Moshe Buchinsky
Department of Economics
UCLA

Economics 203C
Spring 2003

Lecture Note 11

Testing in GMM and ML Estimation

We concentrate here on the tests that are associated with ML and GMM frameworks. The tests are: (1) *Wald* test; (b) *Lagrange Multiplier (LM)* test; and (c) *Likelihood Ratio (LR)* test.

As before we have a population joint density of x and y, conditional on the true parameter vector θ_0, is given by

$$f(y, x; \theta_0) = f(y|x; \theta_0)\, g(x),$$

where $f(y|x; \theta_0)$ represents the *correctly specified* parametric model and $g(x)$ is the density of x, which does not depend on θ_0.

The sample density is given by

$$h(y, x; \theta) = f(y|x; \theta)\, h(x).$$

The goal:

Test the hypothesis

$$H_0\colon\ r(\theta_0) = 0,$$

against the alternative hypothesis

$$H_1\colon\ r(\theta_0) \neq 0,$$

where it is assumed that the function $r(\cdot)$ is continuously differentiable.

Example: Binary logit model

$$f(y|x; \theta_0) = \frac{e^{\theta_{01}x_1 + \theta_{02}x_2}}{1 + e^{\theta_{01}x_1 + \theta_{02}x_2}}.$$

We want to test the hypothesis

$$H_0\colon\ \theta_{02} = 0,$$

that is, in this example

$$r(\theta_0) = \theta_{02}.$$

The function $r(\theta_0)$ can be a scalar, or a vector-valued function, i.e., we may want to test several restrictions on θ_0 simultaneously.

Wald Test:

Note first that if $\widehat{\theta}_n$ is an MLE or a GMM estimator, then we already established that

1. $\widehat{\theta}_n \xrightarrow{p} \theta_0$;

2. $\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \xrightarrow{D} N(0, V)$; and

3. There exists a consistent estimator for V, say $\widehat{V}_n$, that is, $\widehat{V}_n \xrightarrow{p} V$.

The asymptotic distribution for $r(\widehat{\theta}_n)$:

Let $r(\cdot)$ be a $q \times 1$ vector valued function. and let θ be a $K \times 1$ vector of parameters.

We use the *delta method* to obtain the asymptotic distribution for $r(\widehat{\theta}_n)$. By the *Mean Value Theorem* we have that

$$\sqrt{n}r(\widehat{\theta}_n) = \sqrt{n}r(\theta_0) + \frac{\partial}{\partial \theta'}r(\theta_n^*)\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right), \tag{11.1}$$

where θ_n^* is on the line segment connecting $\widehat{\theta}_n$ and θ_0.

We require here that

$$\text{rank}\left(\frac{\partial}{\partial \theta'}r(\theta_0)\right) = q,$$

that is, there is no duplication in the testing, or alternatively, there are no two rows in $r(\cdot)$ that imply the same hypotheses. Also, we denote

$$R(\theta_0)' = \frac{\partial}{\partial \theta'}r(\theta_0).$$

Rewrite (11.1) as

$$\sqrt{n}\left(r(\widehat{\theta}_n) - r(\theta_0)\right) = R(\theta_n^*)'\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right), \tag{11.2}$$

and note that we assume here that

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \xrightarrow{D} N(0, V),$$

so that the right-hand-side of (11.2) converges to a normal distribution, whose covariance matrix is given by

$$A(\theta_0) = R(\theta_0)'VR(\theta_0), \tag{11.3}$$

so that

$$\sqrt{n}\left(r(\widehat{\theta}_n) - r(\theta_0)\right) \xrightarrow{D} N(0, A(\theta_0)). \tag{11.4}$$

Note that we use here the assumption that $R(\theta_0)$ is of rank q, so that the matrix $A(\theta_0)$ is of full non-singular and, hence, can be inverted.

The result in (11.4) is very useful in that it allows us to compute the covariance of any continuously differentiable function of the parameter vector θ_0.

