Moment Inequalities[1]

Julie Holland Mortimer

Economics 853

[1] These notes are adapted from Ariel Pakes' lecture notes from previous years' graduate classes at Harvard, and from Kate Ho's subsequent adaptations.

Outline

Overview

In this lecture we explore recent approaches to estimating models through the use of inequality restrictions, rather than equality restrictions.

The motivation for these approaches comes from:

- Economic problems that cannot be empirically investigated with more standard techniques (e.g., discrete games)
- Sampling methodology or measurement limitations in data (e.g. we don't know an agent's exact income, but we know whether it lies in an interval, so we can get inequalities on income; or we have topcoded income)

The assumptions required in the latter case to ensure that moment inequalities estimators have desirable properties are quite straightforward.

The assumptions required in the former case are more subtle.

Once we investigate these assumptions, we may start questioning the assumptions underlying the estimators of more traditional behavioral problems (particularly those that require discrete actions in a game).

Overview

The appeal of using moment inequalities to analyze a behavioral model is that the set up for the econometric problem is often the same as the set up for a theoretical model.

This similarity makes it easy to interpret the results in a way that is consistent with the theory.

However, dealing with unobservables can be tricky in this context.

Economic Model

- The econometrician observes a set of choices made by various agents.
- Assume agents expect the choices they made to lead to returns that are higher than the returns they would have earned had they made an alternative feasible choice.
- Assume a parametric "return function."
- For each value of θ, compute the difference between the observable part of the actual realized returns and the observable part of returns the agents would have earned had they made the alternative choice.

Economic Model

- Estimator: Accept any value of θ that, on average, makes the observed decisions better (more profitable) than the alternative.
- Question: When do such (possibly set valued) estimators enable us to make valid inferences on the parameters of interest?

There is more than one set of conditions that can be used to justify inequality estimators. We will consider the following two such sets of conditions:

- The set that uses only “structural errors” (i.e., assumptions that are the multiple-agent analog of the assumptions used in discrete choice econometrics).
- The set that incorporates expectational and measurement errors along with the structural errors.

- Ciliberto and Tamer *(Econometrica 2009)* develop the estimator based on the first set of conditions in the context of an entry model.
- Pakes, Porter, Ho and Ishii *(Econometrica 2015)* develop the estimator based on the second set of conditions.
- Pakes *(Econometrica 2010)* provides a comparison of the two sets of conditions.

Conditions Common to the Two Inequality Estimators

Recall the first two conditions that are relevant to discrete games. We covered these in the last lecture.

1. **Nash Condition (C1)**

$$\sup_{d \in D_i, d \neq d_i} \varepsilon\left[\pi(d, \mathbf{d}_{-i}, \mathbf{y}_i, \theta_0) | I_i\right] \leq \varepsilon\left[\pi(d_i, \mathbf{d}_{-i}, \mathbf{y}_i, \theta_0) | I_i\right]$$

   where $D_i \subset D$, for $i = 1, ..., n$.

2. **Counterfactual Condition (C2)**

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

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

Conditions Common to the Two Inequality Estimators

Implications of conditions C1 and C2:

If $d' \in D_i$ and

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

then

$$\varepsilon\left[\Delta\pi(d_i, d', \mathbf{d}_{-i}, \mathbf{z}_i) | I_i\right] \geq 0.$$

To use this inequality directly as a basis for an estimation algorithm, we need the relationships between:

- The expectations underlying agents' decisions ($\varepsilon(.)$) and the expectations of the observed sample moments ($E(.)$),
- $\pi(., \theta)$ and (z_i, d_i, d_{-i}) and their observable analogues.

This is where the two approaches differ. We covered the first set in the last lecture.

Extra Conditions under Full Information and No Specification Error

3. **Expectational Condition (FC3)**

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

$$\forall d \in D_i.$$

4. **Measurement Conditions (FC4)**

$$\pi(., \theta) \text{ is known.}$$

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

$$(\upsilon^f_{2,i}, \upsilon^f_{2,-i})|_{z^o_i, z^o_{-i}} \sim F(.;\theta),\ F(.,\theta) \text{ is known.}$$

Extra Conditions under Full Information and No Specification Error

Implications of conditions FC3 and FC4:

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

$$\forall d \in D_i, \text{ and}$$

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

(Note that there may not be a θ that satisfies these conditions for all vectors of decisions.) To ensure that the model assigns positive probability to the observed decisions for some θ we typically also assume:

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

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

Estimation from Inequality Conditions (Structural Errors Only)

This section follows Ciliberto and Tamer (2009), which studies entry in the airline industry.

