Moment Inequalities Cookbook

Ron Yang

February 16, 2019

This note draws on "Alternative Models for Moment Inequalities" (Pakes 2010) and "A Practical Two-Step Method for Testing Moment Inequalities" (Romano-Shaikh-Wolf 2014). I focus on techniques you need for the problem set. As a heads up, I know the notation in this section is onerous. I have tried my best to make it all consistent while maintaining precision. If you are confused about a definition in this note or Ariel's slides, or if you are interested in further details, greater generality, and proofs, you should read the original papers which are quite readable.

1 Moment Inequalities

1.1 Notation

I introduce notation from Pakes 2010.

- Agent i chooses $d_i \in D_i$ given information set $\mathcal{J}_i$. There are N_j agents in each market j; there are J markets.
- d_{-i} is the set of choices by agents other than i.
- $y_i = y(z_i, d, d_{-i}, \theta)$ is a vector of other variables which affect payoffs. These variables depend on choices (d_i and d_{-i}) and exogeneous variables z_i.
- z_i^o is the subset of exogeneous variables z_i observed by the econometrician.
- $\theta_0 \in \Theta$ is a vector of parameters.
- $\pi(d, d_{-i}, y_i, \theta) = \pi(d, d_{-i}, z_i, \theta)$ is the payoff to agent i of choosing d, given y_i, θ, and that other agents choose d_{-i}.
- $r(d, d_{-i}, y_i, \theta) = r(d, d_{-i}, z_i^o, \theta)$ is the econometrician's specification of the payoff to agent i of choosing d_i, given z_i^o and θ.

1.2 Motivation

Suppose that we are a regulator concerned about ATM operators exiting from rural areas in the UK, and we want to subsidize these ATM operators to retain at least 1 ATM in each rural area. How large would this subsidy have to be? To answer this policy question, we need to measure the incentives for ATM operators to enter or exit a market[1].

In the notation of Pakes 2010, each ATM operator i in a given market chooses $d_i \in D_i = [In, Out]$ to maximize profits $\pi(d, d_{-i}, y_i, \theta)$. The entry/exit decisions of other ATM operators are d_{-i}. Variables y_i which affect profits may include variables which depend on choices (e.g. ATM fees) and exogeneous variables (e.g. local population size). Some set of these exogeneous variables z_i^o are observed by the econometrician.

To measure the profit function, the econometrician specifies a function $r(d, d_{-i}, y_i, \theta)$ which depends on observed decisions (d, d_{-i}), observables y_i, and a vector of parameters θ. For example, $r(\cdot) = (\beta_0 + y_i)d$

[1]For more details, see the following FT article: "Rural areas boosted by banks' move to prevent ATM deserts" (Jan 22, 2019, https://www.ft.com/content/f6a51dda-1e4e-11e9-b2f7-97e4dbd3580d).

for $\theta = (\beta_0)$ is a valid specification where y_i affects the cost of entry. In our policy application, given an estimate of β_0, the policy maker could compute a counterfactual where it raises y_i through a subsidy. The rest of this note will be focused on answering: how do we estimate and make inference on θ?

For clarity, I begin by discussing single-agent discrete choice and single-agent continuous choice using the above notation. I then introduce the Pakes 2010 framework for moment inequalities. I discuss in turn the "Generalized Discrete Choice" approach and the "Profit Inequality" approach.

1.3 Preliminary: Single-Agent Discrete Choice (Logit, Probit)

Suppose that d is discrete, and there is only one agent, so I suppress d_{-i}. As you covered in IO last semester, single-agent discrete choice makes four assumptions:

Behavioral Assumption: Agent i chooses d_i to maximize expected payoffs.

$$d_i = \arg\max_{d \in D_i} \mathcal{E}[\pi(d, z_i, \theta)|\mathcal{J}_i]$$

Informational Assumption: Agents know their payoffs with certainty.

$$\mathcal{E}[\pi(d, z_i, \theta)|\mathcal{J}_i] = \pi(d, z_i, \theta)$$

Measurement Assumption: The econometrician observes payoffs up to an error.

