Notes on Moment Inequalities

Jiding Zhang

November 7, 2017

This writeup is based on Pakes's lecture notes[1], Pakes, Porter, Ho, and Ishii (Econometrica, 2015), Chernozhukov's lecture notes[2], Chernozhukov, Hong and Tamer (Econometrica, 2007), and Bugni (Econometrica, 2010).

1 The PPHI Method

1.1 Notes

For linear moment inequalities, the PPHI method obtains the asymptotic distribution of extreme points of the polyhedron (the set of estimators).

Let $j = 1, \ldots, J$ be markets (?), wherein we observe individual choices.

- revealed preference: $\mathcal{E}\left[\Delta r(d_i, d'(d_i), \theta)\right] \geq 0$.
- moments in market j (approximate $\mathcal{E}$):
$$m(\theta) = \frac{1}{n_j} \sum_i \Delta r^j(d_i^j, d'(d_i^j), \theta) \otimes h(x_i^j).$$
- sample mean across markets:
$$m(\mathbf{P}_J, \theta) = \frac{1}{J} \sum_{j=1}^{J} m(\theta).$$
- estimation and inference (focus on the extreme point of one dimension of θ)
$$\hat{\Theta}_J = \arg\min_\theta \left\| \hat{D}_J^{-1/2} m(\mathbf{P}_J, \theta)_- \right\|$$
where $\hat{D}_J$ is a diagonal matrix (variance), and is a consistent estimator of the true variance of moments. $\hat{D}_J$ could be obtained by a two-step estimation (first obtain a consistent estimator of θ by doing Minimum Distance without weighing, and evaluate the sample variance of moments at the consistent estimator).

[1] https://canvas.harvard.edu/courses/5808/files/1424891

[2] https://ocw.mit.edu/courses/economics/14-385-nonlinear-econometric-analysis-fall-2007/lecture-notes/lecture12.pdf

- asymptotic distribution for the estimator of boundary (say $\underline{\hat{\theta}}_1$): $\sqrt{J}(\underline{\hat{\theta}}_1 - \underline{\theta}_1) \xrightarrow{d} \underline{\hat{\tau}}_1$, where

$$\underline{\hat{\tau}}_1 = \min\left\{\tau_1 : D^0(\underline{\theta})^{-1/2}\Gamma^0(\underline{\theta})\tau + Z^0 \geq 0\right\}.$$

  Here superscript 0 denotes the binding moments $\mathcal{P}m(\underline{\theta}) = 0$. $Z^0 \sim N(0, \Omega^0(\underline{\theta}))$, where $\Omega^0(\underline{\theta})$ is the correlation matrix of moments, and $\Gamma^0(\underline{\theta})$ is the Jacobian matrix. Both could be estimated consistently by evaluating at a consistent estimator $\underline{\hat{\theta}}$. Only the binding moments matter here. The difficulty is we do not know which population moments are binding.

- Procedures (obtaining a point): Get consistent estimators $\hat{\Gamma}_J$ and $\hat{\Omega}_J$. Take $Z^* \sim N(0, \hat{\Omega}_J)$. Consider inequalities:

$$0 \leq \hat{D}^{-1/2}\hat{\Gamma}\tau + Z^* + r_J\left(\hat{D}^{-1/2}m(\mathbf{P}_J, \underline{\hat{\theta}})\right)_+$$

  where $r_J = o(\sqrt{J}/\sqrt{2\ln\ln J})$. If we find the solution, then we minimize τ_1. If not, eliminate moments in the order of $\hat{D}_J^{-1/2}m_j(\mathbf{P}_J, \underline{\hat{\theta}})$ starting from the largest value, until a solution exists. Solve the stochastic LP with the remaining moments (indices denoted by s):

$$\underline{\tau}_1^* = \min\left\{\tau_1 : 0 \leq \hat{D}_s^{-1/2}\hat{\Gamma}_s\tau + Z_s^* + r_J\left(\hat{D}_s^{-1/2}m_s(\mathbf{P}_J, \underline{\hat{\theta}})\right)_+\right\}.$$

Comment: I suspect the convergence rate ($\sqrt{J}$ in the paper and notes) should be $\sqrt{\sum_{i=1}^J n_j}$. The definition of "market" is not clear

1.2 Comments on Ho and Pakes (AER, 2014)

Ho and Pakes (AER, 2014) uses the PPHI method to construct the confidence interval for the parameter (scalar).

