublished thesis; explained in Pakes (2010)

Estimate the costs shoppers assign to driving to a supermarket (important to the analysis of; zoning regulations, public transportation projects,...). 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). Here we show how to turn it into an "ordered" problem, which is the single agent analogue to the problems we face to for many of the investment and product placement problems we consider in I.O.

Assume that the agents' utility functions are additively separable functions of;

- utility from basket of goods bought,
- expenditure on that basket, and
- drive time to the supermarket.

I.e. if $b_i = b(d_i)$ is the basket of goods bought, $s_i = s(d_i)$ is the store chosen, and z_i are individual characteristics

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

where $e(\cdot)$ provides expenditure, $dt(\cdot)$ provides drive time, and I have used the free normalization on expenditure (the cost of drive time are in dollars).

Standard discrete choice. Need to specify the expected utility from each possible choice. Requires
(i) the agent's prior probability for each possible price at each store, and
(ii) the bundle of goods the agent would buy were any particular price vector realized.
(There is a simple reduced form, that I come back to; but not available for interacting agent problems.)

Simplify. Compare the utility from the choice the individual made to that of an alternative feasible choice. Expected difference should be positive. Requires: finding an alternative choice that allows us to isolate the effects of drive time.

For a particular d_i chose $d'(d_i)$ to be the purchase of

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

Note. Need not specify the utility from different baskets of goods; i.e. it allows us to hold fixed the dimension of the choice that generated the problem with the size of the choice set, and investigate the impact of the dimension of interest (travel time) in isolation.

Let $\mathcal{E}(\cdot)$ be the *agent's* expectation operator. Then we assume that

$$\mathcal{E}[\Delta\pi(d_i, d'(d_i), z)] =$$

$$-\mathcal{E}[\Delta e(d_i, d'(d_i))] - \theta_i\, \mathcal{E}[\Delta dt(d_i, d'(d_i))] \geq 0.$$

Note. I have not assumed that the agent's perceptions of prices are "correct" in any sense. I come back to what I need of the agent's subjective expectations.

Case 1: $\theta_i = \theta_0$. More generally all determinants of drive time are captured by variables the econometrician observes and includes in the specification. Assume that

$$N^{-1}\sum_i \mathcal{E}[\Delta e(d_i, d'(d_i))] - N^{-1}\sum \Delta e(d_i, d'(d_i)) \to_P 0,$$

$$N^{-1}\sum_i \mathcal{E}[\Delta dt(d_i, d'(d_i))] - N^{-1}\sum_i \Delta dt(d_i, d'(d_i)) \to_P 0$$

which would be true if, for e.g., agents were correct on average (this is stronger than we need). Then

$$-\mathcal{E}[\Delta e(d_i, d'(d_i))] - \theta\, \mathcal{E}[\Delta dt(d_i, d'(d_i))] \geq 0$$

implies

$$-\frac{\sum_i \Delta e(d_i, d'(d_i))}{\sum_i \Delta dt(d_i, d'(d_i))} \to_p \underline{\theta} \leq \theta_0.$$

If we would have also taken an alternative store which was closer to the individual then

$$-\frac{\sum_i \Delta e(d_i, d'(d_i))}{\sum_i \Delta dt(d_i, d'(d_i))} \rightarrow_p \overline{\theta} \geq \theta_0.$$

and we would have consistent estimates of bounds on θ_0. Note this assumes that there always is an alternative store closer to the individual than the store the agent went to. Below we come back to the adjustment to the procedure needed if this is not the case.

Case 2: $\theta_i = (\theta_0 + \nu_i)$, $\sum \nu_i = 0$. This case allows for a component of the cost of drive times (ν_i) that is known to the agent (since the agent conditions on it when it makes its decision) but not to the econometrician. Then provided $dt(d_i)$ and $dt(d'(d_i))$ are known to the agent

$$\mathcal{E}\Big[\frac{\Delta e(d_i, d'(d_i))}{\Delta dt(d_i, d'(d_i))} - (\theta_0 + \nu_i)\Big] \leq 0,$$

and provided agents expectation on expenditures are not "systematically" biased

$$\frac{1}{N}\sum_i \Big(\frac{\Delta e(d_i, d'(d_i))}{\Delta dt(d_i, d'(d_i))}\Big) \ \to_P \ \underline{\theta} \ \leq \ \theta_0.$$

Notes.

- We did not need to specify (or compute) the utility from all different choices, so there could have been (unobserved or observed) sources of heterogeneity in the $U(b_i)$. Our choice of alternative simply differences them out.

- Case 2 allows for unobserved heterogeneity in the coefficient of interest and does not need to specify what the distribution of that unobservable is. In particular it can be *freely correlated* with the right hand side variable. "Drive time" is a choice variable, so we might expect it to be correlated with the perceived costs of that time (with ν_i).

- If the unobserved determinant of drive time costs (ν_i) is correlated with drive time (dt) then Case 1 and Case 2 estimators should be different, if not they should be the same. So there is a test for whether any unobserved differences in preferences are correlated with the "independent" variable.

Empirical Results.

Data. Neilsen Homescan Panel, 2004 & data on store characteristics from TradeDimensions. Chooses families from Massachusetts.

Discrete Choice Comparison Model. The multinomial model divides observations into expenditure classes, and then uses a typical expenditure bundle for that class to form the expenditure level (the "price index" for each outlet). Other x's are drive time, store characteristics, and individual characteristics. Note that

- the prices for the expenditure class need not reflect the prices of the goods the individual actually is interested in (so there is an error in price, and it is likely negatively correlated with price itself.)

- it assumes that the agents knew the goods available in the store and their prices exactly when they decided which store to choose (i.e. it does not allow for expectational error)

- it does not allow for unobserved heterogeneity in the effects of drive time. We could allow for a random coefficient on drive time, but, then we would need a conditional distribution for the drive time coefficient....

Focus. Median of the drive time coefficient (about forty coefficients; chain dummies, outlet size, employees, amenities...).

• Multinomial Model: median cost of drive time was \$240 (when the median wage in this region is \$17). Also several coefficients have the "wrong" sign or order (nearness to a subway stop, several amenities, and chain dummies).

Inequality estimators. Uses a lot of moments: point estimates, but tests indicated that the model was accepted. Standard errors are very conservative.

- Inequality estimates with

$\theta_i = \theta_0$: .204 [.126, .255]. $\Rightarrow \$4/hour$,

- Inequality estimates with

$\theta_i = \theta_0 + \nu_i$: .544 [.257, .666], $\Rightarrow \$14/hour$

and other coefficients straighten out.

Apparently the unobserved component of the coefficient of drive time is negatively correlated with observed drive time differences.

We now move on to looking for the sets of assumptions that underlie this and other moment inequality models in more detail.

Behavioral Models

We will present the two behavioral models that have been used with moment inequalities. Both models allow for interacting agent and both have four assumptions. Two of these assumptions are the same and two are not.

Assumptions Common To Both Models.

Best Response Condition (C1).

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

$$\leq \mathcal{E}[\pi(d(\mathcal{J}_i), \mathbf{d}_{-i}, \mathbf{y}_i, \theta_0)|\mathcal{J}_i]$$

where $d_i \equiv d(\mathcal{J}_i) \in D_i$ is the agents decision, $D_i \subset \mathcal{D}$ is its choice set, $\mathcal{J}_i$ is its information set, and $\mathcal{E}[\cdot|\mathcal{J}_i]$ takes expectations over $(\mathbf{d}_{-i}, \mathbf{y}_i)$. ♠

Notes.

- No restriction on choice set; could be discrete (a subset of all bilateral contracts, ordered choice ...) or continuous (with corners, non-convexities ...).

- No uniqueness requirement, and equilibrium selection can differ for different observations (we only use "necessary" conditions for an equilibrium)..

- C1 is termed "weak rationality" in Pakes (2010).

Counterfactuals.

To check the best response condition (or the maximization condition in single agent models) we need an approximation to what profits *would have been* had the agent made a choice which in fact it did not make. This is the analogue of what you implicitly do every time you write down a discrete choice model (there is a model for the choice taken and there is a model for the choices not taken). This requires a model of how the agent thinks that $\mathbf{d}_{-i}$ and $\mathbf{y}_i$ are likely to change in response to a change in the agent's decision.

Counterfactual Condition (C2).

$$\mathbf{d}_{-i} = d^{-i}(\mathbf{d}_i, \mathbf{z}_i), \;\; \mathbf{y}_i = y(\mathbf{z}_i, \mathbf{d}_i, \mathbf{d}_{-i}),$$

the distribution of $\mathbf{z}_i$ conditional on $\mathcal{J}_i$ does not depend on d_i. ♠

Exogeneity.

The assumption that the distribution of $\mathbf{z}_i$ conditional on $\mathcal{J}_i$ does not depend on d_i is what we mean by $\mathbf{z}_i$ being an *exogenous* random variable. It is just a more detailed statement of what an "instrument" is when one teaches IV estimation.

• Single agent: no $\mathbf{d}_{-i}$; $\mathbf{y}_i$ often exogenous in this sense.
• Multiple agents, simultaneous moves: $\mathbf{d}_{-i}$ satisfies C2.
• Multiple agents, multi-stage; often a $\mathbf{y}$ which is "endogenous" – its distribution depends on d_i – and then we need a model of that dependence.
• Multiple agent, sequential moves: must

postulate response. We need a model for dynamic games.

