Identification and Inference Based on Moment Inequalities: Applications to Incomplete Models

Hiroaki Kaido

Boston University

May 23
JEA Spring Meeting

Moment Inequalities

Commonly studied economic models give inequality restrictions on moments.

$$E_P[m_j(X_i, \theta)] \leq 0,\ j = 1, \cdots, J$$

where $X_i \in \mathcal{X}$ is a vector of observable variables and $m : \mathcal{X} \times \Theta \to \mathbb{R}^J$ is known up to a finite dimensional parameter $\theta \in \Theta$.

This is often the case when the researcher wants to avoid assumptions that are not based on economic theory but rather made for convenience.

Examples

- Regression models with missing observations or censored variables, where one does not know how censoring occurred.
- Incomplete economic models, which predict multiple outcome values, but one does not know how the selection mechanism works.

Challenges

1. Identification: Point identification of θ is not guaranteed.
$\rightarrow$ Partial Identification

2. Estimation: What does consistency or efficiency mean?

3. Inference: How do we test hypotheses regarding θ?
How do we construct confidence sets?

Identification

If point identification does not hold, we cannot fully recover the parameter of interest from data.

Still, moment restrictions are often informative. They can be used to bound the set of parameter values that are consistent with observations.

When this is the case, θ is said to be partially identified, and the set of observationally equivalent parameter values are called the identified set, denoted by $\Theta_I(P)$.

$$\Theta_I(P) = \big\{\theta \in \Theta : E_P[m_j(X_i, \theta)] \leq 0,\ j = 1, \cdots, J\big\}.$$

Charles Manski pioneered the formal study of partially identified models. For a summary of his work, see Manski (2003).

Ex.1: Entry Game (Tamer, 2003)

Moment inequalities arise naturally in models with multiple equilibria. Consider two firms with the following payoffs:

$$\pi_j = (u_j - \theta_j s_{-j}) s_j, \quad j = 1, 2,$$

where $u_j \sim U[0,1]$ independently, $s_j \in \{0,1\}$ is firm j's entry decision, and s_{-j} is the other firm's entry decision.

The payoffs can be summarized as follows.

		Firm 2	
		0	1
Firm 1	0	$(0,0)$	$(0,u_2)$
	1	$(u_1,0)$	$(u_1-\theta_1, u_2-\theta_2)$

The parameter of interest is the strategic interaction effects: $\theta \equiv (\theta_1, \theta_2) \in [0,1]^2$. Suppose that $u \equiv (u_1, u_2)$ is known to the firms but not to the econometrician.

Ex.1: Entry Game

This is an incomplete model, which has the structure

$$s \in G(u|z, \theta),$$

where the correspondence $G : U \times \mathcal{Z} \times \Theta \to S$ gives the set of outcome values predicted by the model.

Suppose the observed outcome s_i is a pure strategy Nash equilibrium (PSNE). Then, G is given by

$$G\left(u_1, u_2 \mid \theta\right) = \begin{cases} (1,1) & \text{if } u_j > \theta_j \; \forall j \\ (1,0) & u_1 > \theta_1,\, u_2 < \theta_2 \\ (0,1) & u_1 < \theta_1,\, u_2 > \theta_2 \\ \{(1,0), (0,1)\} & \text{if } u_j < \theta_j \; \forall j \end{cases}$$

The model then tells us

$$Pr(s \in A) = P(A) = \int p_u(A) dm_\theta, \;\; p_u(G(u|\theta)) = 1,$$

where m_θ is the distribution of u (known up to parameter θ), and p_u is the conditional distribution of s_i given u (unknown selection mechanism).

Ex.1: Entry Game

Equilibrium selection mechanism is a difficult object to handle empirically. How should one proceed?

Option 1: Complete the model.

- Transform s_i (e.g. $\#$ of entrants) so that the transformed model is complete. (Bresnahan & Reiss, 1991) $\rightarrow$ loss of information.
- Parameterize the selection mechanism. (Bajari, Hong, & Ryan, 2010) $\rightarrow$ The conclusion depends on the additional assumption used to complete the model.

