Identification of Discrete Choice Models Using Moment Inequalities: Theory and Application

Eduardo Morales

Columbia University & Princeton University

April 2, 2012

Topics

- Mapping between statistical and behavioral discrete choice models.
  - Pakes (2010); Pakes et al. (2011); Dickstein and Morales (2012).
- Definition and computation of alternative moment inequalities estimators.
  - Chernozhukov, Hong and Tamer (2007); Andrews and Soares (2010); Pakes et al. (2011).

MAPPING BETWEEN STATISTICAL AND BEHAVIORAL MODELS

Discrete Choice Problem

- Utility of agent i for alternative j is:

$$U_{ij} = \beta x_{ij}^* + \nu_{ij}, \qquad j = 1, \ldots, J, \quad x_{ij}^* \in \mathcal{X} \in R^2.$$

- Choice of individual i is captured in d_i and

$$d_{ij} = \mathbb{1}[U_{ij} \geq U_{ij'}, \text{ for } j' = 1 \ldots, J].$$

- We observe a random sample of vectors

$$(d_i, x_i, z_i) \sim \mathcal{P}, \qquad i = 1, \ldots, N,$$

  with $x_{i1} = x_{i1}^* + \epsilon_{i1}^x$, $z_{i1} = x_{i1}^* + \epsilon_{i1}^z$, $x_{i2} = x_{i2}^*$.
- Assumptions on $(\nu_i, \epsilon_{i1}^x, \epsilon_{i1}^z)$:
  - structural error: $\nu_i \sim F_\nu(\nu_i | x_i^*, \epsilon_{i1}^x, \epsilon_{i1}^z) = F_\nu(\nu_i | x_i^*)$
  - measurement error: $\epsilon_{i1}^x \sim F_{\epsilon^x}(\epsilon_{i1}^x | x_i^*, \epsilon_{i1}^z, \nu_i) = F_{\epsilon^x}(\epsilon_{i1}^x | x_i^*)$.
- Note: we omit subindex i from here onwards.

Example I: Estimation of Structural Parameters

Consumer Choice Problem

- Each individual maximizes a utility function:

$$U_j = \beta_1 \mathbb{E}[x_j|\mathcal{J}] + \beta_2 p_j + \nu_j, \qquad j = 1, \ldots, J.$$

- Instead of agent's i subjective expectations, the econometrician observes both the realized values,

$$x_j = \mathbb{E}[x_j|\mathcal{J}] + \epsilon_j^x,\ \mathbb{E}[\epsilon_j^x|\mathcal{J}] = 0,$$

  and an additional shifter of agent's i expectations:

$$z_j = \mathbb{E}[x_j|\mathcal{J}] + \epsilon_j^z.$$

- The expectational error corresponds to the classical measurement error in an explanatory variable (errors-in-variables).

Example II: Estimation of Structural Parameters

Two-period Entry Problem

- Each firm decides where to locate a plant:

$$U_j = \mathbb{E}[R_j|\mathcal{J}] + \beta_2 F_j + \nu_j, \qquad j = 1, \ldots, J.$$

- Instead of agent's i subjective expectations, the econometrician observes realized values:

$$R_j = \mathbb{E}[R_j|\mathcal{J}] + \epsilon_j^R.$$

  The econometrician also observes Z such that

$$Z_j = \mathbb{E}[R_j|\mathcal{J}] + \epsilon_j^Z,$$

- Similar to previous example but with structural restriction: $\beta_1 = 1$.
- Standard application of the moment inequalities estimator:
  - Ishii (2008), Ho (2009), Holmes (2011), Morales et al. (2011).

Example III: Estimation of Reduced Form Parameters

General Entry Problem

- Each firm decides where to locate a plant. Period t profits of locating plant in location j are:

$$\pi_{jt} = R_{jt} - \sum_{j'=1}^{J} F_{jj'} d_{ij' t-1}$$

- The expected present value of locating plant in location j is:

$$\begin{aligned} U_{jt} &= \mathbb{E}[\pi_{jt} + \sum_{s=t+1}^{\infty} \delta^{s-t} d_{js} \pi_{js} | \mathcal{J}_t, d_{jt} = 1], \\ &= V_{jt} - \sum_{j'=1}^{J} F_{jj'} d_{j' t-1}, \end{aligned}$$

with

$$V_{jt} = \pi_{jt} + \mathbb{E}[\sum_{s=t+1}^{\infty} \delta^{s-t} d_{js} \pi_{js} | \mathcal{J}_t, d_{jt} = 1].$$

Example III: Estimation of Reduced Form Parameters

General Entry Problem (cont.)

- Assume a projection of V_{jt} on a set of observable covariates:

$$V_{jt} = \beta x_{jt} + \nu_{jt}, \qquad (x_t, \nu_t) \in \mathcal{J}_{it}.$$

- Assume a structural form for $F_{jj'}$:

$$F_{jj'} = \theta \| L_j - L_{j'} \|,$$

with $\| \cdot \|$ some measure of distance, and L_j an indicator of location j.

- Therefore:

$$U_{jt} = \beta x_{jt} + \theta \sum_{j'=1}^{J} \| L_j - L_{j'} \| d_{j't-1} + \nu_{jt}.$$

- Additionally, we can allow for measurement error in x_{jt}.