Implication: C1 + C2. After substituting $\mathbf{d}_{-i} = d^{-i}(\mathbf{d}_i, \mathbf{z}_i)$, and $\mathbf{y}_i = y(\mathbf{z}_i, \mathbf{d}_i, \mathbf{d}_{-i})$ into $\pi(\cdot)$, if for $d' \in \mathcal{D}_i$ we let

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

we have

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

To estimate we need the relationships between:

- The expectations underlying agents decisions ($\mathcal{E}(\cdot)$) and the expectations of the observed sample moments ($E(\cdot)$),

- $\pi(\cdot, \theta)$ and (z_i, d_i, d_{-i}) and their observable analogues.

This is where the two approaches differ. One is the natural generalization of standard discrete choice theory to multiple agent settings. The other is an extension of revealed preference arguments. Before we turn to them we need assumptions on the relationship between what we observe, and the models' concepts; a "measurement" model.

General Measurement Model.

Let

$$r(d, d_{-i}, z_i^o, \theta_0)$$

be our *observable* approximation to $\pi(\cdot)$. Then w.l.o.g. we can define the following terms

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

so

$$r(\cdot) = \pi(\cdot) + \nu,$$

and

$$\mathcal{E}[r(\cdot)|\cdot] = \mathcal{E}[\pi(\cdot)|\cdot] + \mathcal{E}[\nu|\cdot].$$

It follows that

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

where

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

and

$$\nu_{1,i,d} \equiv$$

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

Sources of ν_1. Sum of: *expectational error* from incomplete (uncertainty in $\mathbf{z}_i$), and/or asymmetric (uncertainty in $\mathbf{d}_{-i}$) information,

$$\pi(d,\cdot) - \mathcal{E}[\pi(d,\cdot)|\mathcal{J}_i]$$

and *specification and measurement error* or

$$\nu(d,\cdot) - \mathcal{E}[\nu(d,\cdot)|\mathcal{J}_i]$$

(This includes errors that arise from specifying functional forms that involve an approximation error.)

General Points.

- $\mathcal{E}[\nu_{1,i,d}|\mathcal{J}_i] = 0$, by construction. $\mathcal{E}[\nu_{2,i,d}|\mathcal{J}_i] \neq 0$. This distinction is why we need to keep track of two separate disturbances.

• When the left hand side variable (the variable we are trying to explain) is a measure of profits, typically the disturbance is dominated by ν_1 errors, or at least they should not be ignored. When the ν_2 errors can be ignored straightforward moment inequalities based on revealed preference can be used to estimate.

• When the left had side is a control or a decision variable (e.g. investment) then typically the disturbance will be dominated by ν_2 errors or at least we do not want to ignore them. If the ν_1 errors can be ignored we get traditional discrete choice analysis, or generalizations thereof (that I turn to next). No ν_1 error requires certainty about functional forms (often including those that generate an expectation operator)

• Of course both may be present and we may have to deal with that.

Abuses of Assumptions on Disturbances.

ν_2 and selection.

Since $\nu_{2,i} \in \mathcal{J}_i$ and $d_i = d(\mathcal{J}_i)$, d_i will generally be a function of $\nu_{2,i}$ (and perhaps also of $\nu_{2,-i}$). This can generate a selection problem.

Temporarily assume; the agent's expectations (our $\mathcal{E}(\cdot)$) equals the expectations generated by the true data generating process (our $E(\cdot)$), that x is an "instrument" in the sense that $\mathcal{E}[\nu_2|x] = 0$, and that $x \in \mathcal{J}$. Then

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

These expectations *do not* condition on d_i, and any moment which depends on d_i requires properties of the disturbance conditional on d_i. Since d is measurable $\sigma(\mathcal{J})$

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

However since $\nu_2 \in \mathcal{J}$ and

$$\mathcal{E}[\pi(\cdot)|\cdot] = \mathcal{E}[r(\cdot)|\cdot] + \nu_2,$$

if the agent choses d^* then

$$\nu_{2,d^*} - \nu_{2,d} \geq \mathcal{E}[r(\cdot, d)|\cdot] - \mathcal{E}[r(\cdot, d^*)|\cdot]$$

so

$$\mathcal{E}[\nu_{2,d^*}|x, d^*] \neq 0, \;\; and \;\; \mathcal{E}[\nu_{2,d}|x, d] \neq 0.$$

The fact that "x is an instrument" does not "solve" the selection problem.

E.g. **Single agent binary choice.** $d_i \in \{0, 1\}$, with

$$\Delta\pi(d_i, d', \cdot) = \Delta r(d_i, d', \cdot) + \Delta\nu_{2,i} + \Delta\nu^1_{1,i}.$$

Then $d_i = 1$ if

$$\mathcal{E}[\Delta\pi(d_i = 1, d' = 0, \cdot)|\mathcal{J}_i] =$$

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

Assume the $\nu_{2,i}$ were centered at zero. Then

$$\mathcal{E}[\Delta\nu_{2,i}|d_i = 1] =$$

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

which violates our condition.

Expectational or measurement/approximation error.

As noted for there not to be a ν_1 error there would either have to be either;

- no; expectational error, measurement error in rhs variables, or approximation error in the model, or

- a way of analyzing the model with expectational and measurement error.

Expectational Error. The mean of expectational error conditional on variables one knows at the outset should be zero, but we in general do not know its distribution. There is a difference here between single agent and multiple agent models.

- In single agent models if one is willing to make a rational expectations assumption one can sometimes find estimate the characteristics of the needed distribution. Dickstein and Morales (2015) provide a way of using weak semi-parameteric assumptions to overcome expectational problems and show that accounting for expectational errors in a flexible way makes a difference in analyzing exporting decisions.

- In interacting agent models, to compute the distribution of the expectational error we would have to specify what each agent knows about its competitors, and then repeatedly solve for an equilibrium (a process which typically would require additional assumptions as it would require us to select among equilibria).

Measurement and/or Approximation error. Consider measurement (or approximation) error in the simple linear binary choice model, $\Delta U_{d,i,t} = x^*_{i,t}\beta + \nu_2$. Here either x^* is unobserved and what we observe is $x^o = x^* + \nu_1$, or there is a ν_1 error caused by misspecification. The two cases are similar so I deal only with the first. The required choice probability is

$$Pr\{d|x^0, \beta\} = \int_{\nu_1} Pr(\nu_2 \geq x^*\beta)\, dP(x^*|x^0, \beta),$$

and assuming densities exist to carry out the integration we need

$$f(x^*|x^o) = \frac{f(x^0|x^*)f(x^*)}{f(x^o)} = \frac{f_{\nu_1}(\nu_1 = x^o - x^*)f(x^*)}{f(x^0)}.$$

Though we might be willing to assume the distribution of ν_1 has some familiar form, it would be harder to assume a distribution for x^*. To estimate it we would need a de-convolution theorem.

This concludes the discussion of the relevance of assumptions on the disturbances. We now go back to the two modeling frameworks.

I will not go over the first in class in detail, but rather explain what it does. However notes on it are included here. So the reader who is uninterested should be able to go directly to M2 on slide 57.

M1: Generalized Discrete Choice.

This uses the same assumptions as the single agent discrete choice model commonly used in econometrics. Its multiple agent analogue dates to Tamer (*Restud* 2003). More recent econometric implementation; Ciliberto-Tamer (*Econometrica* 2007),

Expectational Condition (FC3): $\forall d \in D_i$.

$$\pi(d, d_{-i}, z_i, \theta_0) = \mathcal{E}[\pi(d, \mathbf{d}_{-i}, \mathbf{z}_i, \theta_0)|\mathcal{J}_i]. \spadesuit$$

FC3: does not allow for any expectational error. It therefore rules out asymmetric and/or incomplete information*.

*Two single agent literatures do allow for ex-

Measurement Conditions (FC4).

$\pi(\cdot,\theta)$ is known.

$z_i = (\nu^f_{2,i}, z^o_i)$, $(d_i, d_{-i}, z^o_i, z^o_{-i})$ observed,

$$(\nu^f_{2,i}, \nu^f_{2,-i})|_{z^o_i, z^o_{-i}} \sim F(\cdot;\theta),$$

$F(\cdot,\theta)$ is known. ♠

FC4 does not allow for specification error (in $\pi(\cdot)$) or measurement error. Some of the z_i are observed by the econometrician (z^o_i) and some are not ($\nu^f_{2,i}$). The agents know $(\nu^f_{2,i}, \nu^f_{2,-i})$ (from FC3).

pectational errors; (i) dynamic discrete choice (Keane and Wolpin, *Review of Economic Dynamics, 2009*), (ii) literature using measures of expectations (see Manski, *Econometrica*, 2004).

Implication FC3 + FC4.

$$\Delta\pi(d_i, d', d_{-i}, z_i^o, \nu_{2,i}^f; \theta_0) \geq 0,$$

$\forall d' \in D_i$, and

$$(\nu_{2,i}^f, \nu_{2,-i}^f)|_{z_i^o, z_{-i}^o} \sim F(\cdot; \theta_0).$$

To insure that the model assigns positive probability to the observed decisions for some θ typically also assume:

$$\pi(d, d_{-i}, z_i^o, \nu_{2,i}^f) = \pi^{as}(d, d_{-i}, z_i^o, \theta_0) + \nu_{2,i,d}^f,$$

and that the distribution $\nu_{2,i}^f$ conditional on $\nu_{2,-i}^f$, has full support.

Notes.

• **Single Agent Problems.** FC3 and FC4 are implicit in the standard single agent discrete choice literature where we observe the choice but not returns (profits or utility)*.

• **Models with Multiple Agents.** Assume now that there is no ν_1 error, and we have a full information equilibrium (or at least that there are common unobservables that affect all agents). Then there is an additional problem. The r.h.s. contains a decision variable, d_{-i}, and by assumption the $-i$ agents know $\nu_{2,i}$ when making their decisions. So we need a different estimation algorithm.