- In `ineq_setup.m`, the moments are not weighed by number of switches.

- In `ineq_run.m`, `varterm3` and `Sigma3` (vcov of moments) are not defined in any files.

- They state that they calculate the vcov (`Sigma3`), before knowing the parameter. It should be a function of the parameter value.

- Technically we can follow the steps and estimate θ consistently, then estimatate vcov, and follow the steps to take draws, evaluate upper bounds and lower bounds, drop moments (if $UB < LB$) to get a distribution. Since this problem is linear, Jacobian matrix is equal to $\Delta p - \Delta p'$.

- For the sequence $o(\sqrt{J}/\sqrt{2\ln\ln J})$, they use $\sqrt{Jtot}/\sqrt{2\ln\ln Jtot}$ where $Jtot$ is the sum (over (h, z)) of the number of patients who can switch to hospital h with instrument z.

- They use $Jtot$ for each component of the moment to obtain the standard error. This is not consistent with the PPHI paper.

2 The CHT Method

2.1 Notes

The CHT method obtains confidence regions for the identified set and the true parameter.

- identified set: $\Theta_I = \{\theta : \mathbb{E}_P[m_i(\theta)] \leq 0\}$.
- criterion function:
$$Q(\theta) \triangleq \left\| (\mathbb{E}_p[m_i(\theta)])' W^{1/2}(\theta) \right\|_+^2 .$$
- empirical analog:
$$Q_n(\theta) \triangleq \left\| \left(\frac{1}{n} \sum_{i=1}^{n} m_i(\theta) \right)' W_n^{1/2}(\theta) \right\|_+^2 .$$
- estimation:
$$\hat{\Theta}_I = C_n(\tilde{c}) = \{\theta : nQ_n(\theta) \leq \tilde{c}\},$$
where $\tilde{c} \geq \mathcal{C}_n$ with probability approaching 1 but $\tilde{c}/n \xrightarrow{p} 0$. $\tilde{c}$ could take the value of $\ln n$. The rate of convergence is approximately $1/\sqrt{n}$. Sometimes we can pick $\tilde{c} = 0$ (Condition C.3 in CHT).
- confidence region:
$$CR = \{\theta : nQ_n(\theta) \leq \hat{c}\}$$
where $\hat{c}$ is the α-th quantile of
$$\mathcal{C}_n = \sup_{\theta \in \Theta_I} nQ_n(\theta).$$
- obtaining $\hat{c}$ from simulation (Theorem 4.2 in CHT): Suppose moment functions has the Donsker property:
$$\sqrt{n} \left(\mathbb{E}_n[m_i(\theta)] - \mathbb{P}[m_i(\theta)] \right) \xrightarrow{d} \Delta(\theta)$$
where $\Delta(\theta)$ is a mean zero Gaussian process. Define
$$\mathcal{C}(\theta) \triangleq \left\| (\Delta(\theta) + \xi(\theta))' W^{1/2}(\theta) \right\|_+^2$$
where the j-th component of $\xi(\theta)$ is $-\infty$ if the j-th population moment is not binding ($\mathbb{E}_P[m_{ij}(\theta)] < 0$) and 0 otherwise ($\mathbb{E}_P[m_{ij}(\theta)] = 0$). Then,
$$\mathcal{C}_n \xrightarrow{d} \mathcal{C} \triangleq \sup_{\theta \in \Theta_I} \mathcal{C}(\theta).$$
- bootstrapping to obtain the distribution of $\hat{c}$:
    1. Take a draw of z_i^*: a n-vector of i.i.d. $N(0, 1)$ variables.

