Example and Monte Carlo From “Moment Inequalities and Their Application.”

A. Pakes, J. Porter, Kate Ho, and Joy Ishii

July 9, 2011

Goals.

• Investigate performance of alternative estimators for the upper and lower bounds for a function of parameters in partially identified models.
  – Take functions equal to the parameters: gives dimension by-dimension extreme points and associated confidence intervals.
  – In problems with 3 or more dimensions it will be difficult to provide a full description of confidence sets; ⇒ need for smaller dimensional statistics. Dimension-by-dimension confidence intervals are familiar from standard practice for point estimation problems.

• Non-parametric ordered choice which allows for structural and non-structural errors (in or not in agent’s information set when decision is made). Also adapt a selection correction suggestion due to Powell for use in moment inequality models.
Preliminary Notation \((\Theta \subset \mathbb{R}^K)\).

\[
\Theta_0 = \{\theta \in \Theta : \mathcal{P}m(z, \theta) \geq 0\},
\]

and the object of interest is

\[
\underline{\theta} = \{\theta \in \Theta_0 : \theta_1 = \arg\min_{\theta \in \Theta_0} \tilde{\theta}_1\}.
\]

Similarly if \(P_Jm(z, \theta) = \frac{1}{J} \sum_{j=1}^{J} m(z^j, \theta)\),

\[
\Theta_J = \arg\min_{\theta \in \Theta} \|(P_Jm(z, \theta))\_\|,
\]

and

\[
\hat{\theta}_1 = \{\theta \in \Theta_0 : \theta_1 = \arg\min_{\tilde{\theta} \in \Theta_J} \tilde{\theta}_1\}.
\]

Direct Estimates of Bounds.

As in the notation above we assume

- non-empty interior and that
- the boundary point is point identified (notation, \(\theta\))
**True Distribution.** For ease of exposition, assume all moments when evaluated at \( \theta \) have unit variance (otherwise we have to standardize them). Split moments up into those that are

- **Binding:** \( \mathcal{P} m_0(z, \theta) = 0 \), and
- **Non-binding:** \( \mathcal{P} m_1(z, \theta) > 0 \).

The true limit distribution for \( \sqrt{J} (\hat{\theta} - \theta) \) converges to that of \( \hat{\tau} \) where if

\[
\hat{\tau} = \arg \min_{[0 \leq \Gamma_0 \tau + Z]} \min \tau_1.
\]

- \( \Gamma_0 \) is the true value of the derivative matrix of the binding moments when evaluated at \( \theta \),

- \( Z \) is a draw from a mean zero normal with covariance matrix equal to the correlation matrix of the binding moments after dividing each by its standard deviation (\( \Sigma_0 \))

**Our Conservative Approximation.** Putting the standard deviations of the moments in explicitly the moments evaluated at \( \theta = \hat{\theta} \), are \( m(z, \hat{\theta}) / \hat{\sigma}(\hat{\theta}) \). Accordingly let

\[
\hat{\Gamma} \equiv \frac{\partial P_j m(z, \hat{\theta}) / \hat{\sigma}(\hat{\theta})}{\partial \theta}, \quad \hat{\Sigma} \equiv \text{Corr}(m(z, \hat{\theta}) / \hat{\sigma}(\hat{\theta})), \quad Z^* \sim N(0, \hat{\Sigma})
\]
Our estimated distribution for $\hat{\theta}$ is obtained by solving, for repeated draws on $Z^*$

$$\min\left[0 \leq \hat{\Gamma}^\tau + Z^* + \frac{\sqrt{J}}{\sqrt{2\ln(ln(J))}} P_Jm(z,\hat{\theta}) + / \hat{\sigma}(\hat{\theta})\right] \tau_1.$$

**Explanation.** First ignore $\sqrt{2\ln(ln(J))}$. Since $\sqrt{JP_Jm_1(z,\hat{\theta})/\hat{\sigma}(\hat{\theta})} \to \infty$, when $J$ is large

$$\{\tau : 0 \leq \hat{\Gamma}^\tau + Z^* + \sqrt{JP_Jm_1(z,\hat{\theta})/\hat{\sigma}(\hat{\theta})}\} \approx \{\tau : 0 \leq \hat{\Gamma}_0^\tau + Z_0^* + \sqrt{JP_Jm_0(z,\hat{\theta})/\hat{\sigma}(\hat{\theta})}\}.$$

Two cases.