*However in the single agent literature the model used . can be derived as a reduced form from a model with ν_1 errors; see Pakes, 2014.

Classic Example: Entry game. Early literature; market specific unobservable clouded the effects of competition on firm value. The number and type of competing firms had a positive effect on firm value on firm value. More profitable markets had more firms and we could not control for sources of market profitability.

Estimation. Ideas date to Tamer (2003). Estimation described here begins with Ciliberto, Murry, and Tamer (2016), interacting agent version of the classic discrete choice literature. The parametric distribution for $(\nu^f_{2,i}, \nu^f_{2,-i})$ does not deliver a likelihood (multiple equilibria).

• Can check whether the conditions of the model are satisfied at the observed (d_i, d_{-i})

for any $(\nu^f_{2,i}, \nu^f_{2,-i})$ and θ, and this, together with $F(\cdot, \theta)$, enable us to calculate the probability of those conditions being satisfied. These are necessary conditions for the choices: $\Rightarrow$ at $\theta = \theta_0$ the probability of satisfying them must be greater then the probability of observing (d_i, d_{-i}) (the necessary conditions deliver an "outer measure")

• Can check whether (d_i, d_{-i}) are the only values of the decision variables to satisfy the necessary conditions for any $(\nu^f_{2,i}, \nu^f_{2,-i})$ and θ; provides a lower bound to the probability of actually observing (d_i, d_{-i}) given θ (provide an "inner measure").

Define

$$\overline{P}\{(d_i, d_{-i}) \, |\theta\} \equiv$$

$Pr\{(\nu^f_{2,i}, \nu^f_{2,-i}) :$ (d_i, d_{-i}) satisfy M1 $|z^o_i, z^o_{-i}, \theta\}$,

$$\underline{P}\{(d_i, d_{-i})\ |\theta\} \equiv$$

$Pr\{(\nu^f_{2,i}, \nu^f_{2,-i}) :$ only(d_i, d_{-i}) satisfy M1$|z^o_i, z^o_{-i}, \theta\}$.

Note that

$$P\{(d_i, d_{-i})|\theta\} \equiv Pr\{(d_i, d_{-i})\ |z^o_i, z^o_{-i}, \theta\},$$

depends on the unknown true equilibrium selection mechanism, but whatever that mechanism

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

which is used as a basis for estimation.

Estimating Equations. If $h(\cdot)$ is a positive function then

$$E(\overline{P}\{(d_i, d_{-i})\ |\theta\} - \{d = d_i, d^{-i} = d_{-i}\})h(z_i^o, z_{-i}^o)$$

$$= (\overline{P}\{(d_i, d_{-i})\ |\theta\} - P\{(d_i, d_{-i})|\theta_0\})h(z_i^o, z_{-i}^o),$$

and

$$E(\{d = d_i, d^{-i} = d_{-i}\} - \underline{P}\{(d_i, d_{-i})\ |\theta_0\})h(z_i^o, z_{-i}^o)$$

$$(P\{(d_i, d_{-i})|\theta\} - \underline{P}\{(d_i, d_{-i})\ |\theta_0\})h(z_i^o, z_{-i}^o)$$

should be non-negative $at\ \theta = \theta_0$.

Ciliberto Tamer's example is entry into airline markets. If i indexes firms and m indexes markets their profit function is

$$\pi_{i,m} = S_m' \alpha + W_{i,m}' \gamma + \sum_{i \neq j} \delta_j^i y_{i,m} + \sum_{j \neq i} Z_{j,m}' \phi_j y_{j,m} +$$

$$+\nu_{2,m}^o + \nu_{2,m}^d + \nu_{2,m}^a + \nu_{2,i,m}$$

where

- i indexes firms m indexes markets (city-pairs)

- S_m are market characteristics.

- and their are destination, origin, airport, and firm-market specific ν_2 disturbances, but no ν_1 disturbances.

Notice that they have left the interactions be firm specific because they are particularly concerned with the differences between the interactions of; low-cost carriers, Southwest (low cost but considered separate), and the majors, and between the majors themselves. A major question is how much does airport dominance restrict entry.

They are especially concerned with the two airports in the Dallas area (Love and Dallas/Ft.Worth), and the effect of the Wright amendment on entry (the Wright amendment restricted the markets that aircraft from Love could service in order to stimulate Dallas/Fort Worth).

This is quite rich (indeed we seem to be playing in a different "league" from the earlier work), but the model is not derived from a demand system and a cost function, and therefore one would assume it has error (an approximation error) which is not accounted for (along with the other issues discussed above).

That is if the real profit or value function from operating in a market were $V(\cdot)$, we could project it down (or take its expectation conditional on) the included (observed and unobserved) variables, and get an equation identical to theirs except that there would be a residual from the projection. The residual would, by construction, be orthogonal to included variables – and so different from the included unobservables (which are determinants of airport choice and hence not orthogonal to them).

Estimation Routine. 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, we employ simulation techniques to obtain an unbiased estimate 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, at the observed (d_i, d_{-i}). Estimate $\overline{P}(d_i, d_{-i}|\theta)$ by the fraction of random draws that satisfy that inequality 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.

Notes

- This can become computationally demanding, particular if every time you need to compute an equilibrium I need to re-evaluate a fixed point (then it can become incredibly computationally demanding).

- The estimator here is based on inequalities, and so under the null we should expect a set of θ values that satisfy it. I.e there is a need for the Econometrics of set estimators (I will come back to this).

- Often sampling error will cause the set to be empty. Take the simple case where we are estimating a single parameter. Then the inequalities set lower and upper bounds for the parameter. If the estimate of the lower bound is the greatest lower bound it will be biased upwards in finite samples. Similarly if the estimate of the upper bound is the least upper bound, then that estimate has a negative bias in finite samples. If they cross there is no point which satisfies all the inequalities.

Tables 3 and 4 in Ciliberto Tamer They report confidence intervals for points. First they find a set such that the probability that the confidence set covers the true point is .95. Then they project each point down onto the axis. Table 3 assumes iid errors. LAR=large airlines. LCC=low cost carriers. WN=Southwest.

- The first column constrains competitive effects to be the same, so as in Berry (1992) there is a unique number of firms. The competitive effects (number of other airlines serving the city-pair) and the airport presence effect are both strong, as in prior work.

- Column 2 allows for heterogeneity in continuation values, but the heterogeneity is just in constant dummies (the presence of one airline affects all other airlines equally). No longer a unique number of firms (depends on who enters). Competitive effects are similar except for the low cost carriers which have a bigger impact, and airport presence stronger yet.

- Column 3 allows one airlines presence to effect different airlines differently. Now Large airlines (LAR) and Southwest (=WN) have strongly negative effects on LCC, and there are smaller differences among the rest.

Table 4. Now errors are allowed to be correlated, and have different variances. Also the effect of the presence of the airline is interacted with its airport presence (do airlines with a greater presence have a bigger deterrence effect on entry into a market?).

- The presence of a low cost carrier or Southwest has a very negative effect on the majors. Also airport presence more generally has a negative effect on others.

- Comparing column 1 to 2 compares iid error specification with a specification which allows for a free covariance structure. There are quite large differences in the coefficients of interest, and they treat the disturbance distribution more carefully than most others.. This is a bit disturbing since there is often little direct evidence on the form of the error distribution, and, since it is a reduced form specification and does contain a reduced form error, they probably have it wrong still. Perhaps it should have been expected because the estimator is a non-linear functional of the error distribution; i.e. it is ϵ's in the tail that will determine entry; it matters if tails are correlated.

- Column 3 takes out the airport presence variables, because one might think they are endogenous, and puts in carrier cost. Coefficients change and the cost effect is particularly sharp for Southwest.

They go through and do various "policy experiments" using these results. They do this by simulating equilibria under different circumstances and giving upper and lower bound probabilities for what would happen were those circumstances in place. In particular they find that elimination of the Wright amendments would have increased entry significantly (and they quantify its various effects).

Remaining Issues: Analysis of Entry Games.

- Partly for the computational reasons introduced above, almost all of the work with these type of models works with "reduced form" profit functions, not derived directly from a rich model of the second stage. The early papers try to circumvent this by using non-parametrics, but that runs into data limitations, as the dimension of the non-parametric function increases as a multiple of: (i) the number of agents, and

(ii) the number of distinguishing characteristics of each agent. It is exactly this issue that the differentiated product demand systems were designed to handle. However that requires a demand system, and a second stage profit function and the latter would have to be recomputed for every value of the parameter vector.

- This is a reduced form, but if we are using approximations to profit functions we might want to allow for errors in our approximations (they are currently absent).

- Lack of True Dynamics; that is who is active this period depends on who was active last period, and on perceptions of profitability in future periods. Note that not only would dynamics get rid of some of the conceptual problems but it would allow you to bring in the information in the sequential nature of most data. That is in many of these data sets we know exactly when each firm entered, so we could condition that firm's entry decision on firms existing in the market at the time they enter. The problems here are a mix of computational and conceptual, and we come back to them after we introduce dynamic games. A lot more can be done now, but the dynamic games solutions still ran into their own problems.

We now go back to the second behavioral model that underlies the use of moment inequalities.

M2: Requirements for Profit Inequalities.

Recall that we already have assumed C1 or the best response (or maximization) condition and C2 or the ability to construct counterfactuals. What we still need is:

- an assumption on the relationship between agents expectations and the expectation operator generated by the DGP, and

- restrictions on the measurement model.

Condition on Agents' Expectations. Let $h(\cdot)$ be a positive valued function, and $x_i \in \mathcal{J}_i$ be observable.

