Graduate Industrial Organization

Lecture: Moment Inequalities Part 1, Theory

Michael J. Dickstein

New York University

December 9, 2020

Part 1: Building Moment Inequalities

Discrete choice applications

- Today's goal: illustrate how to build moment inequalities that identify the parameters of *static single-agent discrete choice models*.
- Comments:
  - In theory, moment inequalities can be used beyond discrete choice settings.
  - Most applied papers that have used moment inequalities (to date) aim to estimate parameters of the utility or production function of agents choosing among a finite set of alternatives.
  - Today, we omit dynamic discrete choice (e.g. Morales, Sheu, and Zahler (2018)) and dynamic games (e.g. Ciliberto and Tamer (2009))

**Thanks to Eduardo Morales for collaborating on the lecture content*

Static Discrete Choice Problem: Utility

- Utility of agent i for alternative j is:

$$U_{ij} = \beta \mathbb{E}[x_{ij} | \mathcal{W}_i] + \nu_{ij}, \qquad j = 1, \ldots, \mathcal{J}_i,$$

  - $\mathbb{E}[\cdot]$ is the expectation operator with respect to the data generating process
  - x_{ij} is a vector of covariates the researcher observes
  - $\mathcal{W}_i$ is the information set that agent i uses to predict the value of $x_i \equiv \{x_{i1}, \ldots, x_{i\mathcal{J}_i}\}$.
- This specification assumes agents have rational expectations: we define agents' expectations with respect to the data generating process.
- If we assume agents have perfect foresight, the utility function simplifies to:

$$U_{ij} = \beta x_{ij} + \nu_{ij}, \qquad j = 1, \ldots, \mathcal{J}_i.$$

Static Discrete Choice Problem: Decision

- Define d_{ij} as an indicator equal to 1 if individual i chooses alternative j. We assume that

$$d_{ij} = \mathbb{1}\{\beta \mathbb{E}[x_{ij}|\mathcal{W}_i] + \nu_{ij} \geq \max_{j' \in \mathcal{J}_i} \beta \mathbb{E}[x_{ij'}|\mathcal{W}_i] + \nu_{ij'}\}.$$

- Denote as ε_{ij} the error that agent i makes when predicting x_{ij}:

$$\varepsilon_{ij} = x_{ij} - \mathbb{E}[x_{ij}|\mathcal{W}_i].$$

- Therefore, we can rewrite d_{ij} as

$$d_{ij} = \mathbb{1}\{\beta x_{ij} + \nu_{ij} - \beta \varepsilon_{ij} \geq \max_{j' \in \mathcal{J}_i} \beta x_{ij'} + \nu_{ij'} - \beta \varepsilon_{ij'}\}.$$

Static Discrete Choice Problem: Unobserved Components

- ν_{ij} is the *structural error*.
  - captures elements of the payoff function the agent knows when making a decision ($\nu_{ij} \subseteq \mathcal{W}_i$) but which the econometrician does not observe.
  - Economic theory generally imposes no restriction on the distribution of the vector $(\nu_{i1}, \ldots, \nu_{i\mathcal{J}_i})$ across individuals. We often impose assumptions on this distribution for identification/convenience purposes.
- ε_{ij} is the *expectational error*.
  - captures elements affecting the payoff relevant variable x_{ij} that the agent does **not** know when making a decision.
  - under the assumption of rational expectations:

$$\mathbb{E}[\varepsilon_{ij}|\mathcal{W}_i] = 0.$$

  - Rational expectations does not imply additional restrictions on the distribution of ε_{ij} conditional on $\mathcal{W}_i$.