Note that (11.4) implies that

$$z_n = \sqrt{n}\,(A(\theta_0))^{-1/2}\left(r(\widehat{\theta}_n) - r(\theta_0)\right) \xrightarrow{D} N(0, I),$$

and consequently

$$z_n' z_n \xrightarrow{D} \chi^2(q).$$

Hence, to test the hypothesis H_0: $r(\theta_0) = 0$,we consider the statistic

$$W_n = nr(\widehat{\theta}_n)'\left(R(\widehat{\theta}_n)'\widehat{V}_n R(\widehat{\theta}_n)\right)^{-1} r(\widehat{\theta}_n), \tag{11.5}$$

where we have

$$R(\widehat{\theta}_n)'\widehat{V}_n R(\widehat{\theta}_n) \xrightarrow{p} A(\theta_0),$$

and under the null hypothesis

$$\sqrt{n}r(\widehat{\theta}_n) \xrightarrow{D} N(0, A(\theta_0)).$$

Hence, by Slutsky's theorem we have that under the null hypothesis

$$W_n \xrightarrow{D} \chi^2(q),$$

where q is the number of restrictions that are imposed by the function $r(\cdot)$.

The statistic W_n is called the *Wald statistic.*

Example: Binary logit model

The null hypothesis is

$$H_0\colon\ \theta_{02} = 0.$$

Let $x' = (x_1, x_2)$. Then

$$V(\widehat{\theta}_n) = \left(E_x\left[\frac{e^{\theta_{01}x_1+\theta_{02}x_2}}{(1+e^{\theta_{01}x_1+\theta_{02}x_2})^2}xx'\right]\right)^{-1}.$$

Note also that based on the previous lecture notes it follows that a consistent estimator for $V(\widehat{\theta}_n)$ is provided by

$$\widehat{V}(\widehat{\theta}_n) = \left(\frac{1}{n} \sum_{i=1}^{n} \frac{e^{\widehat{\theta}_{n1} x_{i1} + \widehat{\theta}_{n2} x_{i2}}}{\left(1 + e^{\widehat{\theta}_{n1} x_{i1} + \widehat{\theta}_{n2} x_{i2}}\right)^2} x_i x_i' \right)^{-1}.$$

Then we have that the Wald statistic is

$$W_n = \frac{n \widehat{\theta}_{n2}^2}{\left\{ \widehat{V}(\widehat{\theta}_n) \right\}_{2,2}},$$

where $\left\{ \widehat{V}(\widehat{\theta}_n) \right\}_{2,2}$ is the (2,2) element of the matrix $\widehat{V}(\widehat{\theta}_n)$. Note that in this case we have

$$W_n \xrightarrow{D} \chi^2(1),$$

or alternatively,

$$T_n = \sqrt{W_n} = \frac{\sqrt{n} \widehat{\theta}_{n2}}{\sqrt{\left\{ \widehat{V}(\widehat{\theta}_n) \right\}_{2,2}}} \xrightarrow{D} N(0,1).$$

The statistic T_n is called the t-ratio, which has a standard normal asymptotic distribution.

Lagrange Multiplier Test:

The setup here is the same as we had before. We need now to also define the *constrained MLE*, that is, the MLE that satisfies the restrictions imposed by the null hypothesis $r(\theta) = 0$. This is defined by

$$\widetilde{\theta}_n = \arg \max_{\theta \in \Theta, \text{ s.t. } r(\theta)=0} \frac{1}{n} \ln L_n(\theta). \tag{11.6}$$

The constrained maximization problem in (11.6) has the usual first-order conditions derived from the Lagrangian,

$$\text{Lagrangian}(\theta) = \frac{1}{n} \ln L_n(\theta) + \lambda' r(\theta),$$

that is,

$$\begin{aligned} \frac{1}{n} \frac{\partial}{\partial \theta} \ln L_n(\widetilde{\theta}_n) + R(\widetilde{\theta}_n) \widetilde{\lambda}_n = 0, \quad \text{and} \\ r(\widetilde{\theta}_n) = 0. \end{aligned} \tag{11.7}$$

