Moment Inequalities - Theory and Applications
Dickstein and Morales (2015)

Michael J. Dickstein
Stanford University

Economics 258

March 4, 2015

Sources of Error in Empirical Models
Structural Error

- (def'n) components of the agent's payoff function that are known to the agent but not in the econometrician's dataset
- typically the single source of error in a structural model
- Economic theory does not generally place restrictions on its distribution
  - Econometrician typically chooses the distribution up to a finite parameter vector
  - Require independence between structural error and observed covariates

Sources of Error in Empirical Economic Models
Expectational Error

- (def'n) mismatch between the agent's expectations and the realized outcome in the future
    - agent seldom knows the benefits he will earn from deciding to enter a market, switch jobs, etc
    - agent forms expectations, and bases his decision on these expectations
- Why is this a problem?— econometricians rarely observe measures of agents ex ante expectations; have only ex post realizations
- Economic theory—i.e. rational expectations—places restrictions on this error

Sources of Error in Empirical Economic Models
Expectational Error

- Implication of rational expectations for expectational error:
  - It is mean independent of the unobserved expectation (analogous to classical error-in-variables)
  - It is correlated with the observed covariate
  - Any variable in the agent's information set is a valid instrumental variable

Example: Daniel's winery in Chile

Daniel's problem — export to Venezuela?

- Daniel is invited to participate in a wine fair in Caracas, Venezuela in June 2001. He must decide whether to participate by Dec 31, 2000.
- Daniel has information on shipping costs and does not face other export costs
- Daniel does NOT know the sales revenue he will earn if he attends the fair
  - forms expectations about export revenues based on market characteristics
  - e.g.: past sales revenue of competitors, current political unrest, nominal exchange rate, etc
- Binary choice: attend fair if expected sales revenue exceed shipping costs

Example: Daniel's winery in Chile

Econometrician's problem

- Chilean customs agency provides dataset on annual country-specific sales revenues obtained by each Chilean wine producer during 1995-2005
- NOT in the data:
  - Exact shipping cost per mile for Daniel to export to Venezuela
  - Daniel's expectation as of Dec 31, 2000 about the revenue he might earn from the fair in 2001

Example: Daniel's winery in Chile

Econometrician's assumptions: for unobserved shipping costs

Shipping costs are the **structural error**

- They are known to Daniel when he made his export decision, but are not observed by the econometrician.
- Assume error is normally distributed around a mean (to estimate); fix variance to an arbitrary constant.

Example: Daniel's winery in Chile

Econometrician's assumptions: for unobserved expectations of revenue

Options:

1. Assume Daniel has perfect foresight. Need:
   - realized revenues = Daniel's expectations
2. Compute Daniel's unobserved expectations. Need:
   - Daniel's information set
   - rational expectations
3. Use ex post measurement of potential revenue as a proxy for the unobserved expectation. Need:
   - rational expectations

*** Option 3 imposes fewer assumptions, but introduces error-in-variables

Goals

- Identify and estimate the index coefficients in a binary choice model, allowing for:
  1. Individual-specific structural errors
  2. Expectational error (error-in-variables)
- Approach: apply moment inequalities

Outline

- Define statistical model
- Introduce moment inequalities
- Dickstein and Morales (2015)
- Inference

Statistical model

Decision rule

- Utility of individual i for any alternative $j \in \{0, 1\}$ is

$$U_j = \beta \mathcal{E}[X_j | \mathcal{J}] + \nu_j = \beta X_j^* + \nu_j.$$

- Decision, d_j, for any j:

$$d_j = \mathbb{1}\{\Delta U_j \geq 0\}, \qquad \Delta U_j = U_j - U_{j'}$$

- Individual revealed preference inequality:

$$d_j \cdot (\beta \Delta X_j^* + \Delta \nu_j) \geq 0, \qquad j \in \{0, 1\}$$

- $\beta \Delta X_j^*$ is the index function, β is the parameter we want to identify and estimate. We assume this index function is **linear in covariates**.

Statistical model

Measurement model

- ν_j not observed
- Expectational error, ε_j:

$$\beta(X_j - \mathcal{E}[X_j|\mathcal{J}]) = \varepsilon_j$$

- Rewrite the payoff function:

$$U_j = \beta X_j + \nu_j - \varepsilon_j.$$

- Impose rational expectations: $\mathcal{E}[\cdot] = \mathbb{E}[\cdot]$

Statistical model

Measurement model (continued)

- Impose rational expectations: $\mathcal{E}[\cdot] = \mathbb{E}[\cdot]$
  - $\mathbb{E}[\varepsilon_j | X_j^*] = 0$
  - $\mathbb{E}[\varepsilon_j | X_j] \neq 0$
- Implications:
  - Rational expectations $\Longrightarrow$ Errors-in-variables assumption.
  - For any $Z \in \mathcal{J}$, $\mathbb{E}[\varepsilon_j | Z] = 0 \Longrightarrow$ any Z in the information set is a valid IV.

