Lecture 6
Testing Nonlinear Restrictions[1]

The previous lectures prepare us for the tests of nonlinear restrictions of the form:

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

In this lecture, we consider Wald, Lagrange multiplier (LM), and quasi-likelihood ratio (QLR) test. The Delta Method will be useful in constructing those tests, especially the Wald test.

1 The Delta Method

The delta method can be used to find the asymptotic distribution of $h(\widehat{\theta}_n)$, suitably normalized, if $d_n(\widehat{\theta}_n - \theta_0) \to_d Z$.

Theorem (Δ-method): *Suppose* $d_n(\widehat{\theta}_n - \theta_0) \to_d Y$ *where* $\widehat{\theta}_n$ *and* Y *are random* k*-vectors,* θ_0 is a non-random k-vector, and $\{d_n : n \geq 1\}$ is a sequence of scalar constants that diverges to infinity as $n \to \infty$. *Suppose* $h(\cdot) : R^k \to R^\ell$ *is differentiable at* θ_0*, i.e.,* $h(\theta) = h(\theta_0) + \frac{\partial}{\partial \theta'} h(\theta_0)(\theta - \theta_0) + o(||\theta - \theta_0||)$ *as* $\theta \to \theta_0$*. Then,*

$$d_n(h(\widehat{\theta}_n) - h(\theta_0)) \xrightarrow{d} \frac{\partial}{\partial \theta'} h(\theta_0) Y.$$

If $Y \sim N(0, V)$, then $\frac{\partial}{\partial \theta'} h(\theta_0) Y \sim N\left(0,\ \frac{\partial}{\partial \theta'} h(\theta_0) V (\frac{\partial}{\partial \theta'} h(\theta_0))'\right)$.

Proof: By the assumption of differentiability of h at θ_0, we have

$$d_n(h(\widehat{\theta}_n) - h(\theta_0)) = \frac{\partial}{\partial \theta'} h(\theta_0) d_n(\widehat{\theta}_n - \theta_0) + d_n o(||\widehat{\theta}_n - \theta_0||).$$

The first term on the right-hand side converges in distribution to $\frac{\partial}{\partial \theta'} h(\theta_0) Y$. So, we have the desired result provided $d_n o(||\widehat{\theta}_n - \theta_0||) = o_p(1)$. This holds because

$$d_n o(||\widehat{\theta}_n - \theta_0||) = ||d_n(\widehat{\theta}_n - \theta_0)|| o(1) = O_p(1) o(1) = o_p(1)$$

and the proof is complete. $\square$

[1] The notes for this lecture is largely adapted from the notes of Donald Andrews on the same topic I am grateful for Professor Andrews' generosity and elegant exposition. All errors are mine.

2 Tests for Nonlinear Restrictions

The R^r-valued function $h(\cdot)$ defining the restrictions and the matrix Ω_0 (defined in Assumption CF) are assumed to satisfy:

Assumption R: (i) *$h(\theta)$ is continuously differentiable on a neighborhood of θ_0.*
(ii) $H = \frac{\partial}{\partial\theta'}h(\theta_0)$ *has full rank r* ($\leq d$, where d is the dimension of θ).
(iii) Ω_0 *is positive definite.*

For example, we might have

$$\begin{aligned} h(\theta) &= \theta_1 - \theta_2, \\ h(\theta) &= \theta_1\theta_2 - \theta_3, \text{ or} \\ h(\theta) &= \begin{pmatrix} \theta_1 \\ \theta_2 + \theta_3 - 1 \end{pmatrix}. \end{aligned} \tag{2}$$

The requirement that Ω_0 is positive definite ensures that the asymptotic covariance matrix, $B_0^{-1}\Omega_0 B_0^{-1}$, of $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ is positive definite.

The Wald statistic for testing H_0 is defined to be

$$\begin{aligned} \mathcal{W}_n &= nh(\widehat{\theta}_n)'(\widehat{H}_n\widehat{B}_n^{-1}\widehat{\Omega}_n\widehat{B}_n^{-1}\widehat{H}_n')^{-1}h(\widehat{\theta}_n), \text{ where} \\ \widehat{H}_n &= \frac{\partial}{\partial\theta'}h(\widehat{\theta}_n) \end{aligned} \tag{3}$$

and $\widehat{B}_n$ and $\widehat{\Omega}_n$ are consistent estimators of B_0 and Ω_0 respectively. See Lecture 5 for the choice of $\widehat{B}_n$ and $\widehat{\Omega}_n$.

