Lecture 8
Inequality Testing and Moment Inequality Models

Inequality Testing

In the previous lecture, we discussed how to test the nonlinear hypothesis $H_0 : h(\theta_0) \leq 0$ when the sample information comes from an asymptotically normal estimator $\hat{\theta}_n$ of θ_0. Now we lay out a more general inequality testing problem.

Let W be our observables. Let F be the true joint distribution of W. Any statistical hypothesis can be thought of as a restriction on F. A generic finite dimensional inequality hypothesis can be written as

$$H_0 : h_F \leq 0, \tag{1}$$

where $h_F = (h_{F,1}, ..., h_{F,r})' \in R^r$ is a parameter determined by F.

To test H_0, one can make use of an estimator $\hat{h}_n$ of h_F based on an i.i.d. sample of W: $W_1, ..., W_n$. Typically, this estimator satisfies

$$\sqrt{n}(\hat{h}_n - h_F) \to_d N(0, \Sigma_F), \tag{2}$$

under F, for some variance matrix Σ_F.

Suppose that we also have an estimator $\widehat{\Sigma}_n$ for Σ_F. Similar to the previous lecture, we can construct a Wald-type test statistic:

$$W_n = \min_{t \leq 0} n(\hat{h}_n - t)'\widehat{\Sigma}_n^{-1}(\hat{h}_n - t). \tag{3}$$

This statistic can be thought of a QLR statistic because

$$W_n = n\left(\min_{t \leq 0} \hat{Q}_n(t) - \min_{t \in R^r} \hat{Q}_n(t)\right), \tag{4}$$

for $\hat{Q}_n(t) = (\hat{h}_n - t)'\widehat{\Sigma}_n^{-1}(\hat{h}_n - t)$.

We would like to choose a critical value $c_n(1-\alpha)$ such that the asymptotic significance level of the test is correct:

$$Asy.Sz := \lim_{n\to\infty}\left[\sup_{F \in \mathcal{F}: h_F \leq 0} \Pr{}_F(W_n > c_n(1-\alpha))\right] \leq \alpha, \tag{5}$$

where $\mathcal{F}$ is the space of F. This space $\mathcal{F}$ is sometimes called the maintained hypothesis because assumptions that defines $\mathcal{F}$ are maintained no matter wether H_0 is true or false.

Some teminologies on general hypothesis tests may be useful and we define them here. When the "$\leq$" holds with equality, the test is called asymptotically size exact. When it holds with strict inequality, the test is asymptotically conservative. The test is called asymptotically similar if it is asymptotically size exact and at the same time

$$\lim_{n\to\infty}\left[\inf_{F\in\mathcal{F}:h_F\leq 0}\Pr{}_F(W_n>c_n(1-\alpha))\right]=\alpha. \tag{6}$$

The Wald, LM and QLR tests for the equality constraints that we discussed before are asymptotically similar. But in inequality testing, asymptotic similarity is too much to ask for.[1] The tests we discussed in the previous lecture and also below are either asymptotically size exact or asymptotically conservative.

Let us consider two types of data-dependent critical values: the PA and the GMS. The PA critical value is defined as the $1-\alpha$ conditional quantile of

$$J_{\widehat{\Sigma}_n,0}:=\min_{t\leq 0}(\widehat{\Sigma}_n^{-1/2}Z_0-t)'\widehat{\Sigma}_n^{-1}(\widehat{\Sigma}_n^{-1/2}Z_0-t), \tag{7}$$

where $Z_0\sim N(0,I_r)$. This quantile can be obtained by drawing S independent draws of Z_0, compute $J_{\widehat{\Sigma}_n,0}$ for each draw and take the $1-\alpha$ quantile of the resulting S-vector.

The GMS critical value is defined as the $1-\alpha$ conditional quantile of

$$J_{\widehat{\Sigma}_n,\varphi_n}:=\min_{t\leq 0}(\widehat{\Sigma}_n^{-1/2}Z_0+\varphi_n-t)'\widehat{\Sigma}_n^{-1}(\widehat{\Sigma}_n^{-1/2}Z_0+\varphi_n-t), \tag{8}$$

where $Z_0\sim N(0,I_r)$ and $\varphi_{n,j}=\begin{cases}0\ if\ \sqrt{n}\hat{h}_{n,j}/\sigma_{n,j}\geq-\kappa_n\\ -\infty\ if\ \sqrt{n}\hat{h}_{n,j}/\sigma_{n,j}<-\kappa_n\end{cases}$ and $\sigma_{n,j}^2$ is the jth diagonal element of $\widehat{\Sigma}_n$.

Now let's derive the asymptotic size of the tests using the PA critical value (the QLR/PA test) and that using the GMS critical value (the QLR/GMS test).

Define the null space of F to be: $\mathcal{F}_0=\{F\in\mathcal{F}:h_F\leq 0\}$.

Assumption 8.1 (a) Under any $F\in\mathcal{F}_0$, $\sqrt{n}(\hat{h}_n-h_F)\to_d N(0,\Sigma_F)$ and $\widehat{\Sigma}_n\to_p\Sigma_F$.