2. For constants $\{c_j\}_{j=1}^J$, calculate $\mathcal{C}_n^*$:

$$\mathcal{C}_n^* = \sup_{\theta\in\hat{\Theta}_I} \left\| (\Delta_n^*(\theta) + \hat{\xi}(\theta))' W_n^{1/2}(\theta) \right\|_+^2 ,$$

where:

$$\hat{\xi}_j(\theta) = \begin{cases} -\infty & \mathbb{E}_n[m_{ij}(\theta)] \leq -c_j\sqrt{\log n/n}, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\Delta_n^*(\theta) = n^{-1/2}\sum_{i=1}^n [m_i(\theta) z_i^*].$$

3. Repeat steps 1 – 2 to obtain a distribution of $\mathcal{C}_n^*$.

4. Take the α-th quantile of $\mathcal{C}_n^*$.

2.2 Comments

- Set estimator: For the simple case, let $W(\theta)$ be a constant matrix. The criterion function could be expressed as

$$Q(\theta) = \|(A\theta)_+\|^2 .$$

  for some matrix A. Note, $\|(A\theta)_+\|$ a convex function (the 2-norm of $(A\theta)_+$, see page 87 of Boyd and Vandenberghe (2004)). Since the norm is non-negative, by definition $Q(\theta)$ is convex. The set estimator is a sublevel set of $Q(\theta)$, hence it is a convex set.

- Computationally this method might be hard, because the objective is a non-smooth convex function.

- Conditon C.3 specifies the degenerate condition. Consider the simple case: moments are linear, i.e., $\mathbb{E}_P[A\theta] \leq 0$. If we could find Θ_n for any n, such that for any $\theta \in \Theta_n$, $\mathbb{E}_n[A\theta] = 0$, then $\{\theta : \mathbb{E}_n[A\theta] = 0\}$ is consistent at the $1/\sqrt{n}$ rate.

3 The Bugni Method

3.1 Notes

The Bugni method obtains confidence regions for the identified set.

- identified set:

$$\Theta_I = \left\{ \theta \in \Theta : \left\{ \mathbb{E}[m_j(Z,\theta)] \leq 0 \right\}_{j=1}^J \right\}.$$

- criterion function:

$$Q(\theta) = G\left(\left\{ \left[\mathbb{E}[m_j(Z,\theta)] \right]_+ \right\}_{j=1}^J \right)$$

  where $G(x) = \sum_{j=1}^J w_j x_j$ or $G(x) = \max\{w_j x_j\}$ for arbitrary positive constants $\{w_j\}_{j=1}^J$ (G is the weighting function).

- set estimator:

$$\hat{\Theta}_I(\tau_n) = \left\{ \theta \in \Theta : \left\{ \mathbb{E}_n m_j(Z,\theta) \le \tau_n/\sqrt{n} \right\}_{j=1}^{J} \right\}$$

where $\tau_n/\sqrt{n} = o(1)$ and $\sqrt{\ln\ln n}/\tau_n = o(1)$.

- bootstrap:

1. calculate for s times for bootstrap samples of size n with replacement from the data:

$$\Gamma_n^* = \begin{cases} \sup_{\theta \in \hat{\Theta}_I(\tau_n)} G\left(\left\{ \sqrt{n}\Big[\mathbb{E}_n^*[m_j(Z,\theta)] - \mathbb{E}_n[m_j(Z,\theta)]\Big]_+ \times \mathbb{I}\Big(|\mathbb{E}_n[m_j(Z,\theta)]| \le \tau_n/\sqrt{n}\Big) \right\}_{j=1}^{J} \right) & \hat{\Theta}_I(\tau_n) \neq \emptyset \\ 0 & \text{o.w.} \end{cases}$$

where $\mathbb{E}_n^*$ is the sample mean in the bootstrap samples.

