Theory and Empirical Work on Imperfectly Competitive Markets.

by Ariel Pakes

(Harvard University).

The Fisher-Schultz Lecture

World Congress
of the
Econometric Society
London, August 2005.

Structure of Talk.

- Overview: recent developments and some outstanding problems.
- Look closer at one problem: Nash equilibria in complex games (either because of choice sets, return functions, or the possiblity of multiple equilibria).

I. Overview.

Context: Build up a set of tools that enable the analysis of market outcomes in oligopolistic situations. For simplicity consider

- "textbook" situation: analyze price and quantity choices in a static framework, and then evolution of state variables in a dynamic one (implicitly assumes static control does not have an independent effect on future costs, demand, or equilibrium choices).

Static Analysis. Conditions on

- the goods marketed (or their characteristics) and their cost functions,
- consumer's preferences over those goods (or over characteristics tuples),
- "institutional" features like; the type of equilibrium, structure of ownership, regulatory rules.

and **computes** equilibria prices, quantities, profits and consumer surplus, as a function of the problem's state variables.

Use. Analyze implications of policy or environmental changes in the "short-run" (conditional on state variables). E.g.'s; Merger analysis, impacts of tariffs, gas prices, ...

Dynamics analyzes how the state variables that are subject to the firms' controls evolve.

- Conceptual Framework: Largely Markov Perfect equilibrium in investment (broadly defined) stategies (Star and Ho,1969; Maskin and Tirole,1988). There is a question of whether weaker notions of equilibria are all that data can throw light on; in particular Self Confirming Equilibria (Fudenberg and Levine, 1997), and what difference it makes.

- Empirical implementation: To match data need to allow for entry, exit, and firm and industry specific sources of uncertainty (Ericson and Pakes 1995; Doraszelski and Satterthwaite,2003).

Use. Longer run analysis of likely impacts of policy or environmental change; e.g. impacts of merger on entry or investments, impact of gas prices on fuel efficiency of capital, ...

Problems induced by possibility of multiple equilibria.

- Estimation. Estimators exist (though

no mle) but there is a nonstandard efficiency issue (see below).

- Policy Analysis. Under some assumptions we can identify the equilibria played (or selection mechanism used) in the past, but no guidance as to what happens after a policy or environmental change. Current options; (i) enumerate equilibria and get bounds on effects of interest, (ii) assume a learning theory and simulate likely equilibria.

Static Analysis.

The analysis requires:

1. a demand system

2. a cost system
3. an equilibrium assumption

Introduce recent work on each of these.

Demand Systems. Recent changes have been

- away from "representative agent" models to models with heterogenous agents.
- and from models set in "product space" to those set in characteristic space (Lancaster, 1971, McFadden, 1974).

Heterogeneous agent models allow us to

- combine data from markets with different distributions of consumer attributes, and

- consider distributional implications of policy and environmental changes (important both per se and to analyze incentives facing regulators and politicians).

Classic analytic work: Houthakker (1955). Simulation estimators (Pakes,1986) have enabled us to aggregate up from the *observed* distribution of consumer characteristics and any functional form and distribution of unobserved characteristics that we might think relevant.

Product vs. Characteristic Space. Product space had two problems

- Too many parameter problem.

- Inability to analyze demand for new goods before they are introduced.

Characteristic space circumvents both these problems but raises two new ones:

- At least for consumer goods need to allow for unobserved characteristics (analogue of unobservable in standard demand analysis). Results in a simultanaeity problem.
- To get realistic cross and own characteristic elasticities (particularly price elasticities) needed to generalize distributional assumptions used in early models (e.g. IIA property of logits).

Berry Levinsohn and Pakes (BLP, 1995) provides practical solutions.

- Early distributional assumptions made to get closed forms for the choice probabilities. Use of simulation and modern computers circumvents this need.
- BLP provides a contraction mapping which, given the parameter vector, generates the unobservable as a linear function of the data. Then can use any technique for controlling for endogeneity used in linear models (e.g. IV).

Production and Cost Functions. Unfortunately cost data are often proprietary, and when not proprietary (as is often the case in regulated industries), of questionnable quality (incentive for not reporting correct data). So very little direct work on cost functions, and when cost information is needed, it is often backed out from the implications of a behavioral model on other decisions (e.g. a Nash inprices in merger analysis).