Maximum Likelihood Estimation

- Model 1:
  - Assume (up to a finite parameter vector) the distributions:

$$\{F_\nu(\nu|x^*), F_\epsilon(\epsilon|x^*), \mathcal{P}_{x^*}(x^*)\}$$

  - The individual i likelihood function is:

$$\begin{aligned}\mathcal{L}(d|x,z) &= \mathbb{P}(d_j = 1|x,z)\\ &= \mathbb{E}(\mathbb{1}\{d_j = 1\}|x,z)\\ &= \mathbb{E}(\mathbb{E}(\mathbb{1}\{d_j = 1\}|x,z,x^*)|x,z)\\ &= \mathbb{E}(\mathbb{E}(\mathbb{1}\{d_j = 1\}|x^*)|x,z)\\ &= \int_x \left[\int_\nu \mathbb{1}\{U_j \geq \max_{j' \in J}\{U_{j'}\}\} dF_\nu(\nu|x^*)\right] dF_{x^*}(x^*|x,z),\end{aligned}$$

with

$$U_j = \beta x_j^* + \nu_j.$$

Maximum Likelihood Estimation

- Model 2:
  - Assume (up to a finite parameter vector) the distributions:

$$\{F_\nu(\nu|x^*), F_\epsilon(\epsilon^x|x^*), \mathcal{P}_{x^*}(x^*)\}$$

    such that:

$$F_\nu(\nu|x^*) = F_\nu(\nu), \quad \text{and} \quad F_\epsilon(\epsilon|x) = F_\epsilon(\epsilon)$$

  - The individual i likelihood function is:

$$\begin{aligned}
\mathcal{L}(d|x) &= \mathbb{P}(d_j = 1|x) \\
&= \mathbb{E}(\mathbb{1}\{d_j = 1\}|x) \\
&= \int_{\nu+\epsilon} \mathbb{1}\{U_j \geq \max_{j' \in J}\{U_{j'}\}\} dF_{\nu+\epsilon}(\nu + \epsilon^x),
\end{aligned}$$

    with

$$U_j = \beta x_j + \nu_j - \beta \epsilon_j^x.$$

Maximum Likelihood Estimation

- Model 3:
  - Assume (up to a finite parameter vector) the distribution
    $$F_\nu(\nu|x^*),$$
    and assume
    $$x = x^*.$$
  - The individual i likelihood function is:
    $$\begin{aligned}\mathcal{L}(d|x) &= \mathbb{P}(d_j = 1|x)\\ &= \mathbb{E}(\mathbb{1}\{d_j = 1\}|x)\\ &= \mathbb{E}(\mathbb{1}\{d_j = 1\}|x^*)\\ &= \int_\nu \mathbb{1}\{U_j \geq \max_{j' \in J}\{U_{j'}\}\} dF_\nu(\nu|x^*),\end{aligned}$$
    with
    $$U_j = \beta x_j^* + \nu_j.$$

Summary of MLE

- In words, dealing with both structural and measurement error in MLE requires:
  - Assuming both the marginal distribution of the unobserved true covariates, and the distribution of both measurement and structural error conditional on these covariates (Model 1); or,
  - Assuming the structural error is independent of the true covariates, and the measurement error is independent of the *observed* covariates (Model 2):
    - Only in this case the difference between structural and measurement error is irrelevant!. We can think of a single error, η, such that:

      $$\eta = \nu + \epsilon$$

      and assume a single distribution $F_\eta(\eta)$; or,
  - Assuming the distribution of the structural error conditional on the true covariates, and assuming that these are measured without error (Model 3).

Moment Inequalities

Introduction

- Given our discrete choice problem, we can derive conditional moment inequalities. A conditional moment inequality is:

$$\mathbb{E}[m(x,z,j,j';\beta_0)|x,z] \geq 0, \text{ where } d_j = 1, \text{ and } j' \neq j.$$

- For simplicity, in these slides we base identification on a set of unconditional moment inequalities derived for each possible value of (x,z,j,j'):

$$\mathbb{E}[m_s(x,z,j,j';\beta_0)] \geq 0, \quad s = 1,\ldots,S.$$

- Moment inequalities will generically lead to set identification. The identified set is:

$$\{\beta \in \mathcal{B} : \int \sum_{s\in S} \sum_{j\in J} \sum_{j'\neq j} (\min\{m_s(x,z,j,j';\beta),0\}^2) d\mathcal{P}(x,z) = 0\}$$

- We denote the identified set as $\mathcal{B}_{\mathcal{M}}(\mathcal{P})$.

Moment Inequalities

Deriving Moment Inequalities from our Discrete Choice Problem: Model 1

- Assumptions:

$$(\epsilon_j^x, \nu_j) = (0,0), \quad \text{for every } j \in J.$$

- Moment inequalities.

$$\mathbb{E}[m_s(x,z,j,j';\beta_0)|x,z] = \mathbb{E}[\mathbb{1}\{d_j = 1\}\chi_s(\Delta x_{jj'})\beta_0 \Delta x_{jj'}|x] \geq 0,$$

where $\chi_s(\Delta x_{jj'})$ has the form:

$$\chi_1(\Delta x_{jj'}) = \mathbb{1}\{\Delta x_{1jj'} \geq 0\}\mathbb{1}\{\Delta x_{2jj'} \geq 0\} \tag{1a}$$

$$\chi_2(\Delta x_{jj'}) = \mathbb{1}\{\Delta x_{1jj'} \geq 0\}\mathbb{1}\{\Delta x_{2jj'} < 0\} \tag{1b}$$

$$\chi_3(\Delta x_{jj'}) = \mathbb{1}\{\Delta x_{1jj'} < 0\}\mathbb{1}\{\Delta x_{2jj'} \geq 0\} \tag{1c}$$

$$\chi_4(\Delta x_{jj'}) = \mathbb{1}\{\Delta x_{1jj'} < 0\}\mathbb{1}\{\Delta x_{2jj'} < 0\}. \tag{1d}$$

- Normalization by scale: (1) $\beta_{01} = 1$; (2) $\|\beta_0\| = 1$.
- Completely deterministic model; very likely rejected by the data.