Static Discrete Choice Problem: Example

- Firm i decides whether to enter a market or stay out: $j = \{1, 0\}$, with

$$\pi_{i0} = 0 \qquad \pi_{i1} = \eta^{-1}\mathbb{E}[r_i|\mathcal{W}_i] - f_i = \eta^{-1}\mathbb{E}[r_i|\mathcal{W}_i] - \beta_0 - \beta_1 dist_i - \nu_i,$$

  - π_{i1} denotes firm i's expectation of the profits upon entry
  - r_i denotes sales revenue conditional on entry
  - f_i denotes fixed entry costs; f_i a function of $dist_i$, the distance between the market and firm i's production.
  - ν_i captures unobserved heterogeneity in fixed entry costs.
- The information set $\mathcal{W}_i$ captures any variable firm i knows and uses to predict its sales revenue upon entry; i.e. any variable firm i uses to predict r_i

Static Discrete Choice Problem: Data and Parameters

- Data. For a random sample of individuals and a *subset* of all choices in $\mathcal{J}_i$, the econometrician observes
  - $x_i = \{x_{i1}, \ldots, x_{i\mathcal{J}_i}\}$,
  - $d_i = \{d_{i1}, \ldots, d_{i\mathcal{J}_i}\}$,
  - $z_i = \{z_{i1}, \ldots, z_{i\mathcal{J}_i}\}$ such that $z_i \subseteq \mathcal{W}_i$.
- Researchers may be interested in performing counterfactuals with respect to
  - x_i; e.g. how does entry change if the potential revenue increases 10%?
  - $\mathcal{W}_i$; e.g. how does entry change if potential entrants become better at predicting ex post revenues?
  - β; e.g. how does entry change if fixed entry costs are subsidized?
- The parameters of the model (a subset of which may be needed to perform these counterfactuals) are:
  - the vector of preference parameters β,
  - the joint distribution of $\nu_i = \{\nu_{i1}, \ldots, \nu_{i\mathcal{J}_i}\}$ across individuals,
  - the information set $\mathcal{W}_i$ and the conditional density $f(x_i|\mathcal{W}_i)$.

Identification Challenge: Unobserved Information Sets

- Researchers generally do not observe agents' information sets, $\{\mathcal{W}_i\}_i$.
- Even if we assume that agents' expectations are rational, we still need to know $\mathcal{W}_i$ to correctly define a proxy for the term $\mathbb{E}[x_{ij}|\mathcal{W}_i]$ entering U_{ij}.
- Therefore, researchers fail to observe two terms in utility: $\mathbb{E}[x_{ij}|\mathcal{W}_i]$ and ν_{ij}
  - Literature on nonparametric identification provides assumptions needed when ν_{ij} is not observed
  - Literature on moment inequalities provides insights when agents' true information sets are not observed.

Identification Challenge: Unobserved Information Sets

Pre-Moment Inequality Approach: Manski (1991)

How shall we handle $\mathbb{E}[x_{ij}|\mathcal{W}_i]$ in U_{ij}?

- Pre-moment inequality approach: assume the researcher observes agents' information sets $\mathcal{W}_i$ *ex post*. That is, the researcher observes all variables *ex post* that the agent used *ex ante*.
- Given this assumption, Manski (1991) introduces a two-step estimator:
  - Step 1: Regress the ex-post realization x_{ij} on the observed information set $\mathcal{W}_i$ to obtain a prediction: $\mathbb{E}\widehat{[x_{ij}|\mathcal{W}_i]}$.
  - Step 2: Use restrictions on the distribution of $(\nu_{i1}, \ldots, \nu_{i\mathcal{J}_i})$ to estimate β given that:

$$d_{ij} = \mathbb{1}\{\beta\mathbb{E}\widehat{[x_{ij}|\mathcal{W}_i]} + \nu_{ij} \geq \max_{j' \in \mathcal{J}_i} \beta\mathbb{E}\widehat{[x_{ij'}|\mathcal{W}_i]} + \nu_{ij'}\}.$$

Motivation for Using Moment Inequalities

- When agents' decisions depend on expectations, we'd like to estimate structural parameters of agents' payoff functions while imposing only weak assumptions on agents' information sets.
- Moment inequalities allow us to identify these structural parameters under the assumption that the researcher observes a **subset of agents' true information sets**.

Motivation for Using Moment Inequalities

- The main objective of using moment inequalities is to relax the assumptions typically needed for point-identification. In today's lecture:
  - Ex. assumptions on agents' information sets
  - Ex. assumptions on agents' consideration sets
- But there are others!
  - Ex. in static entry games with multiple equilibria, want to relax equilibrium selection assumptions (Ciliberto and Tamer, 2009).

Why Not Switch to Moment Inequalities?