By Taylor theorem we have then

$$\frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} \ln L_n(\theta_0) + \frac{1}{n} \frac{\partial^2}{\partial \theta \partial \theta'} \ln L_n(\theta_n^*) \sqrt{n} \left(\widetilde{\theta}_n - \theta_0 \right) + R(\widetilde{\theta}_n) \sqrt{n} \widetilde{\lambda}_n = 0, \tag{11.8}$$

and

$$R(\overline{\theta}_n)'\sqrt{n}\left(\widetilde{\theta}_n - \theta_0\right) = 0. \tag{11.9}$$

Note now that

$$\begin{aligned} &\frac{1}{n}\frac{\partial^2}{\partial\theta\partial\theta'}\ln L_n(\theta_n^*) \xrightarrow{p} -I(\theta_0), \\ &R(\widetilde{\theta}_n) \xrightarrow{p} R(\theta_0), \quad \text{and} \\ &R(\overline{\theta}_n) \xrightarrow{p} R(\theta_0). \end{aligned} \tag{11.10}$$

Substituting the limits from (11.10) into (11.8) gives

$$\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}\ln L_n(\theta_0) - I(\theta_0)\sqrt{n}\left(\widetilde{\theta}_n - \theta_0\right) + R(\theta_0)\sqrt{n}\widetilde{\lambda}_n = 0. \tag{11.11}$$

Multiplying both sides of (11.11) by $R(\theta_0)'I^{-1}(\theta_0)$ gives

$$R(\theta_0)'I^{-1}(\theta_0)\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}\ln L_n(\theta_0) - R(\theta_0)'\sqrt{n}\left(\widetilde{\theta}_n - \theta_0\right) + R(\theta_0)'I^{-1}(\theta_0)R(\theta_0)\sqrt{n}\widetilde{\lambda}_n = 0. \tag{11.12}$$

Note that

$$\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}\ln L_n(\theta_0) \xrightarrow{D} N(0, I(\theta_0)). \tag{11.13}$$

Also, we have that

$$R(\theta_0)'\sqrt{n}\left(\widetilde{\theta}_n - \theta_0\right) = 0,$$

because it is the limit of the first-order conditions in (11.9).

Hence, we have that

$$\sqrt{n}\widetilde{\lambda}_n \xrightarrow{D} N\left(0, \left(R(\theta_0)'I^{-1}(\theta_0)R(\theta_0)\right)^{-1}\right).$$

The *Lagrange multiplier statistic* is defined then as

$$\begin{aligned} LM_n &= \sqrt{n}\widetilde{\lambda}_n'\left(R(\theta_0)'I^{-1}(\theta_0)R(\theta_0)\right)\sqrt{n}\widetilde{\lambda}_n, \\ &= n\widetilde{\lambda}_n'\left(R(\theta_0)'I^{-1}(\theta_0)R(\theta_0)\right)\widetilde{\lambda}_n, \\ &\xrightarrow{D} \chi^2(q). \end{aligned} \tag{11.14}$$

>From the first-order conditions (11.7) we have

$$\sqrt{n}R(\widetilde{\theta}_n)\widetilde{\lambda}_n = -\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}\ln L_n(\widetilde{\theta}_n),$$

so we can rewrite the Lagrange multiplier from (11.14) as

$$LM_n = \frac{1}{\sqrt{n}} \frac{\partial}{\partial\theta} \ln L_n(\widetilde{\theta}_n)' I^{-1}(\theta_0) \frac{1}{\sqrt{n}} \frac{\partial}{\partial\theta} \ln L_n(\widetilde{\theta}_n) \tag{11.15}$$
$$= \frac{1}{n} \frac{\partial}{\partial\theta} \ln L_n(\widetilde{\theta}_n)' I^{-1}(\theta_0) \frac{\partial}{\partial\theta} \ln L_n(\widetilde{\theta}_n).$$

This expression involves the score vector evaluated at the restricted MLE, that is, at $\widetilde{\theta}_n$. This is the reason that the LM_n statistic is sometimes called the *Rao efficient score* statistic.

LM test as a mis-specification test:

Let $f(y|x;\theta)$ be the conditional pdf, corresponding to the original *unrestricted* model. Let also $f(y|x;\theta, r(\theta))$ be the conditional pdf corresponding to the model with parameter restrictions, i.e., the *restricted* model.