As defined, the Wald statistic is a quadratic form in the (unrestricted) estimator $h(\widehat{\theta}_n)$ of the restrictions $h(\theta_0)$. The quadratic form has a positive definite (pd) weight matrix. If H_0 is true, then $h(\widehat{\theta}_n)$ should be close to zero and the quadratic form $\mathcal{W}_n$ in $h(\widehat{\theta}_n)$ also should be close to zero. On the other hand, if H_0 is false, $h(\widehat{\theta}_n)$ should be close to $h(\theta_0) \neq 0$ and the quadratic form $\mathcal{W}_n$ in $h(\widehat{\theta}_n)$ should be different from zero. Thus, the Wald test rejects H_0 when its value is sufficiently large. Large sample critical values are provided by Theorem 6.1 below.

The LM and QLR statistics depend on a restricted extremum estimator of θ_0. The restricted extremum estimator, $\widetilde{\theta}_n$, is an estimator that satisfies

$$\widetilde{\theta}_n \in \Theta,\ h(\widetilde{\theta}_n) = 0, \text{ and } \hat{Q}_n(\widetilde{\theta}_n) \leq \inf\{\hat{Q}_n(\theta) : \theta \in \Theta,\ h(\theta) = 0\} + o_p(n^{-1/2}). \tag{4}$$

The LM statistic is defined by

$$\begin{aligned}\mathcal{LM}_n &= n\frac{\partial}{\partial\theta'}\hat{Q}_n(\widetilde{\theta}_n)\widetilde{B}_n^{-1}\widetilde{H}_n'(\widetilde{H}_n\widetilde{B}_n^{-1}\widetilde{\Omega}_n\widetilde{B}_n^{-1}\widetilde{H}_n')^{-1}\widetilde{H}_n\widetilde{B}_n^{-1}\frac{\partial}{\partial\theta}\hat{Q}_n(\widetilde{\theta}_n), \text{ where} \\ \widetilde{H}_n &= \frac{\partial}{\partial\theta'}h(\widetilde{\theta}_n) \end{aligned} \tag{5}$$

and $\widetilde{B}_n$ and $\widetilde{\Omega}_n$ are consistent estimators under H_0 of B_0 and Ω_0, respectively, that are constructed using the restricted estimator $\widetilde{\theta}_n$ rather than $\widehat{\theta}_n$.

The LM statistic is a quadratic form in the vector of derivatives of the criterion function evaluated at the restricted estimator $\widetilde{\theta}_n$. The quadratic form has a pd weight matrix. If the null hypothesis is true, then the unrestricted estimator should be close to satisfying the restrictions. In this case, the restricted and unrestricted estimators, $\widetilde{\theta}_n$ and $\widehat{\theta}_n$, should be close. In turn, this implies that the derivative of the criterion function at $\widetilde{\theta}_n$ and $\widehat{\theta}_n$ should be close. The latter, $\frac{\partial}{\partial\theta}\hat{Q}_n(\widehat{\theta}_n)$, equals zero by the first-order conditions for unrestricted minimization of $\hat{Q}_n(\theta)$ over Θ. (More precisely, $\frac{\partial}{\partial\theta}\hat{Q}_n(\widehat{\theta}_n)$ is only close to zero, viz., $o_p(n^{-1/2})$, by Assumption EE2, since $\widehat{\theta}_n$ is not required to exactly minimize $\hat{Q}_n(\theta)$ just approximately minimize it.) Then, $\frac{\partial}{\partial\theta}\hat{Q}_n(\widetilde{\theta}_n)$ should be close to zero under H_0 and the quadratic form, $\mathcal{LM}_n$, in $\frac{\partial}{\partial\theta}\hat{Q}_n(\widetilde{\theta}_n)$ also should be close to zero. On the other hand, if H_0 is false, then $\widehat{\theta}_n$ and $\widetilde{\theta}_n$ should not be close, $\frac{\partial}{\partial\theta}\hat{Q}_n(\widetilde{\theta}_n)$ should not be close to zero, and the quadratic form, $\mathcal{LM}_n$, in $\frac{\partial}{\partial\theta}\hat{Q}_n(\widetilde{\theta}_n)$ also should not be close to zero. In consequence, the LM test rejects H_0 when the LM_n statistic is sufficiently large. Asymptotic critical values for $\mathcal{LM}_n$ are determined by Theorem 6.1 below.