$$\pi(d, z_i, \theta) = r(d, z_i^o, \theta) + \nu$$

Distributional Assumption: The error is i.i.d. with known parametric distribution.

$$\nu \sim_{iid} F_\nu(\theta)$$

Given these four assumptions, we can compute the probability of observing a choice d_i.

$$P(d_i|z_i^o, \theta) = P(\nu \text{ s.t. } r(d_i, z_i^o, \theta) \geq r(d', z_i^o, \theta) \quad \forall d' \in D_i \mid y_i, \theta)$$

Taking a product over i, we can form a likelihood and use MLE to estimate θ.

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \prod_i P(d_i|z_i^o, \theta)$$

1.4 Preliminary: Single-Agent Continuous Choice (Moment Equalities)

Suppose that d is continuous, $\pi(d, z_i, \theta)$ is differentiable and concave in d, and there is only one agent, so I suppress d_{-i}. We make two similar assumptions as in discrete choice and two different assumptions.

Behavioral Assumption: Agent i chooses d_i to maximize expected payoffs. At the optimal choice, the agent's first order condition holds.

$$d_i = \arg\max_{d \in D_i} \mathcal{E}[\pi(d, z_i, \theta)|\mathcal{J}_i] \iff \mathcal{E}[\frac{\partial \pi}{\partial d}(d, z_i, \theta)|\mathcal{J}_i] = 0$$

Informational Assumption: Agents are right about their payoffs on average.

$$\mathcal{E}[\pi(d, z_i, \theta)|\mathcal{J}_i] = \mathcal{E}[\pi(d, d_{-i}, z_i, \theta)] \iff \mathcal{E}[\frac{\partial \pi}{\partial d}(d, z_i, \theta)|\mathcal{J}_i] = \mathcal{E}[\frac{\partial \pi}{\partial d}(d, z_i, \theta)]$$

Measurement Assumption: The econometrician observes payoffs up to an error ν.

$$\frac{\partial \pi}{\partial d}(d, z_i, \theta) = \frac{\partial r}{\partial d}(d, z_i^o, \theta) + \nu$$

Distributional Assumption: The error is mean zero conditional on some variables x_i in z_i^o.

$$\mathcal{E}[\nu|x_i] = 0$$

Given these four assumptions, and a positive function $h(x_i)$, we can form moment equalities which hold in expectation.

$$\mathcal{E}[g(d_i, z_i^o|\theta)] = \mathcal{E}[\frac{\partial r}{\partial d}(d_i, z_i^o, \theta) \times h(x_i)] = 0$$

Armed with these moment equalities and a weight matrix W, we can use GMM to estimate θ.

$$\hat{\theta} = \arg\min_{\theta \in \Theta}(\frac{1}{J}\sum_i g(d_i, z_i^o|\theta))' W (\frac{1}{J}\sum_i g(d_i, z_i^o|\theta))$$

1.5 A Framework for Moment Inequalities

When we move to more complex settings, especially with multiple interacting agents, the preliminary single-agent methods become problematic. Pakes 2010 introduces a framework which nests a variety of decision problems to isolate these challenges.

We begin with a behavioral assumption and a counterfactual assumption.

Behavioral Assumption (C1 / Best Response Condition): Agent i chooses d_i to maximize payoffs.

$$\sup_{d \in D_i} \mathcal{E}[\pi(d, d_{-i}, y_i, \theta_0)|\mathcal{J}_i] \leq \mathcal{E}[\pi(d_i = d(\mathcal{J}_i), d_{-i}, y_i, \theta_0)|\mathcal{J}_i]$$

Distributional Assumption (C2 / Counterfactual Condition): Conditional on $\mathcal{J}_i$, z_i are independent of d_i.

$$\begin{aligned} y_i &= y(z_i, d, d_{-i}, \theta) \\ d_{-i} &= d^{-i}(d, z_i, \theta) \\ z_i|(\mathcal{J}_i, d_i = d) &\perp d \end{aligned}$$