*Condition IC*3

$$(1/N)\sum_i \mathcal{E}(\Delta\pi(d_i, d', d_{-i}, z_i)|\mathcal{J}_i) \geq 0 \quad \Rightarrow$$

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

Three progressively weaker conditions. The weakest suffices.

Agents' expectations are

1. correct (Bayesian Nash),

$$\mathcal{E}(\Delta\pi_i(\cdot)|x_i) = E(\Delta\pi_i(\cdot)|x_i),$$

2. or are wrong, but not consistently so

$$(1/N)\sum_i \Big(\mathcal{E}[\Delta\pi(\cdot|x_i] - E[\Delta\pi(\cdot)|x_i]\Big) = 0,$$

3. or are consistently wrong but in an "overly optimistic" way - i.e.

$$(1/N)\sum_i \Big(\mathcal{E}[\Delta\pi(\cdot|x_i] - E[\Delta\pi(\cdot)|x_i]\Big) \ \geq 0.$$

Note. Generalized discrete choice model nested to this: expectations=realizations.

Condition on Measurement Model.

Assume D_i discrete and there is an $x \in \mathcal{J}_i$ and a function $c(\cdot) : D_i \times D_i \to \mathcal{R}^+$, such that we satisfy

Recall that

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

where

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

and

$$\nu_{1,i,d} \equiv$$

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

When we have a comprehensive measure of the profits from the action, it is just that those profits either contain expectational error or are measured with error, are the cases we mostly worry mostly about ν_1 errors.

Sufficient Condition, SIC4: $\nu_2 \equiv 0$.

This is the analogue of no ν_1 error in the generalized discrete choice model. When there are only ν_1 errors in the profit inequalities procedures gives you inequalities by simply averaging the difference in the measured profit function. That is given our three assumptions if

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

then

$$E[r(d, d_{-i}, z_i^o, \theta_0)|\mathcal{J}_i] \geq 0.$$

Which implies that provided $x_i \in \mathcal{J}_i$, at the true θ_0

$$\sum_i r(d, d_{-i}, z_i^o, \theta_0) h(x_i) \rightarrow_{a.s.} \kappa > 0.$$

The case when $\nu_2 \neq 0$: overcoming the selection problem.

The econometrician only has access to $\Delta r(\cdot, \theta)$ and our best response condition is in terms of the conditional expectation of $\Delta \pi(\cdot)$. So we need an assumption which enables us to restrict weighted averages of $\Delta r(\cdot)$ in a way that insures that the expectation of the weighted average of $\Delta r(\cdot, \theta)$ is positive at $\theta = \theta_0$.Here are two ways around it that are frequently used.

PC4a: Differencing. Here there are groups of observations with the same value for the ν_2 error. We end up getting difference in difference inequalities (the difference for one observation contains the same ν_2 error as the difference for the other).

Our supermarket example is a special case of $PC4a$. There $d_i = (b_i, s_i)$,

$$\pi(\cdot) = U(b_i, z_i) - e(b_i, s_i) - \theta_0 dt(s_i, z_i)$$

and $\nu_{2,i,d} \equiv U(b_i, z_i)$. If we measure expenditures up to a $\nu_{1,i,d}$ error,

$$r(\cdot) = -e(b_i, s_i) - \theta_0 dt(s_i, z_i) + \nu_{2,i,d} + \nu_{1,i,d}.$$

We chose a counterfactual with $b_i' = b_i$, so

$$\Delta r(\cdot) = \Delta\pi(\cdot) + \Delta\nu_{1,\cdot}$$

and the utility from the bundle of goods bought is differenced out.

"Matching estimators", i.e. estimators based on differences in outcomes of matched observations, implicitly assume $PC4a$ (no differences in unobservable determinants of the choices made by matched observations).

PC4b: Unconditional Averages and IV's. There is a counterfactual which gives us an inequality that is additive in ν_2 no matter the decision the agent made. The counterfactual may be different for different observations. Then we can form averages which do not condition on d so there is no selection problem.

Assume that $\forall d \in D_i$, there is a $d' \in D_i$ and a $w_i \in \mathcal{J}_i$ such that

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

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

$N^{-1} \sum \nu_{1,i,\cdot} h(x_i) \rightarrow_P 0$ and $N^{-1} \sum \nu_{2,i} h(x_i) \rightarrow 0$

or x is an "instrument" for both ν_2 and ν_1, so provided $h(\cdot) > 0$

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

converges to a positive number.

Case 2 of our supermarket example had two ν_2 components; a decision specific utility from the goods bought, $\nu_{2,i,d} = U(b_i, z_i)$ (like in case 1), and an agent specific aversion to drive time, $\theta_i = \theta_0 + \nu_{2,i}$. As in case 1, taking $d' = (b_i, s_i')$ differenced out the $U(b_i, z_i)$.

Then

$$\Delta r(\cdot) = -\Delta e(\cdot, s_i, s_i') - (\theta_0 + \nu_{2,i})\Delta dt(s_i, s_i', z_i) + \Delta\nu_{1,\cdot\cdot}$$

Set $w_i = [\Delta dt(s_i, s_i', z_i)]^{-1} \in \mathcal{J}_i$, then C1 and C2 $\Rightarrow$

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

This inequality is;
(i) linear in $\nu_{2,i}$, and
(ii) is available for every agent.

So if $E[\nu_2] = 0$, $PC3$ and a law of large numbers insures $N^{-1}\sum_i \nu_{2,i} \rightarrow_P 0$, and

$$\sum_i \Delta e(s_i, s_i', b_i)/\Delta dt(s_i, s_i', z_i) \rightarrow_P \underline{\theta}_0 \leq \theta_0$$

while if $E[\nu_2|x] = 0$ we can use x to form instruments which give us the additional inequalities

$$\frac{\sum_i h(x_i)\frac{\Delta e(s_i,s_i',b_i)}{\Delta dt(s_i,s_i',z_i)}}{\sum_i h(x_i)} \rightarrow_P \underline{\theta_0} \leq \theta_0$$

Notice that $\nu_{2,i}$ can be correlated with $dt(z_i, s_i)$ so this procedure enables us to analyze discrete choice models when a random coefficient affecting tastes for a characteristic is correlated with the characteristics chosen.

General Condition *Condition $IC4$*:

$$\sum_{j \in D_i} \chi\{d_i = j\} c(j, d'(j))(\nu_{2,i,j} - \nu_{2,i,d'(j)})h(x_i) \ \leq 0$$

where $\chi\{d_i = j\}$ is an indicator function.

Notes.

- This is an unconditional average (does not condition on d_i); i.e. for every possible $d \in \mathcal{D}_i$ we specify a $d'(d)$ (*a priori*).

- This average is an average of *differences* in the $\nu_{2,i,j} - \nu_{2,i,d'(j)}$.

- Both (i) the weights, and (ii) the comparison (d'), can vary with j.

- We assumed $x_i \in \mathcal{J}_i$. Could also us an $x_{-i} \in \mathcal{J}_{-i}$ provided x_{-i} is not correlated with $\nu_{1,i}$ which might well be violated in models with asymmetric information.

Summary: Profit Inequality Model.

- Allows for specification errors, incorrect expectations, and incomplete and asymmetric information,

and it does so without requiring the econometrician

- to specify what the agent knows about either its competitors, or about the state of nature

- It requires a restriction on $\{\nu_{2,d}\}$, but given that restriction, there is no need for the distribution of $\{\nu_{2,d}\}$.

Other examples of use of inequalities in I.O.

- Contracting (bargaining) models in vertical markets. A party which accepts a contract must expect to earn more from when the contract was in force then they would have earned were the contract not in force; and if a contract is rejected the opposite must be the case. Enables an analysis of the characteristics of the contracts signed in vertical markets Ho (2009), Crawford and Yorukoglu (2012).

- Product repositioning (see below)

- Ordered choice models and other discrete investments by firms (see below).

Product Repositioning and Short-Run Responses to Environmental Change

- Product repositioning: a change in the characteristics of the products marketed by an incumbent firm.

- Empirical analysis of equilibrium responses to environmental changes typically distinguish between the response of
  - "static" controls (prices or quantities)
  - "dynamic" controls effects (entry, exit, and various forms of investment including in new products).

- Product repositioning generally allocated to dynamics. Dynamics are harder to do formally (especially when there are time constraints, as is often the case when policy decisions must be made) and so often left to informal analysis.

Recent work:

- a number of industries in which firms already in the market can change the characteristics of their products as easily as they can change prices, and

- shows that static analysis that does not take repositioning into account is likely to be misleading, even in the very short run.

- analysis does raise the issue of multiplicity of equilibria (come back to this).

Examples.

- Nosko (2014): Response of the market for CPU's to innovation: easy to change chip performance to lower values than the best performing chips of the current generation.

- Eizenberg (2014): Introduction of the Pentium 4 chip in PC's and notebooks: decisions to stop the production of products with older chips (and lower prices) is easy to implement. Total welfare does not increase, but poorer consumers do better with the low end kept in.