Moment Inequalities

Deriving Moment Inequalities from our Discrete Choice Problem: Model 2

- Assumptions:
  - No structural error: $\nu_j = 0$, for every $j \in J$;
  - Measurement error indep. of true covariate: $F_\epsilon[\epsilon_1^x | x^*] = F_\epsilon[\epsilon_1^x]$
    - No parametric assumption on $F_\epsilon[\epsilon_1^x]$ needed.
  - Additional indicator of the covariate measured with error: z_1.
- Moment inequalities.

$$\begin{aligned}
\mathbb{E}[\mathbb{1}\{d_j = 1\}\chi_s(\Delta z_{1jj'}, \Delta x_{2jj'})\beta_0 \Delta x^*_{jj'} | x, z] &\geq 0, \\
\mathbb{E}[\mathbb{1}\{d_j = 1\}\chi_s(\Delta z_{1jj'}, \Delta x_{2jj'})(\beta_0 \Delta x_{jj'} - \beta_0 \Delta \epsilon^x_{1jj'}) | x, z] &\geq 0, \\
\mathbb{E}[\mathbb{1}\{d_j = 1\}\chi_s(\Delta z_{1jj'}, \Delta x_{2jj'})\beta_0 \Delta x_{jj'} | x, z] &\geq 0,
\end{aligned}$$

with

$$\begin{aligned}
\chi_1(\Delta z_{1jj'}, \Delta x_{2jj'}) &= \mathbb{1}\{\Delta z_{1jj'} \geq 0\}\mathbb{1}\{\Delta x_{2jj'} \geq 0\}, \\
\chi_2(\Delta z_{1jj'}, \Delta x_{2jj'}) &= \mathbb{1}\{\Delta z_{1jj'} \geq 0\}\mathbb{1}\{\Delta x_{2jj'} < 0\}, \\
\chi_3(\Delta z_{1jj'}, \Delta x_{2jj'}) &= \mathbb{1}\{\Delta z_{1jj'} < 0\}\mathbb{1}\{\Delta x_{2jj'} \geq 0\}, \\
\chi_4(\Delta z_{1jj'}, \Delta x_{2jj'}) &= \mathbb{1}\{\Delta z_{1jj'} < 0\}\mathbb{1}\{\Delta x_{2jj'} < 0\}.
\end{aligned}$$

Moment Inequalities

Deriving Moment Inequalities from our Discrete Choice Problem: Model 3

- Assumptions:
  - No structural error: $\nu_j = 0$, for every $j \in J$;
  - Measurement error indep. of true covariate: $F_\epsilon[\epsilon_1^x | x^*] = F_\epsilon[\epsilon_1^x]$;
    - No parametric assumption on $F_\epsilon[\epsilon_1^x]$ needed.
  - Structural restriction: $\beta_{01} = 1$.
- Moment inequalities.

$$\mathbb{E}[\mathbb{1}\{d_j = 1\}\chi_s(\Delta x_{2jj'})\beta_0 \Delta x^*_{jj'} | x] \geq 0,$$
$$\mathbb{E}[\mathbb{1}\{d_j = 1\}\chi_s(\Delta x_{2jj'})(\beta_0 \Delta x_{jj'} - \beta_0 \Delta \epsilon^x_{1jj'}) | x] \geq 0,$$
$$\mathbb{E}[\mathbb{1}\{d_j = 1\}\chi_s(\Delta x_{2jj'})\beta_0 \Delta x_{jj'}] \geq 0,$$

with

$$\chi_1(\Delta x_{2jj'}) = \mathbb{1}\{\Delta x_{2jj'} \geq 0\},$$
$$\chi_2(\Delta x_{2jj'}) = \mathbb{1}\{\Delta x_{2jj'} < 0\}$$

- No need to normalize by scale.

Moment Inequalities

Deriving Moment Inequalities from our Discrete Choice Problem: Model 4

- Assumptions:
  - Structural error organized in nests:

$$\nu_j - \nu_{j'} = \begin{cases} 0 & \text{if } G(j) = G(j'), \\ \mathbb{R} & \text{if } G(j) \neq G(j'), \end{cases}$$

    where $G(j)$ denotes a particular subset of J.
    - No parametric assumption on $F_\nu[\nu|x]$ needed.
  - No measurement error.
- Moment inequalities.

$$\mathbb{E}[\mathbb{1}\{d_j = 1\}\mathbb{1}\{G(j) = G(j')\}\chi_s(\Delta x_{jj'})(\beta_0 \Delta x_{jj'} + \Delta\nu_{jj'})|x] \geq 0,$$
$$\mathbb{E}[\mathbb{1}\{d_j = 1\}\mathbb{1}\{G(j) = G(j')\}\chi_s(\Delta x_{jj'})\beta_0 \Delta x_{jj'}|x] \geq 0,$$

- Trivial to allow for measurement error indep. of the true covariate.
  - use an additional indicator z_1 (as in Model 2); or,
  - impose $\beta_{01} = 1$ as a structural assumption (as in Model 3).

Moment Inequalities

Deriving Moment Inequalities from our Discrete Choice Problem: Model 5

- Assumptions:
  - Ordered choice model:

$$\nu_j = j\eta, \quad \text{and} \quad F_\eta[\eta|x] = F_\eta[\eta],$$

    with $\mathbb{E}[\eta] = 0$.
    - No parametric assumption on $F_\eta[\eta|x]$ needed.
  - No measurement error.
- Moment inequalities.

$$\mathbb{E}[\mathbb{1}\{d_j = 1\}\mathbb{1}\{j - j' = 1\}\chi_s(\Delta x_{jj'})(\beta_0 \Delta x_{jj'} + \Delta\nu_{jj'})|x] \geq 0,$$
$$\mathbb{E}[\mathbb{1}\{d_j = 1\}\mathbb{1}\{j - j' = 1\}\chi_s(\Delta x_{jj'})(\beta_0 \Delta x_{jj'} + \eta)|x] \geq 0,$$
$$\mathbb{E}[\mathbb{1}\{d_j = 1\}\mathbb{1}\{j - j' = 1\}\chi_s(\Delta x_{jj'})\beta_0 \Delta x_{jj'}|x] \geq 0.$$

- Trivial to allow for measurement error indep. of the true covariate.
  - use an "instrument" z_1, and assume $F_\eta[\eta|x^*, \epsilon_1^z] = F_\eta[\eta]$;
  - impose $\beta_{01} = 1$, and assume $F_\eta[\eta|x^*] = F_\eta[\eta]$.