Define the expected payoff of deviating from equilibrium choice d_i to alternate choice d', conditional on z_i and θ_0, as $\Delta\pi(d_i, d', d_{-i}, z_i, \theta_0)$.

$$\Delta\pi(d_i, d', d_{-i}, z_i, \theta_0) = \pi(d_i, d_{-i}, y_i, \theta_0) - \pi(d', d_{-i}(d', z), y_i(z_i, d', d_{-i}), \theta_0)$$

Assumptions C1 and C2 imply that given the agent's information set, the payoff to deviating must be negative.

$$\mathcal{E}[\Delta\pi(d_i, d', d_{-i}, z_i, \theta_0)|\mathcal{J}_i] \geq 0 \quad \forall d' \in D$$

This looks similar to a moment inequality condition that we can estimate. However, the econometrician faces two problems: (1) the econometrician observes $r(\cdot)$, not $\pi(\cdot)$, and (2) the econometrician does not know the agent's information set $\mathcal{J}_i$. To overcome these problems, we need additional assumptions.

These two problems can be summarized in terms of ν, the difference between the actual payoffs and the econometrician's specification.

$$\nu(d_i, d_{-i}, z_i, \theta_0) = \pi_i(d_i, d_{-i}, z_i, \theta_0) - r_i(d_i, d_{-i}, z_i^o, \theta_0)$$

We can decompose this difference into a component known by the agent (ν_2), and a residual component unknown to the agent (ν_1).

$$\nu(d_i, d_{-i}, z_i, \theta_0) = \underbrace{\mathcal{E}[\nu(d_i, d_{-i}, z_i, \theta_0) \mid \mathcal{J}_i]}_{\nu_2} - \underbrace{(\nu(d_i, d_{-i}, z_i, \theta_0) - \mathcal{E}[\nu(d_i, d_{-i}, z_i, \theta_0) \mid \mathcal{J}_i])}_{\nu_1}$$

In our single agent discrete choice setting, we assumed that $\nu_1 = 0$.

$$\mathcal{E}[\pi(d, z_i, \theta)|\mathcal{J}_i] = \pi(d, z_i, \theta), \quad \pi(d, z_i, \theta) = r(d, z_i^o, \theta) + \nu \implies (\nu(d_i, d_{-i}, z_i, \theta_0) - \mathcal{E}[\nu(d_i, d_{-i}, z_i, \theta_0) \mid \mathcal{J}_i]) = 0$$

In our single agent continuous choice setting, we assumed that (conditional on some instruments), $\nu_2 = 0$.

$$\mathcal{E}[\nu|x_i] = 0$$

1.6 Generalized Discrete Choice

We make an informational assumption and a measurement/distributional assumption.

Informational Assumption (DC3): Agents know their payoffs with certainty.

$$\pi_i(d_i; d_{-i}, x) = \mathcal{E}_i[\pi_i(d_i; d_{-i}, x)|\mathcal{J}_i]$$

Measurement/Distributional Assumption (DC4): The ν_2 error is i.i.d. with known parametric distribution.

$$\begin{aligned} r(d, d_{-i}, z_i^o, \theta) &= \pi(d, d_{-i}, z_i, \theta) + \nu_{2,i} \quad \forall d \in D_i \\ z_i &= (\{\nu_{2,i,d}\}_d, z_i^o) \\ (\nu_{2,i}, \nu_{2,-i})|_{d, z_i^o, z_{-i}^o} &\sim F(\cdot, \theta_0) \end{aligned}$$

Since the agent observes their payoffs without error, this yields the condition

$$\Delta\pi_i(d_i, d'; d_{-i}, z_i^o, \nu_{2,i}, \theta_0) \geq 0$$

We can use the known distribution of ν_2 to get an upper and lower bound on the probability of observing (d_i, d_{-i}).