Start with the formulation of profit in Berry (1992):

$$\Pi_{m,k,N} = X_m\beta - \delta \log N + Z_k\alpha + (\rho u_{m0} + \sqrt{1-\rho^2}u_{mk})$$

- $(X_m\beta - \delta \log N)$ is the observable component of profit that is common to all firms in a market
- $Z_k\alpha$ is the observable component of profit that varies across firms within a market.
- The unobserved components of profit also have a common component (ρu_{m0}) and a firm-specific component ($\sqrt{1-\rho^2}u_{mk}$).

Estimation from Inequality Conditions (Structural Errors Only)

Ciliberto and Tamer allow for a more flexible profit function, seen here:

$$\Pi_{m,i} = S'_m\alpha + Z'_{im}\beta_i + W'_{im}\gamma_i + \sum_{j\neq i}\delta^i_j y_{jm} + \sum_{j\neq i} Z'_{jm}\phi^i_j y_{jm} + \epsilon_{im}$$

Definitions:
S_m denotes market characteristics common to all firms
Z_{im} is a vector of firm characteristics that enter into the profits of all firms in the market (e.g., product attributes that consumers value)
W_{im} is a vector of firm characteristics that only affect firm i's profit in market m
y_{jm} is an indicator for the presence of other firms $j \neq i$ in market m
δ^i_j captures the effect of having firm j in market m on firm i's profit
ϕ^i_j captures firm-interaction effects that arise through the Z_m's.
ϵ_{im} has several components: destination, origin, airport, and a firm-market unobservable. This maps to the $\upsilon_{2,i}$ error in our notes, and is assumed to be observed by all players.

Estimation from Inequality Conditions (Structural Errors Only)

We can write down the conditions from the theoretical model (C1 and C2) given this profit function, as well as the conditions that allow us to translate the theoretical model to the data in a perfect information, "structural errors" world (FC3 and FC4).

Note that although the model does specify a parametric distribution for the $(\upsilon^f_{2,i}, \upsilon^f_{2,-i})$ conditional on the observables, it does not deliver a likelihood.

There is no function that takes values of the z^o, υ_2 and θ vectors into uniquely-specified actions, so we cannot construct probabilities for those actions.

This result comes directly from the inequalities of the entry game, and is due to the possibility of multiple equilibria.

Results/Conditions that the Inequalities with FC3 and FC4 Deliver

1. We can check whether the conditions of the model are satisfied at the observed (d_i, d_{-i}) for any $(\upsilon^f_{2,i}, \upsilon^f_{2,-i})$ and θ, and this, together with $F(., \theta)$, enables us to calculate the probability of those conditions being satisfied at any θ.

   These are necessary conditions for the choice to be made: therefore when $\theta = \theta_0$ the probability of satisfying them must be weakly greater than the probability of observing (d_i, d_{-i}).

   I.e. the model delivers an *outer measure* for the actions conditional on θ.

Results/Conditions that the Inequalities with FC3 and FC4 Deliver

2. We can check whether (d_i, d_{-i}) are the only values of the decision variables to satisfy the necessary conditions for any $(\upsilon^f_{2,i}, \upsilon^f_{2,-i})$ and θ, and this can be used to provide a lower bound on the probability of actually observing (d_i, d_{-i}) given θ.

   When $\theta = \theta_0$ the probability that we observe these decisions is weakly greater than the probability that they're the only values that are consistent with the model.

   I.e., the model also delivers an *inner measure* for the action conditional on θ.

