Partial Identification: Testing Many Moment Inequalities via One Sided Thresholded Lasso

Federico Bugni Mehmet Caner Anders Bredahl Kock
Soumendra Lahiri

AUGUST 2015

- Interest centers on many moment inequalities which is linked to partial identification literature.
- The number of many moment inequalities are denoted by p, which will be larger than sample size n.
- Some examples: market structure model of Ciliberto and Tamer (2009), discrete choice model with endogeneity of Chesher-Rosen-Smolinski (2013), dynamic model of imperfect competition of Bajari-Benkard-Levin (2007).
- The first solution to this problem is proposed by Chernozhukov, Chetverikov, Kato (2014).
- This is a major development since the critical values for any test are not developed and face immense difficulty due to high dimensional nature of the problem.

- Let $X_1, \cdots, X_n$ be a sequence of iid random vectors in R^p, $X_i = (X_{i1}, \cdots, X_{ip})'$.
- For $1 \leq j \leq p$, write $\mu_j = E[X_{1j}]$
- Chernozhukov etal (2014) test the null of $H_0 : \mu_j \leq 0$ versus the alternative $H_1 : \mu_j > 0$.

- Define $\hat{\mu}_j = n^{-1}\sum_{i=1}^{n} X_{ij}$, $\hat{\sigma}_j^2 = n^{-1}\sum_{i=1}^{n}(X_{ij} - \hat{\mu}_j)^2$.
- Chernozukov etal (2014) proposed the following test statistic
- $T = \max_{\{1 \le j \le p\}} \frac{\sqrt{n}\hat{\mu}_j}{\hat{\sigma}_j}$. By abusing a bit statistics, we can call their test *maxt*.
- Reject H_0 when $T > cval$, where *cval* is a critical value that is chosen to give certain size.
- The issue is getting *cval* such that we have a certain size in cases when $p > n$, regular central limit theorem type results do not suffice.
- Chernozukov etal proposed two methods: a) self normalization based b) bootstrap based techniques to get such *cval*.

IDEA FOR TESTING:
Chernozukov etal (2014)

- This will be a two step process.
- In the first step, *maxt* test will be conducted to get rid of large inequalities on the left side (negatives). Here a critical value will be used, $cval1$.
- After getting rid of these, critical value of the second step will be formed by use only the inequalities that are larger than the ones in the first step. Lets denote this critical value $cval2$
- In the second step, *maxt* test will again be conducted, and we will use $cval2$ for testing.
- Now we will describe how $cval1, cval2$ are formed.

- First step critical values of Chernozukov etal (2014) are (using moderate deviation inequality for self normalized sums), for $\beta_n \to 0$ when $n \to \infty$.
  Let $\Phi^{-1}(.)$ denote the quantile of the distribution function of standard normal.
- $$cval1 = \frac{\Phi^{-1}(1 - \beta_n / p)}{\sqrt{1 - \Phi^{-1}(1 - \beta_n / p)^2 / n}}.$$
- They form the following set that gets rid of large negative inequalities
- $$\hat{J}_{SN} = \{j \in \{1, \cdots, p\} : \sqrt{n}\hat{\mu}_j / \hat{\sigma}_j > -2cval1\}.$$
- Denote the cardinality of the set $\hat{J}_{SN}$ as $\hat{k}$, so $\hat{k} = |\hat{J}_{SN}|$.

Second Step Critical Value:
a) if $\hat{k} = 0$, $cval2 = 0$.
b) if $\hat{k} \geq 1$, then

$$cval2 = \frac{\Phi^{-1}(1 - (\alpha - 2\beta_n)/\hat{k})}{\sqrt{1 - \Phi^{-1}(1 - (\alpha - 2\beta_n)/\hat{k})^2/n}}.$$

Theorem 4.2 (Chernozhukov etal) (2014). *Suppose that* $\sup_n \beta_n \leq \alpha/3$, *and there exist constants* $0 < c_1 < 1/2$, $C_1 > 0$ *such that*

$$M_{n,3} log^{3/2}(p/\beta_n) \leq C_1 n^{1/2-c_1}, \quad B_n^2 log^2(p/\beta_n) \leq C_1 n^{1/2-c_1}.$$

Then there exist positive constants c, C *depending on* α, c_1, C_1 *such that under* H_0

$$P(T > cval2) \leq \alpha + Cn^{-c},$$

where the result is uniform with respect to common distribution of X_i *where finite second moment and positive variance conditions and the above moment-p tradeoff conditions are verified.*

Note that $M_{n,3} = \max_{1 \leq j \leq p} (E|Z_{1j}|^3)^{1/3}$, and $Z_{1j} = (X_{1j} - \mu_j)/\sigma_j$.
$B_n^2 = (E[\max_{1 \leq j \leq p} Z_{1j}^4])^{1/2}$.