2. Let $\hat{c}_n^B(1-\alpha)$ be the $(1-\alpha)$ quantile of the distribution of Γ_n^*. Then the $(1-\alpha)$ confidence set is

$$\hat{C}_n^B(1-\alpha) = \left\{ \theta \in \Theta : G\left(\left\{ \sqrt{n}\mathbb{E}_n[m_j(Z,\theta)]_+ \right\}_{j=1}^{J} \right) \le \hat{c}_n^B(1-\alpha) \right\}.$$

- parameters to tune: w_j, τ_n.

- The bootstrap might require recalculating moment states, if some of the moment states are obtained from a first-step estimation from data.

- In practice we can pick $\tau_n = \sqrt{\ln n}$.

3.2 Linear Moment Inequalities

- Let $O_{n\times K}$ be the matrix of observations of moment states, where n is the number of observations, and K is the dimension of θ. Let $V_{n\times J}$ be the matrix of IVs, where J is the number of IVs for each observation.

$$\mathbb{E}_n m(Z,\theta) = \frac{1}{n}\sum_{i=1}^{n} m^i(Z^i,\theta) = \frac{1}{n}V'O\theta.$$

- set estimator:

$$\hat{\Theta}_I(\tau_n) = \left\{ \theta \in \Theta : \frac{1}{n}V'O\theta \le \frac{\tau_n}{\sqrt{n}}\mathbf{1} \right\}$$

where $\mathbf{1}$ is a vector of all ones. Note that it is a polyhedron.

- To obtain a bounding box of the set, we solve linear programs: $\max_{\theta\in\Theta_I}\{\theta_j\}$ and $\min_{\theta\in\Theta_I}\{\theta_j\}$ for each j.

3.2.1 $G(x) = \max\{w_j x_j\}$

We assume the weighting function takes the form of $G(x) = \max\{w_j x_j\}$.

- Let V^*, O^* be the bootstrapped IV matrix and matrix of moment states, respectively. Also let w be the vector of weights of each moment. We rewrite the objective function in the optimization. Noting that $\theta \in \hat{\Theta}_I(\tau_n)$ guarantees that
$$\frac{1}{n}V'O\theta \leq \frac{\tau_n}{\sqrt{n}}\mathbf{1},$$
we have
$$\begin{aligned}
&\sup_{\theta\in\hat{\Theta}_I(\tau_n)} G\left(\left\{\sqrt{n}\Big[\mathbb{E}_n^*[m_j(Z,\theta)] - \mathbb{E}_n[m_j(Z,\theta)]\Big]_+ \times \mathbb{I}\Big(|\mathbb{E}_n[m_j(Z,\theta)]| \leq \tau_n/\sqrt{n}\Big)\right\}_{j=1}^J\right) \\
= &\sup_{\theta\in\hat{\Theta}_I(\tau_n)} \max_{j=1,\dots,J}\left\{w_j \max\left\{\frac{\sqrt{n}}{n}\Big((V^*)'O^* - V'O\Big)_j\theta, 0\right\} \times \mathbb{I}\left(\left(\frac{1}{n}V'O\right)_j \theta \geq -\frac{\tau_n}{\sqrt{n}}\right)\right\}.
\end{aligned}$$

- This problem could be solved by solving J LPs: For each j, solve:
$$\begin{aligned}
\max_{\theta} \quad & w_j\frac{\sqrt{n}}{n}\Big((V^*)'O^* - V'O\Big)_j\theta \\
\text{subject to} \quad & \left(\frac{1}{n}V'O\right)_j\theta + \frac{\tau_n}{\sqrt{n}} \geq 0 \\
& \theta \in \hat{\Theta}_I(\tau_n)
\end{aligned}$$
and compare the (non-negative) optimal values. If the first constraint is not satisfied, in the original problem
$$w_j \max\left\{\frac{\sqrt{n}}{n}\Big((V^*)'O^* - V'O\Big)_j\theta, 0\right\} \times \mathbb{I}\left(\left(\frac{1}{n}V'O\right)_j\theta \geq -\frac{\tau_n}{\sqrt{n}}\right) = 0.$$
So here we force the constraint to be satisfied. If the optimal value is negative with the first constraint satisfied, we simply set it to be 0.