Option 2: Use restrictions given by the incomplete model (Tamer, 2003)

- Without further assumptions, the sharp identified set is

$$\Theta_I(P) = \{\theta \in \Theta : P(A) = \int p_u(A) dm_\theta,\ p_u(G(u|\theta)) = 1\}.$$

However, the form of $\Theta_I(P)$ is not tractable as it involves p_u.

Ex.1: Entry Game

One can generate moment inequalities from the model to characterize $\Theta_I(P)$. For example,

$$\begin{aligned} Pr(s = (1,1)) &= (1-\theta_1)(1-\theta_2) \\ Pr(s = (1,0)) &\geq \theta_2(1-\theta_1) \\ Pr(s = (1,0)) &\leq \theta_2. \end{aligned}$$

Note that the probabilities on the left are identified. These restrictions partially identify θ.

Note:

- $Pr(s = (1,1))$ can be written as $E_P[1\{s = (1,1)\}]$. Hence, they are moment equality and inequality restrictions.
- Do these (in)equalities fully characterize the sharp identified set?
  $\rightarrow$ There is a systematic way to generate sharp restrictions.

Ex.1: Entry Game

By Choquet's theorem, (see Galichon and Henry, 2011, Molchanov, 2005) the sharp identification region is

$$\Theta_I(P) = \{\theta \in \Theta : \nu_\theta(A) \leq P(A) \leq \nu^*_\theta(A)\},$$

where

- $\nu_\theta(A) = m_\theta(G(u|\theta) \subseteq A)$ is the lower bound for $p(A)$
  It is called a containment functional or belief function.
- $\nu^*_\theta(A) = m_\theta(G(u|\theta) \cap A \neq \emptyset)$ is the upper bound for $p(A)$
  It is called a capacity functional or the conjugate of the belief function.

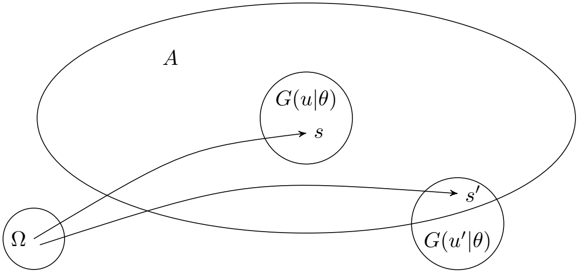
In the previous example, we have

A	$\nu_\theta\left(A\right) \equiv \min \Pr\left(A\right)$	$\nu^*_\theta\left(A\right) \equiv \max \Pr(A)$
$A_1 = \{(1,1)\}$	$(1-\theta_1)(1-\theta_2)$	$(1-\theta_1)(1-\theta_2)$
$A_2 = \{(1,0)\}$	$(1-\theta_1)\theta_2$	θ_2
$A_3 = \{(0,1)\}$	$\theta_1(1-\theta_2)$	θ_1
$A_1^c = A_4 = \{(1,0),(0,1)\}$	$\theta_1+\theta_2-\theta_1\theta_2$	$\theta_1+\theta_2-\theta_1\theta_2$
$A_2^c = A_5 = \{(1,1),(0,1)\}$	$(1-\theta_2)$	$1-\theta_2(1-\theta_1)$
$A_3^c = A_6 = \{(1,1),(1,0)\}$	$(1-\theta_1)$	$1-\theta_1(1-\theta_2)$
$S = \{(1,1),(1,0),(0,1)\}$	1	1

Intuition

Why these inequalities?

The key intuition is that s is a selection of a random set $G(u|\theta)$:
(i.e. s is a random variable taking values within the random set $G(u|\theta)$ with probability 1.)



$$G(u|\theta) \subseteq A \;\; \Rightarrow \;\; s \in A \;\; \Rightarrow \;\; G(u|\theta) \cap A \neq \emptyset$$

These relationships imply the lower and upper bounds (that turn out to be sharp) on $p(A)$.