In contrast there has been alot of recent work on production functions, and it is largely motivated by; (i) access to firm (or plant) level panels (often through governmental agencies), and (ii) a desire to analyze the efficiency impacts of the major changes in the environment (e.g., deregulation, changes in tariffs, privatization ...). Results in focus on particular substantive and technical issues.

Substantive Issues. Micro data provides an ability to distinguish the impacts of the change on

- the efficiency of the output allocation among establishments, from
- the productivity of individual establishments (and their correlates).

E.g. the immediate impact of deregulation of telecoms on the equipment industry was an increase in industry efficiency due to a reallocation of output to more efficient plants (largely resulting from a reallocation of capital; Olley and Pakes, 1996). There was no immediate impact on the productivity of individual establishments.

Technical Issues. Data contain: (i) large serially correlated differences in "productivity" among plants, and (ii) lots of entry and exit (Dunne Roberts and Samuleson, 1982, Davis and Haltwinger, 1986). Leads to worries about

- Simultaneity biases (endogeneity of inputs). Firms' whose productivity is positively effected by the change grow and increase input demand.
- Selection biases (endogeneity of attrition). Firm's whose productivities are negatively effected by the change being studied flounder and often exit.

Problems are accentuated by the fact that we are typically analyzing responses to large structural changes.

Corrections based on statistical models were not rich enough to account for the observed behavior. E.g.: Fixed effects ruled out because changes analyzed typically effects some firms positively and others negatively; Propensity score ruled out because it is a single index model and in fact there is more than one “index” or state variable which determines exit (at least productivity and capital).

Move to corrections based on *economic models* of input and exit choices (Olley and Pakes 1996, Levinsohn and Petrin 2003, Ackerberg Fraser, and Caves 2004, Wooldridge 2004, Ackerberg and Pakes,2005). Recently developed econometric tools (semiparametrics) enable computationally simple estimators that require only minimial assumptions on how input and exit choices are made.

Equilibrium Assumptions. Empirical work relies heavily on a static Nash in prices (or in quantities) assumption. There is a question of what information the agent conditions on when it makes its choice, but despite this the assumption typically does

- surprisingly well at accounting for the cross-sectional variance in prices of consumer goods.

However it does less well

- with shifts in prices over time, and
- in markets with a small number of agents on each side (e.g. vertical markets where "prices" are often implicit in complicated, sometimes proprietary, contracts).

Suggests a need for dynamic pricing models, and notions of equilibrium for "buyer-seller" networks (come back to this below).

Dynamics.

Additional requirements:

- Estimates of entry and exit costs, and the impacts of different types of investments.
- Algorithm for computing equilibrium.

Estimation (logic). Since entry, exit, and investment decisions set likely future states, they are determined by the continuation values from those states.

I.e. if we knew the continuation values and the costs of the decisions up to the parameter vector, we could get the model's predictions for these decisions conditional on that parameter vector.

Estimate by finding that value of the parameters that makes predicted decisions "as close as possible" to observed decisions.

Problem. Continuation value implied by a given value of the parameter vector are hard to compute. Solution: use the nonparametric estimates of continuation values *implicit in the data.* Two procedures

- average continuation values actually earned from a given state (Pakes, Ostrovsky and Berry, 2005), or
- semiparametric estimates of policies from

those states and then use estimated policies to simulate continuation values (Bajari, Benkard, and Levin, 2005).

E.g. Lower bound to entry cost is average of realized continuation values in states where entry was observed, upper bound takes averages in states where there was no entry (application to merger analysis).

Computation. As the number of state variables increase we must face a "curse of dimensionality" in both;

- the size of the state (or the number of points at which we have to calculate continuation values),
- and in the summation or expectation which defines those continuation val-

ues

The burden in both problems grows either geometrically or exponentially in the number of state variables depending on the nature of the states. The increase in speed and memory of computers helps, but it is still the case that computational burden will limit the models we use for some time to come.

Algorithmic innovations designed specifically for these types of problems.

- Applications of AI to dynamic games (Pakes and McGuire,2001; Pakes and Fershtman, 2005). Uses simulation to both; (i) approximate sum over future states needed for decisions (burden becomes linear in the number of state variables), (ii) search for a recurrent class of points (usually linear in number of states, but need not be) and only gets accurate estimates of policies and values on the recurrent class. Works well with dynamic games with assymetric information, but not yet amended to deal with model where observed play depends on behavior off the equilibrium path.