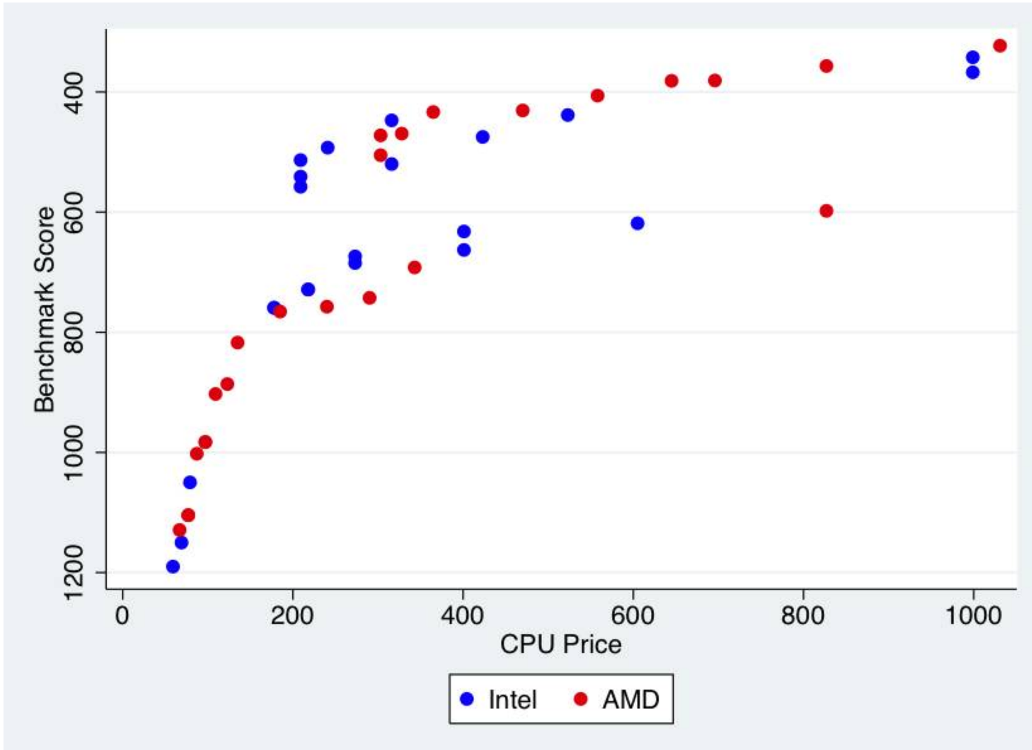
- Wollmann (2016): commercial truck production process is modular (it is possible to connect different cab types to different trailers), so some product repositioning immediate. Considers the bailout of GM and Chrysler, and ask what would have happened had GM and Chrysler been forced to exit the commercial truck market (once allowing for product repositioning and once not), and once with pure exit and once with them being bought out by an existing producer.

Nosko: Intels' Introduction of The Core 2 Duo Generation in Desktops.

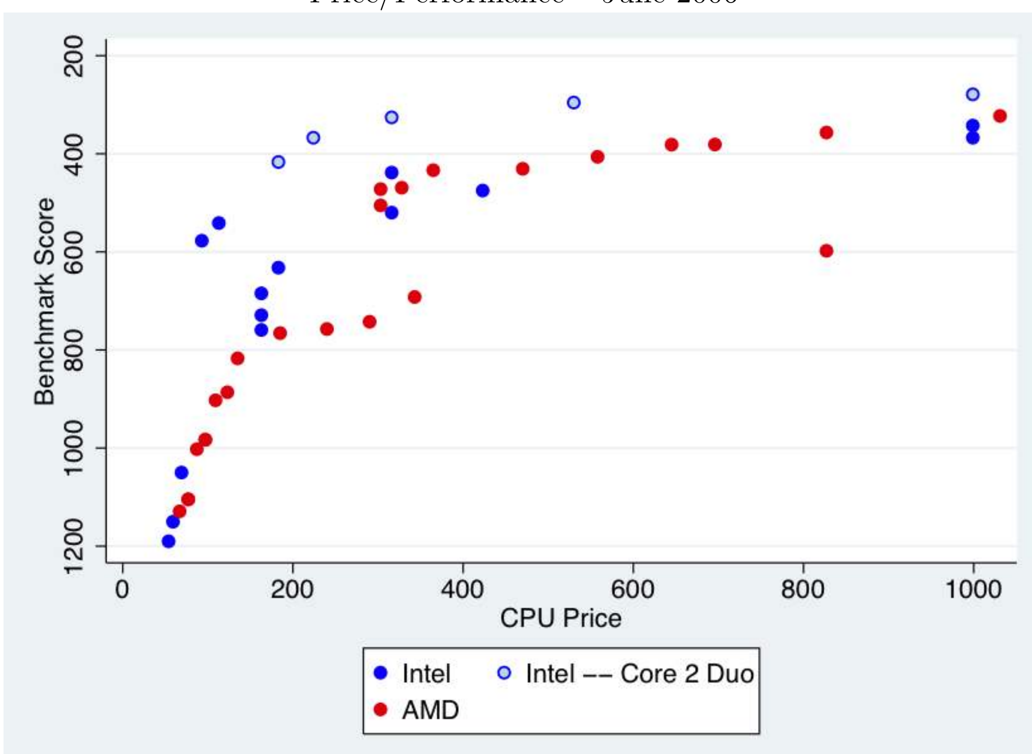
- Chips sold at a given price typically change their characteristics about as often as price changes on a given set of characteristics.

- Figures provide benchmark scores and prices for the products offered at different times.

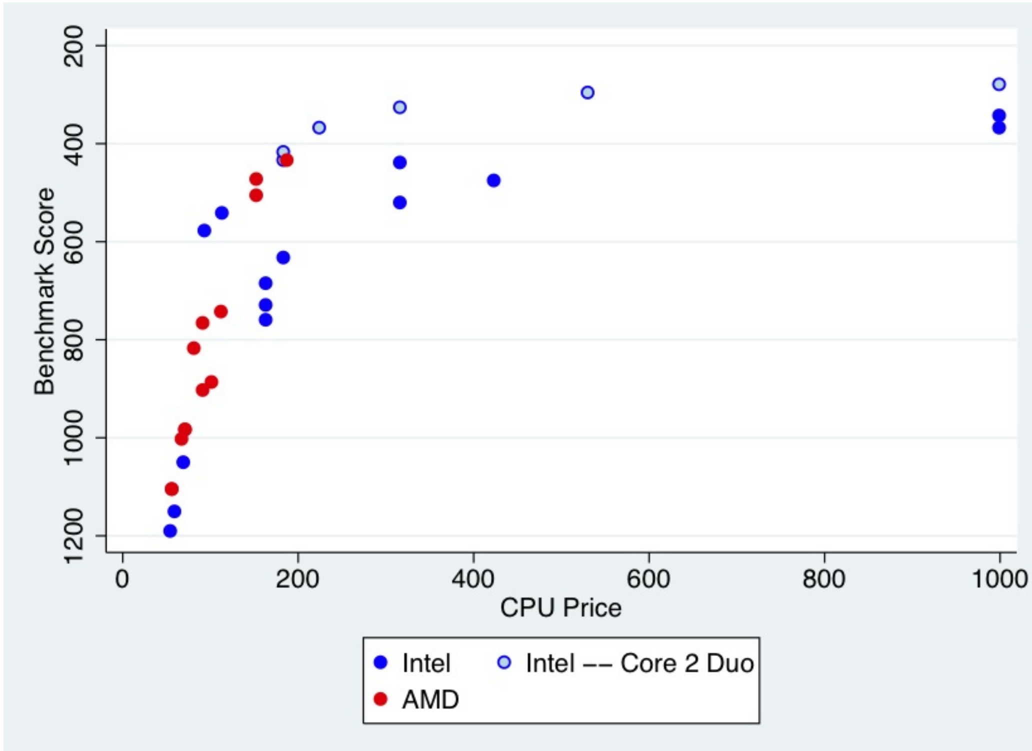
  - June 2006: just prior to the introduction of the Core 2 Duo. The red and blue dots represent AMD's and Intel's offerings. Intense competition for high performance chips with AMD selling the highest priced product at just over $1000: seven



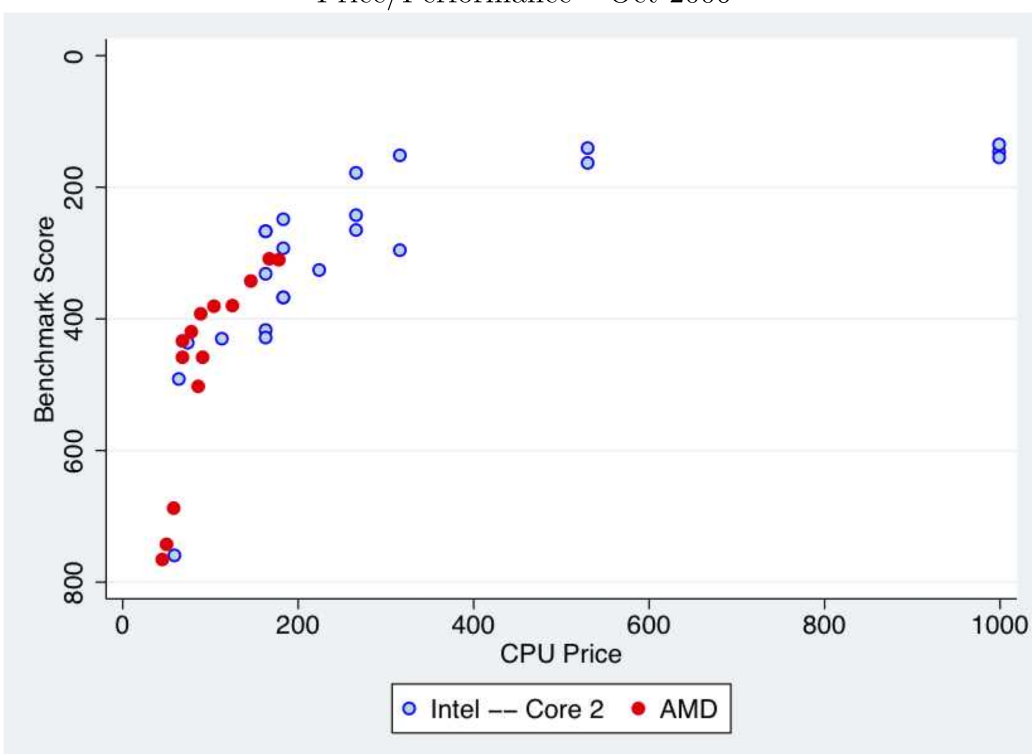
Price/Performance – June 2006



Price/Performance – July 2006



Price/Performance – Oct 2006



Price/Performance – January 2008

sold at prices between \$1000 and \$600.