Moment Inequalities

Deriving Moment Inequalities from our Discrete Choice Problem: Additional Models

See Dickstein and Morales (2012) (*to be posted soon*) for guidance on how to build moment inequalities for the following models:

- Model 6:
  - Assume a parametric distribution for ν_i: $F_\nu[\nu|x;\sigma]$
  - Allow for distribution free measurement error in two cases:
    - multiple indicator assumption;
    - fixed parameter on the variable measured with error.
- Model 7:
  - Assume ν_i is distributed independently of x.
    - No parametric assumption on $F_\nu[\nu|x]$ needed.
  - Allow for distribution free measurement error in two cases:
    - multiple indicator assumption;
    - fixed parameter on the variable measured with error.

Summary of Moment Inequalities

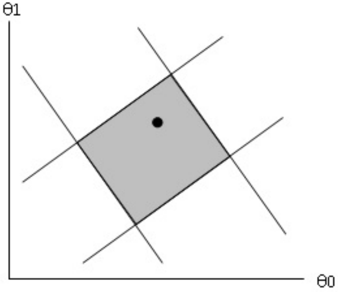
- In words, dealing with both structural and measurement error in Moment Inequalities requires:
  - In order to deal with measurement error, we may apply any of the usual IV solutions to measurement error problems in linear regression models. In particular, no parametric assumption on its distribution function is needed.
  - In order to deal with structural error, one needs to:
    - Assume it away; or,
    - Assume that it is an individual effect common to a subset of choices; or,
    - Assume a parametric distribution on it; or,
    - Assume that it is distributed independently of x (no parametric assumption needed), or,
    - In the special case of an ordered choice model, assume that it increases in an individual specific constant as the choice gets bigger.

ALTERNATIVE MOMENT INEQUALITIES ESTIMATORS: DEFINITION AND COMPUTATION

Identified Set

- Moment inequalities will generically lead to set identification. Given a set S of moment inequalities, the identified set is:

$$\Theta^S = \underset{\theta}{\text{argmin}} \sum_{s=1}^{S} \left(\min \left\{ 0, \mathbb{E}[m_s(Y, X, Z; \theta)] \right\} \right)^2$$



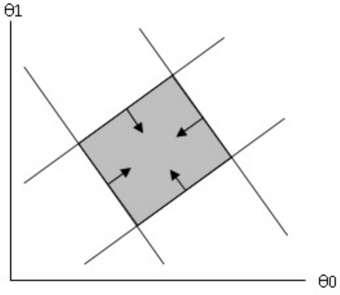
Steps for Estimation

- Step 1: Estimate the identified set given sample moments.
- Step 2: Perform inference on one or more of the following parameters:
  - Interval contained in the identified set: Pakes, Porter, Ho and Ishii (2011).
  - Identified set: Chernozhukov, Hong and Tamer (Econometrica, 2007).
  - True parameter vector: Andrews and Soares (Econometrica, 2010).

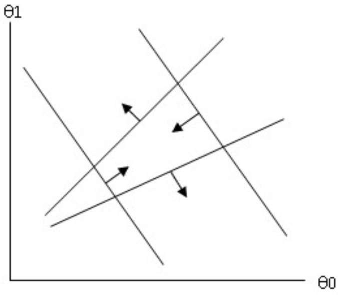
Estimation of the Identified Set

- Estimation is based on the sample analogue of the moment inequalities:

$$\overline{m}_{I,s}(\theta) = \frac{1}{I}\sum_{i=1}^{I} m_s(Y_i, X_i, Z_i; \theta)$$



(c) Case 1



(d) Case 2

Estimation of the identified set

- Two possible criterion functions to define the estimated set:
  - Unweighted criterion function:

$$\hat{\Theta}_I^S = \underset{\theta}{\operatorname{argmin}} \sum_{s=1}^{S} \Big(\min\{0, \overline{m}_{I,s}(\theta)\} \Big)^2$$

  - Weighted criterion function:

$$\hat{\Theta}_I^S = \underset{\theta}{\operatorname{argmin}} \sum_{s=1}^{S} \Big(\min\{0, \Big[\frac{\overline{m}_{I,s}(\theta)}{\hat{\sigma}^2_{I,s}(\theta)}\Big]\} \Big)^2,$$

  with

$$\hat{\sigma}^2_{I,s}(\theta) = \frac{1}{I} \sum_{i=1}^{I} (m_s(Y_i, X_i, Z_i; \theta) - \overline{m}_{I,s}(\theta))^2$$

- They both generate the same estimated set in Case 1. They may generate different estimated sets in Case 2. The weighting lessens the influence of sample moments that have high variance (likely to be further away from their population analogues).

Computation of the estimated set

- We characterize the set $\hat{\Theta}_I^S$ by finding its boundaries along any linear combination of the dimensions of vector θ.
- If the moment functions $\{\overline{m}_{I,s}(\theta) : s = 1, \ldots, S\}$ are linear in θ, use linear programming to find the extremum

$$\begin{aligned} &\max_{\theta} \quad f \cdot \theta \\ &\text{s.t.} \\ &\overline{m}_{I,s}(\theta) \geq 0, \ \text{for } s = 1, ..., S. \end{aligned} \tag{2}$$

- If we want to find the maximum and minimum of our two-dimensional parameter θ, we use:

$$f = \{[1,0], [-1,0], [0,1], [0,-1]\}.$$

Computation of the estimated set

- If there is no value of θ that verifies all the constraints, $\hat{\Theta}_I^S$ will then be a singleton.
- This singleton is the outcome of the following nonlinear optimization problem:
$$\hat{\Theta}_I^S = \underset{\theta}{\operatorname{argmin}} \sum_{s=1}^{S} \Big(\min\{0, \Big[\frac{\overline{m}_{I,s}(\theta)}{\hat{\sigma}_{I,s}^2(\theta)}\Big]\}\Big)^2.$$
- It is suggested to use the *KNITRO* nonlinear optimzation package from Matlab via *ktrlink*.

