Lecture 7
Testing Nonlinear Inequality Restrictions[1]

In Lecture 6, we discussed the testing problems where the null hypothesis is defined by **nonlinear equality restrictions**:

$$H_0 : h(\theta_0) = 0 \text{ versus } H_1 : h(\theta_0) \neq 0. \tag{1}$$

We showed that the Wald, QLR and LM statistics have the same asymptotic distribution under H_0 and the distribution is χ^2_r. The rejection rule for the Wald, QLR and LM tests of significance level α can thus be:

$$\text{reject } H_0 \text{ if } T_n > \chi^2_{r,1-\alpha}, \tag{2}$$

where $T_n = \mathcal{W}_n$, $\mathcal{QLR}_n$ or $\mathcal{LM}_n$ and $\chi^2_{r,1-\alpha}$ is the $1-\alpha$ quantile of χ^2_r. A test defined as such has correct asymptotic size because:

$$\begin{aligned} Asy.SZ \ &:= \lim_{n\to\infty} \sup_{\theta_0\in\Theta:h(\theta_0)=0} \Pr\nolimits_{\theta_0}\left(T_n > \chi^2_{r,1-\alpha}\right) \\ &= \Pr\left(\chi^2_r > \chi^2_{r,1-\alpha}\right) = \alpha. \end{aligned} \tag{3}$$

When h is one-dimensional, these tests are often considered as "two-sided" test as violations of $h(\theta_0) = 0$ on both sides are detected.

In this lecture, we discuss the testing problems where the null hypothesis is defined by **nonlinear inequality restrictions**:

$$H_0 : h(\theta_0) \leq 0 \text{ vs. } H_1 : h(\theta_0) > 0. \tag{4}$$

Testing situations like this arise when, for example, one wishes to test that some matrices are positive semi-definite. They also arise, sometimes in a bit disguise, in specification testing for models defined by inequalities (e.g. moment inequalities).

Notice that, when testing (1), the asymptotic null distributions of $\mathcal{W}_n$, $\mathcal{QLR}_n$ or $\mathcal{LM}_n$ do not depend on θ_0. Test statistics with this property are called **asymptotically pivotal**. With (asymptotically) pivotal test statistics, tests can be constructed easily. The critical values simply can be chosen as the appropriate quantiles of the asymptotic distribution (that is invariant to the true parameter). The asymptotic null rejection probabilities are exactly the significance level we choose, no matter what θ_0 is as long as H_0 holds.

Analogous test statistics for (4) are not asymptotically pivotal. The asymptotic distributions

[1] The note is made up from my head based on Wolak (1991), Andrews and Soares (2010), Gourieroux and Monfort (1995, chapter 21). It does not follow any of those sources closely. Accuracy is not guaranteed. Comments welcome.

of them depend on θ_0, in particular, on whether some or all elements of $h(\theta_0)$ are zero. We define the Wald and QLR statistics first and then discuss the problems caused by the lack of pivotalness.

Throughout, we maintain Assumption QLR. Thus, $\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \to_d N(0, cB_0^{-1}) = N(0, c^2\Omega_0^{-1})$.

The Wald-type statistic is defined as

$$\mathcal{W}_n^* = \min_{t\in[-\infty,0]^r} n\left(h\left(\hat{\theta}_n\right) - t\right)'\left(\hat{H}_n'\hat{c}_n\hat{B}_n^{-1}\hat{H}_n\right)^{-1}\left(h\left(\hat{\theta}_n\right) - t\right), \tag{5}$$

where $\hat{H}_n = \frac{\partial h(\hat{\theta}_n)'}{\partial\theta}$, $\hat{c}_n \to_p c$ and $\hat{B}_n = \frac{\partial^2\hat{Q}_n(\hat{\theta}_n)'}{\partial\theta\partial\theta'}$. Wald statistic defined this way does not require one to compute the inequality constraint estimator $\tilde{\theta}_n$:

$$h\left(\tilde{\theta}_n\right) \leq 0 \text{ and } \hat{Q}_n\left(\tilde{\theta}_n\right) \leq \inf_{\theta\in\Theta:h(\theta)\leq 0}\hat{Q}_n(\theta) + o_p\left(n^{-1/2}\right). \tag{6}$$

This has advantage when the inequality constraints make $\tilde{\theta}_n$ hard to compute.