  We also note that the optimization problem has to be bounded as long as the feasible set is bounded (otherwise it violates the Weierstrass Theorem). The boundedness of the feasible set is guaranteed if $\hat{\Theta}_I(\tau_n)$ is bounded.

- confidence set:
$$\begin{aligned}
\hat{C}_n^B(1-\alpha) &= \left\{\theta \in \Theta : \max_{j=1,\dots,J}\left\{w_j\left[\sqrt{n}\left(\frac{1}{n}V'O\right)_j\theta\right]_+\right\} \leq \hat{c}_n^B(1-\alpha)\right\} \\
&= \left\{\theta \in \Theta : w_j\sqrt{n}\left(\frac{1}{n}V'O\right)_j\theta \leq \hat{c}_n^B(1-\alpha), \forall j\right\}
\end{aligned}$$
Note, the confidence set is a set defined by linear constraints. So it is easy to obtain a bounding box by solving LPs.

3.2.2 $G(x) = \sum_j w_j x_j$

Now we assume the weighting function is linear additive. Let w be the vector of weights.

- Similarly, we rewrite the expression of Γ_n^*. Let $\circ$ be the Hadamard product (pointwise product).

$$\Gamma_n^* = \begin{cases} \sup_{\theta \in \hat{\Theta}_I(\tau_n)} w' \left[\frac{\sqrt{n}}{n} \left[\left((V^*)'O^* - V'O \right) \theta \right]_+ \circ \mathbb{I}\left(\frac{1}{n} V'O\theta \geq -\frac{\tau_n}{\sqrt{n}} \mathbf{1} \right) \right] & \hat{\Theta}_I(\tau_n) \neq \emptyset \\ 0 & \text{o.w.} \end{cases}$$

- The objective function is the sum of J components. We try to reformulate this problem as MILP.

  Step 1: Getting rid of the indicator function.

  Consider component j, which could be expressed as

$$\begin{aligned} & w_j \max\left\{ \frac{\sqrt{n}}{n} \left((V^*)'O^* - V'O \right)_j \theta, 0 \right\} \times \mathbb{I}\left(\left(\frac{1}{n} V'O \right)_j \theta \geq -\frac{\tau_n}{\sqrt{n}} \right) \\ = \quad & w_j \max\left\{ 0, \frac{\sqrt{n}}{n} \left((V^*)'O^* - V'O \right)_j \theta - M y_j \right\} \end{aligned}$$

  where M is a large constant, and y_j is a binary variable:

$$y_j = \begin{cases} 0 & \left(\frac{1}{n} V'O \right)_j \theta \geq -\frac{\tau_n}{\sqrt{n}} \\ 1 & \text{o.w.} \end{cases}$$

  We construct the following constraints:

$$\left(\frac{1}{n} V'O \right)_j \theta + \frac{\tau_n}{\sqrt{n}} - M(1 - y_j) < 0, \tag{1}$$

$$\left(\frac{1}{n} V'O \right)_j \theta + \frac{\tau_n}{\sqrt{n}} + M y_j \geq 0. \tag{2}$$

  If $\left(\frac{1}{n} V'O \right)_j \theta + \frac{\tau_n}{\sqrt{n}} \geq 0$, y_j is forced to be 0 by constraint (1). If $\left(\frac{1}{n} V'O \right)_j \theta + \frac{\tau_n}{\sqrt{n}} < 0$, y_j is forced to be 1 by constraint (2).

  Step 2: Getting rid of $\max\{0, \cdot\}$.