Statistical model

Measurement model (continued)

Notation:

- Split ΔX_j^* into two subvectors:
  - $\Delta X_{1j} = \Delta Z_{1j} = \Delta X_{1j}^* \Longleftarrow Px1$ subvector measured without error
  - $\Delta X_{2j} = \Delta X_{2j}^* + \Delta \varepsilon_j \Longleftarrow (K - P)x1$ subvector measured with error
- Revealed preference inequality becomes:

$$d_j \cdot (\beta_1 \Delta Z_{1j} + \beta_2 \Delta X_{2j} + \Delta \nu_j - \beta_2 \Delta \varepsilon_j) \geq 0.$$

- What is observed?
  - $d_j, \Delta Z_{1j}, \Delta X_{2j}, \Delta Z_{2j}$
- What is not observed?
  - $\Delta \nu_j, \beta_2 \Delta \varepsilon_j$

Statistical model

Assumptions

Assumption 1 *The random variable $\Delta\nu_j$ is independent of the random vector $(\Delta Z_j, \Delta X_j^*)$:*

$$F_\nu(\Delta\nu_j | (\Delta Z_j, \Delta X_j^*)) = F_\nu(\Delta\nu_j).$$

Statistical model

Assumptions

Assumption 1 *The random variable $\Delta\nu_j$ is independent of the random vector $(\Delta Z_j, \Delta X_j^*)$:*

$$F_\nu(\Delta\nu_j|(\Delta Z_j, \Delta X_j^*)) = F_\nu(\Delta\nu_j).$$

- The endogeneity problem is due solely to expectational error
- Observed instruments are independent of the structural error
- Excludes models with random coefficients

Statistical model

Assumptions (continued)

Assumption 2 *The marginal distribution function of* $\Delta\nu_j$ *is known up to a scale parameter, log concave, has mean zero, and, for any* y *in the support of* $\Delta\nu_j$*, verifies the following property:*

$$\frac{\partial^2 \mathbb{E}[\Delta\nu_j | \Delta\nu_j \geq y]}{\partial y^2} \geq 0.$$

Statistical model

Assumptions (continued)

Assumption 2 *The marginal distribution function of $\Delta\nu_j$ is known up to a scale parameter, log concave, has mean zero, and, for any y in the support of $\Delta\nu_j$, verifies the following property:*

$$\frac{\partial^2 \mathbb{E}[\Delta\nu_j | \Delta\nu_j \geq y]}{\partial y^2} \geq 0.$$

- Distribution of structural error known and (a) is log concave and (b) has a right-truncated expectation that is convex in the truncation point
- Includes normal, logistic distribution
- Aside: our model generalizes probit and logit to allow classical measurement error in the covariates

Statistical model

Assumptions

Assumption 3 *The distribution of* $\Delta\varepsilon_j$ *conditional on* $(\Delta X_j^*, \Delta Z_j, \Delta\nu_j)$ *has support* $(-\infty, \infty)$ *and mean zero:*

$$\mathbb{E}[\Delta\varepsilon_j | \Delta X_j^*, \Delta Z_j, \Delta\nu_j] = 0.$$

Statistical model

Assumptions

Assumption 3 *The distribution of* $\Delta\varepsilon_j$ *conditional on* $(\Delta X_j^*, \Delta Z_j, \Delta\nu_j)$ *has support* $(-\infty, \infty)$ *and mean zero:*

$$\mathbb{E}[\Delta\varepsilon_j | \Delta X_j^*, \Delta Z_j, \Delta\nu_j] = 0.$$

- classical error-in-variables assumption
- no parametric assumption (would be necessary for maximum likelihood under probit)
- does not require full independence between expectational error and vector of ν_j
- under rational expectations, need ΔZ_j to be in $\mathcal{J}$