(b) There exists constants C and c such that each element of Σ_F is bounded above by C and the determinant of Σ_F is bounded from below by c for all $F\in\mathcal{F}_0$.

(c) For any sequence $\{F_n\}_{n\geq 1}$ such that $F_n\in\mathcal{F}_0$ for all n and $\Sigma_{F_n}\to\Sigma$, under this sequence of $\{F_n\}$, $\sqrt{n}(\hat{h}_n-h_{F_n})\to_d N(0,\Sigma)$ and $\widehat{\Sigma}_n\to_p\Sigma$.

(d) When the GMS critical value is considered, $\kappa_n\to\infty$.

Theorem 8.1. Under Assumption 8.1, if $\alpha<1/2$, (5) holds with $c_n(1-\alpha)$ being the PA or the GMS critical value.

[1] This doesn't mean that asymptotically similar tests do not exists. They exist but typically have poor power as demonstrated in Andrews (2012) for moment inequality models. For example, a test that always rejects with α probability is an asymptotically similar test, but it also always has trivial power.

Proof.

■

Example: Stochastic Dominance. Let $W = X, Y$. Both X and Y are discrete random variables whose supports are subsets of $x_1, ..., x_r$. Suppose that an interesting null hypothesis is that the distribution of X stochastically dominates that of Y. Let F_X denote the cdf of X and F_Y that of Y, both are marginals of the joint distribution F of W. Then the null hypothesis can be written as

$$H_0 : F_X(x_j) - F_Y(x_j) \leq 0 \text{ for all } j = 1, ..., r. \tag{9}$$

Here $h_F = (F_X(x_1) - F_Y(x_1), ..., F_X(x_r) - F_Y(x_r))'$. Suppose that we have a random sample $\{(X_1, Y_1), ..., (X_n, Y_n)\}$. We can use $\hat{h}_n = (\hat{F}_{X,n}(x_1) - \hat{F}_{Y,n}(x_1), ..., \hat{F}_{X,n}(x_r) - \hat{F}_{Y,n}(x_r))'$, where $\hat{F}_{X,n}(\cdot)$ is the empirical distribution of X and $\hat{F}_{Y,n}(\cdot)$ that of Y. Suppose that X and Y are independent and suppose $(x_1, ..., x_r)$ are in ascending order. Then the first part of Assumption 8.1(a) holds by CLT with the (j,k)th element of Σ_F is $F_X(x_j \wedge x_k)(1 - F_X(x_k \vee x_j)) + F_Y(x_j \wedge x_k)(1 - F_Y(x_k \vee x_j))$ for $j \leq k$. A natural estimator of Σ_F that satisfies Assumption 8.1(a) is the sample analogue of it. Assumption 8.1(b) requires $F_X(x_j)$ or $F_Y(x_j) \geq c_1 \ \forall \ j = 1, ..., r.$ for some constant $c_1 > 0$. Assumptioun 8.1(c) is satisfied without any additional assumption for this particular example, by a triangular array CLT and a triangular array LLN.

Moment Inequality Models

Moment inequality models are a popular type of partially identified models. They come in the form of inequality restrictions on the first moment of a vector of random functions:

$$E_F[m(W, \theta_0)] \geq 0, \tag{10}$$

where W is still our observables, θ_0 is the true value of the model parameter which belongs to a space Θ, m is a known R^k valued function and F is the distribution of W (or in other words, the data generating process). For any data generating process, the inequalities do not necessarily pin down θ_0 to a point. In this sense, the model is potentially partially identified.

A missing data example of moment inequality model has already been given in Lecture 5. Here we give a couple of other examples.

Example: Simple Discrete Game with Multiple Equilibria. Consider a small number (M) of players each with an action space of J elements: $(a_{m,1}, ..., a_{m,J})$. Each of them have a payoff function $\pi_m = f_m(a_1, ..., a_M, X, \theta_0, \epsilon_m)$, where $a_1, ..., a_M$ are an action profile of the players, X is some observed characteristics, and $\epsilon_1, ..., \epsilon_M$ are factors unobserved to the econometrician but common knowledge for the players. Suppose that the players play pure strategy Nash equilibrium.

Then we have a map from $X, \theta_0, \epsilon_1, ..., \epsilon_m$ to a set of equilibria. Parametrize the distribution of $\epsilon_1, ..., \epsilon_M$ by some distribution $G(\cdot, \lambda_0, X)$, then we can compute the max probability of an action profile being played given X and θ_0, λ_0, as well as the min probability of an action profile being played. Denote those by

$$p^{max}(a_1, ..., a_M, X, \theta_0, \lambda_0) \text{ and } p^{min}(a_1, ..., a_M, X, \theta_0, \lambda_0). \tag{11}$$

Then a moment inequality model can be formed:

$$\begin{aligned} E[p^{max}(a_1, ..., a_M, X, \theta_0, \lambda_0) - 1\{(A_1, ..., A_M) = (a_1, ..., a_M)\}] \geq 0 \\ E[1\{(A_1, ..., A_M) = (a_1, ..., a_M)\} - p^{min}(a_1, ..., a_M, X, \theta_0, \lambda_0)] \geq 0 \\ \forall\ (a_1, ..., a_M)' \in \Pi_{m=1}^M \{a_{m,1}, ..., a_{m,J}\}, \end{aligned} \tag{12}$$

where $(A_1, ..., A_M)$ is the actual action profile observable in the data.