- **1st case:** $\text{dim}(m_0) = \text{dim}(\theta)$ (# binding moments = # of parameters). Since we insure $\sqrt{JP_Jm_0(z,\hat{\theta})/\hat{\sigma}(\hat{\theta})} = 0$, provided $(\hat{\theta}, \hat{\Gamma}, \hat{\Sigma}) \to_{a.s.} (\theta, \Gamma, \Sigma)$ as $J \to \infty$, the model without the correction term (without $\sqrt{2\ln(ln(J))}$) yields a consistent parametric bootstrap (Bickel and Friedman).

- **2nd case:** $\text{dim}(m_0) > \text{dim}(\theta)$. Now we insure $\sqrt{JP_Jm_0(z,\hat{\theta})/\hat{\sigma}(\hat{\theta})} = 0$ for only $K$ of the $\text{dim}(m_0)$ moments and insure that the rest are positive. The positive part will cause it to simulate a critical value which is too small, which leads to confidence intervals which undercover. The $\sqrt{2\ln(ln(J))}$ corrects for this.
Relevance of two cases.

- No special structure: expect the first case to be the relevant limiting case.

- However Monte Carlo on typically sized data sets indicates that the binding moments do switch with different sample draws, as would be the case if the second case were relevant.

- So it is the second case which typically better approximates the small sample properties of the estimator, and the correction term, or tuning parameter, is relevant.

Alternative Estimators of Boundaries.

Use alternative estimators introduced for partially identified models and project them onto axis. Estimators used

- Boundaries of confidence sets for a point (CHT); and two modifications (Andrews and coauthors).
  
  - Modified with a prior moment selection step,
  
  - Modified by shifting the mean when finding critical values.

- Boundaries of confidence sets for a set (CHT); and the same two modifications.
Review: Construction of Alternative Estimators.

Point Coverage.

- Divide the parameter space into a grid of points, \( \{\theta_g\}_{g=1}^G \).
- Draw Monte Carlo data sets (indexed by \( d \)).
- For each \( d \), at each \( \theta_g \): simulate normals with mean zero and correlation matrix given by the correlation matrix of the moments at \( \theta_g \) (the same simulation draws are used for all grid points). Calculate the value of the objective function for each simulation draw and find the 95% critical value.
- Accept all points for which the actual moments are less than the critical value.

Details: Moment Selection. Adds an initial step to the routine at each grid point which drops some moments. Included moments differ by grid point. At \( \theta = \theta_g \)

\[
M(\theta_g) = \{ r : \sqrt{J} P_J m_r(z, \theta_g) / \hat{\sigma}(m_r(z, \theta_g)) \leq \sqrt{2 \ln(\ln(J))} \},
\]

where \( \hat{\sigma}(m_r(z, \theta_g)) \) is the sample standard deviation.

Details: Moment Shifting. Before computing simulated critical values at \( \theta_g \) add the interval shift factor

\[
\frac{\sqrt{J}}{\sqrt{2 \ln(\ln(J))}} P_J m_r(z, \theta_g) / \hat{\sigma}(m_r(z, \theta_g)),
\]

to each simulated draw.
Set Coverage.

- For each data set find $\hat{\Theta}_d(\epsilon)$: the set of $\theta_g$ values that generate objective functions which are within $\epsilon$ of the minimized objective function value for that sample.

- Then use simulation to find the value of $c_{\epsilon,\alpha}$ that, under the null that the model is satisfied for all $\theta \in \Theta_d(\epsilon)$, insures that the maximum of the objective function over all values of $\theta \in \hat{\Theta}_d(\epsilon)$ is less than $c_{\epsilon,\alpha}$ with probability $1 - \alpha$ (this uses a different Cholesky transform for each $\theta_g \in \Theta_d$, but the same standard normal draws).

- Finally we go back to the sample of firms, compute the value of the objective function for all $\theta_g$, and accept those values for which $Q_d(\theta_g) \leq c_{\epsilon,\alpha}$.

Moment selection and moment shifting procedures are analogous to those above.

Performance Metrics.

- **Coverage.** For each estimator we computed coverage of:

  \[ \text{true pt, min. over pt.s in id. set, true intervals, true set}. \]

  Coverage equals the fraction of the 1000 Monte Carlo data sets for which the 95% confident interval covered the truth.