Conditional Moment Inequalities

- We first derive two types of moment inequalities conditional on the instrumental variable ΔZ_j.
- *Score Function Moment Inequalities*

$$\mathcal{M}_s(Z, j; \beta) = \mathbb{E}\left[d_j \frac{F_\nu\big(-(\beta_1 \Delta Z_{1j} + \beta_2 \Delta X_j)\big)}{1 - F_\nu\big(-(\beta_1 \Delta Z_{1j} + \beta_2 \Delta X_j)\big)} - d_{j'} \Big| Z \right] \geq 0.$$

- *Revealed Preference Moment Inequalities*

$$\mathcal{M}_r(Z, j; \beta) = \mathbb{E}\Big[d_j(\beta_1 \Delta Z_{1j} + \beta_2 \Delta X_j) + d_{j'} \mathbb{E}\big[\Delta \nu_{j'} | \Delta \nu_{j'} \geq -(\beta_1 \Delta Z_{1j'} + \beta_2 \Delta X_{j'})\big] \Big| Z \Big] \geq 0.$$

- For any Z and j, $\mathcal{M}_s(Z, j; \beta^*) \geq 0$ and $\mathcal{M}_r(Z, j; \beta^*) \geq 0$.

Conditional Moment Inequalities

Aside: Derivation of score function moment inequalities

- $\mathcal{L}(d_j|\Delta X_j^*, \Delta Z_{2j})$, the log of the probability of choosing j conditional on $(\Delta X_j^*, \Delta Z_{2j})$ equals:

$$\mathbb{E}\Big[d_j \log\big(1 - F_\nu(-\beta \Delta X_j^*)\big) + (1-d_j)\log\big(F_\nu(-\beta\Delta X_j^*)\big) | \Delta X_j^*, \Delta Z_{2j}\Big].$$

- The corresponding score function is:

$$\frac{\partial \mathcal{L}(d_j|\Delta X_j^*, \Delta Z_{2j})}{\partial \beta} = 0$$

$$\mathbb{E}\left[d_j \frac{F_\nu\left(-\beta^* \Delta X_j^*\right)}{1 - F_\nu\left(-\beta^* \Delta X_j^*\right)} - (1-d_j)\Big| \Delta X_j^*, \Delta Z_{2j}\right] = 0$$

Conditional Moment Inequalities

Aside: Derivation of score function moment inequalities

- Key: Let the truncation point, y, equal $-\beta\Delta X_j^*$, and let η be the expectational error. Log concavity of $\Delta\nu$'s distribution and Jensen's Inequality ensures that if

$$\mathbb{E}[\eta|\Delta X_j^*, \Delta Z_{2j}] = 0,$$

it holds that

$$\mathbb{E}\Big[\frac{F_\nu(y+\eta)}{1-F_\nu(y+\eta)}\Big|\Delta X_j^*, \Delta Z_{2j}\Big] \geq \frac{F_\nu(y)}{1-F_\nu(y)}.$$

Unconditional Moment Inequalities

- The derive the same two types of unconditional inequalities.
- *Score Function Moment Inequalities*

$$\mathcal{M}_s^q(\beta) = \mathbb{E}\left[\sum_{j\in\{0,1\}} \left\{\Psi_q(\Delta Z_j)\left(d_j \frac{F_\nu\big(-(\beta_1 \Delta Z_{1j} + \beta_2 \Delta X_j)\big)}{1 - F_\nu\big(-(\beta_1 \Delta Z_{1j} + \beta_2 \Delta X_j)\big)} - d_{j'}\right)\right\}\right] \geq 0$$

- *Revealed Preference Moment Inequalities*

$$\begin{aligned}\mathcal{M}_r^q(\beta) = \mathbb{E}\Bigg[\sum_{j\in\{0,1\}} \Bigg\{\Psi_q(\Delta Z_j)\Bigg(& d_j(\beta_1 \Delta Z_{1j} + \beta_2 \Delta X_j) \\ & + d_{j'}\mathbb{E}\big[\Delta\nu_{j'} | \Delta\nu_{j'} \geq -(\beta_1 \Delta Z_{1j'} + \beta_2 \Delta X_{j'})\big]\Bigg)\Bigg\}\Bigg] \geq 0.\end{aligned}$$

- The set of functions $\{\Psi_q(\Delta Z_j),\, q \in Q\}$ groups different values of ΔZ_j into different unconditional moment inequalities. We call them *instrument functions*.