Example: Revealed Preference. (Pakes, Porter, Ho and Ishii, 2012) Consider a single agent decision problem. Suppose an agent tries to choose an action a to maximizes expectation of utility $u = f(a, X, \theta_0) + \epsilon$ given the information provided by variables Z known to the agent at the time of decision making. Suppose that ϵ is mean independent of Z. Then the action that the agent takes (denoted A) should in expectation better than other actions: $E[f(A, X, \theta_0) - f(a, X, \theta_0)|Z] \geq 0$ for any a in the action space $\mathcal{A}$. Suppose that the econometrician observes X then she can use the following moment inequality model to infer about θ_0:

$$E[f(A, X, \theta_0) - f(a, X, \theta_0)|Z = z] \geq 0\ \forall\ a \in \mathcal{A},\ z \in \mathcal{Z}, \tag{13}$$

where $\mathcal{Z}$ is the support of Z. If $\mathcal{Z}$ has finite point, then this model is a moment inequality that we discuss in this lecture. Otherwise, it is a conditional moment inequality model. The key to formulate moment inequality this way is that there are variables that the econometricians observe but the agents did not at the time of the decision making. One such variable, for example, is the price of a product in a Cournot model, which is determined by the quantities chosen by the Cournot players and the demand and the demand is not completely known to the players when they chose the quantities to supply.

Because the moment inequality is not necessarily point identified, there usually is no consistent estimator for θ_0. We have also learned that consistent estimation of the identified set $\Theta_0 = \{\theta \in \Theta : E_F m(W, \theta) \geq 0\}$ is difficult and solutions available are controversial. However, we can still construct confidence sets for θ_0 that has the same interpretation as confidence sets in standard point identified models.

We construct a confidence set CS_n by inverting the test for the null hypothesis: $H_{00} : \theta_0 = \theta$, i.e., a confidence set is the collection of $\theta \in \Theta$ such that a test for this H_{00} fail to reject. The only testable implication of H_{00} is

$$H_0 : E_F m(W, \theta) \geq 0. \tag{14}$$

Thus, testing H_{00} amounts to testing H_0. But testing the latter is a special case of the inequality testing we discribed above with $h_F = E_F m(W, \theta)$. Therefore, a QLR/PA or QLR/GMS test can be constructed and when inverted they leads to the QLR/PA and QLR/GMS confidence sets.

To be specific, the QLR test statistic is

$$T_n(\theta) = n \min_{t \geq 0} (\bar{m}_n(\theta) - t)' \widehat{\Sigma}_n^{-1}(\theta)(\bar{m}_n(\theta) - t), \tag{15}$$

where $\bar{m}(\theta)$ is the sample average of $m(W, \theta)$ and $\widehat{\Sigma}_n(\theta)$ the sample variance. The PA critical value and the GMS critical value can be defined similarly to those on page 2 as well. We denote a generic critical value by $c_n(\theta, 1-\alpha)$. The confidence set of level $1 - \alpha$ is:

$$CS_n(1 - \alpha) = \{\theta \in \Theta : T_n(\theta) \leq c_n(\theta, 1 - \alpha)\} \tag{16}$$

Remember that in the inequality testing context, we would like the asymptotic size of the test to be controled. Here that means, for any $\theta \in \Theta_0$,

$$\lim_{n \to \infty} \sup_{F \in \mathcal{F}_0} \Pr{}_F (T_n(\theta) > c_n(\theta, 1 - \alpha)) \leq \alpha. \tag{17}$$

The asymptotic requirement for the confidence set is slightly stronger:

$$\lim_{n \to \infty} \inf_{F \in \mathcal{F}_0, \theta \in \Theta_0(F)} \Pr{}_F (\theta \in CS_n(1 - \alpha)) \geq 1 - \alpha, \tag{18}$$

i.e.

$$\lim_{n \to \infty} \sup_{F \in \mathcal{F}_0, \theta \in \Theta_0(F)} \Pr{}_F (T_n(\theta) > c_n(\theta, 1 - \alpha)) \leq \alpha. \tag{19}$$

This means, to prove the asymptotic validity of the confidence set, one needs a modified version of Assumption 8.1, where the modification is simply replacing $F \in \mathcal{F}_0$ by $F \in \mathcal{F}_0, \theta \in \Theta_0(F)$ and Σ_F by $\Sigma_F(\theta)$ and in part(c), also requiring θ_n to converge as a sequence. However, this modified Assumption 8.1 typically holds under the same primitive conditions as those required for Assumption 8.1.