Applications of Moment Inequalities

Paul T. Scott

Toulouse School of Economics

ptscott@gmail.com

Empirical IO

Spring 2014

"Alternative Models for Moment Inequalities"
Ariel Pakes (2010)

Two versions of Moment Inequalities

Pakes describes two approaches, both making use of moment inequalities:

1. **Generalized Discrete Choice**
   Extends discrete choice models to multi-agent settings (e.g., entry models). Typically, inequalities result because models don't predict unique outcomes.

2. **Profit Inequalities**
   Makes use of "revealed preference inequalities" which can be constructed with relatively mild assumptions about agents' information sets. Robust to measurement error.

Katz's supermarket choice example I

- An agent's decision $d_i = (b_i, s_i)$ consists of a basket of goods b_i and a supermarket s_i where the goods where purchased.
- An agent with characteristics z_i has utility is given by

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

  where
  - $dt (s_i, z_i)$ is the drive time to the store;
  - $U (b_i, z_i)$ is utility from the bundle;
  - $e (b_i, s_i)$ is the cost of the bundle at the store.
- Assuming the agent chooses d_i to maximize utility, and d' is some other feasible option,

$$\pi (d_i, z_i, \theta) \geq \pi (d', z_i, \theta) .$$

  We can choose particular alternatives for d' to form moment inequalities which may be used for estimation.

Katz's supermarket choice example II

- Consider the option of buying the same bundle at a different store s_i', and write out the utility difference:

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

  where $\mathcal{E}$ is the agent's expectation operator, and $\mathcal{J}_i$ is her information set.

- Define $\nu_{1,i,s,s'}$ as the difference between the expected and realized utility difference:

$$\nu_{1,i,s,s'} \equiv \Delta\pi\left(b, s_i, s_i', z_i\right) - \mathcal{E}\left[\Delta\pi\left(b, s_i, s_i', z_i\right)|\mathcal{J}_i\right]$$

Katz's supermarket choice example III

- Assuming that
$$N^{-1}\sum_i \nu_{1,i,s,s'} \to_P 0,$$
and that s_i' is sampled so that it is always has a longer drive time than s_i,
$$\frac{-\sum_i \Delta e\left(b_i, s_i, s_i'\right)}{\sum_i \Delta dt\left(s_i, s_i', z_i\right)} \to_P \underline{\theta} \leq \theta$$
... why?

Katz's supermarket choice example III

- Assuming that

$$N^{-1}\sum_{i}\nu_{1,i,s,s'} \to_P 0,$$

and that s_i' is sampled so that it is always has a longer drive time than s_i,

$$\frac{-\sum_i \Delta e\left(b_i, s_i, s_i'\right)}{\sum_i \Delta dt\left(s_i, s_i', z_i\right)} \to_P \underline{\theta} \leq \theta$$

... why?

- Similarly, if we sample s_i'' so that it is always a shorter drive time,

$$\frac{-\sum_i \Delta e\left(b_i, s_i, s_i'\right)}{\sum_i \Delta dt\left(s_i, s_i', z_i\right)} \to_P \bar{\theta} \geq \theta$$

Katz's supermarket choice example IV

- Suppose we add an error term in drive time which is known to the agent but unobserved to the econometrician:

$$\theta_i = \theta + \nu_{2,i}$$

- Then, assuming drive time is known to the agent at the time of the decision,

$$N^{-1} \sum_i \Delta dt\left(s_i, s_i', z_i\right)^{-1} \nu_{1,i,s,s'} \rightarrow_P 0$$

- ... and we get a different formula for the bounds on θ:

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

$$N^{-1} \sum_i \left(\frac{\Delta e\left(b_i, s_i, s_i''\right)}{\Delta dt\left(s_i, s_i'', z_i\right)} \right) \rightarrow_P \bar{\theta} \geq \theta$$

Comments on Katz's supermarket choice example

- The ν_2 errors may be correlated with the choices. This may be important if we think there is heterogeneity in drive time aversion which will affect store choice.
- The model is compatible with uncertainty about prices; we only require that agents aren't systematically wrong.
- To use a standard discrete choice model in this setting requires lots of simplifying assumptions. The problem comes from having a huge choice space when we consider different bundles and different stores. However, using revealed preference inequalities allows us to identify some parameters of interest without modeling the whole choice problem.