- **Closeness to true interval.** For each confidence set, compute its projection onto the axis from each data set, and form summary statistics on the end-points of those intervals.
Monte Carlo: An Ordered Choice Model.

Ishii (2008) studies the formation and implications of the ATM networks of banks (abandon surcharges?). Her analysis requires the cost of installing and operating the ATM’s. PPHI analyzes the empirical problem. Current Monte Carlo has a sample design that mimics, to the extent possible, that empirical problem. Basic problem: Firm chooses \( d (= \text{number of ATM’s}) \) to

\[
\max_{d \in \mathbb{D} \subset \mathbb{Z}^+} E[\pi(d, d_{-i}, z_i, \nu_{1,i,d}, \nu_{2,i}) | J_i]
\]

where \( E[\pi(d, d_{-i}, z_i) | J_i] = \text{profits that the firm expects to earn in the second stage if the firm chose } d \text{ and its competitors chose } d_{-i} \text{ in the first stage} \), and \( \mathbb{Z}^+ \) is the non-negative integers.

More detailed notation;
- \( R(d, d_{-i}, z_i) \) is measured net revenue from bank operations
- \( \nu_{1,i,d} \) is expectational and approximation error in those revenues (so \( P[\nu_{1,i,d} | J_i] = 0 \)); and
- \( C(d, \nu_{2,i}; \theta) \) is the cost of purchasing and installing \( d \) ATM’s;
where \( \nu_{2,i} \in J_i \) (and so is a determinant of \( d_i \) and possibly \( d_{-i} \)) but is not observed, and \( \theta \) is to be estimated

\[
\pi(d, d_{-i}, z_i, \nu_{1,i,d}, \nu_{2,i}) = R(d, d_{-i}, z_i) + \nu_{1,i,d} - C(d_i, \nu_{2,i}, \theta)
\]

Note: If (i) \( \nu_{1,i,d} \equiv 0 \equiv d_{-i} \), & (ii) \( \nu_{2,i} \) has a parameteric distribution; standard ordered choice model. We make no assumption on either the distribution of \( \nu_{1,i} \) or \( \nu_{2,i} \). In multiple agent models the distribution of \( \nu_{1,i} \) depends on information sets of agents and equilibrium selection mechanism.
Moment Inequalities.

Necessary conditions for optimal choice include

- Expected increment to returns from the last ATM the bank installed were greater than its cost of an ATM,

- Expected increment to returns from adding one ATM more than the number actually installed was less than that cost.

Assume marginal cost of \( d^{th} \) ATM is

\[
\theta_0 + \theta_1 (d - 1) + \nu_{2,i}
\]

where \( \nu_{2,i} \in \mathcal{J}_i \) and has an unconditional mean of zero, and let

\[
\Delta r_i(\cdot, \theta) = \begin{pmatrix}
R(d_i, d_{i-1}, z_i) - R(d_i - 1, d_{i-1}, z_i) - \theta_0 - \theta_1 (d_i - 1) \\
R(d_i, d_{i-1}, z_i) - R(d_i + 1, d_{i-1}, z_i) + \theta_0 + \theta_1 d_i
\end{pmatrix},
\]

which is observable up to \( \theta \), so that

\[
\Delta \pi_i(\cdot, \theta) = \begin{pmatrix}
\Delta r(d_i, d_i - 1, d_{i-1}, \cdot, \theta) - \nu_{2,i} + \Delta \nu_{1,i,d_i,d_i-1} \\
\Delta r(d_i, d_i + 1, d_{i-1}, \cdot, \theta) + \nu_{2,i} + \Delta \nu_{1,i,d_i,d_i+1}
\end{pmatrix}.
\]

Then our necessary conditions imply that if \( x_i \in \mathcal{J}_i \) is an “instrument” (mean independent of \( \nu_{2,i} \)), and \( h(\cdot) \geq 0 \), then

\[
\mathbb{P}[\Delta m(\cdot, \theta)] = \mathbb{P}\left[ \sum_i \Delta r_i(\cdot, \theta) \otimes h(x_i) \right] \geq 0, \text{ at } \theta = \theta^0.
\]

Note. If \( h(x) \equiv 1 \), and we replace \( \mathbb{P} \) with \( \mathbb{P}_J \) we take an unconditional sample average of the \( \nu_{2,i} \). Since the unconditional mean of \( \nu_{2,i} \) and of \( \Delta \nu_{1,i} \) is zero, \( \mathbb{P}_J(\nu_{2,i} + \Delta \nu_{1,i,d_i,d_i+1}) \to a.s. 0. \)
Boundary Problem

Our upper bound $\Delta R(d_i, d_i - 1, \cdot)$ not available for observations with $d_i = 0$, and if we form averages over the observations with $d_i \neq 0$ we no longer get the unconditional mean. Rather we incur a selection problem; those that had $d_i = 0$ may have higher than average $\nu_{2,i}$, .... Similar problem arises whenever using moment inequalities on censored data (e.g. Tobit).