If one is willing to compute $\tilde{\theta}_n$, there are other ways of defining Wald-type statistics as well:

$$\begin{aligned} \tilde{W}_n^* &= n\left(h\left(\hat{\theta}_n\right) - h\left(\tilde{\theta}_n\right)\right)'\left(\hat{H}_n'\hat{c}_n\hat{B}_n^{-1}\hat{H}_n\right)^{-1}\left(h\left(\hat{\theta}_n\right) - h(\tilde{\theta}_n)\right) \text{ and} \\ W_n^{*h} &= n\left(\hat{\theta}_n - \tilde{\theta}_n\right)'\hat{c}_n^{-1}\hat{B}_n\left(\hat{\theta}_n - \tilde{\theta}_n\right). \end{aligned} \tag{7}$$

The second statistic is a Hausman-Wald type statistic as it has the feature of a Hausman test: measuring the difference between the unrestricted and the restricted estimators.

A QLR test statistic can be defined in the same way as in Lecture 6:

$$\mathcal{QLR}_n = 2n\hat{c}_n^{-1}\left(\hat{Q}_n\left(\tilde{\theta}_n\right) - \hat{Q}_n\left(\hat{\theta}_n\right)\right). \tag{8}$$

Theorem 7.1: Suppose *Assumptions* EE2, CF, R, COV, REE and QLR hold. Then,

(a) under H_0, $\mathcal{W}_n^* = \tilde{\mathcal{W}}_n^* + o_p(1) = \mathcal{W}_n^{*h} + o_p(1) = \mathcal{QLR}_n + o_p(1)$ and

(b) $\mathcal{W}_n^* \to_d J_{\Sigma_0,h_{\theta_0}} = \min_{t\in[-\infty,0]^r}\left(Z_\Sigma + h_{\theta_0} - t\right)'\Sigma_0^{-1}\left(Z_\Sigma + h_{\theta_0} - t\right)$, where

$$\begin{aligned} Z_\Sigma &\sim N(0,\Sigma) \text{ and } h_{\theta_0} = (h_{\theta_0,1}, ..., h_{\theta_0,r})' \\ h_{\theta_0,j} &= \begin{cases} 0 \text{ if } h_j(\theta_0) = 0 \\ -\infty \text{ if } h_j(\theta_0) < 0 \end{cases} \end{aligned}$$

Before giving the (rather tedious) proof, I would like to discuss the implications of part (b). As predicted above, the asymptotic distribution of the test statistics are not parameter free. If we want to construct a test using the same kind of critical value as in the equality constraint case, we have to answer the question: which $J_{\Sigma_0,h_{\theta_0}}$ should we use to take the quantile from? Different $J_{\Sigma_0,h_{\theta_0}}$ have different quantiles. Naturally, we would like to control the size of the test. That is, we want:

$$\sup_{\theta_0:h(\theta_0)\leq 0} \Pr\left(J_{\Sigma_0,h_{\theta_0}} > c_\alpha\right) = \alpha. \tag{9}$$

This idea requires us to find the least favorable $J_{\Sigma_0,h_{\theta_0}}$: the $J_{\Sigma_0,h_{\theta_0}}$ that has the highest $1-\alpha$ quantile. When $r=1$, $J_{\Sigma_0,h_{\theta_0}}$ is one dimensional. The least favorable J_{θ_0} is straightforward: it shoud be the $J_{\Sigma_0,h_{\theta_0}}$ such that $h(\theta_0)=0$, which is $\frac{1}{2}\chi_0^2+\frac{1}{2}\chi_1^2$. When $r>1$, the task is way more complicated, as explained in Wolak (1991).

The type of critical value suggested above is called "fixed critical-value". They are fixed in a sense that they do not vary with data. There are other ways of choosing critical values, which are much more feasible than fixed critical values in the context we are dealing with. We define two types here.

The first type is "plug-in asymptotics (PA)". Literally, it means, we use the asymptotic distribution (not parameter free) and plug in the estimated values of the parameters. In $J_{\Sigma_0,h_{\theta_0}}$, Σ_0 can be consistently estimated under the null. Therefore, we use the estimated value in place of Σ_0. However, h_{θ_0} cannot be consistently estimated because it takes infinite values sometimes. We get around this by using 0 in place of h_{θ_0}. Using 0 instead of $h_{\theta_0}(\leq 0)$ shifts the distribution of $J_{\Sigma_0,h_{\theta_0}}$ to the right and thus tends to make the critical values large. If we only care about controlling the size, this shouldn't cause any problem. To sum up, the PA critical value is taken as the conditional (on data) $1-\alpha$ quantile of $J_{\hat{\Sigma}_n,0}$, where