Moment inequalities: common assumption I

- There are two assumptions involved by both the Profit Inequalities and Generalized Discrete Choice Approach. First, there is an optimality condition.

C1

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

where

- $d_i = d\left(\mathcal{J}_i\right)$ is the observed decision
- $\mathbf{d}_{-i}$ are the actions of other agents (in a multi-agent setting)
- $\mathbf{z}_i$ are some variables affecting profits
- $\mathcal{J}_i$ is the decision maker's information set

- Note: boldface variables are random from the decision maker's prospective.

Moment inequalities: common assumption II

- Next, we have a condition that requires that we can predict responses to deviations

C2

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

- C2 requires a model of how endogenous variables (potentially $\mathbf{y}_i$ and $\mathbf{d}_{-i}$) would respond to a deviation in d_i, and $\mathbf{z}_i$ must be exogenous.
- Note that C2 is much more demanding in games than in single-agent applications. More so in sequential games than simultaneous-move games.

Constructing moment inequalities

- Define

$$\Delta\pi\left(d_i, d', d_{-i}, z_i, \theta_0\right) \equiv \\ \pi\left(d_i, d_{-i}, z_i, \theta_0\right) - \pi\left(d', d^{-i}\left(d', z\right), z_i, \theta_0\right).$$

- C1 and C2 imply the inequality

$$\mathcal{E}\left[\Delta\pi\left(d_i, d', d_{-i}, z_i, \theta_0\right)|\mathcal{J}_i\right] \geq 0 \qquad \forall d' \in D_i$$

- Before we have something we can estimate, we need
  - a measurement model specifying the relationship between the true values of π and our measures of them
  - a relationship between the decion maker's expectation operator and the sample moments we can construct from the data

Measurement models

- The econometrician's measure of profits is r. We assume the following relationship between r and π:

$$r\left(d, d_{-i}, z_i^o, \theta\right) = \pi\left(d, d_{-i}, z_i^o, \theta\right) + \nu\left(d, d_{-i}, z_i^o, z_i, \theta\right).$$

- The agent's decision is based on $\mathcal{E}\left[\pi\left(d, \mathbf{d}_{-i}, \mathbf{z}_i, \theta\right)\right]$, so let's write this another way:

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

where

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

and

$$\nu_{2,i,d} = \mathcal{E}\left[\nu\left(d, \mathbf{d}_{-i}, z_i^o, \mathbf{z}_i, \theta\right) | \mathcal{J}_i\right].$$

Error terms

- ν_1 errors combine expectational error and measurement error – they are mean independent of $\mathcal{J}_i$
  - e.g., the realizations of prices in stores the shopper didn't visit
- ν_2 errors are observed by the agent but not the econometrician
  - e.g., an unobservable variable which shifts an agent's personal disutility of drive time.

Generalized discrete choice: expectations

DC3

$\forall d \in D_i$,

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

which says that there is no uncertainty in exogenous variables $\mathbf{z}_i$ or in competitor's decisions $\mathbf{d}_{-i}$.

Generalized discrete choice: measurement

DC4

$\forall d \in D_i$, $r\left(d, d_{-i}, z_i^o, \theta\right) = \pi\left(d, d_{-i}, z_i, \theta + \nu_{1,i}\right)$ for a known $\pi\left(\cdot, \theta\right)$ and $z_i = \left(\left\{\nu_{2,i,d}\right\}_d, z_i^o\right)$ with $\left(\nu_{2,i,d}, \nu_{2,-i,d}\right)_{d, z_i^o, z_{-i}^o} \sim F\left(\cdot; \theta\right)$ for a known $F\left(\cdot, \theta\right)$.

which says that differences in profits are measured exactly (no d subscript on ν_1. Furthermore, any unobserved components of the profit function have a known distribution.

