Lecture 6
Testing Nonlinear Restrictions

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

$$H_0 : h(\theta_0) = 0 \text{ versus } H_1 : h(\theta_0) \neq 0.$$  \hspace{1cm} (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 $h(\hat{\theta}_n)$, suitably normalized, if $d_n(\hat{\theta}_n - \theta_0) \rightarrow_d Z$.

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

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

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 $\theta_0$, we have

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

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 $d_n o(||\hat{\theta}_n - \theta_0||) = o_p(1)$. This holds because

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

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 $\Omega_0$ (defined in Assumption CF) are assumed to satisfy:

**Assumption R:** (i) $h(\theta)$ is continuously differentiable on a neighborhood of $\theta_0$.
(ii) $H = \frac{\partial}{\partial \theta} h(\theta_0)$ has full rank $r$ ($\leq d$, where $d$ is the dimension of $\theta$).
(iii) $\Omega_0$ is positive definite.

For example, we might have

$$h(\theta) = \theta_1 - \theta_2,$$
$$h(\theta) = \theta_1 \theta_2 - \theta_3,$$
$$h(\theta) = \begin{pmatrix} \theta_1 \\ \theta_2 + \theta_3 - 1 \end{pmatrix}. \quad (2)$$

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

The Wald statistic for testing $H_0$ is defined to be

$$W_n = n h(\hat{\theta}_n)'(\hat{H}_n \hat{B}_n^{-1} \hat{\Omega}_n \hat{B}_n^{-1} \hat{H}_n)'^{-1} h(\hat{\theta}_n),$$
where

$$\hat{H}_n = \frac{\partial}{\partial \theta} h(\hat{\theta}_n) \quad (3)$$

and $\hat{B}_n$ and $\hat{\Omega}_n$ are consistent estimators of $B_0$ and $\Omega_0$ respectively. See Lecture 5 for the choice of $\hat{B}_n$ and $\hat{\Omega}_n$.

As defined, the Wald statistic is a quadratic form in the (unrestricted) estimator $h(\hat{\theta}_n)$ of the restrictions $h(\theta_0)$. The quadratic form has a positive definite (pd) weight matrix. If $H_0$ is true, then $h(\hat{\theta}_n)$ should be close to zero and the quadratic form $W_n$ in $h(\hat{\theta}_n)$ also should be close to zero. On the other hand, if $H_0$ is false, $h(\hat{\theta}_n)$ should be close to $h(\theta_0) \neq 0$ and the quadratic form $W_n$ in $h(\hat{\theta}_n)$ 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 $\theta_0$. The restricted extremum estimator, $\tilde{\theta}_n$, is an estimator that satisfies

$$\tilde{\theta}_n \in \Theta, \ h(\tilde{\theta}_n) = 0, \text{ and } \hat{Q}_n(\tilde{\theta}_n) \leq \inf \{ \hat{Q}_n(\theta) : \theta \in \Theta, \ h(\theta) = 0 \} + o_p(n^{-1/2}). \quad (4)$$
The LM statistic is defined by

$$\mathcal{LM}_n = n \frac{\partial}{\partial \theta} \hat{Q}_n(\tilde{\theta}_n) \tilde{B}_n^{-1} \tilde{H}_n' \tilde{H}_n \tilde{B}_n^{-1} B_n \tilde{H}_n^{-1} \tilde{H}_n' \tilde{H}_n \tilde{B}_n^{-1} \frac{\partial}{\partial \theta} \hat{Q}_n(\tilde{\theta}_n),$$

and $\tilde{B}_n$ and $\tilde{\Omega}_n$ are consistent estimators under $H_0$ of $B_0$ and $\Omega_0$, respectively, that are constructed using the restricted estimator $\tilde{\theta}_n$ rather than $\hat{\theta}_n$.