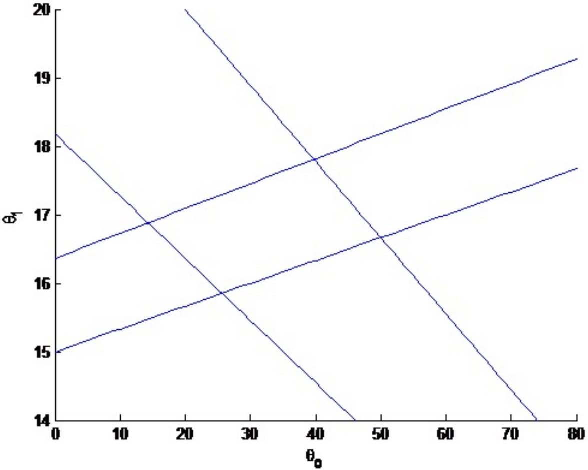
Computation of the estimated set: example

- Sample moments:

$$
\begin{aligned}
900 - \theta_0(-2) - \theta_1(60) &\geq 0 \\
-900 - \theta_0(2) - \theta_1(-55) &\geq 0 \\
200 - \theta_0(1) - \theta_1(9) &\leq 0 \\
-200 - \theta_0(-1) - \theta_1(-11) &\leq 0
\end{aligned}
$$

Computation of the estimated set: example

Vertex, θ0 min		Vertex, θ0 max		Vertex, θ1 min		Vertex, θ1 max	
θ0	θ1	θ0	θ1	θ0	θ1	θ0	θ1
13.0	16.8	51.0	16.6	23.9	15.8	41.3	17.8



Inference: General Intuition

- Consider we want to test the null hypothesis: $H_0 : \theta = \theta_0$.
- We use the following statistic:

$$T_I(\theta_0) = \sum_{s=1}^{S} \Big(\min\{0, \Big[\frac{\overline{m}_{I,s}(\theta_0)}{\hat{\sigma}^2_{I,s}(\theta_0)}\Big]\} \Big)^2.$$

- The finite-sample null distribution of $T_I(\theta_0)$ depends on the degree of *slackness* of the moment inequalities (the population moments). That is, it depends on how much greater than 0 is

$$\mathbb{E}[m_s(Y_i, X_i, Z_i; \theta)], \quad \text{for } s = 1, \dots, S.$$

- The distribution of $T_I(\theta_0)$ is stochastically larger the more moments are equal to 0 at θ_0 (the population moments).
- As the distribution of $T_I(\theta_0)$ gets stochastically larger, the probability of rejecting H_0 decreases.

Inference: General Intuition

- Therefore, it is key to infer whether a population moment binds at a particular value θ_0. This inference is based on slackness factors computed for each moment: $SF_{I,s}(\theta_0)$.
- Three possible slackness factors proposed in the literature:
  - Assume that all the S moments are binding at θ_0: $SF_{I,s} = 0$. This yields the most conservative test.
  - Moment Selection: assume a moment is asymptotically binding if the sample analogue takes a value larger than a threshold.

$$SF_{I,s}^{MS}(\theta_0) = \mathbb{1}\{\sqrt{n}(\frac{\overline{m}_{I,s}(\theta_0)}{\hat{\sigma}_{I,s}(\theta_0)}) \leq \sqrt{2\ln(\ln(n))}\}$$

  - Shifted Mean: shift each moment proportionately to how far away from binding it is in the sample.

$$SF_{I,s}^{SM}(\theta_0) = (\frac{\overline{m}_{I,s}(\theta_0)}{\hat{\sigma}_{I,s}(\theta_0)})(\frac{1}{\sqrt{2\ln(\ln(n))}})\mathbb{1}\{\frac{\overline{m}_{I,s}(\theta_0)}{\hat{\sigma}_{I,s}(\theta_0)} > 0\}$$

Inference for an Interval

- This procedure is based on Pakes, Porter, Ho, and Ishii (2011).
- Objective: build confidence intervals for the vertices of the estimated set, and use the outer bounds to form a unique confidence interval.
- We need four elements for inference:
  - Vertices of the estimated set.
    - We collect these from the computation of the estimated set.
  - Approximation to the asymptotic distribution of all the (weighted) moments recentered at zero.
  - Jacobian of the moments.
  - Slackness factors.
- Once we have this four elements, we perform an optimization very close to that one in equation (**??**).

Est.Set

Inference for an Interval

- Approximation to asymptotic distribution of all the recentered moments.
  - As long as the CLT applies, we just need to draw $r = 1, ..., R$ times from a multivariate normal with zero mean, and covariance equal to the variance of the weighted moments
  - How to do this?
    - Take R standard normal draws.
    - Premultiply each draw by the Cholesky decomposition of the correlation matrix evaluated at the vertex of interest, $\widehat{\Omega}_{I,S}(\hat{\theta})$:

$$\widehat{\Omega}_{I,S}(\hat{\theta}) = diag(\widehat{\Sigma}_{I,S}(\hat{\theta}))^{-\frac{1}{2}}\widehat{\Sigma}_{I,S}(\hat{\theta})diag(\widehat{\Sigma}_{I,S}(\hat{\theta}))^{-\frac{1}{2}}.$$

    - Result:

$$q_r(\hat{\theta}) = chol(\widehat{\Omega}_n(\hat{\theta}))N(0_S, I_S).$$

Inference for an Interval

- Jacobian of the moments.
  - Compute the Jacobian of the unweighted moments, $\overline{m}_{I,s}(\theta)$, and evaluate the result at the vertex of interest:

$$\nabla_\theta \overline{m}_{I,s}(\theta)|_{\hat{\theta}} = \nabla_\theta \left\{ \frac{1}{I} \sum_{i=1}^{I} \sum_{j \in J} \sum_{j' \neq j} \mathbb{1}\{Y_i = j\} h_s(Z_i)(g(X_j;\theta) - g(X_{j'};\theta)) \right\}\Big|_{\hat{\theta}}$$

  - Reweight the elements of the Jacobian using the diagonal elements of the covariance matrix of the unweighted moments, evaluated at the vertex of interest:

$$\widehat{\Gamma}_{I,S}(\hat{\theta}) = diag(\widehat{\Sigma}_{I,S}(\hat{\theta}))^{-\frac{1}{2}} \left(\nabla_\theta \overline{m}_{I,S}(\theta)|_{\hat{\theta}} \right)'$$