Limitations

- (1) While we need only weak assumptions on the distribution of the expectational errors $(\varepsilon_{i1}, \ldots, \varepsilon_{i\mathcal{J}_i})$, we need strong assumptions on the distribution of the structural errors $(\nu_{i1}, \ldots, \nu_{i\mathcal{J}_i})$
  - Exception: Pakes and Porter (2016) impose weak assumptions on $(\nu_{i1}, \ldots, \nu_{i\mathcal{J}_i})$ but assume away expectational errors
- (2) Unclear mapping between assumptions needed for identification and the assumptions needed to perform counterfactuals.
  - Exception (for very specific settings): Dickstein and Morales (2018).
- (3) Severe computational difficulties arise in the estimation of β when the dimensionality of β is relatively large.
  - Standard inference procedures require evaluating a criterion function at each point in a grid covering the parameter space.
  - Alternative: Chen, Christensen, and Tamer (2018) provide a Monte Carlo sampler

Moment Inequalities: Classification

- The applied literature on moment inequalities has so far studied a fairly limited set of models and has resorted to a limited set of "tricks" to derive moment inequalities.
- We will consider two general types of moment inequalities:
  - **revealed-preference moment inequalities**
  - **odds-based moment inequalities**
- Within each general type, the form of the inequalities depends on the assumptions imposed on the distribution of the structural errors $(\nu_{i1}, \ldots, \nu_{i\mathcal{J}_i})$ and the expectational errors $(\varepsilon_{i1}, \ldots, \varepsilon_{i\mathcal{J}_i})$.

Revealed-Preference Inequality: Key Insight

- The key insight behind all revealed-preference moment inequalities is that

$$d_{ij} = \mathbb{1}\left\{\beta x_{ij} + \nu_{ij} - \varepsilon_{ij} \geq \max_{j' \in \mathcal{J}_i} \beta x_{ij'} + \nu_{ij'} - \varepsilon_{ij'}\right\} \tag{1}$$

  implies that, for any $(j, j') \in \mathcal{J}_i$,

$$d_{ij}\left(\beta(x_{ij} - x_{ij'}) + (\nu_{ij} - \nu_{ij'}) + (\varepsilon_{ij} - \varepsilon_{ij'})\right) \geq 0. \tag{2}$$

- In other words, equation (2) is a **necessary** condition for equation (1).
- Equation (2) is also sufficient for equation (1) if and only if the cardinality of the choice set $\mathcal{J}_i$ is equal to two.

Revealed-Preference Inequality: Expectational Error

- The inequality

$$d_{ij}\big(\beta(x_{ij} - x_{ij'}) + (\nu_{ij} - \nu_{ij'}) + (\varepsilon_{ij} - \varepsilon_{ij'})\big) \geq 0,$$

  cannot be used directly for identification of the parameter vector β, as it depends on the unobserved terms ν_{ij}, $\nu_{ij'}$, ε_{ij}, and $\varepsilon_{ij'}$.
- Taking expectations conditional on the true information set $\mathcal{W}_i$, we obtain

$$\mathbb{E}[d_{ij}\big(\beta(x_{ij} - x_{ij'}) + (\nu_{ij} - \nu_{ij'})\big)|\mathcal{W}_i] \geq 0,$$

  as the assumption that agents have rational expectations implies

$$\mathbb{E}[d_{ij}(\varepsilon_{ij} - \varepsilon_{ij'})|\mathcal{W}_i] = 0.$$

- Proof:

$$\begin{aligned}\mathbb{E}[d_{ij}(\varepsilon_{ij} - \varepsilon_{ij'})|\mathcal{W}_i] &= \mathbb{E}[\mathbb{E}[d_{ij}(\varepsilon_{ij} - \varepsilon_{ij'})|\mathcal{W}_i, \nu_i]|\mathcal{W}_i] \\ &= \mathbb{E}[d_{ij}\mathbb{E}[(\varepsilon_{ij} - \varepsilon_{ij'})|\mathcal{W}_i, \nu_i]|\mathcal{W}_i] = \mathbb{E}[d_{ij} \times 0|\mathcal{W}_i] = 0.\end{aligned}$$