**Accounting for Selection in Moment Inequality Problems.** Assume $\nu_{2,i}$ symmetrically distributed (or at least not skewed in the direction of truncation). Then the negative of the bound for the lowest values of $\nu_{2,i}$ becomes an upper bound for the $\nu_{2,i}$ of the missing observations (adapts Powell 1986 to moment inequality models).

**Example.** For notational simplicity work with a model where $\theta_1 = 0$, so that marginal cost is $\theta_0 + \nu_{2,i}$. Then

$$x_i^r \equiv \mathbb{E}[\Delta R(d_i, d_i + 1, \cdot)|\mathcal{J}_i] \leq \theta_0 + \nu_{2,i}$$

for every $i$, while if $d_i > 0$

$$x_i^l \equiv \mathbb{E}[\Delta R(d_i, d_i - 1, \cdot)|\mathcal{J}_i] \geq \theta_0 + \nu_{2,i}$$

Let

$$L = \{i: d_i > 0\}, \quad \text{and} \quad \#L = q_J \times J, \quad q_J \to_{a.s.} q > 1/2.$$ 

The model implies that if $d_i > 0$

$$(x_i^l - \theta) \geq \nu_{2,i}.$$
However for all $i$ we have an underestimate of $\theta + \nu_{2,i}$

$$(x^r_i - \theta) \leq \nu_{2,i} \Rightarrow (\theta - x^r_i) \geq -\nu_{2,i}.$$

Let

$$U = \{ J(1 - q_J) \text{ highest values of } (\theta - x^r_i) \}.$$ 

Then

$$J^{-1} \sum_{i:d_i > 0} (x^l_i - \theta) \geq \int_{\nu_{2,i}:d_i > 0} \nu_2 dF_J(\nu_2) \geq \int_{F^{-1}(1-q)} \nu_2 dF_J(\nu_2) \rightarrow_{J \rightarrow \infty} \int_{F^{-1}(q)} \nu_2 dF(\nu_2).$$

Also

$$J^{-1} \sum_{i \in U} (\theta - x^r_i) \geq \int_{\nu_{2,i}:i \in U} -\nu_2 dF_J(\nu_2) \geq \int_{F^{-1}(1-q)} -\nu_2 dF_J(\nu_2)$$

$$\rightarrow_{J \rightarrow \infty} \int_{F^{-1}(1-q)} -\nu_2 dF(\nu_2) = \int_{F^{-1}(q)} \nu_2 dF(\nu_2).$$

But

$$\int_{F^{-1}(q)} \nu_2 dF(\nu_2) + \int_{F^{-1}(q)} \nu_2 dF(\nu_2) = 0,$$

so as $J \rightarrow \infty$

$$J^{-1} \sum_{i:i \in L} (x^l_i - \theta) + J^{-1} \sum_{i \in U} (\theta - x^r_i) \geq 0$$

or

$$(2q_J - 1)[J^{-1} \sum_{i:i \in L} x^l_i - J^{-1} \sum_{i:i \in U} x^r_i] \geq \theta_0.$$

If we substitute the realizations for the expectations, the inequalities are maintained, and we can develop analogous inequalities for (positive valued) instruments.
Sample Design. 1000 data sets, 873 firms each.

- Observables: random draws on firm’s from Ishii’s data.
- Unobservables: normal draws that reflect: (i) expectational/measurement errors (about 1/2 the cross-sectional variance in $\Delta \pi(\cdot)$), and (ii) an unobservable cost component known to the firm (10% of cross-sectional variance in $\theta_0 + \theta_1(d_i - 1)$)
- $d_i$ is adjusted so that it is negatively correlated with the unobservable cost shock.
- Show results from three samples; (i) four moments and no censoring, (ii) four moments with censoring, and (iii) eight moments, no censoring.