Results/Conditions that the Inequalities with FC3 and FC4 Deliver

More formally, define the probability that the model generated by C1, C2, FC3 and FC4 with the additive separability constraint, a model we will call MF, is satisfied at a particular (d_i, d_{-i}) for a given θ to be:

$$\bar{P}\{(d_i, d_{-i})|\theta\} = \Pr\left\{(\upsilon^f_{2,i}, \upsilon^f_{2,-i}) : (d_i, d_{-i}) \text{ satisfy MF}|z^o_i, z^o_{-i}, \theta\right\},$$

and the analogous lower bound to be:

$$\underline{P}\{(d_i, d_{-i})|\theta\} = \Pr\left\{(\upsilon^f_{2,i}, \upsilon^f_{2,-i}) : \text{only } (d_i, d_{-i}) \text{ satisfy MF}|z^o_i, z^o_{-i}, \theta\right\}.$$

Results/Conditions that the Inequalities with FC3 and FC4 Deliver

Were we to know the equilibrium selection mechanism we could also calculate the actual likelihood of (d_i, d_{-i}) for a given θ or

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

We *do not know* the selection mechanism but we do know that for the true selection mechanism when $\theta = \theta_0$

$$\boxed{\bar{P}\{(d_i, d_{-i})|\theta\} \geq P\{(d_i, d_{-i})|\theta\} \geq \underline{P}\{(d_i, d_{-i})|\theta\}.}$$

Results/Conditions that the Inequalities with FC3 and FC4 Deliver

Let $\{\}$ be the indicator function which takes the value one if the condition inside the brackets is satisfied and zero elsewhere.

Let $h(.)$ be a function which takes only positive values.

Let $E(.)$ provide expectations conditional on the process actually generating the data (including the equilibrium selection process).

Results/Conditions that the Inequalities with FC3 and FC4 Deliver

Then MF implies that

$$E_{\upsilon_2}\left(N^{-1}\sum_i \left(\bar{P}\left\{(d_i, d_{-i})|\theta\right\} - \left\{d = d_i, d^{-i} = d_{-i}\right\}\right) h(z_i^o, z_{-i}^o)\right)$$

$$= \boxed{N^{-1}\sum_i \left(\bar{P}\left\{(d_i, d_{-i})|\theta\right\} - P\left\{(d_i, d_{-i})|\theta\right\}\right) h(z_i^o, z_{-i}^o) \geq 0 \ at \ \theta = \theta_0.}$$

which gives us a moment inequality. An analogous moment inequality can be constructed from the condition on $\underline{P}\{(d_i, d_{-i})|\theta\}$:

$$= \boxed{N^{-1}\sum_i \left(P\left\{(d_i, d_{-i})|\theta\right\} - \underline{P}\left\{(d_i, d_{-i})|\theta\right\}\right) h(z_i^o, z_{-i}^o) \geq 0 \ at \ \theta = \theta_0.}$$

The estimation routine:

- constructs unbiased estimates of $(\underline{P}(\cdot|\theta), \bar{P}(\cdot|\theta))$,
- substitutes them for the true values of the probability bounds into these moments, and
- accepts values of θ for which the moment inequalities are satisfied.

Since typically neither the upper nor the lower bound are analytic functions of θ, we employ simulation techniques to obtain an unbiased estimate of them.

This process can be computationally expensive, often too computationally burdensome to do.

We can get away with fewer function evaluations if we want to rely only on the upper bound probability $\bar{P}(\cdot|\theta)$), as then we can drop as many comparisons as we want, though by dropping inequalities you are likely to widen the set of θ the model accepts.

Data - Ciliberto and Tamer (2009)

The same data used in Berry (1992):

- DB1B data from 2001 (no price data).
- A market is a trip between two airports and a direction, regardless of stops.
- The data covers a sample of markets between the top 100 MSA's; 2742 markets.
- Ciliberto and Tamer focus on strategic interactions between American, Delta, United, and Southwest and pay particular attention to Dallas because of the Wright Amendment.
- They construct Berry's "airport presence" variable; they do not have cost data, but they have plane capacities and they use that to construct an opportunity fixed cost of serving a market.

Results - Ciliberto and Tamer (2009)

Tables 3 and 4 in Ciliberto and Tamer report confidence intervals for points (i.e. the probability that the confidence set covers the identified set is .95). Table 3 assumes iid errors.

- The first column constrains competitive effects of each firm's presence in the market on other firms 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.

Results - Ciliberto and Tamer (2009)

- Column 2 allows competitive effects to vary across firms. 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 4 allows one airline's presence to affect different airlines differently. Now Large airlines (LAR) and Southwest (=WN) have strongly negative effects on LCC, and there are smaller differences among the rest.