- It is difficult to accommodate measurement error here.
- DC3 and DC4 set up the following model:$\forall d' \in D_i$,

$$\Delta\pi\left(d_i, d', d_{-i}, z_i^o, \nu_{2,i}; \theta_0 \geq 0\right); \qquad \left(\nu_{2,i}, \nu_{2,-i}\right)|_{z_i^o, z_{-i}^o} \sim F\left(\cdot; \theta_0\right)$$

Profit inequalities: expectations

PC3

There is a positive valued $h(\cdot)$ and an $x_i \in \mathcal{J}_i$ for which

$$N^{-1} \sum_i \mathcal{E}\left[\Delta\pi\left(d_i, d', \mathbf{d}_{-i}, \mathbf{z}_i, \theta_0\right) |x_i\right] \geq 0$$

$$\Rightarrow E\left(N^{-1} \sum_i \pi\left(d_i, d', \mathbf{d}_{-i}, \mathbf{z}_i, \theta_0\right) h\left(x_i\right)\right) \geq 0$$

which is weaker than DC3 in that it allows for some uncertainty – the actual profit inequalities don't have to *actually* be positive all the time.

- One theory justifying PC3 is correct/rational expectations, which would mean that any x_i in the information set can be used to form valid instruments $h(x_i)$. However Pakes emphasizes that weaker assumptions are compatible with PC3 – e.g., agents can have incorrect expectations, but their mistakes can't be systematically related to x_i.

Profit inequalities: measurement

Pakes considers different versions of PC4, I just want to focus on one:

PC4a (differencing)

Let $g = 1, \ldots, G$ index groups of observations. For each g, there are positive weights $w_{i,g} \in \mathcal{J}_{i,g}$ such that $\sum_{i \in g} w_{i,g} \times \Delta\nu_{2,i,g,d_{i,g},d'_{i,g}} = 0$. Thus,

$$G^{-1} \sum_{g} \sum_{i \in g} w_{i,g} \left(\Delta r \left(d_{i,g}, d'_{i,g}, \cdot; \theta_0 \right) - \mathcal{E} \left[\Delta\pi \left(d_{i,g}, d'_{i,g}, \cdot; \theta_0 \right) | \mathcal{J}_{i,g} \right] \right) \rightarrow_P 0$$

provided that $G^{-1} \sum_g \sum_{i \in g} w_{i,g} \Delta r \left(d_{i,g}, d'_{i,g}, \cdot; \theta_0 \right)$ obeys a law of large numbers.

Comparisons

- Note that PC3 nests DC3 – the profit inequality approach allows for some errors in agents' expectations and it naturally incorporates measurement error.
- The measurement assumptions are not nested. PC4a assumes a particular structure for the ν_2 errors which allows us to elimate them. DC4 assumes a particular distribution for the ν_2 errors which allows us to evaluate the probability of a particular realization if ν_2.

Brief example: Ho and Pakes (2013)

- Ho and Pakes consider a patient i's utility from hospital choice h:

$$W_{i,m,h} = \theta_{p,m} p\left(c_i, h, m\right) + g_{h,m}\left(s_i\right) + f_m\left(d\left(l_i, l_h\right)\right)$$

  where
  - m is the insurance provider;
  - $p\left(c_i, h, \pi\right)$ is the price charged by the hospital to the insurer for condition c_i;
  - $g_{h,m}$ is a hospital-insurer fixed effect for a condition of severity s_i;
  - $f_m\left(\cdot\right)$ is a function capturing the disutility of distance between the patient and hospital's locations
- Ho and Pakes are interested in how different types of contracts between medical care groups and insurers affects price sensitivity (some insurers provide more incentives for doctors to control costs); hence, the dependene of $\theta_{p,m}$. The subscript on f_m also allows the convenience-price tradeoff to vary with the type of insurer.