Simulation Methodology,

- Simulated critical values for the confidence sets and the simulated distribution for the interval estimator were based on 400 draws.
- Nominal coverage: 95%.
- Do not keep track of computational costs: not too much of a burden here because there is no need for a fixed point calculation for each $\theta$—value. In problems where such a calculation is needed, the grid-point methods would be disadvantaged because of need to compute the fixed point at many $\{\theta_g\}_{g=1}^G$. 
Table 1: Average Intervals and Coverage: Four Moments.

<table>
<thead>
<tr>
<th>Estimation Method</th>
<th>([\theta_0;\theta_0])</th>
<th>% Coverage</th>
<th>([\theta_1;\theta_1])</th>
<th>% Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Truncated Sample (True Intervals; (\theta_0 = [14,376;15,298]); (\theta_1 = [1,305;1,321]))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Interval Inference</td>
<td>[13,544;16,162]</td>
<td>95.1</td>
<td>[1,282;1,342]</td>
<td>90.8</td>
</tr>
<tr>
<td>total % relaxation of bounds required for 95% coverage</td>
<td>0.0%</td>
<td></td>
<td>0.3%</td>
<td></td>
</tr>
<tr>
<td>2. Point Inference</td>
<td>[8,411;20,312]</td>
<td>99.4</td>
<td>[825;1,854]</td>
<td>100</td>
</tr>
<tr>
<td>5. Set with Moment Selection</td>
<td>[8,390;20,329]</td>
<td>99.4</td>
<td>[823;1,857]</td>
<td>100</td>
</tr>
<tr>
<td>6. Point with Shifted Mean</td>
<td>[10,861;19,044]</td>
<td>98.3</td>
<td>[949;1,644]</td>
<td>99.4</td>
</tr>
<tr>
<td>7. Set with Shifted Mean</td>
<td>[8,772;19,982]</td>
<td>99.2</td>
<td>[855;1,821]</td>
<td>100</td>
</tr>
<tr>
<td>Truncated Sample (True Intervals; (\theta_0 = [14,342;16,940]); (\theta_1 = [1,288;1,326]))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Interval Inference</td>
<td>[13,519;18,029]</td>
<td>95.7</td>
<td>[1,259;1,350]</td>
<td>92.4</td>
</tr>
<tr>
<td>total % relaxation of bounds required for 95% coverage</td>
<td>0.0%</td>
<td></td>
<td>0.002%</td>
<td></td>
</tr>
<tr>
<td>9. Point Inference</td>
<td>[8,292;22,907]</td>
<td>99.6</td>
<td>[703;1,909]</td>
<td>100</td>
</tr>
<tr>
<td>10. Set Inference</td>
<td>[8,247;22,932]</td>
<td>99.6</td>
<td>[700;1,914]</td>
<td>100</td>
</tr>
<tr>
<td>13. Point with Shifted Mean</td>
<td>[11,444;20,134]</td>
<td>98.3</td>
<td>[960;1,629]</td>
<td>99.5</td>
</tr>
<tr>
<td>14. Set with Shifted Mean</td>
<td>[8,773;22,405]</td>
<td>99.5</td>
<td>[748;1,867]</td>
<td>100</td>
</tr>
</tbody>
</table>
Results: Four Moment Case (“just identified end points”).

Coverage Grid Methods.

- Over cover intervals but not designed to do that.
- Point coverage also over covers $\theta_0$; but if we look at minimum coverage over all $\theta \in \Theta_0$ not bad (ranged from 93.7% to 97.9%)
- Set coverage only mildly conservative for set (ranged from 95.3 to 97.8)

Coverage Interval Estimator.

- Almost exact coverage for $[\theta_0, \bar{\theta}_0]$ but it undercovers for the narrower $[\theta_1, \bar{\theta}_1]$ (90.8% uncensored, 92.4% censored).
- Should we worry? The interval estimates would provide 95% coverage had we accepted estimated intervals that had end-points which in sum had a percentage deviation from true end points that was; $\leq .3\%$ uncensored and $\leq .002\%$ censored.
- Striking result: when the interval estimator misses, it tends to miss by a tiny amount: if we accept intervals whose two end points lie inside the truth in total $\leq 1\%$ the coverage looks conservative (97.5 to 99.5%).
Trade-off: Coverage vs. Closeness to True Values.