Results - Ciliberto and Tamer (2009)

TABLE III
EMPIRICAL RESULTS[a]

	Berry (1992)	Heterogeneous Interaction	Heterogeneous Control	Firm-to-Firm Interaction
Competitive fixed effect	[−14.151, −10.581]			
AA		[−10.914, −8.822]	[−9.510, −8.460]	
DL		[−10.037, −8.631]	[−9.138, −8.279]	
UA		[−10.101, −4.938]	[−9.951, −5.285]	
MA		[−11.489, −9.414]	[−9.539, −8.713]	
LCC		[−19.623, −14.578]	[−19.385, −13.833]	
WN		[−12.912, −10.969]	[−10.751, −9.29]	
LAR on LAR				
LAR: AA, DL, UA, MA				[−9.086, −8.389]
LAR on LCC				[−20.929, −14.321]
LAR on WN				[−10.294, −9.025]
LCC on LAR				[−22.842, −9.547]
WN on LAR				[−9.093, −7.887]
LCC on WN				[−13.738, −7.848]
WN on LCC				[−15.950, −11.608]
Airport presence	[3.052, 5.087]	[11.262, 14.296]	[10.925, 12.541]	[9.215, 10.436]
Cost	[−0.714, 0.024]	[−1.197, −0.333]	[−1.036, −0.373]	[−1.060, −0.508]
Wright	[−20.526, −8.612]	[−14.738, −12.556]	[−12.211, −10.503]	[−12.092, −10.602]
Dallas	[−6.890, −1.087]	[−1.186, 0.421]	[−1.014, 0.324]	[−0.975, 0.224]
Market size	[0.972, 2.247]	[0.532, 1.245]	[0.372, 0.960]	[0.044, 0.310]
WN			[0.358, 0.958]	
LCC			[0.215, 1.509]	

(*Continues*)

Results - Ciliberto and Tamer (2009)

TABLE III—*Continued*

	Berry (1992)	Heterogeneous Interaction	Heterogeneous Control	Firm-to-Firm Interaction
Market distance	[4.356, 7.046]	[0.106, 1.002]	[0.062, 0.627]	[−0.057, 0.486]
WN			[−2.441, −1.121]	
LCC			[−0.714, 1.858]	
Close airport	[4.022, 9.831]	[−0.769, 2.070]	[−0.289, 1.363]	[−1.399,−0.196]
WN			[1.751, 3.897]	
LCC			[0.392, 5.351]	
U.S. center distance	[1.452, 3.330]	[−0.932, −0.062]	[−0.275, 0.356]	[−0.606, 0.242]
WN			[−0.357, 0.860]	
LCC			[−1.022, 0.673]	
Per capita income	[0.568, 2.623]	[−0.080, 1.010]	[0.286, 0.829]	[0.272, 1.073]
Income growth rate	[0.370, 1.003]	[0.078, 0.360]	[0.086, 0.331]	[0.094, 0.342]
Constant	[−13.840, −7.796]	[−1.362, 2.431]	[−1.067, −0.191]	[0.381, 2.712]
MA			[−0.016, 0.852]	
LCC			[−2.967, −0.352]	
WN			[−0.448, 1.073]	
Function value	1756.2	1644.1	1627	1658.3
Multiple in identity	0.837	0.951	0.943	0.969
Multiple in number	0	0.523	0.532	0.536
Correctly predicted	0.328	0.326	0.325	0.308

[a] These set estimates contain the set of parameters that cannot be rejected at the 95% confidencet level. See Chernozhukov, Hong, and Tamer (2007) and the Supplemental Material for more details on constructing these confidence regions.

Results - Ciliberto and Tamer (2009)

TABLE IV
VARIABLE COMPETITIVE EFFECTS

	Independent Unobs	Variance–Covariance	Only Costs
Fixed effect			
AA	[−9.433, −8.485]	[−8.817, −8.212]	[−11.351, −9.686]
DL	[−10.216, −9.255]	[−9.056, −8.643]	[−12.472, −11.085]
UA	[−6.349, −3.723]	[−4.580, −3.813]	[−10.671, −8.386]
MA	[−9.998, −8.770]	[−7.476, −6.922]	[−11.906, −10.423]
LCC	[−28.911, −20.255]	[−14.952, −14.232]	[−11.466, −8.917]
WN	[−9.351, −7.876]	[−6.570, −5.970]	[−12.484, −10.614]
Variable effect			
AA	[−5.792, −4.545]	[−4.675, −3.854]	
DL	[−3.812, −2.757]	[−3.628, −3.030]	
UA	[−10.726, −5.645]	[−8.219, −7.932]	
MA	[−6.861, −4.898]	[−7.639, −6.557]	
LCC	[−9.214, 13.344]		
WN	[−10.319, −8.256]	[−11.345, −10.566]	
Airport presence	[14.578, 16.145]	[10.665, 11.260]	
Cost	[−1.249, −0.501]	[−0.387, −0.119]	
AA			[−0.791, 0.024]
DL			[−1.236, 0.069]
UA			[−1.396, −0.117]
MA			[−1.712, 0.072]
LCC			[−17.786, 1.045]
WN			[−0.802, 0.169]