- Core 2 Duo introduced in July. By October; (i) AMD no longer markets any high priced chips (ii) there are no chips offered between \$1000 and \$600 dollars.

- November 2006: Only Core 2 Duo's at the high end.

- Nosko goes on to explain

  - that the returns from the research that went into the Core 2 Duo came primarily from the markups Intel was able to earn as a result of emptying out the space of middle priced chips and dominating the high priced end of the spectrum.

– how a similar phenomena would likely occur if AMD were to merge with Intel.

Analytic Framework Used in these Papers.

- Two-period sub-game perfect model (backward induction)
  - product offerings set in the first stage and
  - prices set in the second.
- Two-period model ignores effect on subsequent periods. Come back to correct this.
- Even for two-period model, need
  - Estimates of the fixed costs of adding and of deleting products.

– A way of dealing with the multiplicity problem if we compute counterfactuals (and all papers do).

Estimates of Fixed Costs (F);

The three examples use

- Estimates of demand and cost as a function of product characteristics (use either BLP or the pure characteristics model in Berry and Pakes, 2007).

- An assumption on the pricing (or quantity setting) in the "but for" world in which; (i) one the products that was offered was not, and (ii) one that was not offered was offered (use Nash pricing equilibrium).

- The profit inequality approach proposed in Pakes, Porter, Ho, and Ishii (2015) and Pakes (2010).

Constant F case.

- x_j be a vector of 1's and 0's; 1 when the product is offered. Say e_z is vector with one in the "z" spot and zero elsewhere.

- Assume z had been added. Compute the the implied profits had the product not been added (unilateral deviation in a simultaneous move game).

- Let $\Delta\pi_j(x_j, x_j - e_z, x_{-j}) \equiv \pi_j(x_j, x_{-j}) - \pi_j(x_j - e_z, x_{-j})$.

- $\mathcal{I}_j$ is the agent's information set. z_j added because

$$E[\Delta\pi_j(x_j, x_j - e_z, x_{-j})|\mathcal{I}_j] \geq F.$$

- Average over all the products introduced and assume agents' expectations are unbiased. $\Rightarrow$ a consistent lower bound for F.

- If z is a feasible addition that was not offered and $\Delta\pi_j(x_j, x_j + e_z, x_{-j}) \equiv \pi_j(x_j, x_{-j}) - \pi_j(x_j + e_z, x_{-j})$, then

$$E[\Delta\pi_j(x_j, x_j + e_z, x_{-j})|\mathcal{I}_j] \leq F.$$

  which gives us an upper bound to F.

Complications: Non-constant F.

- If the fixed costs are a function of observed characteristics of the product all we need is more complicated moment inequality estimators.

- Allowance for unobservable fixed cost differences that were known to the agents when they made their product choices implies that the products provided may have been partially selected on the basis of having lower than average unobservable fixed costs (and vice versa for those that were not selected). Need a way of dealing with ν_2 errors.

- In addition to the suggestions above, you could assume a bounded support as in Manski (2003); for an application which combines them see Eizenberg (2014).

Complications: Sunk (in contrast to Fixed) Costs.

- Find a z that was not marketed, and assume that the firm could have marketed it and commit to withdrawing it in the next period *before* competitors next period decisions are taken.

- Then our behavioral assumption implies that the difference in value between, (i) adding this z and then withdrawing it in the next period, and (ii) the value from just marketing the products actually marketed, would be less than zero. I.e.

  $$E[\pi_j(x_j+e_z, x_{-j})-\pi_j(x_j, x_{-j})|\mathcal{I}_j] \leq F+\beta W,$$

$W \geq 0$ is the cost of withdrawing and β is the discount rate.

- Lower bounds require further assumptions, but the upper bound ought to be enough for examining extremely profitable repositioning moves following environmental changes (like those discussed in Nosko (2014)).

Discrete Investment Choices by A Firm.

This application is due to Ishii (thesis and PPHI). It is about analyzing choices of a number of ATM's but as will become obvious similar analysis could be used for at least some types of entry games.

Ishii analyzes how ATM networks affect market outcomes in the banking industry. The part of her study we consider here is the choice of the number of ATMs. General issue: techniques that can be used to empirically analyze "lumpy" investment decisions, or investment decisions subject to adjustment costs which are not convex for some other reason*, in market environments.

*Actually Ishii's problem has two sources of non-convexities. One stems from the discrete nature of the number of ATM choice, the other from the fact that network effects can generate increasing returns to increasing numbers of ATMs

Ishii uses a two-period model with simultaneous moves in each period.

- 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.

- Second period interest rates are set conditional on the ATM networks in existence and consumers chose banks.

Note that there are likely to be many possible Nash equilibria to this game so again there is a multiplicity problem.

Getting the second stage profit function?
- Estimate a demand system for banking

services (discrete choice model among a finite set of banks with consumer and bank specific unobservables; as in BLP).and

- an interest rate setting equation.

Both conditional on the number of ATMs of the bank and its competitors, i.e. on (d_i, d_{-i}). Interest rates set in a simultaneous move Nash game.

Note. We need to know what interest rates would be and where consumers would go were there a different network of ATMs to get the counterfactuals. Need to assume that the solution to the second stage is unique; or at least that you are calculating the one all participants agree would occur. Come back to the realism of this below.

The ATM Choice Model. To complete the analysis of ATM networks Ishii requires estimates of the cost of setting up and running ATMs. Crucial to the analysis of the implications of existing network (is there over or under investment, are ATM networks allowing for excessive concentration and excessively low interest on customer accounts,...) and of what the network is likely to result from alternative institutional rules (of particular interest is the analysis of systems that do not allow surcharges, as suggestions to eliminate surcharges have been part of the public debate for some time).

We infer what cost must have been for the network actually chosen to be optimal. So we model choice network size; of $d_i \in \mathcal{D} \subset \mathcal{Z}^{+}$, the non-negative integers. We assume a simultaneous move gain. The agent forms a perception on the distribution of actions of its competitors and of likely values of the variables that determine profits in the next period, and chooses the d_i that maximizes expected profits. So this is a multiple agent ordered choice model.

Formally

$$\mathcal{E}[\pi(y_i, d_i, d_{-i}, \theta)|\mathcal{J}_i] = \mathcal{E}[r(z_i, d_i, d_{-i})|\mathcal{J}_i] - (\theta + \nu_{2,i})d_i, \quad (1)$$

where

- $\mathcal{J}_i$ is the information known by the agents when the decisions on the number of ATM's must be made,

- θ is average cost of an ATM, and the $\nu_{2,i}$ capture the effects of cost differences among banks that are unobserved to the econometrician but known to the agent. What we know is there are a set of instruments such that $E[\nu_{2,i}|x_i] = 0$

Clearly a necessary condition for an optimal choice of d_i is that:

- expected profits from the observed d_i is greater than the expected profits from $d_i - 1$

- expected profits from the observed d_i is greater than the expected profits from $d_i + 1$.

Since we can calculate what the bank would earn in income in both those situations, these two differences provide inequalities that the costs of ATMs must satisfy, and when we average them over banks, they provide an inequality estimator of θ.[†]

The inequality for the first case is[‡]

$$0 \leq \mathcal{E}[\pi(z_i, d_i, d_{-i}, \theta)|\mathcal{J}_i] - \mathcal{E}[\pi(z_i, d_i - 1, d_{-i}, \theta)|\mathcal{J}_i] =$$

$$\mathcal{E}[r(z_i, d_i, d_{-i})|\mathcal{J}_i] - \mathcal{E}[r(z_i, d_i - 1, d_{-i})|\mathcal{J}_i] - (\theta + \nu_{2,i})$$

[†]These conditions will also be sufficient if the expectation of $\pi(\cdot)$ is (the discrete analogue of) concave in d_i for all values of d_{-i}, a condition which works out to be almost always satisfied at the estimated value of θ.

[‡]More formally to get this we use $PC4$ substituting

$$h(j, d'(j), \cdot) = 1 \;\; if j = d_i; \;\; h(j, d'(j), \cdot) = -1 \;\; if j = d_i - 1,$$

and $h(j, d'(j), \cdot) = 0$ elsewhere.

This will give us are upper bound for θ. I will let you work out the second case. It gives us our lower bound.

A few points are worthy of note.

- Note we have chosen $d'(d_i)$ in a way that insures we keep a $\nu_{2,i}$ for every agent (there is no selection).

- To do this we need to solve out for the returns that would be earned were there a different ATM network (for $r(y_i, d_i - 1, d_{-i})$, etc.) $\Rightarrow$ we have to solve out for the interest rates that would prevail were the alternative networks chosen. This is why you need the structural static model; i.e. we need approximations to counterfactuals.

- The expectation is conditional on information known when the decisions are made. It is over any component of y_i not known at the time decisions are made, and over the actions of the competitors (over d_{-i}). Note that we do not need to specify what that information set is.