The grid methods are making a steep trade-off for conservativeness. For a minimal coverage increase they are giving much wider confidence intervals. Illustrate with non-truncated. Lengths of the average intervals in the $\theta_0$ and $\theta_1$ dimensions were

- Interval estimates: $[2,618; 60]$.

and the results are similar in the truncated sample.

Understanding the Differences.

Figure 1: Minimum point is $\approx \theta_0$. Oblong shaped area where the function does rise, but only minimally, and it remains below the critical value (which also rises but very slowly).

How different are the objective function values for the points in the identified sets which generate the end-points?

Index the end-points of the confidence regions generated by the point (interval) estimators by $p$ ($i$) and let

$$\theta(\vartheta_0(p)) \equiv \arg\min \{G(\theta) : \theta_0 = \vartheta_0(p), \theta_1 \in [\theta_1(p), \vartheta_1(p)]\},$$

so that $\theta(\vartheta_0(p))$ generates the minimum value of the average of $G(\theta)$ over data sets when the first component of $\theta$ is set equal to $\vartheta_0(p)$ and its second component is allowed to vary over its confidence interval.
Table 2: From Distribution Across MC Data Sets*

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$\Delta \bar{G}(\theta)$</th>
<th>$\Delta \bar{G}(\theta)/\sqrt{V(\Delta G_d(\theta))}$</th>
<th>$VST(\theta)$</th>
<th>$CST(\theta)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. $\theta(\theta_0(p))$</td>
<td>7.07</td>
<td>2.85</td>
<td>6.95</td>
<td>6.41</td>
</tr>
<tr>
<td>2. $\theta(\bar{\theta}_0(p))$</td>
<td>7.77</td>
<td>3.00</td>
<td>6.03</td>
<td>6.41</td>
</tr>
<tr>
<td>3. $\theta(\theta_1(p))$</td>
<td>7.47</td>
<td>2.47</td>
<td>6.84</td>
<td>6.41</td>
</tr>
<tr>
<td>4. $\theta(\bar{\theta}_1(p))$</td>
<td>6.81</td>
<td>2.55</td>
<td>6.95</td>
<td>6.41</td>
</tr>
<tr>
<td>5. $\theta = [\theta_0(p), \theta_1(p)]$</td>
<td>238</td>
<td>4.85</td>
<td>6.86</td>
<td>6.41</td>
</tr>
<tr>
<td>6. $\theta = [\bar{\theta}_0(p), \bar{\theta}_1(p)]$</td>
<td>189</td>
<td>4.82</td>
<td>6.98</td>
<td>6.41</td>
</tr>
<tr>
<td>7. $\theta(\theta_0(i))$</td>
<td>5.09</td>
<td>1.16</td>
<td>4.64</td>
<td>6.32</td>
</tr>
<tr>
<td>8. $\theta(\bar{\theta}_0(i))$</td>
<td>4.37</td>
<td>1.39</td>
<td>4.56</td>
<td>6.28</td>
</tr>
<tr>
<td>9. $\theta(\theta_1(i))$</td>
<td>.73</td>
<td>.35</td>
<td>4.62</td>
<td>6.31</td>
</tr>
<tr>
<td>10. $\theta(\bar{\theta}_1(i))$</td>
<td>2.47</td>
<td>1.48</td>
<td>4.54</td>
<td>6.27</td>
</tr>
<tr>
<td>11. $\theta = [\theta_0(i), \theta_1(i)]$</td>
<td>15.4</td>
<td>1.80</td>
<td>4.57</td>
<td>6.29</td>
</tr>
<tr>
<td>12. $\theta = [\bar{\theta}_0(i), \bar{\theta}_1(i)]$</td>
<td>20.6</td>
<td>2.37</td>
<td>4.50</td>
<td>6.25</td>
</tr>
</tbody>
</table>

Table notation:
- $\Delta \bar{G}(\theta) = \bar{G}(\theta) - \bar{G}(\theta_m)$ ($\theta_m$ minimizes $\bar{G}(\theta)$ over $\Theta$),
- $V(\cdot)$ provide the variance of its argument over samples.
- $VTS(\theta)$ is the variance of the simulated objective function used to determine the critical value for the pointwise test and
- $CST(\theta)$ provides the critical value of that test (both averaged over data sets).
Results.