Results - Ciliberto and Tamer (2009)

TABLE IV–*Continued*

Wright	[−17.800, −16.346]	[−16.781, −15.357]	[−14.284, −10.479]
Dallas	[0.368, 1.323]	[0.839, 1.132]	[−5.517, −2.095]
Market size	[0.230, 0.535]	[0.953, 1.159]	[1.946, 2.435]
WN	[0.260, 0.612]	[0.823, 1.068]	
LCC	[−0.432, 0.507]		
Market distance	[0.009, 0.645]	[0.316, 0.724]	[−0.039, 1.406]
WN	[−3.091, −1.819]	[−2.036, −1.395]	
LCC	[−1.363, 1.926]		
Close airport	[−0.373, 0.422]	[0.400, 1.433]	[3.224, 6.717]
WN	[1.164, 3.387]	[2.078, 2.450]	
LCC	[1.059, 3.108]	[1.875, 2.243]	
U.S. center distance	[−9.271, 0.506]	[0.015, 0.696]	[2.346, 3.339]
WN	[0.276, 1.008]	[0.668, 1.097]	
LCC	[−0.930, 0.367]		
Per capita income	[0.929, 1.287]	[0.824, 1.052]	[1.416, 2.307]
Income growth rate	[0.136, 0.331]	[0.151, 0.316]	[1.435, 2.092]
Constant	[−0.522, 0.163]	[−0.827, −0.523]	[−12.404, −10.116]
MA_m	[0.664, 1.448]	[0.279, 0.747]	
LCC	[−1.528, −0.180]	[−0.233, 0.454]	
WN	[1.405, 2.215]	[1.401, 1.659]	
Function value	1616	1575	1679
Multiple in identity	0.9538	0.9223	0.9606
Multiple in number	0.6527	0.3473	0.0728
Correctly predicted	0.3461	0.3375	0.3011

The Profit Inequality Approach

Pakes, Porter, Ho and Ishii (2010); Pakes (2010).

This approach allows for differences between the primitives that underly agents' decisions and the econometrician's constructs for those primitives.

We start with the measurement model that provides the relationship between these two objects.

The Profit Inequality Approach

Let our *observable* approximation to $\pi(.)$

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

Define the error in this approximation as:

$$\upsilon(d, d_{-i}, z_i^o, z_i, \theta_0) = r^I(d, d_{-i}, z_i^o, \theta_0) - \pi(d, d_{-i}, z_i).$$

The Profit Inequality Approach

Then by definition:

$$r^I(d, d_{-i}, z_i^o, \theta_0) = \varepsilon\left[\pi(d_i, \mathbf{d}_{-i}, \mathbf{z}_i)|I_i\right] + \upsilon^I_{2,i,d} + \upsilon^I_{1,i,d},$$

where

$$\upsilon^I_{2,i,d} = \varepsilon\left[\upsilon(d, d_{-i}, z_i^o, z_i, \theta_0)|I_i\right]$$

and

$$\upsilon^I_{1,i,d} = \left(\pi(d, .) - \varepsilon\left[\pi(d, .)|I_i\right]\right) + \left(\upsilon(d, ;) - \varepsilon\left[\upsilon(d, .)|I_i\right]\right).$$

We will discuss the sources and interpretation of υ^I_1 and υ^I_2 in the following slides.

The Profit Inequality Approach

Note: For all $d \in D_i$, $\varepsilon\left[\upsilon^{I}_{1,i,d}|I_i\right] = 0$, by construction, while $\varepsilon\left[\upsilon^{I}_{2,i,d}|I_i\right] \neq 0$. It is this distinction that forces us to keep track of two separate disturbances and consider their relative importance.