Revealed-Preference Inequality: Structural Error

- The resulting moment inequality is therefore:

$$\mathbb{E}[d_{ij}\beta(x_{ij} - x_{ij'})|\mathcal{W}_i] + \mathbb{E}[d_{ij}(\nu_{ij} - \nu_{ij'})|\mathcal{W}_i] \geq 0.$$

- What is the second term in this inequality? A selection correction. Intuitively, even if we observe that i prefers j over j', we cannot conclude that $\beta\mathbb{E}[x_{ij}|\mathcal{W}_i] > \beta\mathbb{E}[x_{ij'}|\mathcal{W}_i]$ because j may be preferred over j' because $\nu_{ij} >> \nu_{ij'}$.
- For simplicity in the notation, let's define $s_{jj'}(W_i;\beta)$ as

$$s_{jj'}(\mathcal{W}_i;\beta) \equiv \mathbb{E}[d_{ij}(\nu_{ij} - \nu_{ij'})|\mathcal{W}_i].$$

Revealed-Preference Inequality: Structural Error

- Even if the distribution of ν_{ij} is independent of $\mathcal{W}_i$ and identical across j, $s_{jj'}(\mathcal{W}_i;\beta) \neq 0$. Why? d_{ij} is a function of $(\nu_{ij} - \nu_{ij'})$.
- Furthermore, if the distributions of ν_{ij} and $\nu_{ij'}$ conditional on $\mathcal{W}_i$ are identical, $s_{jj'}(\mathcal{W}_i;\beta) \geq 0$.
- Therefore, it could be that, at the true value of the parameter vector β,

$$\mathbb{E}[d_{ij}\beta(x_{ij} - x_{ij'})|\mathcal{W}_i] + \mathbb{E}[d_{ij}(\nu_{ij} - \nu_{ij'})|\mathcal{W}_i] \geq 0,$$

  but

$$\mathbb{E}[d_{ij}\beta(x_{ij} - x_{ij'})|\mathcal{W}_i] < 0.$$

- We therefore need to consider the term $s_{jj'}(\mathcal{W}_i;\beta)$.

Revealed-Preference Inequality: Structural Error

Two ways researchers handle the selection correction term $s_{jj'}(\mathcal{W}_i;\beta)$:

- (1) Impose assumptions on the distribution of the vector ν_i conditional on $\mathcal{W}_i$ such that $s_{jj'}(\mathcal{W}_i;\beta) = 0$.
- (2) Impose assumptions on the distribution of the vector ν_i conditional on $\mathcal{W}_i$ that allow the researcher to derive a function $\bar{s}_{jj'}(\mathcal{W}_i;\beta)$ such that

$$\bar{s}_{jj'}(\mathcal{W}_i;\beta) \geq s_{jj'}(\mathcal{W}_i;\beta).$$

Revealed Preference Inequalities: "No Selection"

- One may impose different assumptions on the distribution of the vector $\nu_i = (\nu_{i1}, \ldots, \nu_{i\mathcal{J}_i})$ conditional on $\mathcal{W}_i$ such that $s_{jj'}(\mathcal{W}_i; \beta) = 0$.
- We note four cases here:
  1. No unobserved heterogeneity;
  2. Group-of-choices fixed effects;
  3. Group-of-individuals fixed effects;
  4. Ordered-choice model.

Revealed Preference Inequalities: "No Selection"

(1) No unobserved heterogeneity

- Implies $\nu_{ij} = 0$ for all $j \in \mathcal{J}_i$.
- Applied in Holmes (2011): Walmart choosing where to open stores.
- Discussed in Pakes (2010) and Pakes et al. (2015).
- In most empirical applications, this assumption is too restrictive.

Revealed Preference Inequalities: "No Selection"

(2) Group-of-choices fixed effects

- Implies ν_{ij} is common to a subset of choices:

$$\nu_{ij} - \nu_{ij'} = 0 \qquad \text{if} \qquad g(j) = g(j'),$$

  where the function $g(\cdot)$ creates a partition of the set of potential choices.
- Applied in Morales et al. (2018): exporters deciding which markets to enter.
- Discussed in Pakes (2010) and Pakes et al. (2015).
- We can only exploit inequalities that compare potential choices j and j' such that $g(j) = g(j')$.