Our behavioral assumptions imply.

$$E\Big(r(z_i,d_i,d_{-i})-r(z_i,d_i-1,d_{-i})-(\theta_0+\nu_{2,i})\Big) \geq 0$$

and

$$E\Big(r(z_i,d_i,d_{-i})-r(z_i,d_i+1,d_{-i})+(\theta_0+\nu_{2,i})\Big) \leq 0,$$

with $\sum \nu_{2,i} = 0$ by construction. If we had an instrument (an x which is the in the agents' information set when it made its decision) that was orthogonal to $\nu_{2,i}$ and $h(\cdot)$ was a positive value function, our behavioral assumptions would also imply

$$\sum_i E\Big(r(z_i,d_i,d_{-i})-r(z_i,d_i-1,d_{-i})-(\theta_0+\nu_{2,i})\Big)h(x_i) \geq$$

Simplest Estimator. Let $\Delta\bar{r}_L$ be the sample average of the returns made from the last ATM installed, and $\Delta\bar{r}_R$ be the sample average of the returns that would have been made if one more ATM had been installed. Then

$$\Delta\overline{r}_L - \theta \geq 0 \quad (i.e.\ \ \Delta\overline{r}_L \geq \theta),$$

and

$$-\Delta\overline{r}_R + \theta \geq 0 \quad (i.e.\ \ \theta \geq \overline{r}_R).$$

Assuming $|\Delta\overline{r}_R| \leq \Delta\overline{r}_L$

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

Notes. With more instruments the *lower bound* for θ_0 is the *maximum* of a finite number of moments, each of which distribute (approximately) normally. So actual lower bound has a positive bias in finite samples. The estimate of the upper bound is a minimum, so the estimate will have a negative bias. $\Rightarrow \hat{\Theta}_J$ may well be a point even if Θ_0 is an interval. Importance of test.

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 and have the same sign for all i. In this case, weights proportional to $1/|t_i|$ satisfy Assumption 3.

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).

We start out now with slight more detailed notation, allowing for a different structural error for every d_i, d_{i+t}, say $\nu_{2,i,d_i,d_i+t} = t\eta_i$ in the ATM model (so η_i is now the firm specific unobserved cost of the ATM). By definition of the parameter θ_0, $\mathcal{E}\eta_i = 0$. To deal with the boundary problem, we make an additional assumption that

Assume 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 un-truncated 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 choice set in the ATM model is $d_i \geq 0$ 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 order statistic notation, i.e.

$$\eta_{(1)} \leq \eta_{(2)} \leq \cdots \leq \eta_{(n)}.$$

Let

$$L_\eta = \{i : \eta_i \leq \eta_{(n_L)}\} \text{ and } U_\eta = \{i : \eta_i \geq \eta_{(n_L+1)}\}.$$

Similarly, let $\Delta r_i^+ = \Delta r(\mathbf{d}_i, \mathbf{d}_i + 1, \mathbf{d}_{-i}, \mathbf{z}_i)$ and

$$\Delta r_{(1)}^+ \leq \Delta r_{(2)}^+ \leq \cdots \leq \Delta r_{(n)}^+$$

while

$$U_r = \{i : \Delta r_i^+ \geq \Delta r_{(n_L+1)}^+\}.$$

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

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

and form

$$\sum_i \sum_{d' \in \mathcal{D}_i} h^i(d'; \mathbf{d}_i, \mathcal{J}_i, \mathbf{x}_{-i}) \Delta r(\mathbf{d}_i, d', \mathbf{d}_{-i}, \mathbf{z}_i^o, \theta_0)$$

$$= \frac{1}{n} \sum_{i \in L} \Delta r(\mathbf{d}_i, \mathbf{d}_i - 1, \mathbf{d}_{-i}, \mathbf{z}_i^o, \theta_0) + \frac{1}{n} \sum_{i \in U_r} \Delta r(\mathbf{d}_i, \mathbf{d}_i + 1, \mathbf{d}_{-i}, \mathbf{z}_i^o, \theta_0)$$

$$\geq \frac{1}{n}\sum_{i\in L}\Delta r(\mathbf{d}_i,\mathbf{d}_i-1,\mathbf{d}_{-i},\mathbf{z}_i^o,\theta_0)+\frac{1}{n}\sum_{i\in U_\eta}\Delta r(\mathbf{d}_i,\mathbf{d}_i+1,\mathbf{d}_{-i},\mathbf{z}_i^o,\theta_0)$$

$$=\frac{1}{n}\sum_{i\in L}\{\mathcal{E}[\Delta\pi(\mathbf{d}_i,\mathbf{d}_i-1,\mathbf{d}_{-i},\mathbf{z}_i)|\mathcal{J}_i]-\nu_{2,i,\mathbf{d}_i,\mathbf{d}_i-1}\}$$

$$+\frac{1}{n}\sum_{i\in U_\eta}\{\mathcal{E}[\Delta\pi(\mathbf{d}_i,\mathbf{d}_i+1,\mathbf{d}_{-i},\mathbf{z}_i)|\mathcal{J}_i]-\nu_{2,i,\mathbf{d}_i,\mathbf{d}_i+1}\}$$

$$\geq -\frac{1}{n}\left\{\sum_{i\in L}\nu_{2,i,\mathbf{d}_i,\mathbf{d}_i-1}+\sum_{i\in U_\eta}\nu_{2,i,\mathbf{d}_i,\mathbf{d}_i+1}\right\}.$$

The first inequality holds by the definition of U_r and noting $\Delta r(\mathbf{d}_i,\mathbf{d}_i+1,\mathbf{d}_{-i},\mathbf{z}_i^o,\theta_0)=\Delta r_i^+ +\theta_0$. The second follows from the fact that $\mathcal{E}[\Delta\pi(\mathbf{d}_i,\mathbf{d}_i-1,\mathbf{d}_{-i},\mathbf{z}_i)|\mathcal{J}_i]>0$ for $i\in L$ and $\mathcal{E}[\Delta\pi(\mathbf{d}_i,\mathbf{d}_i+1,\mathbf{d}_{-i},\mathbf{z}_i)|\mathcal{J}_i]>0$ for all i.

Now note that

$$-\frac{1}{n}\left\{\sum_{i\in L}\nu_{2,i,\mathbf{d}_i,\mathbf{d}_i-1}+\sum_{i\in U_\eta}\nu_{2,i,\mathbf{d}_i,\mathbf{d}_i+1}\right\}=\frac{1}{n}\left\{\sum_{i\in L}\eta_i-\sum_{i\in U_\eta}\eta_i\right\}$$

$$\geq \quad \frac{1}{n}\left\{\sum_{i\in L_\eta}\eta_i-\sum_{i\in U_\eta}\eta_i\right\}=\frac{1}{n}\left\{\sum_{i=1}^{n_L}\eta_{(i)}-\sum_{i=n_L+1}^{n}\eta_{(i)}\right\}.$$

Under the assumption that η_i are i.i.d. and symmetrically distributed about zero, the last term above

has mean zero. So, $\mathcal{E}\left[-n^{-1}\sum_{i\in L}\nu_{2,i,\mathbf{d}_i,\mathbf{d}_i-1}-n^{-1}\sum_{i\in U_\eta}\nu_{2,i,\mathbf{d}_i,\mathbf{d}_i+1}\right]$
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 to PPHI shows that we can use instruments along with a symmetry assumption to generate more moment inequalities for the lower bound.

Inequality Method, ATM Costs*

	θ_J	95% CI for θ LB	UB
1. $h(x) \equiv 1, d \geq 1$ u.b. $\widehat{\theta}$	[24,452, 25,283]	20,544	29,006
2. $h(x) \equiv 1, d \geq 0$	[24,452, 26,444]	20,472	30,402
$h(x) =$ Inst.			
3. $d \geq 1$ for u.b. $\widehat{\theta}$	19,264	16,130	23,283
4. $d \geq 0$	20,273	17,349	24,535
$\{d : \|d - d_i\| = 1, 2\}$, $h(x) = 1$			
5.$\{d : \|d - d_i\| = 1, 2\}$; $d \geq 1$ u.b. $\widehat{\theta}$	[24,452, 25,283]	20,691	28,738
6.$\{d : \|d - d_i\| = 1, 2\}$; $d \geq 0$	[24,452, 26,644]	20,736	29,897
F.O.C (Hansen & Singleton,1982)			
7. h(x)=1	28,528	23,929	33,126
8. h(x)=IV	16,039	11,105	20,262

* There are 291 banks in 10 markets. The IV are 1,pop, # Banks in Mkt, # Branches of Bank). 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.

Results (see table).

- First two rows just use a constant and you can see that when you do the selection correction (second row) the upper bound goes up a bit. There is mediocre precision.

- When we add instruments we get a point estimate, but it is just outside the bounds and a formal test marginally rejects the instruments.

- Adding equations for $|d - d_i| = 2$ does not do much, as it shouldn't if the profit function is concave.

- An alternative procedure is Hansen and Singleton's F.O.C. estimator. It gets a number which is about the upper bound of our c.i. and would be rejected if we accepted the c.i. of the IV estimator.

- Works out to $4,500 per ATM per month). Quite a bit larger than prior estimates which do not take into account all aspects of costs.

Implications. Ishii (thesis). Large banks subsidize their ATM networks in order to gain customers (whom they pay lower interest rates to). The question of whether to force equal access to all ATMs and a central surcharge was considered in congress. She considers a counterfactual with the same number of ATMs, imposes a universal ATM user fee that would just cover ATM costs, and recalculates equilibrium. A centralized surcharge would reallocate profits from large to small banks and decrease concentration markedly. Welfare effects (conditional on the network) not as obvious because of costs of ATMs. She also show that investment in ATMs is suboptimal; so one might want to make the ATMs endogenous and see what happens, but then we get faced with, among other things, the issue of multiplicity of equilibria.

Digression: Multiple Equilibria and Counterfactual's in Ishii's game.

Selection of Equilibria for Counterfactuals. Possibilities that have been used.