$$\begin{aligned} \overline{P}((d_i, d_{-i})|\theta) &= P(\{\nu_{2,i}, \nu_{2,-i}\}|\Delta\pi_i(d_i, d'; d_{-i}, z_i^o, \nu_{2,i}, \theta_0) \geq 0 \text{ holds at } (d_i, d_{-i})\} \\ \underline{P}((d_i, d_{-i})|\theta) &= P(\{\nu_{2,i}, \nu_{2,-i}\}|\Delta\pi_i(d_i, d'; d_{-i}, z_i^o, \nu_{2,i}, \theta_0) \geq 0 \text{ holds at only } (d_i, d_{-i})\} \end{aligned}$$

Given estimates of these bounds, and some positive function $h(\cdot, \cdot)$, we can form moment inequalities which will hold at $\theta = \theta_0$.

$$\begin{aligned} g_u(d_i, d_{-i}, \theta_0) &= (\overline{P}((d_i, d_{-i})|\theta_0) - P((d_i, d_{-i})|\theta)) \times h(z_i^o, z_{-i}^o) \geq 0 \\ g_l(d_i, d_{-i}, \theta_0) &= (P((d_i, d_{-i})|\theta) - \underline{P}((d_i, d_{-i})|\theta_0)) \times h(z_i^o, z_{-i}^o) \geq 0 \end{aligned}$$

1.7 Profit Inequalities

We make an informational assumption and a measurement/distributional assumption.

Informational Assumption (PC3): Agents are right about their payoffs on average.

$$\begin{aligned} \exists h(\cdot) \geq 0,\ x_i \in \mathcal{J}_i \ s.t. \ \frac{1}{N_j}\sum_i \mathcal{E}[\Delta\pi(d_i, d', d_{-i}, z_i, \theta_0)|x_i] \geq 0 \\ \implies E[\frac{1}{N_j}\sum_i \Delta\pi(d_i, d', d_{-i}, z_i, \theta_0)h(x_i)] \geq 0 \end{aligned}$$

What does this get us? Take the following condition from C1 & C2.

$$\mathcal{E}[\Delta\pi|\mathcal{J}_i] = \mathcal{E}[\Delta r|\mathcal{J}_i] - \Delta\nu_2 \geq 0$$

With PC3, take sample averages of this expression over markets.

$$\frac{1}{N_j}\sum \mathcal{E}[\Delta\pi|\mathcal{J}_i] \geq 0 \implies E[\frac{1}{N_j}\sum_i \Delta r h(x) - \Delta\nu_2 h(x)] \geq 0$$

If $E[\Delta\nu_2 h(x)] = 0$, then we have the moment inequalities

$$E[\frac{1}{N_j}\sum_i \Delta r(\cdot|\theta)h(x)] \geq 0$$

This is a market-level condition: in a given market j, the average over firms i of individual-level moments is non-negative. Put another way, we are sampling markets j, not firms i. Asymptotic results will therefore be taken as $J \to \infty$, not as $N_j \to \infty$. The sample analogue of this expression is, for i.i.d. J markets and N_j firms in each market:

$$\frac{1}{J}\sum_j \frac{1}{N_j}\sum_i \Delta r(\cdot|\theta)h(x)$$

If there does not exist $h(x)$ for $\Delta\nu_2$ such that $E[\Delta\nu_2 h(x)] = 0$, Pakes discusses two alternative approaches:

Measurement/Distributional Assumption (PC4a / Differencing): There exist two choices d and d' with the same value of ν_2.

Suppose we have groups G and weights $w_{i,g} \in \mathcal{J}_{i,g}$ such that

$$\sum_{i\in g} w_{i,g} \times \Delta\nu_{2,i,g,d_{i,g},d'_{i,g}} = 0$$

That is, within group, ν_2 differences weighted by $w_{i,g}$ are zero. Then, comparing choices within group will have no ν_2, so we construct moment inequalities. For intuition,

Measurement/Distributional Assumption (PC4a / IV & Reference Choice): There exists a choice whose payoff we observe precisely, with no ν_2, and we have an instrument for ν_2.