Inference for an Interval

- Evaluate the slackness factor at the vertex of interest and normalize by $\sqrt{I}$.
  - We could use either $SF^{MS}_{I,s}$ or $SF^{SM}_{I,s}$.
    The option described in Pakes, Porter, Ho, and Ishii (2011) is Shifted Mean:

$$SF^{SM}_{I,s}(\hat{\theta})\sqrt{I} = \left(\frac{\overline{m}_{I,s}(\hat{\theta})}{\hat{\sigma}_{I,s}(\hat{\theta})}\right)\left(\frac{1}{\sqrt{2\ln(\ln(n))}}\right)\mathbb{1}\left\{\frac{\overline{m}_{I,s}(\hat{\theta})}{\hat{\sigma}_{I,s}(\hat{\theta})} > 0\right\}\sqrt{I}$$

Inference for an Interval

- Compute the following linear programing problem for each draw r and each vertex $\hat{\theta}$:

$$
\begin{aligned}
\theta_r = \max_{\theta} \quad & f \cdot \sqrt{I}(\hat{\theta} - \theta) \\
\text{s.t.} & \\
& \widehat{\Gamma}_{I,S}(\hat{\theta})\sqrt{I}(\hat{\theta} - \theta) + q_r(\hat{\theta}) + SF^{SM}_{I,S}(\hat{\theta})\sqrt{I} \geq 0
\end{aligned}
\tag{3}
$$

- As before, if we want to find the maximum and minimum of our two-dimensional parameter θ, we use:

$$f = \{[1,0], [-1,0], [0,1], [0,-1]\}.$$

In equation (**??**), we should be careful to use the estimated vertex $\hat{\theta}$ that corresponds to each vector f.

- We obtain R draws of the asymptotic distribution of each of the estimated vertices of the estimated set.

Inference for an Interval

- For each pair of vertices corresponding to a given dimension d of θ.
  - For the min vertex, take the $\alpha/2$ quantile of the set of simulated vertices, θ_r, $r = 1, \ldots, R$. Denote this number:

$$\underline{\theta}_{d,\alpha/2}.$$

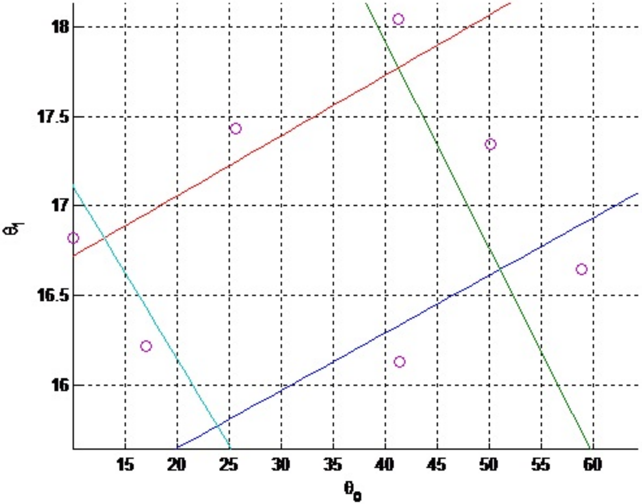
  - For the max vertex, take the $(1 - \alpha/2)$ quantile of the set of simulated vertices, θ_r, $r = 1, \ldots, R$

$$\overline{\theta}_{d,1-\alpha/2}.$$

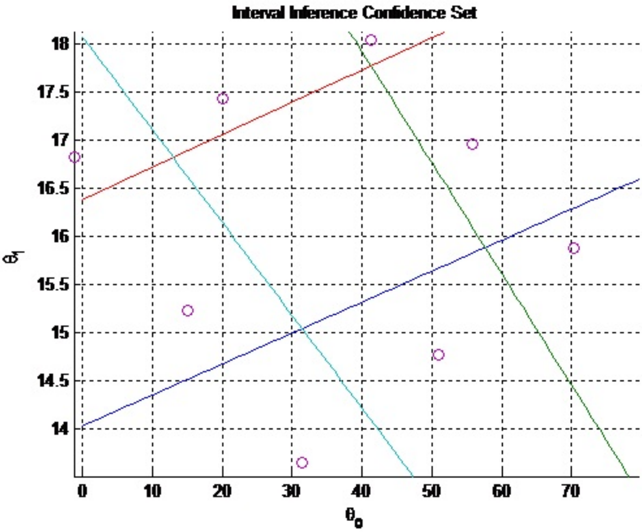
- The confidence interval for θ in the dimension d with significance level α is:

$$(\underline{\theta}_{d,\alpha/2}, \overline{\theta}_{d,\alpha/2}).$$

Inference for an Interval: Example



Inference for an Interval: Example

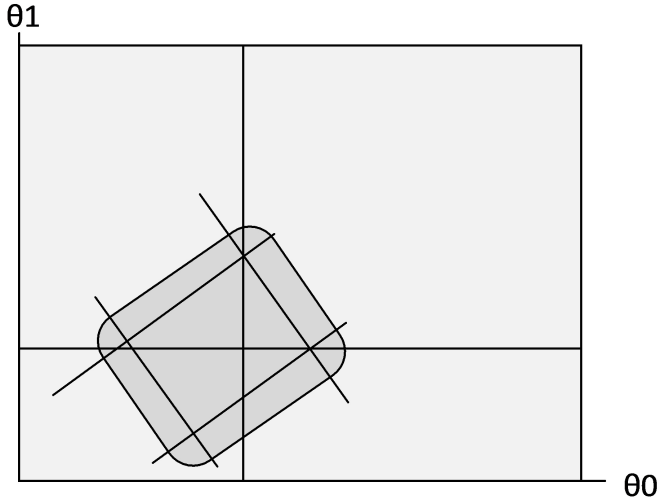


Set/Point Inference: General Intuition

- The estimation of confidence intervals for the true parameter and for the identified set are based on the inversion of an Anderson-Rubin T statistic.
- General steps in the algorithm:
  1. Define θ grids, $\widehat{\Theta}_I^{Grid}$ and $\widehat{\Theta}_I^{\epsilon}$, where $\widehat{\Theta}_I^{\epsilon} \subset \widehat{\Theta}_I^{Grid}$.
  2. Calculate $T_r(\theta)$, at a set of points in either $\widehat{\Theta}_I^{Grid}$ or $\widehat{\Theta}_I^{\epsilon}$ depending on whether the focus of inference is the identified set or the true value of the parameter.
  3. Determine a critical value as a quantile of $T_r(\theta)$ for $r = 1, ..., R$
  4. Calculate $T^{obs}(\theta)$ at each $\theta \in \widehat{\Theta}_I^{Grid}$ with the observed data for all moments.
  5. Define the confidence set as those θ points where $T^{obs}(\theta)$ falls below the critical value.