The LM statistic is a quadratic form in the vector of derivatives of the criterion function evaluated at the restricted estimator $\tilde{\theta}_n$. 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, $\tilde{\theta}_n$ and $\hat{\theta}_n$, should be close. In turn, this implies that the derivative of the criterion function at $\tilde{\theta}_n$ and $\hat{\theta}_n$ should be close. The latter, $\frac{\partial}{\partial \theta} \hat{Q}_n(\tilde{\theta}_n)$, equals zero by the first-order conditions for unrestricted minimization of $\hat{Q}_n(\theta)$ over $\Theta$. (More precisely, $\frac{\partial}{\partial \theta} \hat{Q}_n(\tilde{\theta}_n)$ is only close to zero, viz., $o_p(n^{-1/2})$, by Assumption EE2, since $\tilde{\theta}_n$ is not required to exactly minimize $\hat{Q}_n(\theta)$ just approximately minimize it.) Then, $\frac{\partial}{\partial \theta} \hat{Q}_n(\tilde{\theta}_n)$ should be close to zero under $H_0$ and the quadratic form, $\mathcal{LM}_n$, in $\frac{\partial}{\partial \theta} \hat{Q}_n(\tilde{\theta}_n)$ also should be close to zero. On the other hand, if $H_0$ is false, then $\tilde{\theta}_n$ and $\hat{\theta}_n$ should not be close, $\frac{\partial}{\partial \theta} \hat{Q}_n(\tilde{\theta}_n)$ should not be close to zero, and the quadratic form, $\mathcal{LM}_n$, in $\frac{\partial}{\partial \theta} \hat{Q}_n(\tilde{\theta}_n)$ also should not be close to zero. In consequence, the LM test rejects $H_0$ when the $\mathcal{LM}_n$ 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 $\tilde{\theta}_n$. 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) $\tilde{c}_n \rightarrow_p c$ for some sequence of non-zero random variables $\{\tilde{c}_n : n \geq 1\}$.

For example, in the ML example with correctly specified model, the information matrix equality yields $\Omega_0 = B_0$, so Assumption QLR holds with $c = \tilde{c}_n = 1$. 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

$$\tilde{c}_n = \frac{1}{n} \sum_{i=1}^n (Y_i - g(X_i, \tilde{\theta}_n))^2.$$
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 \(c = \hat{c} = 1\), since \(\Omega_0 = B_0 = \Gamma_0' V_0^{-1} \Gamma_0\).

The QLR statistic is defined by

\[ QLR_n = 2n(\hat{Q}_n(\bar{\theta}_n) - \hat{Q}_n(\bar{\theta}_n))/\hat{c}_n. \]  

(7)

When \(H_0\) is true, the restricted and unrestricted estimators, \(\tilde{\theta}_n\) and \(\hat{\theta}_n\), should be close and, hence, so should the value of the criterion function evaluated at these parameter values. Thus, under \(H_0\), the statistic \(QLR_n\) should be close to zero. Under \(H_1\), \(QLR_n\) 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 \(\theta_0\) of \(Q(\theta)\), should leave \(\hat{Q}_n(\tilde{\theta}_n)\) noticeably larger than \(\hat{Q}_n(\hat{\theta}_n)\). In consequence, the QLR test rejects \(H_0\) when \(QLR_n\) 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 \(LM_n\) statistic simplifies to

\[ LM_n = n \frac{\partial}{\partial \theta'} \hat{Q}_n(\bar{\theta}_n) \bar{B}_n^{-1} \frac{\partial}{\partial \theta} \hat{Q}_n(\bar{\theta}_n)/\hat{c}_n \text{ or} \]

\[ LM_n = n \frac{\partial}{\partial \theta'} \hat{Q}_n(\bar{\theta}_n) \bar{\Omega}_n^{-1} \frac{\partial}{\partial \theta} \hat{Q}_n(\bar{\theta}_n) \]  

(8)

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

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

**Assumption REE:** \(\bar{\theta}_n \overset{p}{\to} \theta_0\) under \(H_0\).

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

For the Wald and LM tests, we assume that consistent estimators under \(H_0\) of \(B_0\) and \(\Omega_0\) are used in constructing the test statistics:

**Assumption COV:** *For the Wald statistic, \(\tilde{B}_n \overset{p}{\to} B_0\) and \(\tilde{\Omega}_n \overset{p}{\to} \Omega_0\) under \(H_0\). For the LM statistic, \(\tilde{B}_n \overset{p}{\to} B_0\) and \(\tilde{\Omega}_n \overset{p}{\to} \Omega_0\) under \(H_0\).*

Estimators that satisfy this assumption are discussed in Lecture 5.
Theorem 6.1: (a) Under Assumptions EE2, CF, R, and COV,

\[ W_n \xrightarrow{d} \chi^2_r \text{ under } H_0, \]

where \( \chi^2_r \) denotes a chi-square distribution with \( r \) degrees of freedom.

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

\[ LM_n \xrightarrow{d} \chi^2_r \text{ under } H_0. \]

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

\[ QLR_n \xrightarrow{d} \chi^2_r \text{ under } H_0. \]

One rejects \( H_0 \) using the Wald test if \( W_n > k_{\alpha}^r \), where \( k_{\alpha}^r \) is the \( 1 - \alpha \) quantile of the \( \chi^2_r \) distribution. The LM and QLR tests are defined analogously. These tests have significance level approximately equal to \( \alpha \) 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 \( \alpha \). 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 \( h(\hat{\theta}_n) \) about \( \theta_0 \) give

\[
\sqrt{n}h(\hat{\theta}_n) = \sqrt{n}h(\theta_0) + \frac{\partial}{\partial \theta} h(\theta^*_n)\sqrt{n}(\hat{\theta}_n - \theta_0) \\
H \sqrt{n}(\hat{\theta}_n - \theta_0) + o_p(1) \\
\xrightarrow{p} Z_0 \sim N(0, HB_0^{-1}\Omega_0B_0^{-1}H')
\]

where \( \theta^*_n \) lies between \( \hat{\theta}_n \) and \( \theta_0 \) (and, hence, satisfies \( \theta^*_n \rightarrow_p \theta_0 \) by Assumption EE2(i)), the second equality holds because \( h(\theta_0) = 0 \) under \( H_0 \) and \( \frac{\partial}{\partial \theta} h(\theta^*_n) \rightarrow_p \frac{\partial}{\partial \theta} h(\theta_0) = H \) using Assumption R, and the convergence in distribution uses Theorem 5.1.
By Assumptions COV, EE2, and R,

\[
(\hat{H}_n \tilde{B}_n^{-1} \hat{\Omega}_n \tilde{B}_n^{-1} \hat{H}_n)^{-1} \overset{p}{\to} (HB_0^{-1} \Omega_0 B_0^{-1} H')^{-1}.
\] (10)

Combining (9) and (10) gives

\[
W_n = \sqrt{n}h(\hat{\theta}_n)'(\hat{H}_n \tilde{B}_n^{-1} \hat{\Omega}_n \tilde{B}_n^{-1} \hat{H}_n)^{-1} \sqrt{n}h(\hat{\theta}_n) \overset{d}{\to} Z'_0'(HB_0^{-1} \Omega_0 B_0^{-1} H')^{-1} Z_0,
\] (11)

where \( Z = (HB_0^{-1} \Omega_0 B_0^{-1} H')^{-1/2} Z_0 \sim N(0, I_r) \) and \( Z' Z \sim \chi^2_r \).

Next, we establish part (b). Element-by-element mean value expansions of \( \frac{\partial}{\partial \theta} \hat{Q}_n(\hat{\theta}_n) \) and \( h(\hat{\theta}_n) \) about \( \theta_0 \) yield

\[
\sqrt{n} \frac{\partial}{\partial \theta} \hat{Q}_n(\theta) = \sqrt{n} \frac{\partial}{\partial \theta} \hat{Q}_n(\theta_0) + \frac{\partial^2}{\partial \theta \partial \theta^T} \hat{Q}_n(\theta_0) \sqrt{n}(\theta_n - \theta_0) \text{ and }
\]