- Continuous Time Models (Doraszelski and Judd, 2005). Reduces the burden of the summation over future states to be linear, in the number of state variables. Cost; after a change in firm A's state occurs, the distribution of firm B's state no longer depends on its prior investments but rather just on its investments after the change.

Also possible to use parametric approximation techniques used in other fields (see Judd,2000), but they have not worked well for problems of the type we have been computing in IO.

II. Nash Equilibria and Estimation.

Basic idea and details in Pakes, Porter, Ho and Ishii, 2005. Briefly:

- 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 different choices from a *known* set of alternatives.

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

- Estimator: accept any value of θ that makes that difference non-negative.

- Question: Under what conditions will such (possibly set valued) estimators enable us to make valid inferences on the parameters of interest?

- Empirical Applications; (i) cost estimates from complex markets (ATM networks, electric utilities), (ii) characterizing contracts in buyer-seller networks (e.g. HMO/hospital).

Assumption 1 : Nash Condition.

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

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

where $D(d_i) \subset \mathcal{D}$, *for* $i = 1, \ldots n$. ♠

Notes.

- No restriction on choice set.
- No uniqueness requirement, and can allow for imperfect choices.

E.g.'s choice sets: a subset of a discrete space (all exclusive deals or all bilateral contracts); ordered choice (a number of stores or ATM's); continuous choice (allows for corners and non-convexities, and generates more information than first order conditions alone)...

Definition. If $(d, d') \in \mathcal{D}^2$ are two different feasible choices

$$\Delta\pi(d, d', d_{-i}, \mathbf{y}_i, \theta_0) \equiv$$

$$\pi(d, d_{-i}, \mathbf{y}_i, \theta_0) - \pi(d', d_{-i}, \mathbf{y}_i, \theta_0).$$

Assumption 2 : Functional Form and

Stochastic Assumptions.

$$\mathbf{y}_i \equiv (y_i, \nu_{1,i}, \nu_{2,i})$$

where $y_i \in Y$ *is observed by the econometrician, and* $(\nu_{1,i}, \nu_{2,i})$ *are unobservables that are now defined.*

Define $\nu_{1,i}$*:*

$$\Delta\pi(d, d', d_{-i}, \mathbf{y}_i, \theta_0) =$$

$$\mathcal{E}[\Delta\pi(d, d', d_{-i}, \mathbf{y}_i, \theta_0)|\mathcal{J}_i] + \nu_{1,i},$$

so that

$$\mathcal{E}[\nu_{1,i}|\mathcal{J}_i] = 0.$$

Define $\nu_{2,i}$*:*

$$\mathcal{E}[\Delta\pi(d, d', d_{-i}, \mathbf{y}_i, \theta_0)|\mathcal{J}_i] =$$

$$\mathcal{E}[\Delta r(d, d', d_{-i}, y_i, \theta_0)|\mathcal{J}_i] \;+\; \nu_{2,i},$$

where $\Delta r(\cdot)$ *is a known function, and*

$$\mathcal{E}[\nu_{2,i}|x_i] = 0$$

for an observable "instrument" $x_i \subset \mathcal{J}_i$. ♠

- ν_1. Typically the only disturbance in theory models. Generated by uncertainty in y_i or in d_{-i} (by asymmetric or incomplete information) and/or from measurement error. Note d **not** a function of ν_1.

- ν_2. Typically the only disturbance in econometric models. *Is* known to the agent when decisions are made, but *is not* known by the econometrican. So d **can be** a function of ν_2.

If we do not condition on the decision both

ν_1 and ν_2 are mean independent of x. However the conditional mean of ν_2, conditional on a decision, will depend on x (if we observe the decision and $\Delta r(\cdot, x) < 0$ then $\nu_2 > 0$), while that of ν_1 will not.

So far we have not specified

- the joint distribution of the $\{\nu_2\}$ or of the $\{\nu_1\}$, or
- what agent i knows about $\nu_{2,-i}$ or y_{-i}, (could be full info, signals, no info)

and in addition to multiple interacting agents with assymetric and/or incomplete information, we allowed for

- discrete choice sets, and *endogenous regressors.*

However in many (though not all) examples we need an additional condition (condition 1 in paper). Either

$$\boldsymbol{\nu_2} \equiv \mathbf{0},$$

or

$$\boldsymbol{\nu_1} \equiv \mathbf{0},$$

suffices. In many cases we can allow for both but do not have necessary conditions for that.

For now assume $\nu_2 \equiv 0$, and then show how to allow for ν_2 in our examples (pretty generic problems).