Multiplier Bootstrap Based Critical Values(Chernozhukov etal, 2014):

Step 1. Generate cval1 according to the following algorithm.

i) Draw independent std normal revs $\epsilon_1, \cdots, \epsilon_n$ independent of $X's$.

ii) Construct the multiplier bootstrap test statistic

$$W^{MB} = \max_{1 \leq j \leq p} \frac{n^{-1/2} \sum_{i=1}^{n} \epsilon_i (X_{ij} - \hat{\mu}_j)}{\hat{\sigma}_j}.$$

iii) Calculate cval1 as the conditional $1 - \beta_n$ quantile of W^{MB} given the data.

Step 2: Delete large negative inequalities:

- Define the following set:
-

$$\hat{J}_{MB} = \{j \in 1, \cdots, p : \sqrt{n}\hat{\mu}_j / \hat{\sigma}_j > -2cval1\}. \tag{1}$$

Bootstrap Algorithm to create *cval*2:

- Generate independent std normal revs $\epsilon_1, \cdots, \epsilon_n$ independent of data X.
- Construct the multiplier bootstrap test statistic

$$W_{\hat{J}_{MB}} = \max_{j \in \hat{J}_{MB}} \frac{n^{-1/2} \sum_{i=1}^{n} \epsilon_i (X_{ij} - \hat{\mu}_j)}{\hat{\sigma}_j}.$$

- If $\hat{J}_{MB}$ is empty, set $cval2 = 0$
- Otherwise $cval2$ is the $(1 - (\alpha - 2\beta_n))$ conditional quantile of $W_{\hat{J}_B}$ given X.

- *Assumption: Suppose there exist positive constants* $0 < c_1 < 1/2$ *and* $C_1 > 0$ *such that*

$$(M_{n,3}^3 \cup M_{n,4}^2 \cup B_n)^2 log^{7/2}(pn) \leq C_1 n^{1/2-c_1}.$$

- **Theorem 4.4. Chernozhukov etal 2014**. (Validity of two step MB method). *Suppose that above assumption is satisfied, and* $\sup_n \beta_n < \alpha/2$, $log(1/\beta_n) \leq C_1 logn$. *Then there exist positive constants* c, C *depending only on* c_1, C_1 *such that under* H_0

$$P(T > cval2) \leq \alpha + Cn^{-c}.$$

- Note that the results are uniform over the distribution of X_i for which finite second moments, and positive variance are verified with the assumption above.

Difference Between Two Methods

- Our method uses the same test statistic as Chernozukov etal (2014).
- But we choose the critical value in step 1 ($cval1$) differently.
- Chernozhukov etal (2014) use the same max test to eliminate the large negative inequalities in step 1.
- We will use one sided thresholded lasso or thresholded least squares to eliminate large negative as well as moderate negative inequalities in our step 1.
- Instead of pretesting we are estimating in the first step.
- Our aim is to have the same size, but gain from power by adjusting critical values in step 1 differently.

Slight Generalization of The Previous Model:

$$H_0 : \mu_j \leq 0, \forall j = 1, \cdots, p \text{ and } \mu_j = 0 \forall j = p+1, \cdots, k.$$

$$H_1 : \mu_j > 0, \text{ for some } j = 1, \cdots, p \text{ or } \mu_j \neq 0 \text{ for some } j = p+1, \cdots, k.$$

Test statistics: Simple Generalization of *maxt* test of Chernozhukov etal (2014)

$$T_n = max\{ \max_{j=1,\cdots,p} \frac{\sqrt{n}\hat{\mu}_j}{\hat{\sigma}_j}, \max_{s=p+1,\cdots,k} \frac{\sqrt{n}|\hat{\mu}_s|}{\hat{\sigma}_s} \}.$$

Step 1: Involves lasso estimation

1. $$\hat{\mu}_L = argmin_{t \in R^p}\{(\hat{\mu} - t)'\hat{W}(\hat{\mu} - t) + \lambda_n \|\hat{W}^{1/2}t\|_1\}, \qquad (2)$$

   where
2. $\hat{\mu} = (\hat{\mu}_1, \cdots, \hat{\mu}_j, \cdots, \hat{\mu}_p)'$. Remember that $\hat{\mu}_j = n^{-1}\sum_{i=1}^n X_{ij}$.
3. $\hat{W} \equiv diag\{1/\hat{\sigma}_j^2\}_{j=1}^p$, which is a $p \times p$ diagonal matrix.
4. λ_n is a positive tuning parameter, that converges to zero.
5. Form the following set:

$$\hat{J}_L \equiv \{j = 1, \cdots, p : \hat{\mu}_{j,L}/\hat{\sigma}_j \geq -\lambda_n\}.$$

So inequalities in that set will be kept in forming the critical value for the test in the second step.