Forming the Grids: $\widehat{\Theta}_I^{Grid}$ and $\widehat{\Theta}_I^{\epsilon}$

$\widehat{\Theta}_I^{\epsilon} \subset \widehat{\Theta}_I^{Grid}$



Inference for the Identified Set

- This procedure is based on Chernozhukov, Hong and Tamer (Econometrica, 2007).
- Steps of the procedure:
  - (1) At $\theta \in \widehat{\Theta}_I^{\varepsilon}$, compute R draws $\{q^r(\theta); r = 1, \ldots, R\}$ such that:

    $$q_r(\theta) = chol(\widehat{\Omega}_{I,S}(\theta))N(0_S, I_S),$$

    with

    $$\widehat{\Omega}_{I,S}(\hat{\theta}) = diag(\widehat{\Sigma}_{I,S}(\hat{\theta}))^{-\frac{1}{2}}\widehat{\Sigma}_{I,S}(\hat{\theta})diag(\widehat{\Sigma}_{I,S}(\hat{\theta}))^{-\frac{1}{2}}.$$

    Note that we are taking draws from the asymptotic distribution of the normalized recentered moments, evaluated at each point θ.

Inference for the Identified Set

- Steps of the procedure (cont.)
  - (2) Compute one of the following T-statistic for each value of θ and draw r:

$$T_r^N(\theta) = \sum_{s=1}^{S} (\min\{0, q_{r,s}(\theta)\})^2$$

$$T_r^{MS}(\theta) = \sum_{s=1}^{S} \{(\min\{0, q_{r,s}(\theta)\})^2 \times SF_{I,s}^{MS}(\theta)\}$$

$$T_r^{SM}(\theta) = \sum_{s=1}^{S} \{(\min\{0, q_{r,s}(\theta)\})^2 + SF_{I,s}^{SM}(\theta)\}$$

  - (3) For each draw r, take the maximum across θ:

$$T_r^{\max} = \max_{\theta \in \widehat{\Theta}_n^{\varepsilon}} T_r^k(\theta), \quad k = \{N, MS, SM\}.$$

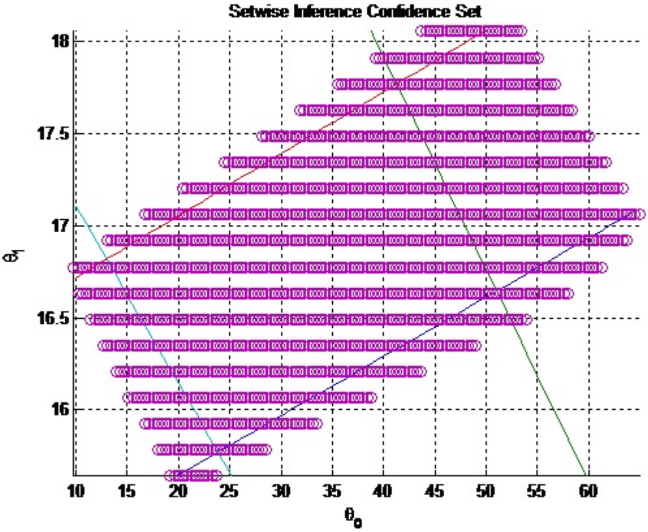
Inference for the Identified Set

- Steps of the procedure (cont.)
  - (4) Compute the critical value c_α as the $1-\alpha$ quantile of the distribution of $\{T_r^{\max}; r = 1, \ldots, R\}$.
  - (5) Return to the larger grid of theta points, $\widehat{\Theta}_n^{Grid}$, and calculate $T^{obs}(\theta)$ at each candidate value $\theta \in \widehat{\Theta}_n^{Grid}$:

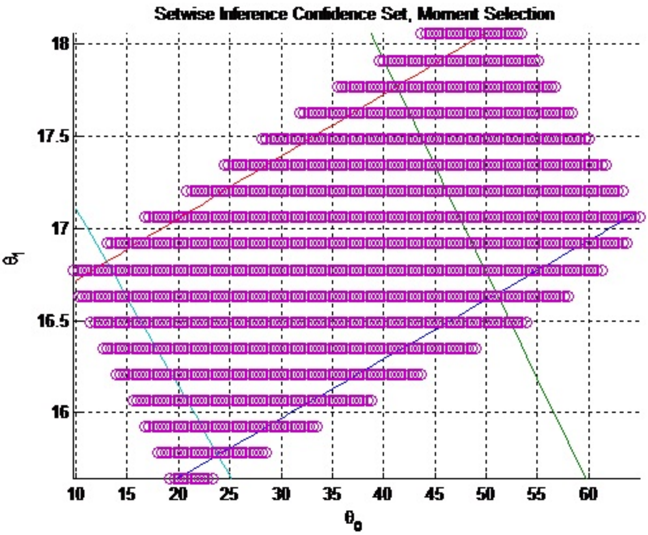
$$T^{obs}(\theta) = \sum_{s=1}^{S} (\min\{0, \frac{\overline{m}_{I,s}(\theta)}{\hat{\sigma}_{I,s}(\theta)}\})^2$$

  - (6) Compare $T^{obs}(\theta)$ against c_α and accept θ into the confidence set whenever $T^{obs}(\theta) < c_\alpha$.