Instrument Functions

- Example: When ΔZ_j is a 2×1 vector, define 4 instrument functions:

$$\Psi_1(\Delta Z_j) = \mathbb{1}\{\Delta Z_{1j} \geq 0\}\mathbb{1}\{\Delta Z_{2j} \geq 0\},$$
$$\Psi_2(\Delta Z_j) = \mathbb{1}\{\Delta Z_{1j} \geq 0\}\mathbb{1}\{\Delta Z_{2j} < 0\},$$
$$\Psi_3(\Delta Z_j) = \mathbb{1}\{\Delta Z_{1j} < 0\}\mathbb{1}\{\Delta Z_{2j} \geq 0\},$$
$$\Psi_4(\Delta Z_j) = \mathbb{1}\{\Delta Z_{1j} < 0\}\mathbb{1}\{\Delta Z_{2j} < 0\}.$$

- In empirical application, we define instruments using the level of ΔZ_j

Goals

- Theory-testing
- Measurement
- Methodology

Assuming an Information Set

- The researcher rarely observes $\mathcal{J}_{ij}$.
- Standard assumption - $\mathcal{J}_{ij}$ is a set of variables observed by the researcher:

$$\mathcal{J}_{ij} = Z_{ij}.$$

- Then, the export probability conditional on Z_j is:

$$\mathcal{P}(d_{ij} = 1 | Z_{ij}) = \Phi\big(\sigma_{\nu}^{-1}\big(\eta^{-1}\mathbb{E}[r_{ij}|Z_{ij}] - \beta_0 - \beta_1 dist_j\big)\big).$$

- Can estimate $\mathbb{E}[r_{ij}|Z_{ij}]$ non-parametrically (approach in WR, 1979).

Potential for Bias

- If firms' true information sets differ from observed vector of covariates and

$$\mathbb{E}[r_{ij}|Z_{ij}] = \mathbb{E}[r_{ij}|\mathcal{J}_{ij}] + \xi_{ij}, \qquad \xi_{ij} \neq 0,$$

  then the true export probability conditional on Z_j is

$$\mathcal{P}(d_{ij} = 1|Z_{ij}) =$$

$$\int_{k\xi+\nu} \mathbb{1}\{k\mathbb{E}[r_{ij}|Z_{ij}] - \beta_0 - \beta_1 dist_j - \nu - k\xi \geq 0\} f(k\xi + \nu|Z_j) d(k\xi + \nu).$$

- The estimates of β_0 and β_1 will be biased unless

$$f(k\xi + \nu|Z_j) = \phi(\nu).$$

- This equality holds if and only if $\mathbb{E}[r_{ij}|Z_{ij}] = \mathbb{E}[r_{ij}|\mathcal{J}_{ij}]$.

Information Set: Options

Export probability:

$$\mathcal{P}_{ij} = \mathcal{P}(d_{ij} = 1 | \mathcal{J}_{ij}) = \Phi\big(\sigma_\nu^{-1}\big(\eta^{-1}\mathbb{E}[r_{ij}|\mathcal{J}_{ij}] - \beta_0 - \beta_1 dist_j\big)\big).$$

(1) Perfect foresight

- $\mathbb{E}[r_{ij}|Z_{ij}] = r_{ij}$

(2) Two-step approach of Willis and Rosen (1979)

- Estimate $\mathbb{E}[r_{ij}|Z_{ij}]$ nonparametrically in a first stage

Information Set: Options

(2) Two-step approach of Willis and Rosen (1979), continued

- In model, export revenues are a function of:
  - domestic sales of firm i: r_{ih},
  - total domestic sales of all firms exporting to j: R_{hj},
  - total aggregate exports from h to j: R_j,
- Information set that exporters might have at t:
  - lagged own domestic sales: r_{iht-1},
  - lagged total aggregate exports from h to j: R_{jt-1},
  - distance from h to j (as proxy for R_{hj}): $dist_j$.

Information Set: Options

(3) Partially observed information sets

- We assume firms are likely to know (in addition to other variables):
  - lagged own domestic sales: r_{iht-1},
  - lagged aggregate exports from home country to each destination: R_{jt-1},
  - distance from home country: $dist_j$.
- We introduce two new types of moment inequalities,
  - odds-based moment inequalities
  - generalized revealed-preference moment inequalities
- We show how to perform counterfactuals and to test $H_0 : Z_{ij} \in \mathcal{J}_{ij}$.