Estimation and Inference.

Take $x_i \in \mathcal{J}_i$. Provided $d' \in \mathcal{D}(d_i)$ the theory implies

$$\mathcal{E}\left[\Delta\pi(d_i, d', d_{-i}, \mathbf{y}_i, \theta_0)|x_i\right] \geq 0.$$

So if $h : X \rightarrow \mathcal{R}^+$ is a vector of positive functions of x_i,

$$\mathcal{E}_x\left[\mathcal{E}\left[\Delta\pi(d_i, d', d_{-i}, \mathbf{y}_i, \theta_0)|x_i\right] \otimes h(x_i)\right] \geq 0.$$

$\Delta\pi(\cdot, \theta)$ is not observed, but $\Delta r(\cdot, \theta)$ is. Moreover

$$\Delta\pi(\cdot) = \Delta r(\cdot) + \nu, \text{ and } \mathcal{E}[\nu|x] = 0.$$

So

$$\mathcal{E}_x \left[\mathcal{E} \left[\Delta\pi(d_i, d', d_{-i}, \mathbf{y}_i, \theta_0) | x_i \right] \otimes h(x_i) \right] =$$

$$\mathcal{E}_x \left[\Delta r(d_i, d', d_{-i}, y_i, \theta_0) \otimes h(x_i) \right] \geq 0.$$

Estimator. Take the sample analogue of these inequalities, form a norm that penalize values of θ that don't satisfy them, and then minimize that norm, or minimize

$$G_N(\theta) \equiv$$

$$\| \frac{1}{N} \sum_i \left[\Delta r(d_i, d', d_{-i}, y_i, \theta_0) \otimes h(x_i) \right]_- \|,$$

Note that the minimum, say $\hat{\Theta}$, and its population analogue, Θ_0, can be **set** valued.

Inference.

- Provide a test of the null that there exists a value of θ for which all the inequalities are satisfied. Nonstandard distribution but easy to simulate.
- Assume $\hat{\Theta}$ is convex. Provide upper and lower bounds for smooth functions of the parameters, and confidence intervals for those bounds.

Example 1: ATM Networks.

Ishii (2004) requires cost of setting up and running ATMs for analysis of likely impacts of changing rules governing payment for use of network (surcharges). Two period model.

- Second period. Standard static equilibrium (BLP demand system and Nash in interest rates equilibrium). Provides the returns each agent earns given its own, and its competitors, number of ATM's.
- First period. Choice of the number of ATM's. Multiple agent ordered choice problem (each agent choses $d_i \in \mathcal{Z}_+$).

Behavioral Assumption. Each agent expects its choice to lead to higher returns than alternatives. Returns are $r(y_i, d, d_{-i})$ (obtainable from first stage estimates) minus costs of ATMs, or

$$\pi(y_i, d, d_{-i}, \theta) = r(y_i, d, d_{-i}) - \theta d + \nu_{1,d,i}.$$

Simplest Implication.

$$\mathcal{E}[r(y_i, d_i, d_{-i}) - r(y_i, d_i - 1, d_{-i}) | \mathcal{J}_i] - \theta \geq 0$$

and

$$\mathcal{E}[r(y_i, d_i, d_{-i}) - r(y_i, d_i + 1, d_{-i}) | \mathcal{J}_i] + \theta \geq 0.$$

Simplest Estimator. Let $\Delta \overline{r}_L$ be the sample average of the returns made from the last ATM installed, and $\Delta \overline{r}_R$ 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 $\Delta\bar{r}_R \leq \Delta\bar{r}_L$

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

Notes. (1) Can extend the model to allow for a ν_2 as well as a ν_1, i.e.

$$ATM\ cost = (\theta + \nu_2)d, \ \ \nu_2 \in \mathcal{J},$$

since we take averages of all the ν_2 regardless of the decision the agent made, the d choice does not induce a selection in the ν_2, and the the (set) estimator remains consistent (provided $h(x)$ is formed from instruments). However if there is a ν_2, d is not an IV ($\Rightarrow$ test for ν_2).

(2) 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.

(3) Standard ordered choice model assumes $\nu_1 \equiv 0$, a parametric distribution for ν_2, and uses MLE (ordered probit or logit). Since there are realizations of $\Delta r_R > \Delta r_L$ in the data, this has likelhood of $-\infty$ for all θ.

Results (see table).

- $h(x)$ = constant $\Rightarrow$ interval, $h(x) =$

all $\Rightarrow$ a point.