Suppose $\forall d \in D_i$, we have $d' \in D_i$ and $w_i \in \mathcal{J}_i$ such that

$$w_i\Delta r = w_i\mathcal{E}[\Delta\pi|\mathcal{J}_i] + \nu_{2,i} + \Delta\nu_{1,i}$$

Then, if $x_i \in \mathcal{J}_i$, $E[\nu_{2,i}|x_i] = 0$, and $h(\cdot) > 0$,

$$\frac{1}{N}\sum_i w_i\Delta r h(x_i) \to_p \frac{1}{N}\sum_i w_i\mathcal{E}[\Delta\pi|\mathcal{J}_i]h(x_i) \geq 0$$

2 Inference

2.1 Notation

I introduce some notation from RSW. This deviates from Ariel's slides, but is internally consistent and clarifies the confidence set algorithm.

- W_i for $i \in [1 \dots n]$ is an i.i.d. sequence of random K-dimensional vectors on $\mathbb{R}^K$ with true distribution $P \in \mathbf{P}$.
- $\hat{P}_n$ is the empirical distribution of W_i for $i \in [1 \dots n]$.
- $\mu(P)$ is the K-dimensional mean of P. $\mu_j(P)$ is the j-th dimension of $\mu(P)$.
- $\bar{W}_n = \mu(\hat{P}_n)$ and $\bar{W}_{j,n} = \mu_j(\hat{P}_n)$.
- $\Sigma(P)$ is the K by K covariance matrix of P. $\sigma_j^2(P)$ is the variance of the j-th component of P.
- $\Omega(P)$ is the K by K correlation matrix of P.
- $\hat{\Omega}_n = \Omega(\hat{P}_n)$ and $S_{j,n}^2 = \sigma_j^2(\hat{P}_n)$. $S_n^2 = \text{diag}(S_{1,n}^2 \dots S_{k,n}^2)$.

2.2 Motivation

From part 1, we obtained moment conditions (our W_i's). For every value of θ, we have a corresponding sequence of W_i's.

$$\mathcal{E}[g(d_i, d_{-i}, X_i, \theta)] = \mathcal{E}[W_i] \leq 0$$

We are interested in estimating the **identified set**, $\Theta_0(P)$, which contains those parameter vectors θ such that the moment conditions hold.

$$\Theta_0(P) = \{\theta \in \Theta \mid \mathcal{E}_P[g(d_i, d_{-i}, X_i, \theta)] \leq 0\}$$

This can be estimated by constructing the analogous object for the empirical distribution, $\Theta_0(\hat{P}_n)$.

$$\Theta_0(\hat{P}_n) = \{\theta \in \Theta \mid \mathcal{E}_{\hat{P}_n}[g(d_i, d_{-i}, X_i, \theta)] \leq 0\}$$

For inference, we are interested in a **confidence set** $\mathcal{C}_n \subseteq \Theta$ such that for any distribution P, the probability that any θ in the identified set is greater than $1 - \alpha$.

$$\lim_{n \to \infty} \inf_{P \in \mathbf{P}} \inf_{\theta \in \Theta_0(P)} P(\theta \in \mathcal{C}_n) \geq 1 - \alpha$$

RSW construct their proposed confidence set using hypothesis tests of the following null for each θ.

$$H_\theta : \mathcal{E}_P[g(d_i, d_{-i}, X_i, \theta)] \leq 0$$

We need to choose a **test statistic** T_n such that evidence against H_θ implies a high T_n. In practice, RSW consider test statistics which are functions of the normalized empirical means and the empirical correlation matrix.

$$T_n = T(\frac{\sqrt{n}\bar{W}_n}{S_n}, \hat{\Omega}_n)$$

Two intuitive statistics are based the distance between the moments and 0.

$$T_n^{max} = \max_{j \in [1,K]} \frac{\sqrt{n}\bar{W}_{j,n}}{S_{j,n}}$$

$$T_n^{mmm} = \sum_{j=1}^{K} (\frac{\sqrt{n}\bar{W}_{j,n}}{S_{j,n}})^2 \cdot \times \mathbb{1}\{\bar{W}_{j,n} > 0\}$$

RSW discuss several alternate test statistics. For details and discussion, see their paper.

2.3 One-Step Confidence Set ("Least Favorable Set")

Discussion It is easier to motivate the One-Step Confidence Set as a special case of the Two-Step Confidence Set, than to explain the Two-Step as a generalization of the One-Step.