Brief example: Ho and Pakes (2013)

- Let $S_m(h, h', s)$ be the set of patients from plan m with severity s who chose hospital h but had h' in their choice set.
- Let $\Delta x(i, h, h') = x_{i,h} - x_{i,h'}$ for $x \in \{W, p\}$
- Let $\Delta f_m(l_i, l_h, l_{h'}) = f_m(d(l_i, l_h)) - f_m(d(l_i, l_{h'}))$
- Choosing $i \in S_m(h, h', s)$ and $i' \in S_m(h', h, s)$,

$$\Delta W(i, h, h') + \Delta W(i', h', h) =$$

$$\theta_{p,m}[\Delta p(c_i, h, h') + \Delta p(c_{i'}, h', h)] + \Delta f_m(l_i, l_h, l_{h'}) + \Delta f_m(l_{i'}, l_{h'}, l_h)$$

- Then, estimation is based on

$$E\left[\Delta W(i, h, h') + \Delta W(i', h', h)\right] \geq 0$$

Brief example: Morales et al. (2011)

- Morales, Sheu, and Zahler consider a dynamic decision problem in which a firm i decides which countries j to export to in period t. There are fixed and sunk costs associated with entering an export market.

- Net profits of exporting to j depends on the set of countries $b_{i,t-1}$ exported to in the previous period:

$$\pi_{ijt}\left(b_{t-1}\right) = v_{ijt} - fc_{ijt} - I\left\{j \notin b_{t-1}\right\} sc_{ijt}$$

  where v_{ijt} depends on the prices and costs for the firm's sector in country j in period t (think of v as data, or see the paper for details).

- There is also a "basic cost" (startup cost) to becoming an exporter. Profits to exporting are:

$$\pi_{it}\left(b_t, b_{t-1}\right) = \sum_{j \in b_t} \pi_{ijt} - I\left\{b_{t-1} = \emptyset\right\} bc_{it}$$

Brief example: Morales et al. (2011)

- Morales et al consider one-period deviations from the observed (optimal) path of export decisions. If the observed path for a firm is

$$\{b_1, \ldots, b_t, b_{t+1}, \ldots, b_T\},$$

  they consider

$$\{b_1, \ldots, b_t', b_{t+1}, \ldots, b_T\}.$$

- Note that this involves either delaying or bringing forward some sunk costs and net profits. Only profits in periods t and $t+1$ are affected.

Brief example: Morales et al. (2011)

- They define inequalities:

  $\Delta\pi_{it}\left(b_{it}, b'\right) =$

  $\pi_{it}\left(b_{it}, b_{it-1}\right) - \pi_{it}\left(b'_t, b_{it-1}\right) + \delta\left(\pi_{it+1}\left(b_{it+1}, b_{it}\right) - \pi_{it+1}\left(b_{it+1}, b'\right)\right)$

- Then estimation is based on

$$E\left(\Delta\pi_{it}\left(b_{it}, b'\right)\right) \geq 0$$

- The crucial assumption is that agents make the correct decisions on average, from the perpsective of their dynamic problem. Theory only delivers that $\mathcal{E}\left(\Delta\pi_{it}\left(b_{it}, b'\right)|\mathcal{J}_{it}\right) \geq 0$.