Note that $\mathcal{LM}_n$ can be computed without computing $\widehat{\theta}_n$. This can have computational advantages if the criterion function is simpler when the restrictions $h(\theta)$ are imposed than when unrestricted. For example, if the unrestricted model is a nonlinear regression model and the restricted model is a linear model, then the LS estimator has a closed-form solution under the restrictions, but not otherwise.

Next, we consider the QLR statistic. It has the proper asymptotic null distribution only when the following assumption holds:

Assumption QLR: (i) $\Omega_0 = cB_0$ *for some scalar constant* $c \neq 0$.
(ii) $\widehat{c}_n \to_p c$ *for some sequence of non-zero random variables* $\{\widehat{c}_n : n \geq 1\}$.

For example, in the ML example with correctly specified model, the information matrix equality yields $\Omega_0 = B_0$, so Assumption QLR holds with $c = \widehat{c}_n = 1$. In the LS example with conditionally homoskedastic errors (i.e., $E(U_i|X_i) = 0$ a.s. and $E(U_i^2|X_i) = \sigma^2$ a.s.), Assumption QLR holds with $c = \sigma^2$ and

$$\widehat{c}_n = \tfrac{1}{n}\textstyle\sum_{i=1}^n (Y_i - g(X_i, \widehat{\theta}_n))^2. \tag{6}$$

In the GMM, MD, and TS examples with optimal weight matrix (e.g., $A'A = V_0^{-1}$ for the GMM and MD estimators, see Lecture 5), Assumption QLR holds with $c = \widehat{c} = 1$, since $\Omega_0 = B_0 = \Gamma_0' V_0^{-1} \Gamma_0$.

The QLR statistic is defined by

$$\mathcal{QLR}_n = 2n(\hat{Q}_n(\widetilde{\theta}_n) - \hat{Q}_n(\widehat{\theta}_n))/\widehat{c}_n. \tag{7}$$

When H_0 is true, the restricted and unrestricted estimators, $\widetilde{\theta}_n$ and $\widehat{\theta}_n$, should be close and, hence, so should the value of the criterion function evaluated at these parameter values. Thus, under H_0, the statistic $\mathcal{QLR}_n$ should be close to zero. Under H_1, $\mathcal{QLR}_n$ should be noticeably larger than zero, since the minimization of $\hat{Q}_n(\theta)$ over the restricted parameter space, which does not include the minimum value θ_0 of $Q(\theta)$, should leave $\hat{Q}_n(\widetilde{\theta}_n)$ noticeably larger than $\hat{Q}_n(\widehat{\theta}_n)$. In consequence, the QLR test rejects H_0 when $\mathcal{QLR}_n$ is sufficiently large. Asymptotic critical values are provided by Theorem 6.1 below.

For the ML example, the QLR statistic equals the standard likelihood ratio statistic, viz., minus two times the logarithm of the likelihood ratio.

When Assumption QLR holds, the $\mathcal{LM}_n$ statistic simplifies to

$$\begin{aligned} \mathcal{LM}_n &= n \frac{\partial}{\partial \theta'} \hat{Q}_n(\widetilde{\theta}_n) \widetilde{B}_n^{-1} \frac{\partial}{\partial \theta} \hat{Q}_n(\widetilde{\theta}_n)/\widehat{c}_n \text{ or} \\ \mathcal{LM}_n &= n \frac{\partial}{\partial \theta'} \hat{Q}_n(\widetilde{\theta}_n) \widetilde{\Omega}_n^{-1} \frac{\partial}{\partial \theta} \hat{Q}_n(\widetilde{\theta}_n) \end{aligned} \tag{8}$$

(since in this case one can take $\widetilde{\Omega}_n = \widehat{c}_n \widetilde{B}_n$ and $\frac{\partial}{\partial \theta} \hat{Q}_n(\widetilde{\theta}_n) = -\widetilde{H}_n' \widetilde{\lambda}_n$ for some vector $\widetilde{\lambda}_n$ of Lagrange multipliers).

For the LM and QLR tests, we use the following assumption.

Assumption REE: $\widetilde{\theta}_n \xrightarrow{p} \theta_0$ under H_0.