Given a test statistic T_n, RSW suggest a **critical value** for comparison.

$$\hat{c}_n(1 - \alpha) = J_n^{-1}(1 - \alpha, \hat{P}_n)$$

where

$$J_n(x, P) = P(T(\frac{\sqrt{n}(\bar{W}_n - \mu(P))}{S_n}, \hat{\Omega}_n) \leq x)$$

$$J_n^{-1}(1-\alpha,P) = x \text{ s.t. } P(T(\frac{\sqrt{n}(\bar{W}_n-\mu(P))}{S_n},\hat{\Omega}_n) \leq x) = 1-\alpha$$

To compute this critical value, we could bootstrap. For this class, we will instead simulate from an asymptotic normal distribution.

Algorithm

1. Choose a level $1-\alpha$ for $0<\alpha<1$.
2. For a specific $\theta \in \Theta$, compute the sequence of $W_i = g(d_i, d_{-i}, X_i, \theta)$.
3. Estimate the empirical mean $\mu(\hat{P}_n)$, covariance $\Sigma(\hat{P}_n)$, and correlation $\hat{\Omega}_n$ of W_i.
4. Use your chosen test statistic function and compute $T_n = T(\frac{\sqrt{n}\bar{W}_n}{S_n}, \hat{\Omega}_n)$.
5. We could take draws from $W \sim N(\mu(\hat{P}_n), \Sigma(\hat{P}_n))$. However, we are subtracting out the $\mu(\hat{P}_n)$ anyways (i.e. we have set the means to zero), so take $NSIM$ draws from $W-\mu(\hat{P}_n) \sim N(0, \Sigma(\hat{P}_n))$.
6. For each draw of $W-\mu(\hat{P}_n)$, use your chosen test statistic function and compute $T(\frac{\sqrt{n}(\bar{W}_n-\mu(P))}{S_n}, \hat{\Omega}_n)$. This gives you a vector of $NSIM$ test statistics.
7. Take the $1-\alpha$ quantile of your vector of test statistics, $J_n^{-1}(1-\alpha, \hat{P}_n)$. This value is your critical value, $\hat{c}_n(1-\alpha)$.
8. Accept θ in the confidence set if $T_n \leq \hat{c}_n(1-\alpha)$.
9. Repeat 2-8 for $\theta \in \Theta$.

2.4 Two-Step Confidence Set ("RSW")

Discussion Define a version of our previous J_n function with shifted means λ.

$$J_n(x,\lambda,P) = P(T(\frac{\sqrt{n}(\bar{W}_n-\mu(P))}{S_n} + \frac{\sqrt{n}\lambda}{S_n}, \hat{\Omega}_n) \leq x)$$

If we shift by the means of the true distribution $\mu(P)$, we attain our probability of interest: the probability that our empirical test statistic is below a critical value.

$$J_n(x,\mu(P),P) = P(T(\frac{\sqrt{n}(\bar{W}_n-\mu(P))}{S_n} + \frac{\sqrt{n}\mu(P)}{S_n}, \hat{\Omega}_n) \leq x) = P(T_n \leq x)$$

Unfortunately, we do not observe $\mu(P)$. However, we can use a first-step confidence set for $\mu(P)$ to inform our choice of λ. Specifically, we pick a level β (RSW suggest as a heuristic choosing $\beta \approx \frac{\alpha}{10}$) and construct a confidence set

$$M_n(1-\beta) = \{\mu \in \mathbb{R}^k \mid \max_{j\in[1,K]} \frac{\sqrt{n}(\mu_j - \bar{W}_{j,n})}{S_{j,n}} \leq K_n^{-1}(1-\beta, \hat{P}_n)\}$$

where

$$K_n(x,P) = P(\max_{j\in[1,K]} \frac{\sqrt{n}(\mu_j(P) - \bar{W}_{j,n})}{S_{j,n}} \leq x)$$

Then, we look at all candidate values of $\mu(P)$ in this region and select the candidate value which yields maximal critical value.