"The Welfare Effects of Bundling in Multichannel Television Markets"
Crawford and Yurukoglu (2013)

Overview

- The welfare effecgs of bundling are ambiguous in theory, calling for empirical work.
- Crawford and Yurukoglu quantify the effects of forced unbundling (à la carte pricing) of TV stations. This involves many moving parts:
  - Consumer demand for stations and bundles.
  - Bargaining between distributors (e.g., cable services) and content providers (conglomerates of channels)
  - Optimal pricing by distributors
  - Entry and exit of distributors or channels (ignored)

Main findings

- Estimated surplus change from -1.7% and 6.0%.
- If we were to take input costs as fixed (the prices content providers provide distributors) there would be large gains from à la carte as consumers would save money buying only a smaller set of channels.
- but the input costs rise dramatically in the new equilibrium, canceling out the potential gains.

Viewing

- Consumer's utility function:

$$v_{ij}\left(t_{ij}\right)=\sum_{c\in C_j}\gamma_{ic}\log\left(1+t_{ijc}\right),$$

where t_{ijc} is the number of hours household i watches channel c conditional on having access to bundle j. Note: this specification is compatible with corner solutions for some channels. They also include the outside option of not watchig tv as $c=0$.

- We can then define the utility of a given bundle as

$$v_{ij}^{*}\left(\gamma_i, C_j\right)=\sum_{c\in C_j}\gamma_{ic}\log\left(1+t_{ijc}^{*}\right)$$

where t_{ijc}^{*} is the optimal amount of time to watch channel c if subscribed to j

Bundle choice

- In addition to heterogeneous viewing tastes, there is heterogeneity in price sensititivity: $\alpha_i = \alpha + \pi_p y_i$, where y_i is income.
- BLP-like aggregation to market shares:

$$s_{jndm} = \int \frac{\exp\left(\delta_{jndm} + \mu_{ijndm}\right) dF^n\left(i\right)}{1 + \sum_{j'} \exp\left(\delta_{j'ndm} + \mu_{ij'ndm}\right)}$$

where

  - $\delta_{jdnm} = z'_{jndm}\psi + \alpha p_{jndm} + \xi$
  - $\mu_{ijndm} = v^*_{ijndm} + \pi_p y_i p_{jndm}$

Demand estimation: viewing

- To estimate the distribution of γ_i, they assume

$$\gamma_i = \chi_i \circ (\Pi o_i + \nu_i)$$

where o_i are household-level demographics, and

$$\chi_{ic} = \begin{cases} 0 & \text{with prob } \rho_{o_i c} \\ 1 & \text{with prob } 1 - \rho_{o_i c} \end{cases}$$

- They the estimate ρ, together with parameterized disrtibutions for the ν's (allowing for a some correlation).

Demand estimation: selection

- Note the potential for selection into bundles with channels that selected households have strong preferences for. To deal with this, they match conditional moments.
- That is, what they can construct with data are moments which involve the time spent on channels conditional on bundle selection, and the estimation is based on simulation of these same moments.
- This means they estimate the viewing parameters and bundle choice parameters jointly.
- To estimate bundle price sensitivity α_i, the assume that the bundle-market shocks ξ are independent of non-price characteristics and the prices in nearby markets (Hausman instruments).
  - What does this assumption require economically?

Distributors

- Distributors engage in Nash-Bertrand competition.
- Variable profits for a distributor:

$$\Pi_{fndm}\left(\mathbf{b}_{ndm}, \mathbf{p}_{ndm}\right) = \sum_{j \in \mathbf{b}_{fndm}} \left(p_{jndm} - \sum_{c \in C_{jndm}} \tau_{fc} \right) s_{jndm}\left(\mathbf{b}_{ndm}, \mathbf{p}_{ndm}\right)$$

- Implicit in this is assumption that contracts with channels involve costs-per subcribed τ_{fc}
- Given demand parameters and prices of each bundle, can recover $\hat{mc}_{fndm} = \sum_{c \in C_{jndm}} \tau_{fc}$

Bargaining between distributors and channels

- Let K represent a conglomerate of channels. They assume each distributor and conglomerate negotiate separately and arrive at the Nash bargaining solution.