Odds-Based Inequalities

- If $Z_{ij} \subset \mathcal{J}_{ij}$, then

$$\mathcal{M}(Z_{ij}; (\beta_0, \beta_1, \sigma_\nu)) = \mathbb{E}\left[\begin{array}{c} m_l(d_{ij}, r_{ij}, dist_j; (\beta_0, \beta_1, \sigma_\nu)) \\ m_u(d_{ij}, r_{ij}, dist_j; (\beta_0, \beta_1, \sigma_\nu)) \end{array} \middle| Z_{ij} \right] \geq 0,$$

with

$$m_l(\cdot) = d_{ij} \frac{1 - \Phi\big(\sigma_\nu^{-1}\big(kr_{ij} - \beta_0 - \beta_1 dist_j\big)\big)}{\Phi\big(\sigma_\nu^{-1}\big(kr_{ij} - \beta_0 - \beta_1 dist_j\big)\big)} - (1 - d_{ij}),$$

$$m_u(\cdot) = (1 - d_{ij}) \frac{\Phi\big(\sigma_\nu^{-1}\big(kr_{ij} - \beta_0 - \beta_1 dist_j\big)\big)}{1 - \Phi\big(\sigma_\nu^{-1}\big(kr_{ij} - \beta_0 - \beta_1 dist_j\big)\big)} - d_{ij},$$

where $(\beta_0, \beta_1, \sigma_\nu)$ denotes the true value of the parameter vector.

Generalized Revealed-Preference Inequalities

- If $Z_{ij} \subset \mathcal{J}_{ij}$, then

$$\mathcal{M}^r(Z_{ij};(\beta_0,\beta_1,\sigma_\nu)) = \mathbb{E}\left[\begin{array}{c} m_l^r(d_{ij}, r_{ij}, dist_j;(\beta_0,\beta_1,\sigma_\nu)) \\ m_u^r(d_{ij}, r_{ij}, dist_j;(\beta_0,\beta_1,\sigma_\nu)) \end{array} \middle| Z_{ij} \right] \geq 0,$$

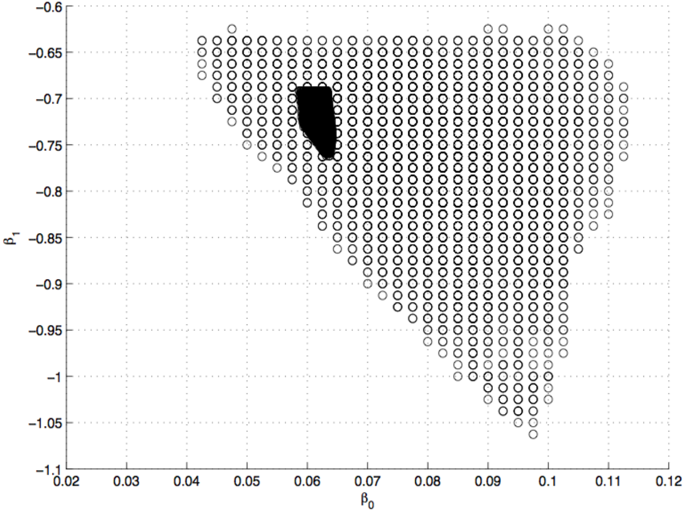
with

$$m_l^r(\cdot) = -(1-d_{ij})\big(kr_{ij} - \beta_0 - \beta_1 dist_j\big) + d_{ij}\sigma_\nu \frac{\phi\big(\sigma_\nu^{-1}(kr_{ij} - \beta_0 - \beta_1 dist_j)\big)}{\Phi\big(\sigma_\nu^{-1}(kr_{ij} - \beta_0 - \beta_1 dist_j)\big)},$$

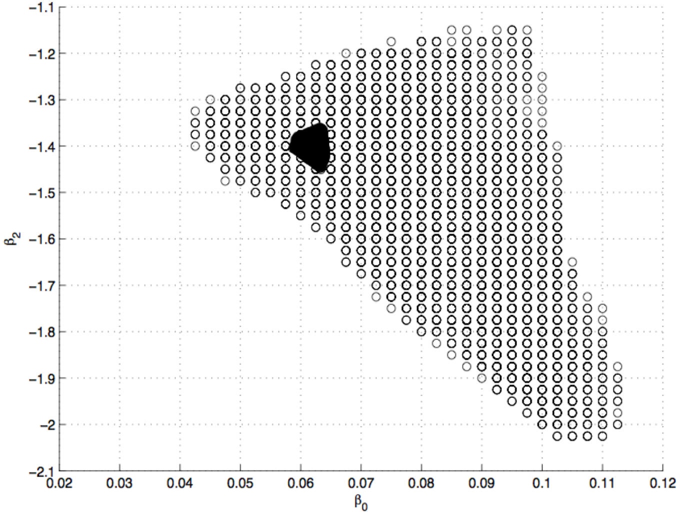
$$m_u^r(\cdot) = d_{ij}\big(kr_{ij} - \beta_0 - \beta_1 dist_j\big) + (1-d_{ij})\sigma_\nu \frac{\phi\big(\sigma_\nu^{-1}(kr_{ij} - \beta_0 - \beta_1 dist_j)\big)}{1-\Phi\big(\sigma_\nu^{-1}(kr_{ij} - \beta_0 - \beta_1 dist_j)\big)},$$

where $(\beta_0, \beta_1, \sigma_\nu)$ denotes the true value of the parameter vector.