Moment (In)equalities

In many incomplete models, one can characterize the (sharp) identified set $\Theta_I(P)$ as:

$$\Theta_I(P) = \big\{\theta \in \Theta : E_P[m_j(X_i,\theta)] \le 0,\ j = 1,\cdots,J_1$$
$$E_P[m_j(X_i,\theta)] = 0,\ j = J_1+1,\cdots,J_2\big\}.$$

Comments:

- Games with mixed strategy Nash equilibria or correlated equilibria can also be analyzed using conditional moment inequalities (Beresteanu, Molchanov, Molinari, 2011)
- Incomplete models can occur in a variety of contexts:
  - Generalized IV models (Chesher & Rosen, 2012)
  - Auctions (Haile & Tamer, 2003)
- If the sampling process reveals P asymptotically, $\Theta_I(P)$ is the best object one can recover from the observations.

Estimation

How can one estimate $\Theta_I(P)$ from data?

Form a population criterion function based on moments.
For example, for a positive definite matrix W, define

$$Q(\theta) \equiv E_P[m(X_i,\theta)]'_+ W(\theta) E_P[m(X_i,\theta)]_+,$$

where $y_+ = \max\{y, 0\}$. This criterion function is minimized on $\Theta_I(P)$.

Construct an estimator of $\Theta_I(P)$ using a sample analog.

$$Q_n(\theta) \equiv (\frac{1}{n}\sum_{i=1}^{n} m(X_i,\theta))'_+ W(\theta)(\frac{1}{n}\sum_{i=1}^{n} m(X_i,\theta))_+$$

Note: If moment equalities are present, one can add a GMM-type criterion function to Q_n.

Consistency

Define a level-set estimator by

$$\hat{\Theta}_n(c) \equiv \{\theta : nQ_n(\theta) \leq c\}.$$

Chernozhukov, Hong, and Tamer (2007) show that, if $\Theta_I(P)$ is well-behaved (compact, non-empty interior, boundary does not involve "thin parts"), the level-set estimator $\hat{\Theta}_n(0)$ (argmin of Q_n) is consistent in the following sense:

$$d_H(\hat{\Theta}_n(0), \Theta_I(P)) \xrightarrow{p} 0,$$

where $d_H(A, B) \equiv \max\{\sup_{a\in A} \inf_{b\in B} \|a - b\|, \sup_{b\in B} \inf_{a\in A} \|a - b\|\}$ is the Hausdorff distance between two sets.

In more general cases (e.g. potentially overidentified by moment (in)equalities), one can use a sequence of levels $c = c_n$ that slowly diverges e.g. $c_n = \ln n$.

Inference

Consider unconditional moment inequalities:

$$E_P[m_j(X_i,\theta)] \le 0,\ j = 1, \cdots, J_1.$$

How can one construct a confidence set (CS)?

What to cover?

There are two types CS_{1n}, CS_{2n} of confidence sets:

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\theta\in\Theta_I(P)} P(\theta \in CS_{1n}) \ge 1-\alpha$$

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} P(\Theta_I(P) \subseteq CS_{2n}) \ge 1-\alpha.$$

- CS_{1n} covers each $\theta \in \Theta_I(P)$ (hence the true parameter value) with probability $1-\alpha$ asymptotically.
- CS_{2n} covers the whole identified set $\Theta_I(P)$ with probability $1-\alpha$ asymptotically.

It is also desirable for CSs to ensure coverage uniformly over a reasonable class of DGPs.

CS for θ: (Andrews & Soares, 2010)

Consider testing the hypothesis $H_0 : \theta \in \Theta_I(P)$. One can collect all parameter values that pass the test and form a confidence region.

Using the sample criterion function Q_n, define a test statistic:

$$T_n(\theta) \equiv nQ_n(\theta).$$

Construct a CS by

$$CS_n \equiv \{\theta \in \Theta : T_n(\theta) \leq c_n(\theta)\},$$

where the critical value $c_n(\theta)$ must be chosen in such a way that CS_n covers θ with probability $1 - \alpha$ asymptotically.

A challenge: The (null) limiting distribution of T_n changes depending on (θ, P).