- Formally, τ_{fK} mazimizes

$$\left[\Pi_f\left(\tau_{fK};\Psi_{-fK}\right)-\Pi_f\left(\infty;\Psi_{-fK}\right)\right]^{\zeta_{fK}}\left[\Pi_K\left(\tau_{fK};\Psi_{-fK}\right)-\Pi_K\left(\infty;\Psi_{-fK}\right)\right]^{1-\zeta_{fK}}$$

  where Ψ_{-fK} are the input prices set in all other bilateral contracts, $\Pi_f=\sum_{ndm}\Pi_{fndm}$, and

$$\Pi_K\left(\tau_{fK};\Psi_{-fK}\right)=\sum_{c\in K}\left(\sum_f \tau_{fc}Q_{fc}\left(\Psi\right)+r_c^{ad}t_c\left(\Psi\right)\right)$$

  where the rates of advertising payments r_c^{ad} are assumed to be exogenous.

- Note that the ζ's allow for asymmetric bargaining power.

Rationalizations of Nash bargaining

- Rubinstein (1982) gives a rationalization of the Nash bargaining solution in a two-agent setting.
  - Agents take turns making offers which the other agent can accept or reject
  - If agents are sufficiently patient, they will immediately agree on the Nash bargaining solution
- Collard-Wexler, Gowrisankaran, and Lee (2013) consider a multi-agent extension which can rationalize the use of Nash bargaining in a setting like Crawford and Yurukoglu's.

Cost estimation

- Viewing and bundle demand estimated jointly. This allows them to infer marginal costs for each bundle, which are sums of the channel-specific costs.

$$\hat{mc}_{fndm} = \sum_{c \in C_{jndm}} \tau_{fc}$$

- They then parameterized a model of $\hat{\tau}_{fc}(\eta, \varphi)$ based on observed average input costs, distributor size, and ownership shares (some distributor's companies own some cable channels).
- To estimate, they build moments matching the backed-out $\hat{mc}_{fndm}$ to $\sum_{c \in C_{jndm}} \tau_{fc}(\eta, \varphi)$

Cost estimation I

- Viewing and bundle demand estimated jointly. This allows them to infer marginal costs for each bundle, which are sums of the channel-specific costs.

$$\hat{mc}_{fndm} = \sum_{c \in C_{jndm}} \tau_{fc}$$

- They then parameterized a model of $\hat{\tau}_{fc}(\eta, \psi)$ based on observed average input costs, distributor size, and ownership shares (some distributor's companies own some cable channels).
- To estimate, they build moments matching the backed-out $\hat{mc}_{fndm}$ to $\sum_{c \in C_{jndm}} \tau_{fc}(\eta, \psi)$

Cost estimation II

- They also use moment inequalities to assist with cost estimation. Note that A distributor could add or subtract channels from the bundles it offers (conditional on having a contract with the channel providers).
- Following Pakes's (2010) notation,

$$\Pi_{fndm}\left(\mathbf{b}_{fndm}, \mathbf{b}_{-fndm}, \mathbf{p}_{fndm}, \mathbf{p}_{-fndm}\right) =$$

$$r_{fndm}\left(\mathbf{b}_{fndm}, \mathbf{b}_{-fndm}, \mathbf{p}_{fndm}, \mathbf{p}_{-fndm}\right) + \nu_{fndmb,1} + \nu_{fndmb,2}$$

where r is the constructed function.

Cost estimation III

- Assuming the ν_2 error are constant for all bundles for a given distributor and market, they will cancel. Then, assuming that the ν_1 errors are zero on average, they can use

$$E\left[\Delta r_{fndm}\left(b,b'\right)+\Delta r_{fndm'}\left(b',b\right)\right]\geq 0$$

for estimation. Specifically:

$$\min\left\{0, J^{-1}\sum_j \Delta r_{fndm}\left(b_{jndm},b';\eta,\varphi\right)+\Delta r_{fndm'}\left(b',b_{jndm};\eta,\varphi\right)\right\}=0$$