• $VST(\theta)$ associated with pointwise end-points are $\geq 50\%$ higher than those for the interval end-points. This difference is not used in generating the c.i.’s for the interval procedure but is used for those from the pointwise procedure. Interestingly the critical values (and hence tail of the distribution) are relatively invariant to $\theta$.

• $\theta(\theta_0(p)), \theta(\bar{\theta}_0(p)), \theta(\theta_1(p)), \theta(\bar{\theta}_1(p))$, all typically in pointwise procedures confidence set. So there are typically points in that set where $\Delta \overline{G}(\theta)/V(\Delta G_d(\theta))^{1/2} > 3$.

• $[\theta_0(p), \theta_1(p)], [\theta_0(p), \theta_1(p)]$ are not typically in the pointwise procedures c.i., and for them $\Delta \overline{G}(\theta)/V(\Delta G_d(\theta))^{1/2} \approx 5$.

• For the interval procedure the max of $\Delta \overline{G}(\theta)/V(\Delta G_d(\theta))^{1/2}$ among points in the box formed from the intervals is $\approx 2.4$.

• Difference in the maximum value of $\Delta \overline{G}(\cdot)/V(\Delta G_d(\cdot))^{1/2}$ between points typically in the pointwise identified set and points the box formed from the interval estimates exist, but are small relative to the differences of this statistic at the end points from the two boxes.

• Latter difference is clearly related to the negative correlation between the values of $\theta_0$ and $\theta_1$ that lie on the boundaries of the pointwise procedures confidence sets, and so the extent of this difference is likely to vary across applied problems. However negatively correlated distributions of parameters is not unusual in applied work.
Behavior in samples which generate point estimates.

- Interval for $\theta_1$ only 1% of actual value (for $\theta_0$ it is 10%); helps explain coverage and raises the question of how the alternative estimators behaves when there are point estimates.

- 4 moment case: point estimates only .4% for the base case, and not at all in the truncated sample. Double the number of instruments (add the number of banks that operated in the market, and the number of markets the bank operated in); 16.7% of our runs on the eight moment example resulted in point estimates.

Results.

- The extra moments do not affect the identified intervals. In this case the c.i.’s for the interval estimates *must* narrow; but can go either way for those generated from the grid points. It increases length of intervals for all but the moment shifting procedure.

- All the points stressed for that case are even more salient in the eight moment case.
Table 3: Average Intervals and Coverage: Eight Moments

<table>
<thead>
<tr>
<th>Estimation Method</th>
<th>$[\theta_0; \theta_1]$</th>
<th>% Coverage</th>
<th>$[\hat{\theta}_0; \hat{\theta}_1]$</th>
<th>% Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Truncated Sample: True Intervals; $\theta_0 = [14,376;15,298]$; $\theta_1 = [1,305;1,321]$</td>
<td>$[13,655;15,974]$</td>
<td>80.50</td>
<td>$[1,284;1,341]$</td>
<td>84.30</td>
</tr>
<tr>
<td>Interval Inference</td>
<td>$[6,810;19,842]$</td>
<td>100</td>
<td>$[889;1,981]$</td>
<td>100</td>
</tr>
<tr>
<td>total % relaxation of bounds required for 95% coverage</td>
<td>2.4%</td>
<td></td>
<td>0.7%</td>
<td></td>
</tr>
<tr>
<td>Point Inference</td>
<td>$[6,810;19,844]$</td>
<td>100</td>
<td>$[889;1,981]$</td>
<td>100</td>
</tr>
<tr>
<td>Set Inference</td>
<td>$[6,815;19,840]$</td>
<td>100</td>
<td>$[889;1,981]$</td>
<td>100</td>
</tr>
<tr>
<td>Point with Moment Selection</td>
<td>$[6,810;19,844]$</td>
<td>100</td>
<td>$[889;1,981]$</td>
<td>100</td>
</tr>
<tr>
<td>Set with Moment Selection</td>
<td>$[6,810;19,844]$</td>
<td>100</td>
<td>$[889;1,981]$</td>
<td>100</td>
</tr>
<tr>
<td>Point with Shifted Mean</td>
<td>$[10,762;17,569]$</td>
<td>99.3</td>
<td>$[1,097;1,634]$</td>
<td>100</td>
</tr>
<tr>
<td>Set with Shifted Mean</td>
<td>$[9,916;17,865]$</td>
<td>99.6</td>
<td>$[1,073;1,707]$</td>
<td>100</td>
</tr>
</tbody>
</table>