Sources of υ_1^I:

By construction υ_1^I is the sum of:

1. Expectational error

$$\pi(d,;) - \varepsilon\left[\pi(d,.)|I_i\right]$$

   and

2. Specification and measurement error

$$\upsilon(d,;) - \varepsilon\left[\upsilon(d,.)|I_i\right]$$

The Profit Inequality Approach

The expectational error is caused by:

i uncertainty in $\mathbf{z}_i$, and/or

ii asymmetric information which is uncertainty in $\mathbf{d}_{-i}$

So to compute its distribution (as we'd need to if we were using MLE, for example), we'd have to specify what each agent knew about its competitors, and then repeatedly solve for an equilibrium (a process which would typically require us to select among equilibria).

Since we'd have to compute returns from a counterfactual, the specification error would probably be non-trivial. The inequalities methodology does not require this step.

Sources of v_2^I and Selection:

v_2^I is that part of profits that the agent does condition on when making its decision but the econometrician has not included in the profit specification.

Since $v_{2,i}^I \in I_i$ and $d_i = d(I_i)$, d_i will generally be a function of $v_{2,i}^I$ (and perhaps also of $v_{2,-i}^I$).

This can generate a selection problem.

For example, temporarily ignore any difference between the agent's expectations (our $\varepsilon(.)$) and the expectations generated by the true data generating process (our $E(.)$).

Assume that x is an "instrument" in the sense that $\varepsilon(\upsilon_2^I|x) = 0$, and, in addition, that $x \in I$. Then:

$$\varepsilon(\upsilon_1^I|x) = \varepsilon(\upsilon_2^I|x) = 0.$$

However, these expectations do not condition on d_i

Any moment which depends on d_i (ie on the choice made by agent i - all our inequalities depend on this) requires properties of the disturbance conditional on d_i.

Since d is measurable given the information set I,

$$\varepsilon(\upsilon_1^I | x, d) = 0.$$

However, since $\upsilon_2 \in I$, and

$$\varepsilon(\pi(.)|.) = \varepsilon(r(.)|.) - \upsilon_2,$$

if the agent chooses d^* then

$$\upsilon_{2,d} - \upsilon_{2,d^*} \geq \varepsilon(r(.,d)|.) - \varepsilon(r(.,d^*)|.)$$

so

$$\boxed{\varepsilon(\upsilon_{2,d}|x,d) \neq 0.}$$

More intuitively: firms are choosing d^* based on the observed v_{2,d^*}, so it makes sense that there's a selection issue, ie that $\varepsilon(\upsilon_{2,d^*}|x,d^*) \neq 0$.

This result implies that the statement that "*x* is an instrument" does not "solve" the selection problem. Formally

$$\varepsilon\left[\Delta\pi(d_i,.)|x_i,d_i\right] = \varepsilon\left[\Delta r(d_i,.)|x_i,d_i\right] - \varepsilon\left[\upsilon^I_{2,i,d_i} - \upsilon^I_{2,i,d'}|x_i,d_i\right].$$

Theory gives us $\varepsilon\left[\Delta\pi(d_i,.)|x_i,d_i\right] \geq 0$, but this only implies

$$\varepsilon\left[\Delta r(d_i,.)|x_i,d_i\right] \geq 0$$

if

$$\varepsilon\left[\upsilon^I_{2,i,d_i} - \upsilon^I_{2,i,d'}|x_i,d_i\right] \geq 0,$$

and *this requires additional conditions on the measurement model.* We will come back to these conditions shortly. (See IC4.)

The Additional Requirements for Profit Inequalities

Recall that we need an assumption on the relationship between agents' expectations and the expectation operator generated by the data generating process, plus restrictions on the measurement model.

3. **Agents' Expectations (IC3)**
   Let $h(.)$ be a positive valued function. There is a known subset of the observed variables, say $x_i \in I_i$, that satisfy:

$$\frac{1}{N}\sum_i \varepsilon\left[\Delta\pi(d_i, d', d_{-i}, z_i)|x_i\right] \geq 0 \Rightarrow$$

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

Correct Expectations are Sufficient.