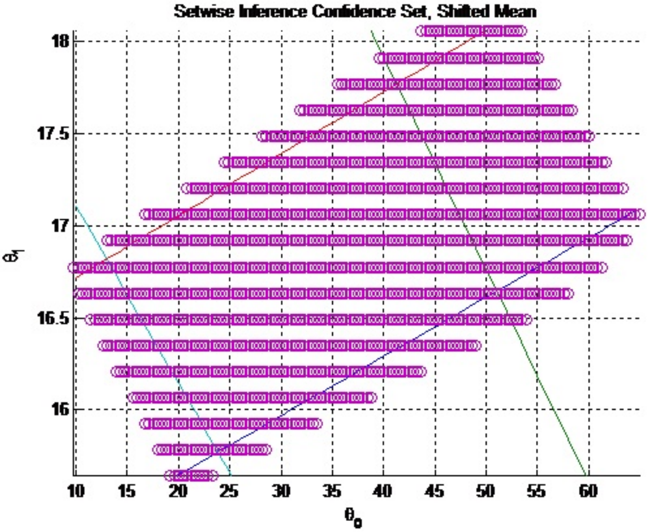
Inference for the Identified Set: Example



Inference for the Identified Set: Example



Inference for the Identified Set: Example



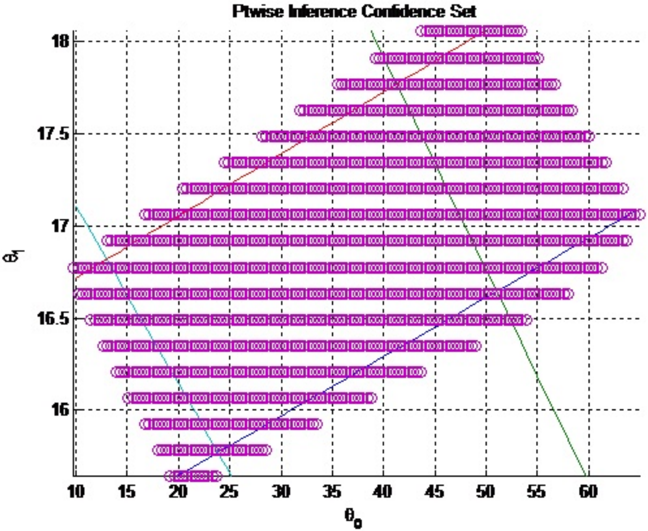
Inference for the True Parameter

- This procedure is based on Andrews and Soares (Econometrica, 2010).
- Steps of the procedure:
  - (1) At every $\theta \in \widehat{\Theta}_n^{Grid}$, calculate $\{q_r(\theta);\ r = 1, \ldots, R\}$:

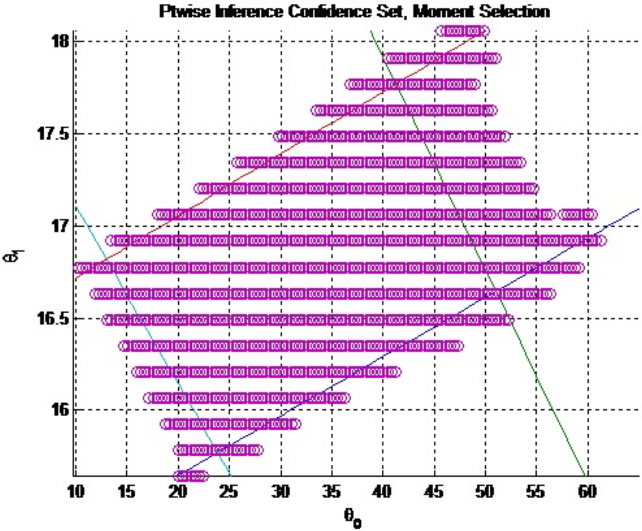
$$q_r(\theta) = chol(\widehat{\Omega}_{I,S}(\theta))N(0_S, I_S)$$

  - (2) For each of these θ and r, calculate: $T_r(\theta)$, $T_r^{MS}(\theta)$, or $T_r^{SM}(\theta)$.
  - (3) For each θ, calculate the $(1-\alpha)$ quantile. This the critical value, $c(\alpha, \theta)$.
  - (4) Calculate $T^{obs}(\theta)$ at each candidate value $\theta \in \widehat{\Theta}_n^{Grid}$.
  - (5) Compare $T^{obs}(\theta)$ against $c(\alpha, \theta)$ and accept θ into the confidence set whenever $T^{obs}(\theta) < c(\alpha, \theta)$.

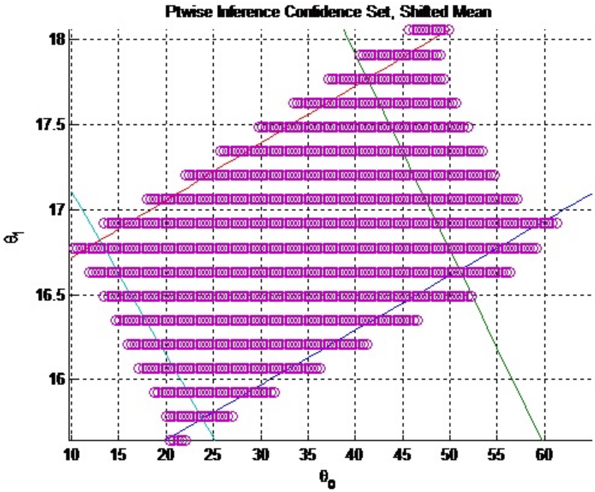
Inference for the Identified Set: Example



Inference for the Identified Set: Example



Inference for the Identified Set: Example



Comparison of Inference Procedures

Table 1: Estimated Confidence Intervals				
	θ0 min	θ0 max	θ1 min	θ1 max
True Θ Interval	14,376	15,298	1,305	1,321
Average Confidence Intervals From				
PPHI Procedure	13,573	16,170	1,283	1,343
Ptwise Procedure	8,500	20,307	832	1,850
Setwise Procedure	8,485	20,321	830	1,852
Moment Selection, Ptwise Procedure	8,514	20,280	832	1,850
Moment Selection, Setwise Procedure	8,485	20,321	830	1,852
Shifted Mean, Ptwise Procedure	10,775	19,272	927	1,658
Shifted Mean, Setwise Procedure	8,822	20,014	858	1,820