This assumption can be verified using the results of Lecture 3 with the parameter space Θ replaced by the restricted parameter space $\widetilde{\Theta} = \{\theta \in \Theta : h(\theta) = 0\}$.

For the Wald and LM tests, we assume that consistent estimators under H_0 of B_0 and Ω_0 are used in constructing the test statistics:

Assumption COV: *For the Wald statistic,* $\widehat{B}_n \xrightarrow{p} B_0$ *and* $\widehat{\Omega}_n \xrightarrow{p} \Omega_0$ *under* H_0*. For the* LM *statistic,* $\widetilde{B}_n \xrightarrow{p} B_0$ *and* $\widetilde{\Omega}_n \xrightarrow{p} \Omega_0$ *under* H_0.

Estimators that satisfy this assumption are discussed in Lecture 5.

Theorem 6.1: (a) *Under Assumptions* EE2, CF, R, *and* COV,

$$\mathcal{W}_n \xrightarrow{d} \chi^2_r \text{ under } H_0,$$

where χ^2_r *denotes a chi-square distribution with* r *degrees of freedom.*

(b) *Under Assumptions* EE2, CF, R, COV, *and* REE,

$$\mathcal{LM}_n \xrightarrow{d} \chi^2_r \text{ under } H_0.$$

(c) *Under Assumptions* EE2, CF, R, COV, REE, *and* QLR,

$$\mathcal{QLR}_n \xrightarrow{d} \chi^2_r \text{ under } H_0.$$

One rejects H_0 using the Wald test if $W_n > k^r_\alpha$, where k^r_α is the $1-\alpha$ quantile of the χ^2_r distribution. The LM and QLR tests are defined analogously. These tests have significance level approximately equal to α for large n.

Under sequences of local alternatives to $H_0 : h(\theta) = 0$, the Wald, LM, and QLR test statistics have noncentral chi-squared distributions with the same noncentrality parameter, see next Lecture Thus, the three tests, Wald, LM, and QLR, have the same large sample power functions. One cannot choose between these tests based on (first order) large sample power. The choice can be made on computational grounds. It also can be made based on the closeness of the true finite sample size of a test to its nominal asymptotic size α. The best test according to the latter criterion depends on the particular testing context. The folklore (backed up by various simulation studies), however, is that the Wald test over-rejects in many contexts and the LM and QLR tests often are preferable in consequence.

Proof of Theorem 6.1: First we establish part (a). Element-by-element mean value expansions of $h(\widehat{\theta}_n)$ about θ_0 give

$$\begin{aligned} \sqrt{n}h(\widehat{\theta}_n) &= \sqrt{n}h(\theta_0) + \frac{\partial}{\partial\theta'}h(\theta^*_n)\sqrt{n}(\widehat{\theta}_n - \theta_0) \\ &= H\sqrt{n}(\widehat{\theta}_n - \theta_0) + o_p(1) \\ &\xrightarrow{p} Z_0 \sim N(0,\ HB_0^{-1}\Omega_0 B_0^{-1}H'), \end{aligned} \tag{9}$$

where θ^*_n lies between $\widehat{\theta}_n$ and θ_0 (and, hence, satisfies $\theta^*_n \to_p \theta_0$ by Assumption EE2(i)), the second equality holds because $h(\theta_0) = 0$ under H_0 and $\frac{\partial}{\partial\theta'}h(\theta^*_n) \to_p \frac{\partial}{\partial\theta'}h(\theta_0) = H$ using Assumption R, and the convergence in distribution uses Theorem 5.1.

By Assumptions COV, EE2, and R,

$$(\widehat{H}_n \widehat{B}_n^{-1} \widehat{\Omega}_n \widehat{B}_n^{-1} \widehat{H}_n)^{-1} \xrightarrow{p} (H B_0^{-1} \Omega_0 B_0^{-1} H')^{-1}. \tag{10}$$

Combining (9) and (10) gives

$$\mathcal{W}_n = \sqrt{n} h(\widehat{\theta}_n)' (\widehat{H}_n \widehat{B}_n^{-1} \widehat{\Omega}_n \widehat{B}_n^{-1} \widehat{H}_n)^{-1} \sqrt{n} h(\widehat{\theta}_n) \xrightarrow{d} Z_0' (H B_0^{-1} \Omega_0 B_0^{-1} H')^{-1} Z_0, \tag{11}$$

where $Z = (H B_0^{-1} \Omega_0 B_0^{-1} H')^{-1/2} Z_0 \sim N(0, I_r)$ and $Z'Z \sim \chi_r^2$.