- Standard condition - each agent knows:
  - i the other agents' strategies ($\mathbf{d}_{-i}(I_{-i})$), and
  - ii the joint distribution of other agents' information sets and the primitive sources of uncertainty (of $(I_{-i}, \mathbf{z}_{-i})$) conditional on I_i, and regularity conditions.
- Weaker condition: agents' conditional expectations of the profit difference are correct.
  - This does not require knowledge of other agents' strategies, or the distribution of $(\mathbf{d}_{-i}, \mathbf{z}_i)$ conditional on I_i.
  - Agent uncertainty is permitted and we do not need to fully specify how the agent forms its expectations.

The Additional Requirements for Profit Inequalities

Incorrect Expectations are Possible.

All we need is the average of

$$\varepsilon\left[\Delta\pi(d_i, d', d_{-i}, z_i)|x_i\right] - E\left[\Delta\pi(d_i, d', d_{-i}, z_i)|x_i\right] \geq 0$$

Relevant cases:

- Agents' beliefs are not exactly right but the difference between agents' expectations on $\Delta\pi(., \theta_0)$ and the expectation of the data generating process are mean-zero conditional on x (in that case the expression is satisfied with equality). Or
- Agents can be "consistently overly optimistic" about the relative profits from the decisions they make.

The Additional Requirements for Profit Inequalities

4. **Condition on the Measurement Model (IC4)** This final condition is designed to deal with the selection problem noted above. Assume D_i is discrete. $\exists$ observed $x \in I_i$ and a function $c(.) : D_i \times D_i \rightarrow R^+$ such that we satisfy:

$$E\left[\sum_{j \in D_i} \chi\{d_i = j\} c(j, d'(j)) \left(\upsilon^I_{2,i,j} - \upsilon^I_{2,i,d'(j)}\right) h(x_i)\right] \geq 0$$

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

Notes regarding IC4

- This expectation is an unconditional average (does not condition on d_i); for every possible $d \in D_i$ we specify a $d'(d)$.
- This average is an average in the *differences* in the $\upsilon^I_{2,i,j} - \upsilon^I_{2,i,d'(j)}$.
- Both (i) the weights and (ii) the comparison (d') can vary with j.

Satisfying IC4

- $\forall d, \upsilon_{2,i,d} = \upsilon_{2,i}$.
  - This is Hansen and Singleton's (1982) classic article, but can allow for discreteness in choice sets, choices which are on the boundaries of the choice set, and interacting agents.

Satisfying IC4 (continued)

- $\upsilon_{2,i,d}$ can vary across decisions, but the same value of $\upsilon^{l}_{2,i,d}$ appears in more than one of them (so there are "group" effects).
  - Examples:
    - Entry models with location-specific fixed effects,
    - Social interaction models with group effects,
    - Panel data discrete choice models with choice-specific fixed effects, and
    - Cross-sectional discrete choice models where the same $\upsilon^{l}_{2,i,d}$ appear in more than one choice.
  - For each choice $d_i = j$, pick an alternative $d'(d)$ such that the two choices have the same value of $\upsilon_{2,i,d}$
    - e.g. pick an alternative in the same market; the same group; a firm making the same choice in a different time period.
  - The υ_2 terms will difference out when we generate the inequality.

Satisfying IC4 (continued)

- Ordered choice models (including the vertically-differentiated demand model used in I.O.).
  - E.g. the firm is buying a discrete number of units, so $d_i \in Z_+$ and $\upsilon_{2,i}$ is a cost component known to the agent but not the econometrician.
  - Take $d'(j) = j + 1$. The difference in profits will always contain υ_2, now interpreted as the cost savings from not purchasing the additional unit.
  - We will take an unconditional average across this cost (not conditional on the choice).
  - This is the method used in Ishii (2005).

Satisfying IC4 (continued)

- Contracting models in which υ_2 is interpreted as a component of the contract that the agents know but the econometrician does not.
  - The cost is a profit to the seller if the contract is established and a saving to the buyer if the contract is not accepted.
  - We can difference it out by adding together the inequality of the buyer and that of the seller.
  - This is an extension of the model in Ho (2009); the original paper takes account of υ_1 but ignores υ_2.

The Additional Requirements for Profit Inequalities

Satisfying IC4 (continued)

- Models for micro data where a variable needed for an inequality is unobserved (or is measured with error) at the micro level but is observed at a higher level of aggregation (say because of the availability of Census data).