The limiting behavior of the moments

At each $\theta \in \Theta_I(P)$ (and under a fixed P):

$$\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n} m_j(X_i,\theta))_+$$

$$= \Big(\underbrace{\sqrt{n}\{\frac{1}{n}\sum_{i=1}^{n} m_j(X_i,\theta) - E_P[m_j(X_i,\theta)]\}}_{Z_{j,n}(\theta)} + \underbrace{\sqrt{n}E_P[m_j(X_i,\theta)]}_{h_{P,j,n}(\theta)}\Big)_+$$

$$\xrightarrow{d} \begin{cases} Z_j(\theta)_+, & \text{if } E_P[m_j(X_i,\theta)] = 0 \text{ (binding)} \\ 0 & \text{if } E_P[m_j(X_i,\theta)] < 0 \text{ (slack)}, \end{cases}$$

The (pointwise) limiting distribution of the moments changes discontinuously depending on whether the j-th constraint binds at θ or not.

In general, along a sequence (θ_n, P_n) such that $Z_{j,n}(\theta_n) \xrightarrow{d} Z_j$ and $h_{P_n,j,n}(\theta_n) \to h_j \in [-\infty, 0], j = 1, \cdots, J_1$, one has

$$T_n \xrightarrow{d} T = (Z+h)_+' W (Z+h)_+,$$

where $h = (h_1, \cdots, h_J)$ measures the slackness of the constraints.

Bootstrap Critical Value: AS (2010)

Step 1: At each $\theta \in \Theta$, select moments that are nearly binding.

$$\varphi_j(\theta) = \begin{cases} 0 & \text{if } \frac{1}{\sqrt{n}} \sum_{i=1}^n m_j(X_i, \theta) \geq -\kappa_n \\ -\infty & \text{otherwise.} \end{cases} \quad j = 1, \cdots, J_1.$$

where κ_n is a tuning parameter such as $\kappa_n = \sqrt{\ln n}$. The GMS (generalized moment selection) function $\varphi(\theta) \equiv (\varphi_1(\theta), \cdots, \varphi_J(\theta))$ approximates $h_{P,n}(\theta)$ conservatively.

Step 2: Bootstrap T_n^* based on selected & re-centered moments.

$$T_n^*(\theta) = [\sqrt{n}\big(\frac{1}{n}\sum_{i=1}^n m(X_i^*, \theta) - \frac{1}{n}\sum_{i=1}^n m(X_i, \theta)\big) + \varphi_n(\theta)]_+' W(\theta)$$
$$\times [\sqrt{n}\big(\frac{1}{n}\sum_{i=1}^n m(X_i^*, \theta) - \frac{1}{n}\sum_{i=1}^n m(X_i, \theta)\big) + \varphi_n(\theta)]_+$$

Step 3: Let $c_n^*(\theta)$ be the $1-\alpha$ quantile of $T_n^*(\theta)$ and construct $CS_n = \{\theta \in \Theta : T_n(\theta) \leq c_n^*(\theta)\}$. This CS covers each $\theta \in \Theta_I(P)$ with probability $1-\alpha$ asymptotically.

Applications:

Ciliberto & Tamer (2009) study airline markets using a general version of Example 1. The profit function for airline k in market i is

$$\pi_k(s_i, x_i, u_i; \theta) = \{v_i' \alpha_k + z_{ik}' \beta_k + w_{ik}' \gamma_k + \sum_{j \neq k} \delta_j^k s_{ij} + \sum_{j \neq k} z_{ij}' \phi_j^k s_{ij} + u_{ik}\} s_{ik},$$

- v_i: market characteristics
- $z_i = (z_{i1}, \cdots, z_{iK})$: firm characteristics that enter the profits of all firms in the market
- $w_i = (w_{i1}, \cdots, w_{iK})$: w_{ik} enters only into firm k's profit in market i.

Let $x_i = (v_i, z_i', w_i')'$ be the vector of covariates.

The parameter vector θ includes $\beta_k, \gamma_k, \{\delta_j^k, \phi_j^k\}_{j \neq k}, k = 1, \cdots, K$. δ_j^k and ϕ_j^k capture the (fixed and variable) impacts of firm j's entry on firm k's profit. These impacts are called the strategic interaction effects.