- Test: $d_i \notin$ IV accepts, $d_i \in IV$ rejects.
- CI pretty tight, and pretty stable across specifications ($\approx \$4,500$ per ATM per month).

Implications. Ishii (2004): Current structure of ATM networks allow large network banks to pay lower interest rates **and** have more customers (and profits). A centralized system would reallocate profits from large to small banks and decrease concentration markedly. Moreover there is overinvestment in ATM's.

Example 2: Buyer-Seller Networks.

Analyze the nature of contracts emanating from a market with a small number of **both** buyers and sellers.

Empirical example, **Ho (2004)**. Contracts between HMO's and hospitals.

Setup: Two period game.

Second period. Given hospital networks, HMOs engage in a premium setting game. Static Nash equilibrium yields estimates of primitives. Can calculate profits for *each* HMO conditional on *any* configuration of HMO networks.

First period. Sellers (hospitals) make take it or leave if offers of two part tariffs (fixed fee plus per patient transfer) to buyers (HMO's).

Empirical Goal. Reduced form relationship between contract parameters and observables.

Moment Inequalities.

Necessary conditions for equilibrium include:

- If HMO accepts contract: HMO *expects* to earn more profits from a network which includes the hospital than from the same network without the hospital.
- If HMO rejects contract: HMO *expects* to earn more from the network that does not include the hospital.

- Hospitals expect to increase their *total* returns from accepted contracts.

Results: Table 2.

- Lower cost hospitals keep about half of their cost savings.
- Markups are markedly higher at capacity constrained hospitals,
- Hospitals in systems earn about $4-600 more a month per patient which is 25-33% higher markups.

Implications. Their may be an antitrust reason for looking into hospital systems, and a disconnect between socially and privately optimal investment in hospitals.

Robustness Questions.

- Might we expect contracts with these characteristics to emanate from a Nash equilibrium?
- Robustness of econometric results to
  - the existence of ν_2 variance?
  - assumptions on nature of game?

Numerical Analysis. Compute equilibria in markets with charactersitic distributions similar to those in Ho's data, but scaled down to have two hospitals and two HMO's in each market. Leaves only a small number of inequalities per market.

Table 2. Project markups onto variables of interest; the residual becomes the fixed fee. Similar results to real data but

- $\approx$ 30% of variance in markup, $\Leftrightarrow$ 8% of variance in transfers, due to ν_2.

More variables: interesting economic results, but still significant variance in ν_2.

Implications of ν_2. If there was a ν_2 error then the realized returns of a buyer would be, $\Delta r^B(\cdot) - \nu_2$, and a contract would accepted *iff*

$$\mathcal{E}[\Delta r^B(\cdot)|\mathcal{J}] \geq \nu_2,$$

so the mean of ν_2, *conditional on the decision*, will depend on any variable that determines returns.

ν_2 **variance** $\Rightarrow$ lower bound is too high, and the upper bound too low, so our bounds are not really bounds at all.

Allowing for ν_2. Need model for ν_2, then form moments which either

- do not condition on d,
- or do not involve ν_2.

Model. Assume ν_2 is in the transfers (which is where we lack information). Then the increment to

- the seller's profit when contracting, and
- the buyer's savings when not

both include ν_2, so the moment

$$\chi_{b,s}\Delta\pi^S(\cdot) + (1-\chi_{b,s})\Delta\pi^B(\cdot)$$

$$= \chi_{b,s}\Delta r^S(\cdot) + (1-\chi_{b,s})\Delta r^B(\cdot) + \nu_2.$$

is linear in ν_2 *no matter the outcome.*

Also the sum of the increment in buyers and sellers profits does not contain the transfers (and hence not ν_2) and must have positive expectation if contracting.

Robustness results. Ho's data

- increase number of variables,
- change game (both sides expect to make profits from contracts in force, and no gains to trade from couples not contracting),
- allowing for ν_2 errors
  - just hospital effects (difference in difference inequalities)
  - unrestricted ν_2 as above.

Signs same, but imprecise estimates.

Robustness results. Simulated data.

- No good instrument for $N_{j,k}$.
- ν_2 disturbances: Move LB up and UB down (as predicted). However neither jumps over true value; so the ν_2 **variance causes the estimator to be quite accurate.**
- Robust inequalities. Lose information on upper bound, just as we do when we can not use $N_{j,k}$ as instrument.

End Talk.