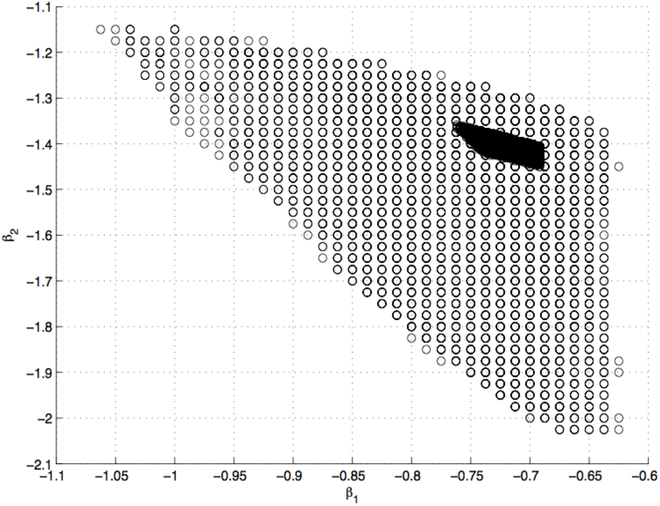
Identified Set and Confidence Set



Identified Set and Confidence Set



Identified Set and Confidence Set



Empirical Application: $(r_{iht-1}, R_{jt-1}, dist_j) \in \mathcal{J}_{ijt}$

Parameter estimates for fixed costs specification (in $000s)				
Method		sigma	constant	distance
Under perfect foresight	Estimate	1,074.0	760.9	1,180.1
	Std error	46.7	36.7	53.2
Under two-step approach	Estimate	701.9	502.2	812.3
	Std error	24.3	20.2	30.0
Under partial knowledge	Lower bound	311.7	218.3	428.3
of information sets	Upper bound	341.0	245.8	479.0

Comparison of Different Methods.

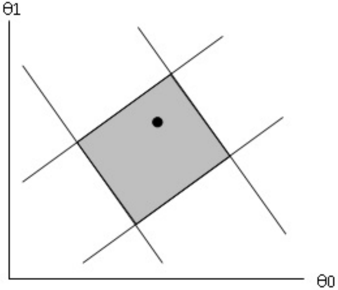
	Entry cost estimates ($000s)			
	Via moment inequalities			
Destination country	Lower bound	Upper bound	Percentage change relative to perfect foresight	Percentage change relative to two-step approach
Argentina	269.99	298.15	-67% to -70%	-50% to -55%
Japan	977.63	1,061.96	-62% to -65%	-44% to -49%
United States	592.55	632.03	-64% to -66%	-46% to -50%

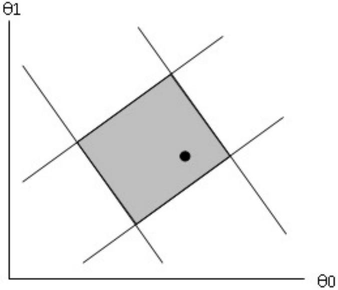
ALTERNATIVE MOMENT INEQUALITIES ESTIMATORS: DEFINITION AND COMPUTATION

Identified Set

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

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





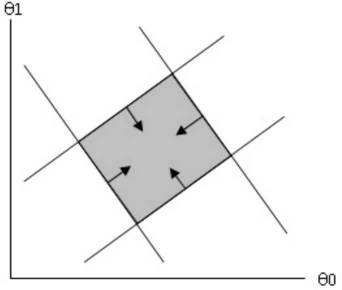
Steps for Estimation

- 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 (Econometrica, 2014).
  - Identified set: Chernozhukov, Hong and Tamer (Econometrica, 2007).
  - True parameter vector: Andrews and Soares (Econometrica, 2010).