- Enumerate all possible (or at least all relevant) equilibria (used in Eizenberg, 2014).
  - Seems like there may be many, but investment history limits what can be supported. (see Lee and Pakes, 2009, for an example).

- Use a learning model to select among equilibria (used in Wollmann, 2016).

– Eg.s: best response, fictitious play (Fudenberg and Levine, 1998, for a discussion of alternatives.).

– Will settle down at a Nash equilibrium. Repeat and get a probability distribution of possible equilibria.

– Probably not suitable for major changes that induce experimentation (Doraszelski, Lewis, and Pakes, 2016).

This is taken from Lee and Pakes (2009, *Economic Letters*). Take Ishii's information on Pittsfield, Massachusetts and analyze the likely impact of a change in Pittsfield's banking environment (a hypothetical merger and unexpected shock to Pittsfield's economy which changes the costs of running an ATM).

There were eight banks before the merger, so we examine the actions of the seven remaining banks in the market. We assume the merged bank has a profit function which consists of the sum of the profits from the two banks which merged and starts with their ATMs, giving us an initial allocation of ATMs to the seven banks of $(9, 0, 3, 1, 0, 0, 1)$. Note that, as is often the case in empirical work, there is significant heterogeneity across the firms inherited from past actions and events (the banks differ in the number and locations of their branches, in the amenities they provide customers...). We are assuming that these characteristics of the banks *do not* change.

The realized costs of agent i if it uses n_i ATMs in period t are given by:

$$C(n_i, t) = [b_{0,i} + b_{1,i,t}]n_i + b_2 n_i^2$$

where $(b_{0,i}, b_2)$ are known constants and $b_{1,i,t}$ is the random draw on the cost shock. These are iid draws from a normal distribution with mean μ and variance σ^2 that is common across firms. For simplicity, we assume switching costs and fixed costs of each machine to be 0; we only focus on the per-period operational costs.

Firms do not know their future cost shocks before they chose the number of ATMs they operate in the next period, and we focus on Nash Equilibria in expected costs. In the first period after the merger, each firm receives its own realization of the cost shock $b_{1,i,t}$. As firms realize that their costs have changed, each firm will use an average over cost draws after the switch in regimes to form their expectation of costs for the next period (μ). There are no dynamics other than that induced by learning about the likely value of the cost shocks and the likely play of competing firms.

Number and Nature of Equilibria

The first part of the analysis proceeds by simply enumerating the "limiting equilibria": i.e., the Nash equilibria when all firms know the expected value of the cost shock. Since banks are asymmetric, there are $170,544$ different allocations of up to 15 ATMs among seven banks. Table 1 lists all equilibrium allocations when firms know the expected value of the cost shock for different values of μ.

Results.

- initial post merger allocation is (9,0,3,1,0,0,1) does not constitute a best response for any of our cost specifications.

- the number of equilibria is always *strikingly small* in comparison to the number of total possible allocations.

- within a specification for costs, the different equilibria are quite similar to each other (no two equilibria for the same cost specification in which one firm differs in its number of ATMs by more than one ATM,...)

- "comparative statics"; if an allocation which had been an equilibrium is no longer an equilibrium when we lower the cost, this former equilibrium was always the equilibrium with the least number of ATMs at the higher cost. If an allocation becomes an equilibrium allocation when it had not been one at the higher cost, the new equilibrium allocation always has a larger total number of ATMs then the equilibria that are dropped out (and those that are dropped are always the equilibria with the lowest number of ATMs).

Possible Equilibria for Four Mean Cost Specifications

	Mean Cost (μ)	20,000	15,000	10,000	0
ATM Allocation	# of ATMs	Is Allocation An Equilibrium?			
(4,0,4,0,0,1,1)	10	Yes	No	No	No
(5,0,3,0,0,1,1)	10	Yes	No	No	No
(4,0,4,0,0,1,2)	11	No	Yes	No	No
(4,0,4,0,1,1,1)	11	No	Yes	Yes	No
(5,0,3,0,1,1,2)	12	Yes	Yes	Yes	Yes

Equilibrium Selection through Belief Formulation.

Investigate the implications of different processes for forming beliefs about competitors' play.

- Best response; each firm believes its competitors' will play the same strategy in the current period as they did in the prior period

- Fictitious play; each firm believes the next play of its competitors will be a random draw from the set of tuples of plays observed since the regime change.

Note: here we consider forming beliefs about competitor's actions. An alternative would have been to consider "learning" about the outcomes of one's actions (that is I have a perception of the returns to different actions and update those beliefs). We return to this second formulation when we come to reinforcement learning later on in the course.

Each run is stopped when we have converged to a single allocation, where convergence is defined as having remained in the same allocation state for 50 iterations.

This location was viewed as a “rest point” of the process. Note that *all* rest points are Nash equilibria of the game where each agent knows its mean costs. Table 2 provides the fraction of rest points at various equilibria for the different cost specifications. We tried different mean cost-shocks and different coefficient variations for those shocks.

Fraction of Rest Points at Alternative Equilibria

Mean (μ)	20,000			15,000			10,000			
CV $(\sigma/\mu)^a$	1	.5	.25	1	.5	.25	1	.5	.25	1
Best Reply										
4040011	.89	.87	.82							
5030011	.10	.10	.13							
4040012				.27	.14	.01				
4040111				.40	.21	.02	.04[b]	.00	.00	
5030112	.01	.03	.06	.33	.65	.97	.94	1.0	1.0	1.
Fictitious Play.										
4040011	.47	.41	.41							
5030011	.34	.44	.30							
4040012				.00	.00	.00				
4040111				.10	.01	.00	.00	.00	.00	
5030112	.15	.15	.29	.90	.99	1.0	1.0	1.0	1.0	1.

The initial condition is (9,0,3,1,0,0,1) for all runs and is never an equilibrium based on true expected costs.

[a] CV is the coefficient of variation of the cost shock. For the base specification where $\mu = 0$, the variance of the cost shocks were set to be the same as when $\mu = 20,000$.

[b] In this specification under Best Reply, approximately 2% of trials resulted in "cycling."

Note that

- The variance in the cost shocks can cause a *distribution* of rest points from a given initial condition.

- Apparently there is a dependence of the distribution of the equilibria on belief formulation process. This is troubling because of the lack of evidence on the empirical relevance on how one forms beliefs.

- On the brighter side, it appears that the distribution of the number of ATMs from the lower cost specifications always stochastically dominated those from the higher cost specifications.

References

- Andrews, D. and G. Soares "Inference for Models Defined by Moment Inequalities Using Generalized Moment Selection Procedures," *Econometrica*, 2010, 78(1): 119-157.

- Berry, S. and A. Pakes "The Pure Characteristics Demand Model," *International Economic Review,* 2007, 48(4): 1193-1225.

- Ciliberto, F., C. Murry, and E. Tamer "Market Structure and Competition in Airline Markets, Harvard University working paper.

- Crawford, G. and A. Yurukoglu "The Welfare Effects of Bundling in Multichannel Television Markets" *American Economic Review*, 2012, 102(2): 643-685.

- Dickstein, M. and E. Morales, "What do Exporters Know?", Working Paper, NYU, Stern, 2016.

- Doraszelski, U., G. Lewis, and A. Pakes "Just Starting Out: Learning and Equilibrium in a New Market," 2016, Harvard University working paper.

- Eizenberg, A. "Upstream Innovation and Product Variety in the United States Home PC Market, *Review of Economic Studies*, (2014) 81(3): 10031045.

- Fudenberg, D. and D. Levine "Learning in Games: Where Do We Stand? *European Economic Review*, 1998, 42(3-5): 631-639.

- Ho, K. "Barriers to Entry of a Vertically Integrated Health Insurer: An Analysis of Welfare and Entry Costs," *Journal of Economics and Management Strategy*, 2009, 18(2): 487-545.

- Ishii, J. "Compatibility, Competition, and Investment in Network Industries: ATM Networks in the Banking Industry," Harvard University PhD thesis, 2007.

- Katz, M. "Essays in Industrial Organization and Technological Change," Harvard University PhD dissertation, 2007, 1-157.

- Keane, M. and K. Wolpin "Empirical applications of discrete choice dynamic programming models," *Review of Economic Dynamics*, 2009, 12(1): 1-22.

- Lee, R. and A. Pakes "Multiple Equilibria and Selection by Learning in an Applied Setting," *Economic Letters*, 2009, 104(1): 13-16.

- Manski, C. *Partial Identification of Probability Distributions,* New York: Springer-Verlag, 2003.

- Manski, C. "Measuring Expectations," *Econometrica*, 2004, 72(5): 1329-1376.

- Nosko, C. "Competition and Quality Choice in the CPU Market," *Competition and Quality Choice in the CPU Market*, 2014, University of Chicago working paper.

- Pakes, A. "Alternative Models for Moment Inequalities," *Econometrica*, 2010, 78(6): 1783-1822.

- Pakes A. "Behavioral and Descriptive Forms of Choice Models," *International Economic Review*, 2014, 55(3): 603-624.

- Pakes, A., J. Porter, K. Ho, and J. Ishi, "Moment Inequalities and Their Application," *Econometrica*, 2015, 83(1): 315-334.

- Powell, J. "Symmetrically Trimmed Least Squares Estimation of Tobit Models, Econometrica, 1986, 54(6): 1435-1460.

- Savage, L. *The Foundations of Statistics*, 1954, New York, Wiley.

- Tamer, E. "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria" *Review of Economic Studies*, 2003, 70(1): 117-167.

- Wollmann, T. "Trucks Without Bailouts: Equilibrium Product Characteristics for Commercial Vehicles," forthcoming *AER*.