Duality between Lasso and Least squares in our case:

- $$\hat{\mu}_{L,j} = sgn(\hat{\mu}_j)[|\hat{\mu}_j| - \hat{\sigma}_j \frac{\lambda_n}{2}]_+. \quad (3)$$
- This is shown by Buhlmann-van de Geer (2009), equation (2.5).
- We can write the set for binding moments, (instead of $\hat{J}_L$) via least squares thresholding
- $$\hat{J}_{LS} \equiv \{j = 1, \cdots, p : \hat{\mu}_j / \hat{\sigma}_j \geq -\frac{3}{2}\lambda_n\}.$$
- This is a major computational advantage over lasso based set.

Step 2: Now we repeat the same bootstrap procedure of Chernozukov etal (2014) with our set of inequalities and equalities

1. Generate independent std normal revs $\epsilon_1, \cdots, \epsilon_p$ independent of the data X.
2. Construct the multiplier bootstrap test statistic, which is a simple extension of Chernozhukov etal (2014) in form, but use different set of inequalities,

$$W_L^{MB} = max\{\max_{j \in \hat{J}_L} \frac{n^{-1/2} \sum_{i=1}^n \epsilon_i (X_{ij} - \hat{\mu}_j)}{\hat{\sigma}_j}, \max_{s=p+1,\cdots,k} \frac{n^{-1/2} \sum_{i=1}^n \epsilon_i (X_{is} - \hat{\mu}_s)}{\hat{\sigma}_s}\}.$$

3. Calculate the conditional $(1-\alpha)$ quantile of W_L^{MB}, and call it $cval_{lasso}$.

Theorem: *Under comparable standard regularity assumptions on moments as in Chernozukov etal (2014) and under H_0,*

$$P(T_n > cval_{lasso}) \leq \alpha + o(1),$$

uniformly in the distribution of the data that satisfies regularity assumptions.

Our choice of λ_n is guided by theory. We put a simple plug in estimate of λ_n as

$$\lambda_n = \frac{C}{n^{1/2}} \left(\frac{M_{n,3}^2}{n^{1/3}} - \frac{1}{n} \right)^{-1/2},$$

where $C \geq 4$, and we use an estimate $\max_{1 \leq j \leq p} (n^{-1} \sum_{i=1}^n |X_{ij}|^3)^{1/3}$ for $M_{n,3}$.

SETUP:

- we tried 1000 monte carlo iterations, sample size is $n = 400$ as in Chernozhukov etal (2014), $p = 200, 500, 1000$ as in their case.
- Same setup as theirs

$$X_i = \mu + A'\epsilon_i,$$

where $\Sigma = A'A$, $\epsilon_i = (\epsilon_{i1}, \cdots, \epsilon_{ip})'$ are iid random variables with $E\epsilon_i = 0_p$, and variance of these errors is 1.

- Here as they did we consider $\epsilon_{i,j} \sim t_4/\sqrt{2}$, we also did uniform random errors, and this is in paper now.
- In the slides we show four specific designs. Number of bootstrap iterations is also 1000.

DESIGNS:

1. 1. $\mu_j = -0.8$ for 90% of p (specifically $j > p * 0.1$) and the rest 10% of the moments are zero. Null is true, and $\Sigma = \rho^{|j-k|}$, ρ takes values 0, 0.5, 0.9. Design 4 in Chernozukov etal (2014).
2. 2. $\mu_j = -0.3$ for 0% of p (specifically $j > p * 0.1$) and the rest 10% is $\mu_j = 0.05$ which is violation of the null. New design. Same correlation structure as 1.
3. 3. $\mu_j = -0.5$ for 0% of p (specifically $j > p * 0.1$) and the rest 10% is $\mu_j = 0.05$ which is violation of the null. New design. Same correlation structure as 1.
4. 4.$\mu_j = -0.75$ for 0% of p (specifically $j > p * 0.1$) and the rest 10% is $\mu_j = 0.05$ which is violation of the null. Design 8 in Chernozhukov etal (2014). Same correlation structure as 1.

Table 1: Design 1, Size of the test: 5% nominal size

p	ρ	**MB Lasso**	*MB*	*MBH*	*MB2S*
	0	4.80	0.20	4.60	4.60
200	0.5	4.90	0.80	4.50	4.50
	0.9	4.80	0.60	4.70	4.70
	0	4.50	0.60	4.30	4.30
500	0.5	5.60	0.60	5.50	5.50
	0.9	5.20	0.70	5.20	5.20
	0	5.90	0.60	5.30	5.30
1000	0.5	5.10	0.60	4.80	4.80
	0.9	5.00	0.70	4.90	4.90

Note: MB Lasso is our technique which uses lasso in step 1, and then uses multiplier bootstrap, MB, MBH, MB2S are techniques of Chernozhukov etal (2014) which uses multiplier bootstrap only 1 step procedure, and self normalization in step 1, multiplier bootstrap in step 2, and both multiplier bootstraps in steps 1-2 respectively.