$$J_{\hat{\Sigma}_n,0} = \min_{t\in[-\infty,0]^r}\left(Z_{\hat{\Sigma}_n}-t\right)'\left(\hat{\Sigma}_n\right)^{-1}\left(Z_{\hat{\Sigma}_n}-t\right), \tag{10}$$

where $\hat{\Sigma}_n = \hat{c}_n\hat{H}_n'\hat{B}_n^{-1}\hat{H}_n$. You can show that $J_{\hat{\Sigma}_n,0}\to_d J_{\Sigma_0,0}$, and the $1-\alpha$ quantile of $J_{\hat{\Sigma}_n,0}$ converges in probability to that of $J_{\Sigma_0,0}$, which is no smaller than that of $J_{\Sigma_0,h_{\theta_0}}$, at any θ_0 under the null. This implies that the size of the test is well controlled asymptotically.

The second type is "generalized moment selection (GMS)".[2] The PA procedure replaces h_{θ_0} by zero and can potentially produce critical values that are too large. Large critical values affects the power of the test. To improve power, GMS procedure uses the data to decide whether h_{θ_0} is infinite or not. The judgement is not perfect, but it helps reducing the critical value while at the same time preserving the size property of the test. The GMS procedure replaces h_{θ_0} by a moment

[2] We use the the term "moment" even though $h(\theta_0)$ are not necessarily moment of anything. But the term is developed in moment inequality literature. It's convenient to follow the tradition.

selection function:

$$\xi_n = \begin{cases} -\infty \text{ if } \varkappa_n h\left(\hat{\theta}_n\right) < -1 \\ 0 \text{ if } \varkappa_n h\left(\hat{\theta}_n\right) > -1 \end{cases},$$

where $\varkappa_n = o(\sqrt{n})$ is a sequence of positive numbers. The GMS critical value is taken to be the conditional $1-\alpha$ quantile of $J_{\hat{\Sigma}_n,\xi_n}$. It can be shown that the sizes of the GMS tests are well controlled asymptotically and the tests have better power than PA tests.

Notice that PA and GMS tests are developed only recently because they require simulating the quantiles of non-regular distributions. Even though they are much easier to use than fixed critical value tests, they could not be more feasible than the latter without powerful computers.

Now, let's prove Theorem 7.1.

Proof of Theorem 7.1: First we establish part (a). We start with the asymptotic equivalence between $\mathcal{W}_n^*$ and $\tilde{\mathcal{W}}_n^*$. Let t_n^q be the solution to the quadratic minimization problem in (5). It suffices to show that

$$\sqrt{n}\left(h\left(\hat{\theta}\right) - h\left(\tilde{\theta}_n\right)\right) = \sqrt{n}\left(h\left(\hat{\theta}\right) - t_n^q\right) + o_p(1). \tag{11}$$

We show this by comparing the Kuhn-Tucker conditions for the quadratic minimization problem and those for $\inf_{\theta\in\Theta:h(\theta)\leq 0}\hat{Q}_n(\theta)$. By the Kuhn-Tucker theorem, θ_n^q satisfies:

$$\begin{aligned} (\text{KT}^q): \qquad & \left(\hat{H}_n'\hat{c}_n\hat{B}_n^{-1}\hat{H}_n\right)^{-1}\left(h\left(\hat{\theta}_n\right) - t_n^q\right) = \lambda_n^q \\ & h\left(\theta_n^q\right) \leq 0,\ \lambda_n^q \geq 0 \\ & h\left(\theta_n^q\right)'\lambda_n^q = 0, \end{aligned} \tag{12}$$

where the λ_n^q are the Lagrange multipliers. Eliminating λ_n^q, we have

$$\begin{aligned} (\text{KT}^q): \qquad & \left(\hat{H}_n'\hat{c}_n\hat{B}_n^{-1}\hat{H}_n\right)^{-1}\left(h\left(\hat{\theta}_n\right) - t_n^q\right) \geq 0 \\ & t_n^q \leq 0 \\ & t_n^{q\prime}\left(\hat{H}_n'\hat{c}_n\hat{B}_n^{-1}\hat{H}_n\right)^{-1}\left(h\left(\hat{\theta}_n\right) - t_n^q\right) = 0, \end{aligned} \tag{13}$$