Next, we establish part (b). Element-by-element mean value expansions of $\frac{\partial}{\partial \theta} \hat{Q}_n(\widetilde{\theta}_n)$ and $h(\widetilde{\theta}_n)$ about θ_0 yield

$$\begin{aligned} \sqrt{n} \tfrac{\partial}{\partial \theta} \hat{Q}_n(\widetilde{\theta}_n) &= \sqrt{n} \tfrac{\partial}{\partial \theta} \hat{Q}_n(\theta_0) + \frac{\partial^2}{\partial \theta \partial \theta'} \hat{Q}_n(\theta_n^*) \sqrt{n}(\widetilde{\theta}_n - \theta_0) \text{ and} \\ 0 = \sqrt{n} h(\widetilde{\theta}_n) &= \sqrt{n} h(\theta_0) + \frac{\partial}{\partial \theta'} h(\theta_n^+) \sqrt{n}(\widetilde{\theta}_n - \theta_0) \\ &= \tfrac{\partial}{\partial \theta'} h(\theta_n^+) \sqrt{n}(\widetilde{\theta}_n - \theta_0), \end{aligned} \tag{12}$$

where $\theta_n^* \to_p \theta_0$ and $\theta_n^+ \to_p \theta_0$. Let $B_n(\theta) = \frac{\partial^2}{\partial \theta \partial \theta'} \hat{Q}_n(\theta)$ and $H(\theta) = \frac{\partial}{\partial \theta'} h(\theta)$. Then, pre-multiplication of the first equation of (12) by $H(\theta_n^+) B_n(\theta_n^*)^{-1}$ gives

$$\begin{aligned} & H(\theta_n^+) B_n(\theta_n^*)^{-1} \sqrt{n} \frac{\partial}{\partial \theta} \hat{Q}_n(\widetilde{\theta}_n) \\ = \; & H(\theta_n^+) B_n(\theta_n^*)^{-1} \sqrt{n} \frac{\partial}{\partial \theta} \hat{Q}_n(\theta_0) + H(\theta_n^+) \sqrt{n}(\widetilde{\theta}_n - \theta_0) \\ = \; & H(\theta_n^+) B_n(\theta_n^*)^{-1} \sqrt{n} \frac{\partial}{\partial \theta} \hat{Q}_n(\theta_0) \\ \xrightarrow{p} \; & Z_0 \sim N(0, H B_0^{-1} \Omega_0 B_0^{-1} H') \end{aligned} \tag{13}$$

using the second equation of (12) and Assumptions CF(iii), CF(iv), REE, and R.

Combining (13), the fact that $\sqrt{n} \frac{\partial}{\partial \theta} \hat{Q}_n(\widetilde{\theta}_n) = O_p(1)$ by (17) below, $H(\theta_n^+) \to_p H$, $\widetilde{H}_n \to_p H$, $B_n(\theta_n^*) \to_p B_0$, and $\widetilde{B}_n \to_p B_0$ gives

$$\widetilde{H}_n \widetilde{B}_n^{-1} \sqrt{n} \frac{\partial}{\partial \theta} \hat{Q}_n(\widetilde{\theta}_n) \xrightarrow{d} Z_0 \sim N(0, H B_0^{-1} \Omega_0 B_0^{-1} H'). \tag{14}$$

By Assumptions COV, REE, and R,

$$(\widetilde{H}_n \widetilde{B}_n^{-1} \widetilde{\Omega}_n \widetilde{B}_n^{-1} \widetilde{H}_n')^{-1} \xrightarrow{p} (H B_0^{-1} \Omega_0 B_0^{-1} H')^{-1}. \tag{15}$$

Combining (14) and (15) gives the desired result

$$\mathcal{LM}_n \xrightarrow{d} Z_0' (H B_0^{-1} \Omega_0 B_0^{-1} H')^{-1} Z_0 = Z'Z \sim \chi_r^2 \tag{16}$$

under H_0, where Z is as above.