Table 2: Design 2, Power of the test: $\mu_j = -0.3$ for 90% of moments 10% of the moments, $\mu_j = 0.05$

p	ρ	**MB Lasso**	*MB*	*MBH*	*MB2S*
200	0	51.30	14.20	14.00	14.00
	0.5	45.70	12.70	12.60	12.60
	0.9	30.10	8.30	8.20	8.30
500	0	61.30	16.20	15.90	15.90
	0.5	55.10	14.60	14.50	14.50
	0.9	36.50	11.00	10.90	10.90
1000	0	65.10	19.30	18.90	18.90
	0.5	61.60	18.20	17.80	17.80
	0.9	44.10	13.70	13.40	13.50

Note: MB Lasso is our technique which uses lasso in step 1, and then uses multiplier bootstrap, MB, MBH, MB2S are techniques of Chernozhukov etal (2014) which uses multiplier bootstrap only 1 step procedure, and self normalization in step 1, multiplier bootstrap in step 2, and both multiplier bootstraps in steps 1-2 respectively.

Table 3: Design 3, Power of the test: $\mu_j = -0.5$ for 90% of moments 10% of the moments, $\mu_j = 0.05$

p	ρ	**MB Lasso**	*MB*	*MBH*	*MB2S*
	0	52.00	14.20	28.40	31.20
200	0.5	46.40	12.70	26.60	30.20
	0.9	33.10	8.30	16.10	20.70
	0	61.70	16.20	30.70	34.20
500	0.5	55.80	14.60	27.40	29.90
	0.9	38.60	11.00	17.20	21.00
	0	65.20	19.30	31.50	35.50
1000	0.5	61.90	18.20	29.00	33.10
	0.9	44.70	13.70	19.50	22.90

Note: MB Lasso is our technique which uses lasso in step 1, and then uses multiplier bootstrap, MB, MBH, MB2S are techniques of Chernozhukov etal (2014) which uses multiplier bootstrap only 1 step procedure, and self normalization in step 1, multiplier bootstrap in step 2, and both multiplier bootstraps in steps 1-2 respectively.

Table 4: Design 4, Power of the test: $\mu_j = -0.75$ for 90% of moments 10% of the moments, $\mu_j = 0.05$

p	ρ	**MB Lasso**	*MB*	*MBH*	*MB2S*
	0	52.00	14.20	51.30	51.40
200	0.5	46.40	12.70	45.40	45.40
	0.9	33.10	8.30	32.70	32.70
	0	61.70	16.20	61.10	61.00
500	0.5	55.80	14.60	54.50	54.70
	0.9	38.60	11.00	37.90	37.90
	0	65.30	19.30	64.40	64.30
1000	0.5	61.90	18.20	61.20	61.30
	0.9	44.70	13.70	44.00	44.20

Note: MB Lasso is our technique which uses lasso in step 1, and then uses multiplier bootstrap, MB, MBH, MB2S are techniques of Chernozhukov etal (2014) which uses multiplier bootstrap only 1 step procedure, and self normalization in step 1, multiplier bootstrap in step 2, and both multiplier bootstraps in steps 1-2 respectively.

Final thoughts after Simulations:

- Chernozukov etal (2014) critically depends on choice of β_n for the size and the power.
- Their second step uses $(1 - \alpha + 2\beta_n)$ quantile of the distribution (either self normalization based, or bootstrap based).
- So if β_n is small, then this helps in power thru full coverage, but small β_n lets also go far left in the first step by adjusting critical values far to the left and keeping a lot of large negative inequalities and reducing the power.
- So β_n choice is a tradeoff between first and second step critical values.

Final Thoughts after Simulations:

- Our method relies on λ_n choice.
- First step involves thresholded least squares estimation, so no testing there.
- We use $(1 - \alpha)$ quantiles instead of $(1 - \alpha + 2\beta_n)$ of Chernozhukov etal (2014) method in step 2, so we gain from power there.
- Small λ_n helps us to truncate near 0, so we can only keep equalities and some small negative inequalities. This increases power.
- But small λ_n in lasso also causes overfit, may create artificial negative inequalities, and that may reduce power.
- However, we show that via thresholded lasso we can prevent overfit asymptotically, and come up with a λ_n expression that ties moments and sample size to λ_n.