\[
0 = \sqrt{n}h(\hat{\theta}_n) = \sqrt{n}h(\theta_0) + \frac{\partial}{\partial \theta} h(\theta_0) \sqrt{n}(\theta_n - \theta_0)
\] (12)

where \( \theta_n^* \to_p \theta_0 \) and \( \theta_n^+ \to_p \theta_0 \). Let \( B_n(\theta) = \frac{\partial^2}{\partial \theta \partial \theta^T} \hat{Q}_n(\theta) \) 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

\[
H(\theta_n^*)B_n(\theta_n^*)^{-1} \sqrt{n} \frac{\partial}{\partial \theta} \hat{Q}_n(\theta_n) = H(\theta_n^*)B_n(\theta_n^*)^{-1} \sqrt{n} \frac{\partial}{\partial \theta} \hat{Q}_n(\theta_0) + H(\theta_n^+) \sqrt{n}(\theta_n - \theta_0)
\] (13)

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

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

\[
\tilde{H}_n \tilde{B}_n^{-1} \sqrt{n} \frac{\partial}{\partial \theta} \hat{Q}_n(\theta_n) \overset{d}{\to} Z_0 \sim N(0, H B_0^{-1} \Omega_0 B_0^{-1} H')
\] (14)

By Assumptions COV, REE, and R,

\[
(\hat{H}_n \tilde{B}_n^{-1} \hat{\Omega}_n \tilde{B}_n^{-1} \hat{H}_n)^{-1} \overset{p}{\to} (HB_0^{-1} \Omega_0 B_0^{-1} H')^{-1}.
\] (15)

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

\[
\mathcal{LM}_n \overset{d}{\to} Z'_0'(HB_0^{-1} \Omega_0 B_0^{-1} H')^{-1} Z_0 = Z' Z \sim \chi^2_r
\] (16)
under $H_0$, where $Z$ is as above.

We now show that
\[
\sqrt{n} \frac{\partial}{\partial \theta} \dot{Q}_n(\tilde{\theta}_n) = O_p(1).
\] (17)

With probability that goes to one as $n \to \infty$, $\tilde{\theta}_n$ is in the interior of $\Theta$ and there exists a random vector $\tilde{\lambda}_n$ of Lagrange multipliers such that
\[
\frac{\partial}{\partial \theta} \dot{Q}_n(\tilde{\theta}_n) + H(\tilde{\theta}_n)'\tilde{\lambda}_n = o_p(n^{-1/2}).
\] (18)

Equations (13) and (18) combine to give
\[
H(\theta_n^+)B_n(\theta_n^*)^{-1} H(\tilde{\theta}_n)'\sqrt{n}\tilde{\lambda}_n = O_p(1).
\] (19)

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

Next, we establish part (c). A two-term Taylor expansion of $\dot{Q}_n(\tilde{\theta}_n)$ about $\tilde{\theta}_n$ gives
\[
\text{QLR}_n = 2n(\dot{Q}_n(\tilde{\theta}_n) - \dot{Q}_n(\tilde{\theta}_n))/\tilde{c}_n
\]
\[
= 2n\frac{\partial}{\partial \theta} \dot{Q}_n(\tilde{\theta}_n)(\tilde{\theta}_n - \tilde{\theta}_n)/\tilde{c}_n + n(\tilde{\theta}_n - \tilde{\theta}_n)'\frac{\partial^2}{\partial \theta \partial \theta} \dot{Q}_n(\theta_n^*)(\tilde{\theta}_n - \tilde{\theta}_n)/\tilde{c}_n
\]
\[
= o_p(1) + n(\tilde{\theta}_n - \tilde{\theta}_n)'\frac{\partial^2}{\partial \theta \partial \theta} \dot{Q}_n(\theta_n^*)(\tilde{\theta}_n - \tilde{\theta}_n)/\tilde{c}_n,
\] (20)