Applications:

In general, the model does not uniquely determine the probability of observing a specific equilibrium outcome due to multiple equilibria (recall Example 1).

CT use (conditional) moment inequalities with singleton events: $A = \{\tilde{s}\}$

$$\nu_\theta(\{\tilde{s}\}|x_i) \leq P(s_i = \tilde{s}|x_i) \leq \nu_\theta^*(\{\tilde{s}\}|x_i), \ \ \tilde{s} \in \{0,1\}^K.$$

where

- $\nu_\theta(\{\tilde{s}\}|x_i)$ is the conditional probability of $\tilde{s}$ being a unique equilibrium. (Note: CT call this function H_1.)
- $\nu_\theta^*(\{\tilde{s}\}|x_i)$ is the conditional probability of $\tilde{s}$ being a unique equilibrium + the conditional probability of $\tilde{s}$ being an element of the set of equilibria (and being always selected).
  (Note: CT call this function H_2.)

Both ν_θ, ν_θ^* can be calculated from the model using simulation methods.
$P(s_i = \tilde{s}|x)$ can be estimated from data.

Applications:

A sample criterion function can be defined as

$$Q_n(\theta) \equiv \sum_{\tilde{s}\in\{0,1\}^K} w_{\tilde{s}}(\theta)(\frac{1}{n}\sum_{i=1}^{n}\nu_\theta(\tilde{s}|x_i) - \hat{P}_n(s_i = \tilde{s}|x_i))_+^2$$

$$+ \sum_{\tilde{s}\in\{0,1\}^K} w_{\tilde{s}}(\theta)\frac{1}{n}\sum_{i=1}^{n}(\hat{P}_n(s_i = \tilde{s}|x_i) - \nu_\theta^*(\tilde{s}|x_i))_+^2.$$

Application

- CT compute a confidence region and project it to each coordinate of the parameter.
- # of markets (trip between two major airports) was 2,742.
- Firms are American, Delta, United, Southwest, other medium airlines, and LCCs.

  CT find that the LCCs' impacts tend to be larger in magnitude (CI=[-19.623, -14.578]) relative to those of big firms such as American (CI=[-10.914, -8.822]).

Challenges & Recent Developments:

Tuning parameters:
$\rightarrow$ Andrews & Barwick (2012) provide suggestions on how the tuning parameter values for models with up to 10 inequalities. Epstein, Kaido, & Seo (2015) study a CLT that does not require tuning parameters for some incomplete models.

Computation:
$\rightarrow$ In some examples, one can exploit additional structures, e.g. convexity of the identified set: Beresteanu & Molinari (2008), Kaido & Santos (2014)

Inference on a single parameter or subvector:
$\rightarrow$ Sub-vector inference: Romano & Shaikh (2008), Bugni, Canay, & Shi (2014), Kaido, Molinari, Stoye (2015)

More general moment restrictions:
$\rightarrow$ Conditional moment inequalities and many moment inequalities: Andrews & Shi (2013), Chernozhukov, Lee, & Rosen (2013), Chernozhukov, Chetverikov, & Kato (2014), Menzel (2014)

Summary & Future Directions

The partial identification approach is a way to be agnostic about a part of an economic model. This provides a benchmark.

Alternative ways to study incomplete models

- Complete the model by making a particular assumption on the selection mechanism.
- Sensitivity analysis?
  e.g. parameterize the selection mechanism in such a way that one extreme is the complete model above (e.g. i.i.d. selection) and the other extreme is the fully agnostic model.
- Be agnostic about the selection mechanism.

The researcher can assess the sensitivity of the conclusion with respect to the assumption made on the selection mechanism. This can be done by filling the gap between the benchmark and a particular way to complete the model.

Summary & Future Directions

Again, Incomplete models arise in many empirical examples:

- Entries of airlines (Ciliberto & Tamer, 2009)
- English auctions (Haile & Tamer, 2003)
- Mergers in the banking industry (Uetake & Watanabe, 2012)
- Discrete choice under social interactions (Soetvent & Kooreman, 2007)
- Network formations (Sheng, 2014, Miyauchi, 2014)

There have been developments on both econometric theory and applications.