Revealed Preference Inequalities: “No Selection”

(3) Group-of-individuals fixed effects

- Implies that ν_{ij} is common to a subset of individuals:

$$\nu_{ij} - \nu_{i'j} = 0 \qquad \text{if} \qquad g(i) = g(i'),$$

  where the function $g(\cdot)$ creates a partition of the set of individuals.

- Applied in Ho and Pakes (2014): patients deciding which hospital to visit.
- It requires double-differencing. Find two individuals i and i' such that $d_{ij} = d_{i'j'} = 1$ and build

$$\mathbb{E}[\beta(x_{ij} - x_{ij'}) + (\nu_{ij} - \nu_{ij'})|d_{ij} = 1, d_{i'j'} = 1, \mathcal{W}_i, \mathcal{W}_{i'}] \geq 0,$$
$$\mathbb{E}[\beta(x_{i'j'} - x_{i'j}) + (\nu_{i'j'} - \nu_{i'j})|d_{ij} = 1, d_{i'j'} = 1, \mathcal{W}_i, \mathcal{W}_{i'}] \geq 0.$$

  As long as $g(i) = g(i')$, we can sum these two inequalities and obtain:

$$\mathbb{E}[\beta(x_{ij} - x_{ij'}) + \beta(x_{i'j'} - x_{i'j})|d_{ij} = 1, d_{i'j'} = 1, \mathcal{W}_i, \mathcal{W}_{i'}] \geq 0.$$

Revealed Preference Inequalities: "No Selection"

(4) Ordered-choice model

- Ordered-choice model means $\nu_{ij} = j\eta_i$.
- Additionally, assume that $\mathbb{E}[\eta_i|\mathcal{W}_i] = 0$ and $\mathcal{J}_i = \mathcal{J}_{i'} = \mathcal{J}$ for all i and i'.
- Applied in Ishii (2008): banks deciding how many ATMs to install.
- Discussed in Pakes (2010) and Pakes et al. (2015).
- We must build inequalities such that, for every individual i, the alternative choice j' is one unit below the actual choice of i; i.e. $j' = j - 1$. Therefore

$$\mathbb{E}[d_{ij}\beta(x_{ij} - x_{i(j-1)})|\mathcal{W}_i] + \mathbb{E}[d_{ij}(\nu_{ij} - \nu_{i(j-1)})|\mathcal{W}_i] \geq 0,$$

and, summing these inequalities for all j in the choice set $\mathcal{J}$,

$$\sum_{j\in\mathcal{J}} \mathbb{E}[d_{ij}\beta(x_{ij} - x_{i(j-1)})|\mathcal{W}_i] + \sum_{j\in\mathcal{J}} \mathbb{E}[d_{ij}(\nu_{ij} - \nu_{i(j-1)})|\mathcal{W}_i] \geq 0.$$

Revealed Preference Inequalities: "No Selection"

(4) Ordered-choice model (cont.)

- If each individual i chooses j such that $j' = j - 1$ is in $\mathcal{J}$, then

$$
\begin{aligned}
\sum_{j\in\mathcal{J}} \mathbb{E}[d_{ij}(\nu_{ij} - \nu_{i(j-1)})|\mathcal{W}_i] &= \sum_{j\in\mathcal{J}} \mathbb{E}[d_{ij}(j\eta_i - (j-1)\eta_i)|\mathcal{W}_i] \\
&= \sum_{j\in\mathcal{J}} \mathbb{E}[d_{ij}\eta_i|\mathcal{W}_i] \\
&= \mathbb{E}[\sum_{j\in\mathcal{J}} d_{ij}\eta_i|\mathcal{W}_i] \\
&= \mathbb{E}[\eta_i \sum_{j\in\mathcal{J}} d_{ij}|\mathcal{W}_i] \\
&= \mathbb{E}[\eta_i \times 1|\mathcal{W}_i] \\
&= \mathbb{E}[\eta_i|\mathcal{W}_i] = 0.
\end{aligned}
$$

Revealed Preference Inequalities: "No Selection"

(4) Ordered-choice model (cont.)