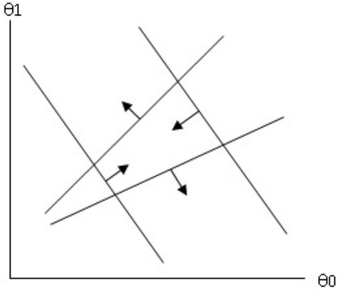
Estimation of the Identified Set

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

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



(c) Case 1



(d) Case 2

Estimation of the identified set

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

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

  - Weighted criterion function:

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

  with

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

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

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

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

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

- Apply simplex routine in Matlab via *linprog*

Computation of the estimated set

- If there is no value of θ that verifies all the constraints, $\hat{\Theta}_n^S$ will be a singleton.
- This singleton is the outcome of a nonlinear optimization problem:

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

- Use a nonlinear optimization package, like *KNITRO* (in Matlab via *ktrlink* with license).

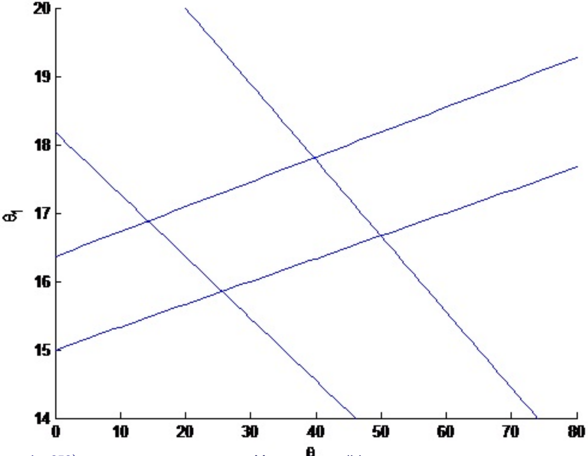
Computation of the estimated set: example

- 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:

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

- The finite-sample null distribution of $T_n(\theta_0)$ depends on the degree of *slackness* of the population moments–i.e. how much greater than 0 is:

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

Inference: General Intuition

- Key: need to infer whether a population moment binds at a particular value θ_0.
- Compute slackness factor, $SF_{n,s}(\theta_0)$
  - accounts for whether moment is likely to be binding
  - moments likely to be nonbinding asymptotically –i.e. $\overline{m}_{n,s}(\theta) >>> 0$–will have larger slackness factors

Inference: General Intuition

- Three slackness factors proposed in the literature:
  - Assume that all the S moments are binding at θ_0: $SF_{I,s} = 0$.
    - yields the most conservative test
  - Moment Selection:

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

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

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

Inference for an Interval: PPHI (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.
  - Approximation to the asymptotic distribution of all the (weighted) moments recentered at zero.
  - Jacobian of the moments.
  - Slackness factors.

Inference for an Interval

- Approximation to asymptotic distribution of all the recentered moments.
  - Draw $r = 1, ..., R$ times from a multivariate normal with zero mean, and covariance equal to the variance of the weighted moments
    - Take R standard normal draws.
    - Premultiply each draw by the Cholesky decomposition of the correlation matrix evaluated at the vertex of interest, $\widehat{\Omega}_{n,S}(\hat{\theta})$:

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

    - Result:

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

Inference for an Interval

- Jacobian of the moments.
  - Compute the Jacobian of the sample unweighted moments, $\overline{m}_{n,s}(\theta)$, and evaluate the result at the vertex of interest:
  - When the moments are linear in θ, the derivative matrix multiplied by θ simply equals the mean of the weighted moments:

$$\widehat{\Gamma}_{n,s}(\theta) * \theta = \frac{1}{n}\left[\sum_{i=1}^{n} \frac{\Delta x_{i,s}}{\widehat{\sigma}_{n,s}(\theta)}, \sum_{i=1}^{n} \frac{\Delta y_{i,s}}{\widehat{\sigma}_{n,s}(\theta)}\right] * \begin{pmatrix}\theta_0 \\ \theta_1\end{pmatrix} = \frac{m_{n,s}(\theta)}{\widehat{\sigma}_{n,s}(\theta)}$$

- evaluate the weights, $\widehat{\sigma}_{n,s}(\theta)$, at θ values equal to the relevant vertex.