Further developments can be made through close interactions between empirical researchers and econometricians.

Thank you. Any comments are welcome.
A copy of the slides will be available from:
http://stat.econ.osaka-u.ac.jp/~suryo/

References I

D. W. K. Andrews and P. J. Barwick. Inference for parameters defined by moment inequalities: A recommended moment selection procedure. *Econometrica*, 80(6):2805–2826, 2012.

D. W. K. Andrews and X. Shi. Inference based on conditional moment inequalities. *Econometrica*, 81(2):609–666, 2013.

D. W. K. Andrews and G. Soares. Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, 78(1):119–157, 2010.

P. Bajari, H. Hong, and S. Ryan. Identification and estimation of a discrete game of complete information. *Econometrica*, 78(5):1529–1568, 2010.

A. Beresteanu, I. Molchanov, and F. Molinari. Sharp identification regions in models with convex moment predictions. *Econometrica*, 79(6):1785–1821, 2011.

References II

- T. Bresnahan and P. Reiss. Empirical models of discrete games. *Journal of Econometrics*, 48(1):57–81, 1991.
- F. A. Bugni. Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set. *Econometrica*, 78(2):735–753, 2010.
- F. A. Bugni, I. A. Canay, and X. Shi. Inference for functions of partially identified parameters in moment inequality models. Working Paper CWP22/14, CeMMAP, 2014.
- V. Chernozhukov, D. Chetverikov, and K. Kato. Gaussian approximation of suprema of empirical processes. 2014.
- V. Chernozhukov, H. Hong, and E. Tamer. Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, 75(5):1243–1284, 2007.

References III

- V. Chernozhukov, S. Lee, and A. Rosen. Intersection bounds: estimation and inference. *Econometrica*, 81(2):667–737, 2013.
- A. Chesher and A. M. Rosen. Simultaneous equations models for discrete outcomes: coherence, completeness, and identification. Working Paper CWP21/12, CeMMAP, 2012.
- F. Ciliberto and E. Tamer. Market structure and multiple equilibria in airline markets. *Econometrica*, 77(6):1791–1828, 2009.
- L. Epstein, H. Kaido, and K. Seo. Robust confidence regions for incomplete models. cemmap Working Papers, CWP20/15, 2015.
- A. Galichon and M. Henry. Set identification in models with multiple equilibria. *Review of Economic Studies*, 78:1264–1298, 2011.
- P. A. Haile and E. Tamer. Inference with an incomplete model of english auctions. *Journal of Political Economy*, 111(1), 2003.

References IV

- H. Kaido, F. Molinari, and J. Stoye. Inference for projections of identified sets, 2013. manuscript in preparation.
- H. Kaido and A. Santos. Asymptotically efficient estimation of models defined by convex moment inequalities. *Econometrica*, 82(1):387–413, 2014.
- C. F. Manski. *Partial Identification of Probability Distributions*. Springer Verlag, New York, 2003.
- K. Menzel. Consistent estimation with many moment inequalities. *Journal of Econometrics*, 182(2):329–350, 2014.
- Y. Miyauchi. Structural estimation of a pairwise stable network with nonnegative externality. 2014.
- I. Molchanov. *Theory of Random Sets*. Springer-Verlag, Berlin, 2005.

References V

- J. P. Romano and A. M. Shaikh. Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference*, 138:2786–2807, 2008.
- S. Sheng. A structural econometric analysis of network formation games. 2014.
- A. R. Soetevent and P. Kooreman. A discretechoice model with social interactions: with an application to high school teen behavior. *Journal of Applied Econometrics*, 22(3):599–624, 2007.
- E. Tamer. Incomplete simultaneous discrete response model with multiple equilibria. *Review of Economic Studies*, 70(1):147–165, 2003.
- K. Uetake and Y. Watanabe. Entry by merger: Estimates from a two-sided matching model with externalities. 2012.