$$\hat{c}_n(1-\alpha+\beta) = \sup_{\lambda \in M_n(1-\beta)\cap\mathbb{R}^k} J_n^{-1}(1-\alpha+\beta, \lambda, \hat{P}_n)$$

RSW show that the maximal λ takes the following form, where K_n^{-1} can be computed from a process analogous to that for J_n^{-1} in the One-Step Confidence Set. Intuitively, RSW shift the mean down for moments where the empirical moments are low relative to the simulated critical value, i.e. "far from binding."

$$\lambda_j^* = \min\{\bar{W}_{j,n} + \frac{S_{j,n}K_n^{-1}(1-\beta, \hat{P}_n)}{\sqrt{n}}, 0\}$$

Finally, the RSW Two-Step critical value is

$$\hat{c}_n(1-\alpha+\beta) = J_n^{-1}(1-\alpha+\beta, \lambda^*, \hat{P}_n)$$

The One-Step Confidence Set can be motivated from the Two-Step Confidence Set in two ways. First, if the econometrician sets $\beta = 0$, then the first stage is mechanically suppressed. Second, suppose that all empirical moments are high relative to the simulated critical value, i.e. "very close to binding."

$$\bar{W}_{j,n} + \frac{S_{j,n}K_n^{-1}(1-\beta, \hat{P}_n)}{\sqrt{n}} \geq 0 \quad \forall j$$

This causes $\lambda_j^* = 0$ for all j. This is a "least favorable" case where there is nothing to learn from the negative moments.

Algorithm

1. Choose levels $1-\beta$ and $1-\alpha+\beta$ for $0 < \beta < \alpha$.
2. For a specific $\theta \in \Theta$, compute the sequence of $W_i = g(d_i, d_{-i}, X_i, \theta)$.
3. Estimate the empirical mean $\mu(\hat{P}_n)$, covariance $\Sigma(\hat{P}_n)$, and correlation $\hat{\Omega}_n$ of W_i.
4. Use your chosen test statistic function and compute $T_n = T(\frac{\sqrt{n}\bar{W}_n}{S_n}, \hat{\Omega}_n)$.
5. Take $NSIM$ draws from $W - \mu(\hat{P}_n) \sim N(0, \Sigma(\hat{P}_n))$.
6. For each draw of $W - \mu(\hat{P}_n)$, compute $\frac{\sqrt{n}(W-\mu_j(\hat{P}_n))}{S_{j,n}}$. This gives you a vector of $NSIM$ test statistics.
7. Take the $1-\beta$ quantile of your vector of test statistics, $K_n^{-1}(1-\beta, \hat{P}_n)$. Compute $\lambda_j^* = \min\{\bar{W}_{j,n} + \frac{S_{j,n}K_n^{-1}(1-\beta,\hat{P}_n)}{\sqrt{n}}, 0\}$ for each j.
8. Take $NSIM$ draws from $W - \mu_j(\hat{P}_n) + \lambda^* \sim N(\lambda^*, \Sigma(\hat{P}_n))$.
9. For each draw of $W - \mu_j(\hat{P}_n) + \lambda^*$, compute $T(\frac{\sqrt{n}(\bar{W}_n - \mu(P))}{S_n} + \frac{\sqrt{n}\lambda}{S_n}, \hat{\Omega}_n)$. This gives you a vector of $NSIM$ test statistics.
10. Take the $1-\alpha+\beta$ quantile of your vector of test statistics, $J_n^{-1}(1-\alpha+\beta, \lambda^*, \hat{P}_n)$. This is your critical value, $\hat{c}_n(1-\alpha+\beta)$.
11. Accept θ in the confidence set if $T_n \leq \hat{c}_n(1-\alpha+\beta)$.
12. Repeat 2-11 for $\theta \in \Theta$.

3 Tips

1. Be consistent about your choice of test statistic / norm.
2. Don't confuse $\hat{\Omega}$ and $\hat{\Sigma}$.
3. Mind the appropriate scaling factors. If your confidence sets look off, check that you have $\sqrt{n}$'s in the correct places.