Inference for an Interval

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

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

Inference for an Interval

- Compute the following linear programing problem for each draw $r = 1, ..., R$ and each vertex $\hat{\theta}$: (total of $2xdxR$ optimizations)

$$
\begin{aligned}
\theta_r = \max_{\theta} \quad & f \cdot \sqrt{n}(\hat{\theta} - \theta) \\
\text{s.t.} \quad & \\
& \widehat{\Gamma}_{n,S}(\hat{\theta})\sqrt{n}(\hat{\theta} - \theta) + q_r(\hat{\theta}) + SF^{SM}_{n,S}(\hat{\theta})\sqrt{n} \geq 0
\end{aligned} \tag{2}
$$

- As before, 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 (2), 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

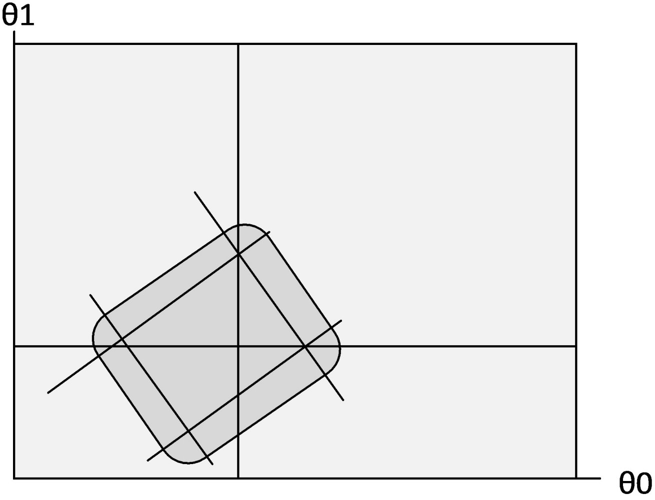
- For each pair of vertices corresponding to a given dimension d of θ.
  - For the min vertex, take the $\alpha/2$ quantile of the set of simulated vertices, θ_r, $r = 1, \ldots, R$. Denote this number:
$$\underline{\theta}_{d,\alpha/2}.$$
  - For the max vertex, take the $(1-\alpha/2)$ quantile of the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

Set/Point Inference: General Intuition

- Based on the inversion of an Anderson-Rubin T statistic.
- General steps in the algorithm:
  1. Define θ grids, $\widehat{\Theta}_n^{Grid}$ and $\widehat{\Theta}_n^{\epsilon}$, where $\widehat{\Theta}_n^{\epsilon} \subset \widehat{\Theta}_n^{Grid}$.
  2. Calculate $T_r(\theta)$, at a set of points in either $\widehat{\Theta}_n^{Grid}$ or $\widehat{\Theta}_n^{\epsilon}$ 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 $\theta \in \widehat{\Theta}_n^{Grid}$ 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 $\widehat{\Theta}_I^{\epsilon}$

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



Inference for the Identified Set

Chernozhukov, Hong and Tamer (Econometrica, 2007)

- Steps of the procedure:
  - (1) At $\theta \in \widehat{\Theta}_n^{\varepsilon}$, compute R draws $\{q^r(\theta); r = 1, \ldots, R\}$ such that:

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

    with

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

    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:

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

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

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

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

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

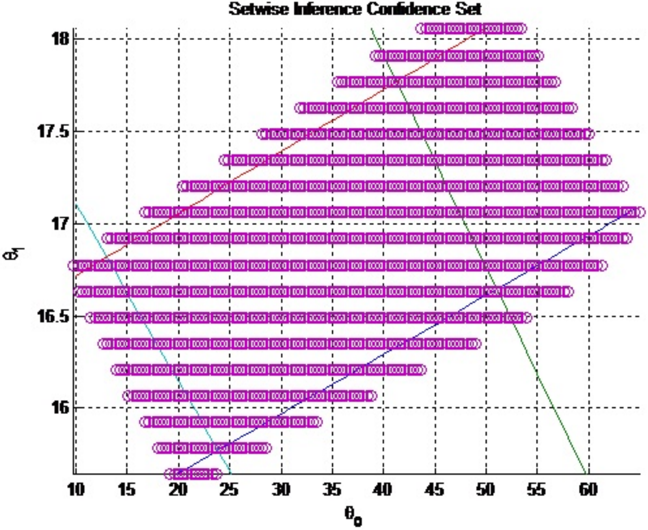
Inference for the Identified Set

- 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 $\theta \in \widehat{\Theta}_n^{Grid}$:

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

  - (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 True Parameter

Andrews and Soares (2010)

- Steps of the procedure:
  - (1) At every $\theta \in \widehat{\Theta}_n^{Grid}$, calculate $\{q_r(\theta);\ r = 1, \ldots, R\}$:

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

  - (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 $T^{obs}(\theta) < c(\alpha, \theta)$.