Note that in order to construct the LM_n statistic we only need the restricted estimate for θ_0, i.e., $\widetilde{\theta}_n$. However, $\partial \ln L_n(\theta)/\partial\theta$ and $I(\theta)$ are for the unrestricted *original* model. If we consider the original model as an extension of the restricted model, then the LM test can be used to test the significance of this extension. The *extended* model may be so complicated that it will be really difficult that we may not be feasible to obtain an estimate. However, we should be able to compute its score and information matrix. To do that we only need the restricted estimate from the simpler restricted model.

Tests of this nature of extensions of a maintained model under the null hypothesis are called *mis-specification tests*. They can be seen as a statistical procedure that allows us to perform sensitivity checks of the model.

Likelihood Ratio Test:

Again, the setup here is the same as we had before.

Let $\widehat{\theta}_n$ be the unrestricted MLE for θ_0, and let $\widetilde{\theta}_n$ be the restricted MLE for θ_0.

Consider the statistic

$$LR_n = -2 \ln \left(\frac{L_n(\widetilde{\theta}_n)}{L_n(\widehat{\theta}_n)} \right), \tag{11.16}$$
$$= 2 \left[L_n(\widehat{\theta}_n) - L_n(\widetilde{\theta}_n) \right].$$

Using Taylor's theorem we can expand $L_n(\widetilde{\theta}_n)$ around $\widehat{\theta}_n$ as follows

$$L_n(\widetilde{\theta}_n) = L_n(\widehat{\theta}_n) + \frac{\partial}{\partial\theta'} L_n(\widehat{\theta}_n) \left(\widetilde{\theta}_n - \widehat{\theta}_n\right) + \frac{1}{2} \left(\widetilde{\theta}_n - \widehat{\theta}_n\right)' \frac{\partial^2}{\partial\theta\partial\theta'} L_n(\theta_n^*) \left(\widetilde{\theta}_n - \widehat{\theta}_n\right), \tag{11.17}$$

where θ_n^* is on the line segment connecting $\widetilde{\theta}_n$ and $\widehat{\theta}_n$.

Note that the second term of (11.17) is zero, because it is merely the first order conditions for the unrestricted MLE. Substitution of $L_n(\widetilde{\theta}_n)$ from (11.17) into (11.16) gives

$$\begin{aligned} LR_n &= -\sqrt{n}\left(\widetilde{\theta}_n - \widehat{\theta}_n\right)' \sqrt{n}\left(\widetilde{\theta}_n - \widehat{\theta}_n\right), \\ &= \sqrt{n}\left(\widetilde{\theta}_n - \widehat{\theta}_n\right)' \left(-\frac{1}{n}\frac{\partial^2}{\partial\theta\partial\theta'}L_n(\theta_n^*)\right) \sqrt{n}\left(\widetilde{\theta}_n - \widehat{\theta}_n\right). \end{aligned} \tag{11.18}$$

Using Taylor's theorem again to expand the score evaluated at the restricted MLE we have

$$\begin{aligned} \frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta'}L_n(\widetilde{\theta}_n) &= \frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta'}L_n(\widehat{\theta}_n) + \frac{1}{n}\frac{\partial^2}{\partial\theta\partial\theta'}L_n(\theta_n^*)\sqrt{n}\left(\widetilde{\theta}_n - \widehat{\theta}_n\right), \\ &= \frac{1}{n}\frac{\partial^2}{\partial\theta\partial\theta'}L_n(\theta_n^*)\sqrt{n}\left(\widetilde{\theta}_n - \widehat{\theta}_n\right), \end{aligned} \tag{11.19}$$

where θ_n^* is on the line segment connecting $\widetilde{\theta}_n$ and $\widehat{\theta}_n$, and the second equality follows because $\partial L_n(\widehat{\theta}_n)/\partial\theta' = 0$.

Note also that

$$\frac{1}{n}\frac{\partial^2}{\partial\theta\partial\theta'}L_n(\theta_n^*) \xrightarrow{p} -I(\theta_0). \tag{11.20}$$

Substitution of the results in (11.19) and (11.20) into (11.18) give then that

$$LR_n \xrightarrow{D} \frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta'}L_n(\widetilde{\theta}_n)' I^{-1}(\theta_0)\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta'}L_n(\widetilde{\theta}_n)$$

The limiting distribution of $LR_n$is then the same as that for the Wald and LM tests, that is

$$LR_n \xrightarrow{D} \chi^2(q).$$

Conclusion: The Wald test, the LM test, and the LR test all have the same asymptotic distribution. That is, under the null hypothesis all tests are *asymptotically equivalent*.