- The resulting moment inequality is therefore

$$\sum_{j \in \mathcal{J}} \mathbb{E}[d_{ij}\beta(x_{ij} - x_{i(j-1)})|\mathcal{W}_i] \geq 0.$$

- Similarly, we can build inequalities for which the alternative choice j' is one unit above the actual choice of i, i.e. $j' = j + 1$. In this case

$$\sum_{j \in \mathcal{J}} \mathbb{E}[d_{ij}(\nu_{ij} - \nu_{i(j+1)})|\mathcal{W}_i] = -\mathbb{E}[\eta_i|\mathcal{W}_i] = 0,$$

  and the resulting moment inequality becomes

$$\sum_{j \in \mathcal{J}} \mathbb{E}[d_{ij}\beta(x_{ij} - x_{i(j+1)})|\mathcal{W}_i] \geq 0.$$

- If some individuals choose j such that either $j - 1$ or $j + 1$ do not belong to $\mathcal{J}$, selection issues arise. See Pakes (2010) for a discussion.

Revealed Preference Inequalities: Bounded Selection

- The usefulness of deriving an upper bound $\bar{s}_{jj'}(\mathcal{W}_i;\beta)$ on the term

$$s_{jj'}(\mathcal{W}_i;\beta) \equiv \mathbb{E}[d_{ij}(\nu_{ij} - \nu_{ij'})|\mathcal{W}_i]$$

is that, if it is true that the following inequality holds at the true value of the parameter vector β,

$$\mathbb{E}[d_{ij}\beta(x_{ij} - x_{ij'})|\mathcal{W}_i] + \mathbb{E}[d_{ij}(\nu_{ij} - \nu_{ij'})|\mathcal{W}_i] \geq 0, \tag{3}$$

then it will also be true that, at the true value of the parameter vector β,

$$\mathbb{E}[d_{ij}\beta(x_{ij} - x_{ij'})|\mathcal{W}_i] + \bar{s}_{jj'}(\mathcal{W}_i;\beta) \geq 0. \tag{4}$$

- The cost of using equation (4) for identification (instead of equation (3)) is that the identified set will be larger; i.e. the set of values of β consistent with this inequality will be larger.

Revealed Preference Inequalities: Bounded Selection

Dickstein and Morales (2018)

- One can construct such upper bound $\bar{s}_{jj'}(\mathcal{W}_i;\beta)$ as long as the assumed distribution of ν_i conditional on $\mathcal{W}_i$ is such that for any pair of alternatives j and j', the distribution of $\nu_{ij} - \nu_{ij'}$ conditional on $\mathcal{W}_i$ verifies:
  - it is known up to a finite parameter vector
  - it has mean zero
  - its truncated expectation is convex in the truncation point.
- Using mathematical notation, for all j and j' included $\mathcal{J}_i$, it must hold that

$$\mathbb{E}[\nu_{ij} - \nu_{ij'}|\mathcal{W}_i] = 0, \quad \text{and} \quad \mathbb{E}[\nu_{ij} - \nu_{ij'}|\mathcal{W}_i, \nu_{ij} - \nu_{ij'} \geq \lambda] \tag{5}$$

  is a known function of λ and convex in λ.
- Both the normal and the logistic distribution verify these restrictions.

Odds-Based Inequalities: Assumptions

- Dickstein and Morales (2018) introduce the odds-based inequality for the special case of single-agent binary choice models.
- In order to derive odds-based inequalities in Dickstein and Morales (2018), the distribution of $\nu_{ij} - \nu_{ij'}$ conditional on $\mathcal{W}_i$ must be:
  - known up to a finite parameter vector; and,
  - log-concave.
- The log-concavity of the distribution of $\nu_{ij} - \nu_{ij'}$ conditional on $\mathcal{W}_i$ implies:

$$\frac{F_{\nu|\mathcal{W}}(\lambda)}{1 - F_{\nu|\mathcal{W}}(\lambda)} \quad \text{and} \quad \frac{1 - F_{\nu|\mathcal{W}}(\lambda)}{F_{\nu|\mathcal{W}}(\lambda)}$$

  are convex in the index λ, where $F_{\nu|\mathcal{W}}(\cdot)$ is the CDF of $\nu_{ij} - \nu_{ij'}$ conditional on the information set $\mathcal{W}_i$.
- Both the normal and the logistic distribution are log-concave.