We now show that

$$\sqrt{n}\frac{\partial}{\partial\theta}\hat{Q}_n(\widetilde{\theta}_n) = O_p(1). \tag{17}$$

With probability that goes to one as $n \to \infty$, $\widetilde{\theta}_n$ is in the interior of Θ and there exists a random vector $\widetilde{\lambda}_n$ of Lagrange multipliers such that

$$\frac{\partial}{\partial\theta}\hat{Q}_n(\widetilde{\theta}_n) + H(\widetilde{\theta}_n)'\widetilde{\lambda}_n = o_p(n^{-1/2}). \tag{18}$$

Equations (13) and (18) combine to give

$$H(\theta_n^+)B_n(\theta_n^*)^{-1}H(\widetilde{\theta}_n)'\sqrt{n}\widetilde{\lambda}_n = O_p(1). \tag{19}$$

Since $H(\theta_n^+)B_n(\theta_n^*)^{-1}H(\widetilde{\theta}_n)' \to_p HB_0^{-1}H'$ and $HB_0^{-1}H'$ is nonsingular, (19) implies $\sqrt{n}\widetilde{\lambda}_n = O_p(1)$. This and (18) imply (17).

Next, we establish part (c). A two-term Taylor expansion of $\hat{Q}_n(\widetilde{\theta}_n)$ about $\widehat{\theta}_n$ gives

$$\begin{aligned}
\mathcal{QLR}_n &= 2n(\hat{Q}_n(\widetilde{\theta}_n) - \hat{Q}_n(\widehat{\theta}_n))/\widehat{c}_n \\
&= 2n\frac{\partial}{\partial\theta}\hat{Q}_n(\widehat{\theta}_n)(\widetilde{\theta}_n - \widehat{\theta}_n)/\widehat{c}_n + n(\widetilde{\theta}_n - \widehat{\theta}_n)'\frac{\partial^2}{\partial\theta\partial\theta'}\hat{Q}_n(\theta_n^*)(\widetilde{\theta}_n - \widehat{\theta}_n)/\widehat{c}_n \\
&= o_p(1) + n(\widetilde{\theta}_n - \widehat{\theta}_n)'\frac{\partial^2}{\partial\theta\partial\theta'}\hat{Q}_n(\theta_n^*)(\widetilde{\theta}_n - \widehat{\theta}_n)/\widehat{c}_n,
\end{aligned} \tag{20}$$

where θ_n^* lies between $\widetilde{\theta}_n$ and $\widehat{\theta}_n$ (and, hence, satisfies $\theta_n^* \to_p \theta_0$) and the third inequality holds because $n^{1/2}\frac{\partial}{\partial\theta}\hat{Q}_n(\widehat{\theta}_n) = o_p(1)$ by EE2(ii) and $n^{1/2}(\widetilde{\theta}_n - \widehat{\theta}_n) = O_p(1)$ by (22) below.

Element-by-element mean-value expansions of $(\partial/\partial\theta)\hat{Q}_n(\widetilde{\theta}_n)$ about $\widehat{\theta}_n$ gives

$$\begin{aligned}
\sqrt{n}\frac{\partial}{\partial\theta}\hat{Q}_n(\widetilde{\theta}_n) &= \sqrt{n}\frac{\partial}{\partial\theta}\hat{Q}_n(\widehat{\theta}_n) + \frac{\partial^2}{\partial\theta\partial\theta'}\hat{Q}_n(\theta_n^+)\sqrt{n}(\widetilde{\theta}_n - \widehat{\theta}_n) \\
&= o_p(1) + (B_0 + o_p(1))\sqrt{n}(\widetilde{\theta}_n - \widehat{\theta}_n),
\end{aligned} \tag{21}$$

where $\theta_n^+ \to_p \theta_0$. Since B_0 is nonsingular, (21) implies

$$\sqrt{n}(\widetilde{\theta}_n - \widehat{\theta}_n) = B_0^{-1}\sqrt{n}\frac{\partial}{\partial\theta}\hat{Q}_n(\widetilde{\theta}_n) + o_p(1). \tag{22}$$