By the Kuhn-Tucker Theorem, $\tilde{\theta}_n$ satisfies

$$\begin{aligned}(\mathrm{KT}^c): \quad & \frac{\partial \hat{Q}_n\left(\tilde{\theta}_n\right)}{\partial \theta} + \tilde{H}_n \lambda_n^c = o_p\left(n^{-1/2}\right) \\ & h\left(\tilde{\theta}_n\right) \leq 0,\ \lambda_n^c \geq 0 \\ & h\left(\tilde{\theta}_n\right)' \lambda_n^c = 0. \end{aligned} \tag{14}$$

Mean-value expansion of $\frac{\partial \hat{Q}_n(\tilde{\theta}_n)}{\partial \theta}$ around $\hat{\theta}_n$ gives:

$$\begin{aligned}\tilde{H}_n \lambda_n^c &= o_p\left(n^{-1/2}\right) - \frac{\partial \hat{Q}_n\left(\hat{\theta}_n\right)}{\partial \theta} + B_n^*\left(\hat{\theta}_n - \tilde{\theta}_n\right) \\ &= o_p\left(n^{-1/2}\right) + B_n^*\left(\hat{\theta}_n - \tilde{\theta}_n\right), \end{aligned} \tag{15}$$

where $B_n^* = \frac{\partial^2 \hat{Q}_n(\theta_n^*)}{\partial \theta \partial \theta'}$ for some θ_n^* lying between $\hat{\theta}_n$ and $\tilde{\theta}_n$, and the second equality holds by Assumption EE2(ii). Let H_n^+ be the matrix that satisfies $h\left(\hat{\theta}_n\right) - h\left(\tilde{\theta}_n\right) = H_n^+\left(\hat{\theta}_n - \tilde{\theta}_n\right)$. Such a matrix always exists by the mean-value theorem and $H_n^+ \to_p H_0$ under our assumptions (R(i), EE2(i), REE, CF(iv)). Premultiplying $H_n^+ (B_n^*)^{-1}$ to both sides of the equation above and we have:

$$H_n^+ (B_n^*)^{-1} \tilde{H}_n \lambda_n^c = h\left(\hat{\theta}_n\right) - h\left(\tilde{\theta}_n\right) + o_p\left(n^{-1/2}\right). \tag{16}$$

Thus

$$\lambda_n^c = \left(H_n^+ (B_n^*)^{-1} \tilde{H}_n\right)^{-1}\left(h\left(\hat{\theta}_n\right) - h\left(\tilde{\theta}_n\right) + o_p\left(n^{-1/2}\right)\right). \tag{17}$$

Using this to eliminate λ_n^c in (14), we have

$$\begin{aligned}(\mathrm{KT}^c): \quad & \left(H_n^+ (B_n^*)^{-1} \tilde{H}_n\right)^{-1}\left(h\left(\hat{\theta}_n\right) - h\left(\tilde{\theta}_n\right) + o_p\left(n^{-1/2}\right)\right) \geq 0 \\ & h\left(\tilde{\theta}_n\right) \leq 0, \\ & h\left(\tilde{\theta}_n\right)' \left(H_n^+ (B_n^*)^{-1} \tilde{H}_n\right)^{-1}\left(h\left(\hat{\theta}_n\right) - h\left(\tilde{\theta}_n\right) + o_p\left(n^{-1/2}\right)\right) = 0 \end{aligned} \tag{18}$$

The systems (KT^q) and (KT^c) uniquely pin down t_n^q and $h\left(\tilde{\theta}_n\right)$ respectively. The two systems are asymptotically equivalent. Therefore, $\sqrt{n}\left(h\left(\tilde{\theta}_n\right) - h(\theta_0)\right) = \sqrt{n}\left(t_n^q - h(\theta_0)\right) + o_p(1)$.[3] Because $\sqrt{n}\left(h\left(\hat{\theta}_n\right) - h(\theta_0)\right) \to_d N(0, H_0 c B_0^{-1} H_0)$, we have (11).

[3] More rigorous argument uses the facts that (1) either system implicitly defines a continuous mappings from elements with known asymptotic distributions to $\sqrt{n}\left(h\left(\tilde{\theta}_n\right) - h(\theta_0)\right)$ or $\sqrt{n}\left(t_n^q - h(\theta_0)\right)$, respectively, (2) The continuous mappings defined in the two systems are the same; and (3) the elements with known asymptotic distributions in both systems have the same asymptotic distributions.