Estimation from Profit Inequalities

C1 and C2 imply that for each i

$$0 \leq \sum_j \varepsilon \left(\chi\{d_i = j\}\, c(j, d'(j))\Delta\pi(j, d'(j), .)\right) h(x_i).$$

Average over i. IC3 implies that this average is less than or equal to

$$E\left[\frac{1}{N}\sum_i \sum_j \left(\chi\{d_i = j\}\, c(j, d'(j))\Delta\pi(j, d'(j), .)\right) h(x_i)\right],$$

which from IC4 is less than

$$E\left[\frac{1}{N}\sum_i \sum_j \left(\chi\{d_i = j\}\, c(j, d'(j))\Delta r(j, d'(j), .)\right) h(x_i)\right].$$

Since this last inequality is in terms of *observable* moments, we can use its sample analogue as a basis for estimation.

Estimation from Profit Inequalities - Katz (2007)

Katz's problem was to estimate the costs shoppers assign to driving to a supermarket.

These costs are important for the choice of supermarket locations and, as a result, for the analysis of the impact of zoning regulations.

They have been 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 supermarkets).

Estimation from Profit Inequalities - Katz (2007)

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

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

Since utilities are only defined up to a monotone transformation, there is a free normalization for each individual. Katz normalizes the coefficient on expenditure for each individual to equal one.

He allows for heterogeneity in the cost of drive time that is known to the agents when they make their decision but unobserved by the econometrician. This will be one component of $\upsilon_{2,i}$.

Possible counterfactuals: Purchase of any bundle of goods at any store.

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

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

This choice of $d'(d_i)$ allows us to difference out the impact of the basket of goods chosen on utility.

I.e. if $e(d)$ and $dt(d)$ provide the expenditure and the drive time for store choice d, and $(\theta + \upsilon_{2,i})$ is agent i's cost of drive time (in units of expenditure),

$$\varepsilon \left[\sum_j \chi \{d_i = j\} \Delta\pi(j, d'(j), z_i) | I_i \right] =$$

$$\varepsilon \left[\sum_j \chi \{d_i = j\} \left(e(j) - e(d'(j)) + (\theta + \upsilon_{2,i}) \left(dt(j) - dt(d'(j))\right)\right) | I_i \right] \geq 0,$$

at $\theta = \theta_0$.

Estimation from Profit Inequalities - Katz (2007)

Assuming, as seems reasonable, that $(dt\,(d_i), dt\,(d'(d_i))) \subset I_i$, this together with the fact that $dt(j) - dt(d'(j)) < 0$ by choice of alternative implies that:

$$\varepsilon \left[\sum_j \chi \{ d_i = j \} \left(\frac{e(j) - e(d'(j))}{dt(d'(j)) - dt(j)} - (\theta + \upsilon_{2,i}) \right) | I_i \right] \leq 0.$$

Estimation from Profit Inequalities - Katz (2007)

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

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

Estimation from Profit Inequalities - Katz (2007)

This result provides a lower bound to θ.

Were we to consider a second alternative in which the bundle of goods purchased was the same as in the actual choice but the counterfactual store required *less drive time*, we would also get an upper bound to θ_0.

Katz (2007) shows that these bounds are quite informative and provides a range for the average cost of drive time which accords with auxiliary information, while more standard discrete choice estimators do not.

To obtain these inequalities we chose an alternative that allowed us to difference out the impact of the bundle of goods chosen on utility (differencing out our ”group” effect).

We then rearranged these differences to form a moment which was linear in the remaining source of υ_2 variance no matter d_i (the source being differences in the costs of travel time).

If we were interested in the impact of a particular good purchased on utility, we would have considered baskets of goods which differed only in that good and goods which had cross partials with that good in the utility function, at the *same* supermarket (thus differencing out the effects of travel time and other components of utility).

A lot more options would present themselves were we to have data on multiple shopping trips for each household.

Full Information - No Error Model

- Does not allow for:
    - Specification or measurement error
    - Asymmetric or incomplete information
    - Incorrect expectations

  except to the extent that these details do not cause differences in the profits earned from different choices.

- Requires a parametric assumption on the distribution of υ_2.

Summary

Profit Inequalities Model

- Allows for specification errors, incorrect expectations, and incomplete and asymmetric information.
- The econometrician does not need to specify what the agent knows about either its competitors, or about the state of nature.
- However, it requires a (sometimes tricky) restriction on υ_2.
  - Given that restriction, there is no need for a distributional assumption here.