  Note,

$$\begin{aligned} & \max \quad w_j \max\left\{ 0, \frac{\sqrt{n}}{n} \left((V^*)'O^* - V'O \right)_j \theta - M y_j \right\} \\ \Leftrightarrow \quad & \max \quad w_j v_j \left[\frac{\sqrt{n}}{n} \left((V^*)'O^* - V'O \right)_j \theta - M y_j \right] \end{aligned}$$

  where v_j is a binary variable.

  Now let's construct a new variable z_j, such that

$$z_j = \begin{cases} \frac{\sqrt{n}}{n} \left((V^*)'O^* - V'O \right)_j \theta - M y_j & v_j = 1 \\ 0 & v_j = 0 \end{cases}$$

  We again construct constraints, where L is a large constant:

$$z_j \leq \frac{\sqrt{n}}{n} \left((V^*)'O^* - V'O \right)_j \theta - M y_j + (1 - v_j) L, \tag{3}$$

$$z_j \leq v_j L. \tag{4}$$

If $v_j = 1$, z_j can be as large as $\frac{\sqrt{n}}{n}\Big((V^*)'O^* - V'O\Big)_j \theta - My_j$ according to constraint (3). If $v_j = 0$, z_j can be as large as 0 by constraint (4).

Finally, we reformulate the optimization problem:

$$\begin{aligned}
\max_{\theta,y,v,z} \quad & \sum_{j=1}^{J} w_j z_j \\
\text{subject to} \quad & \left(\frac{1}{n}V'O\right)_j \theta + \frac{\tau_n}{\sqrt{n}} - M(1-y_j) < 0 \quad \text{for } j = 1,\dots,J \\
& \left(\frac{1}{n}V'O\right)_j \theta + \frac{\tau_n}{\sqrt{n}} + My_j \geq 0 \quad \text{for } j = 1,\dots,J \\
& z_j \leq \frac{\sqrt{n}}{n}\Big((V^*)'O^* - V'O\Big)_j \theta - My_j + (1-v_j)L \quad \text{for } j = 1,\dots,J \\
& z_j \leq v_j L \quad \text{for } j = 1,\dots,J \\
& y_1,\dots y_J, v_i, \dots v_J \in \{0,1\} \\
& \theta \in \hat{\Theta}_I(\tau_n)
\end{aligned}$$

- confidence set:

$$\begin{aligned}
\hat{C}_n^B(1-\alpha) &= \left\{\theta \in \Theta : G\left(\left\{\sqrt{n}\mathbb{E}_n[m_j(Z,\theta)]_+\right\}_{j=1}^{J}\right) \leq \hat{c}_n^B(1-\alpha)\right\} \\
&= \left\{\theta \in \Theta : w'\left(\frac{\sqrt{n}}{n}\Big[\Big((V^*)'O^* - V'O\Big)\theta\Big]_+\right) \leq \hat{c}_n^B(1-\alpha)\right\}
\end{aligned}$$

The confidence set is convex.

To obtain a bounding box, we reformulate the constraint set.

Each component in the inequality takes the form of $w_j \max\{A_j\theta, 0\}$. We have the following:

$$\max\{A_j\theta, 0\} = y_j \Leftrightarrow \max\{A_j\theta, 0\} \leq y_j, \max\{A_j\theta, 0\} \geq y_j.$$

Also note,

$$\max\{A_j\theta, 0\} \leq y_j \Leftrightarrow \begin{cases} A_j\theta \leq y_j \\ 0 \leq y_j \end{cases}$$

$$\max\{A_j\theta, 0\} \geq y_j \Leftrightarrow \begin{cases} A_j\theta \leq Mz_j \\ 0 \leq A_j\theta + M(1-z_j) \\ A_j\theta + M(1-z_j) \geq y_j \\ Mz_j \geq y_j \\ z_j \in \{0,1\} \end{cases}$$

Thus we can find the bounding box by solving MILPs.