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

Element-by-element mean-value expansions of $\frac{\partial}{\partial \theta} \dot{Q}_n(\tilde{\theta}_n)$ about $\tilde{\theta}_n$ gives
\[
\sqrt{n} \frac{\partial}{\partial \theta} \dot{Q}_n(\tilde{\theta}_n) = \sqrt{n} \frac{\partial}{\partial \theta} \dot{Q}_n(\tilde{\theta}_n) + \frac{\partial^2}{\partial \theta \partial \theta} \dot{Q}_n(\theta_n^*)(\tilde{\theta}_n - \tilde{\theta}_n)
\]
\[
= o_p(1) + (B_0 + o_p(1))\sqrt{n}(\tilde{\theta}_n - \tilde{\theta}_n),
\] (21)

where $\theta_n^* \to_p \theta_0$. Since $B_0$ is nonsingular, (21) implies
\[
\sqrt{n}(\tilde{\theta}_n - \tilde{\theta}_n) = B_0^{-1} \sqrt{n} \frac{\partial}{\partial \theta} \dot{Q}_n(\tilde{\theta}_n) + o_p(1).
\] (22)

Substitution of (22) into (20) gives
\[
\text{QLR}_n = o_p(1) + n \frac{\partial}{\partial \theta} \dot{Q}_n(\tilde{\theta}_n)'B_0^{-1}B_n(\theta_n^*)B_0^{-1} \frac{\partial}{\partial \theta} \dot{Q}_n(\tilde{\theta}_n)/\tilde{c}_n
\]
\[
= o_p(1) + n \frac{\partial}{\partial \theta} \dot{Q}_n(\tilde{\theta}_n)'B_n(\theta_n^*)B_0^{-1} \frac{\partial}{\partial \theta} \dot{Q}_n(\tilde{\theta}_n)/\tilde{c}_n
\]
\[
= o_p(1) + \mathcal{L}\mathcal{M}_n,
\] (23)
where the second equality holds by \( (17) \) and \( B_n^{-1}B_n(\theta_n^*)B_0^{-1} = \tilde{B}_0^{-1} + o_p(1) \) and the third inequality holds by the simplified expression for \( \mathcal{L} \mathcal{M}_n \) given in \( (8) \). Part (c) now holds by \( (23) \) and part (b). □

3 One-Sided Hypotheses

Sometimes, one-sided hypotheses are of interest:

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

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

\[
W_n = \frac{n \left[ h(\hat{\theta}_n) \right]^2}{H_n B_n^{-1} \Omega_n B_n^{-1} H_n^T}, \quad \text{where} \quad [a]_+ = \max\{0, a\}. \tag{25}
\]

As defined, the Wald statistic is a quadratic form in the positive part of the (unrestricted) estimator \( h(\hat{\theta}_n) \) of the restrictions \( h(\theta_0) \). The quadratic form has a positive definite (pd) weight matrix. If \( H_0 \) is true, then \( \left[ h(\hat{\theta}_n) \right]^2 \) should be close to zero and the quadratic form \( W_n \) also should be close to zero. On the other hand, if \( H_0 \) is false, \( h(\hat{\theta}_n) \) should be close to \( h(\theta_0) > 0 \) and the quadratic form \( W_n \) 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 6.2 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

\[
\chi^2_+ = \left[N(0, 1)\right]^2_+ \tag{26}
\]

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

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

\[
W_n \left\{ \begin{array}{l}
= 0 \text{ wpa1 if } h(\theta_0) < 0 \\
\rightarrow_d \chi^2_+ \text{ if } h(\theta_0) = 0
\end{array} \right.
\]

**Proof of Theorem 6.2**: When \( h(\theta_0) < 0 \), \( h(\hat{\theta}_n) < 0 \) wpa1 because \( h(\hat{\theta}_n) \rightarrow_p h(\theta_0) \) by Assump-
tions R(i) and EE2(i). Therefore, \( W_n = 0 \) 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:

\[
W_n = \sqrt{n}h(\hat{\theta}_n)^2 \to_d \frac{[Z_0]^2}{H B_0^{-1} \Omega_0 B_0^{-1} H'} = \left[ \sqrt{H B_0^{-1} \Omega_0 B_0^{-1} H' Z_0} \right]^2 = \chi^2_+. \tag{27}
\]