Substitution of (22) into (20) gives

$$\begin{aligned}
\mathcal{QLR}_n &= o_p(1) + n\frac{\partial}{\partial\theta}\hat{Q}_n(\widetilde{\theta}_n)'B_0^{-1}B_n(\theta_n^*)B_0^{-1}\frac{\partial}{\partial\theta}\hat{Q}_n(\widetilde{\theta}_n)/\widehat{c}_n \\
&= o_p(1) + n\frac{\partial}{\partial\theta}\hat{Q}_n(\widetilde{\theta}_n)'\widetilde{B}_n^{-1}\frac{\partial}{\partial\theta}\hat{Q}_n(\widetilde{\theta}_n)/\widehat{c}_n \\
&= o_p(1) + \mathcal{LM}_n,
\end{aligned} \tag{23}$$

where the second equality holds by (17) and $B_n^{-1}B_n(\theta_n^*)B_0^{-1}=\widetilde{B}_0^{-1}+o_p(1)$ and the third inequality holds by the simplified expression for $\mathcal{LM}_n$ given in (8). Part (c) now holds by (23) and part (b). $\square$

3 One-Sided Hypotheses

Sometimes, one-sided hypotheses are of interest:

$$H_0 : h(\theta_0) \leq 0 \text{ vs. } H_1 : h(\theta_0) > 0. \tag{24}$$

We only consider one-dimensional hypotheses, i.e., those with real-valued $h(\cdot)$.Testing multiple dimensional inequality constraints is much trickier, as explained in Wolak (1991).

We use the same assumptions as above. We consider Wald Statistic only. Deriving the asymptotic properties of LM and QLR statistics requires deriving those for inequality-constraint extremum estimators, which is not covered by the previous lectures. The Wald statistic for testing H_0 is defined to be

$$\mathcal{W}_n = \frac{n\left[h(\widehat{\theta}_n)\right]_+^2}{\widehat{H}_n\widehat{B}_n^{-1}\widehat{\Omega}_n\widehat{B}_n^{-1}\widehat{H}_n'}, \text{ where } [a]_+ = \max\{0,a\}. \tag{25}$$

As defined, the Wald statistic is a quadratic form in the positive part of the (unrestricted) estimator $h(\widehat{\theta}_n)$ of the restrictions $h(\theta_0)$. The quadratic form has a positive definite (pd) weight matrix. If H_0 is true, then $\left[h(\widehat{\theta}_n)\right]_+$ should be close to zero and the quadratic form $\mathcal{W}_n$ in $\left[h(\widehat{\theta}_n)\right]_+$ also should be close to zero. On the other hand, if H_0 is false, $h(\widehat{\theta}_n)$ should be close to $h(\theta_0)>0$ and the quadratic form $\mathcal{W}_n$ in $\left[h(\widehat{\theta}_n)\right]_+$ should be different from zero. Thus, the Wald test rejects H_0 when its value is sufficiently large. Large sample critical values are provided by Theorem 6.2 below.

Under the one-sided null hypothesis, there are two cases: $h(\theta_0) < 0$ and $h(\theta_0) = 0$. The asymptotic distributions of the Wald statistic are different in the two cases. Theorem 6.2 summarize the result. Define the mixed chi-squared distribution

$$\chi_+^2 = [N(0,1)]_+^2. \tag{26}$$

Let "wpa1" denote "with probability approaching one".

Theorem 6.2: (a) *Under Assumptions* EE2, CF, R, *and* COV,

$$\mathcal{W}_n \begin{cases} = 0 \text{ wpa1 if } h(\theta_0) < 0 \\ \to_d \chi_+^2 \text{ if } h(\theta_0) = 0 \end{cases}.$$

Proof of Theorem 6.2: When $h(\theta_0) < 0$, $h(\widehat{\theta}_n) < 0$ wpa1 because $h(\widehat{\theta}_n) \to_p h(\theta_0)$ by Assumptions R(i) and EE2(i). Therefore, $\mathcal{W}_n = 0$ wpa1. When $h(\theta_0) = 0$, the proof is the same as that for Theorem 6.1(a) except for (11). The one-sided counterpart of (11).is:

$$\mathcal{W}_n = \frac{\left[\sqrt{n}h(\widehat{\theta}_n)\right]_+^2}{\widehat{H}_n\widehat{B}_n^{-1}\widehat{\Omega}_n\widehat{B}_n^{-1}\widehat{H}_n} \xrightarrow{d} \frac{[Z_0]_+^2}{HB_0^{-1}\Omega_0 B_0^{-1}H'} = \left[\sqrt{HB_0^{-1}\Omega_0 B_0^{-1}H'}Z_0\right]_+^2 =_d \chi_+^2. \tag{27}$$