Next we establish the asymptotic equivalence between $\tilde{\mathcal{W}}_n^*$ and $\mathcal{W}_n^{*h}$. Combining (15) and (17), we have

$$\left(\hat{\theta}_n - \tilde{\theta}_n\right) = (B_n^*)^{-1} \tilde{H}_n \left(H_n^+ (B_n^*)^{-1} \tilde{H}_n\right)^{-1} \left(h\left(\hat{\theta}_n\right) - h\left(\tilde{\theta}_n\right) + o_p\left(n^{-1/2}\right)\right). \tag{19}$$

Plug that in $\mathcal{W}_n^{*h}$, we have

$$\begin{aligned}
\mathcal{W}_n^{*h} &= n\left(h\left(\hat{\theta}_n\right) - h\left(\tilde{\theta}_n\right) + o_p\left(n^{-1/2}\right)\right)' \left(H_n^+ (B_n^*)^{-1} \tilde{H}_n\right)^{-1} \tilde{H}_n' (B_n^*)^{-1} \hat{c}_n^{-1} \hat{B}_n \cdot \\
&\quad (B_n^*)^{-1} \tilde{H}_n \left(H_n^+ (B_n^*)^{-1} \tilde{H}_n\right)^{-1} \left(h\left(\hat{\theta}_n\right) - h\left(\tilde{\theta}_n\right) + o_p\left(n^{-1/2}\right)\right) \\
&= n\left(h\left(\hat{\theta}_n\right) - h\left(\tilde{\theta}_n\right)\right)' \left(\hat{H}_n \hat{c}_n \left(\hat{B}_n\right)^{-1} \hat{H}_n\right)^{-1} \left(h\left(\hat{\theta}_n\right) - h\left(\tilde{\theta}_n\right)\right) + o_p(1) \\
&= \tilde{\mathcal{W}}_n^* + o_p(1).
\end{aligned} \tag{20}$$

The second equality holds because Assumption COV holds and $\sqrt{n}(h\left(\hat{\theta}_n\right) - h\left(\tilde{\theta}_n\right)) = O_p(1)$, the later of which holds by the arguments used to establish (11).

Next we establish the asymptotic equivalence between $\mathcal{QLR}_n$ and $\mathcal{W}_n^{*h}$. Using second-order Taylor expansion of $\hat{Q}_n$ around $\hat{\theta}_n$, we have

$$\begin{aligned}
\mathcal{QLR}_n &= 2n\hat{c}_n^{-1}\left[\frac{\partial \hat{Q}_n(\hat{\theta}_n)}{\partial \theta'}\left(\hat{\theta} - \tilde{\theta}_n\right) + \frac{1}{2}\left(\hat{\theta} - \tilde{\theta}_n\right)' \frac{\partial^2 \hat{Q}_n(\theta_n^*)}{\partial\theta\partial\theta}\left(\hat{\theta} - \tilde{\theta}_n\right)\right] \\
&= 2n\hat{c}_n^{-1}\left[o_p\left(n^{-1}\right) + \frac{1}{2}\left(\hat{\theta} - \tilde{\theta}_n\right)' \hat{B}_n \left(\hat{\theta} - \tilde{\theta}_n\right)\right] \\
&= \left(\hat{\theta} - \tilde{\theta}_n\right)' \hat{c}_n^{-1} \hat{B}_n \left(\hat{\theta} - \tilde{\theta}_n\right) + o_p(1) \\
&= \mathcal{W}_n^{*h} + o_p(1),
\end{aligned} \tag{21}$$

where the second equality holds by Assumption EE2(ii), COV and $\hat{\theta} - \tilde{\theta}_n = O_p(n^{-1/2})$ (which holds by (19) and $\sqrt{n}(h\left(\hat{\theta}_n\right) - h\left(\tilde{\theta}_n\right)) = O_p(1)$).

Now we establish part (b). Define

$$S\left(m, \Sigma\right) = \min_{t \in [-\infty, 0]^r} \left(m - t\right)' \Sigma \left(m - t\right), \tag{22}$$

The proof is a direct application of the continuous mapping theorem once we observe that

$$\mathcal{W}_n^* = S\left(\sqrt{n}(h\left(\hat{\theta}_n\right) - h\left(\theta_0\right)) + \sqrt{n}h\left(\theta_0\right), \left(\hat{H}_n' \hat{c}_n \hat{B}_n^{-1} \hat{H}_n\right)^{-1}\right), \tag{23}$$

and show that S is continuous in both of its arguments on $(R \cup \{-\infty\})^r \times \Xi$, where Ξ is the space of positive definite matrices. The proof of this is left as an exercise and a version of it can be found in Andrews and Soares (2010). Note that we need the continuity to hold on the extended real space because we want to allow h_{θ_0} to take infinite values.