Application of the Wald, LM, and LR to GMM:

Wald Test:

The Wald test applies directly to testing restrictions using the GMM estimator.

Recall that for the GMM estimator $\widehat{\theta}_n$ we have

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \xrightarrow{D} N\left(0, \Lambda(\theta_0)\right),$$

where

$$\Lambda(\theta_0) = \left(A(\theta_0)V^{-1}A(\theta_0)'\right)^{-1} A(\theta_0)V^{-1}W(\theta_0)V^{-1}A(\theta_0)' \left(A(\theta_0)V^{-1}A(\theta_0)'\right)^{-1}, \tag{11.21}$$

where

$$\begin{aligned} W(\theta_0) &\equiv E_0\left[\varphi(y,x;\theta_0)\varphi(y,x;\theta_0)'\right], \\ A(\theta_0) &\equiv E_0\left[\frac{\partial}{\partial\theta}\varphi\left(y_i,x_i;\widehat{\theta}_n\right)\right], \end{aligned}$$

and V is the limit of V_n, i.e., $V_n \xrightarrow{p} V$.

Recall also that if $V = W(\theta_0)$, then $\Lambda(\theta_0)$ in (11.21) simplifies to

$$\Lambda(\theta_0) = \left(A(\theta_0)V^{-1}A(\theta_0)'\right)^{-1}. \tag{11.22}$$

Hence, the Wald statistic is given by

$$W_n = nr(\widehat{\theta}_n)'\left(R(\widehat{\theta}_n)'\widehat{\Lambda}_n R(\widehat{\theta}_n)\right)^{-1} r(\widehat{\theta}_n), \tag{11.23}$$

where $\widehat{\Lambda}_n$ is a consistent estimator for $\Lambda(\theta_0)$.

Under the null hypothesis we have as before, for the MLE, that

$$W_n \xrightarrow{D} \chi^2(q),$$

where q is the number of restrictions that are imposed by the function $r(\cdot)$.

Lagrange Multiplier Test:

The constrained GMM estimator is defined as

$$\widetilde{\theta}_n = \arg\min_{\theta\in\Theta,\ r(\theta)=0} m_n(\theta)'V_n^{-1}m_n(\theta).$$

The first-order conditions for the minimum are

$$2\frac{\partial}{\partial\theta}m_n(\widetilde{\theta}_n)V_n^{-1}m_n(\widetilde{\theta}_n) + R\left(\widetilde{\theta}_n\right)\lambda_n = 0, \tag{11.24}$$

$$r\left(\widetilde{\theta}_n\right) = 0. \tag{11.25}$$

Using the Taylor's series expansion of (11.24) around θ_0 we get

$$2\frac{\partial}{\partial\theta}m_n(\widetilde{\theta}_n)V_n^{-1}\sqrt{n}m_n(\theta_0) + 2\frac{\partial}{\partial\theta}m_n(\widetilde{\theta}_n)V_n^{-1}\frac{\partial}{\partial\theta'}m_n\left(\theta_n^*\right)\sqrt{n}(\widetilde{\theta}_n - \theta_0) + R(\widetilde{\theta}_n)\sqrt{n}\lambda_n = 0, \tag{11.26}$$

and using the Taylor's series expansion of (11.25) around θ_0 we get

$$R\left(\overline{\theta}_n\right)\sqrt{n}(\widetilde{\theta}_n - \theta_0) = 0. \tag{11.27}$$

Under the same assumption we used in Lecture Note 10 (on the GMM estimator) we have the following:

$$\begin{aligned}
&\frac{\partial}{\partial\theta} m_n(\widetilde{\theta}_n) \xrightarrow{p} A(\theta_0), \\
&\frac{\partial}{\partial\theta} m_n\left(\theta_n^*\right) \xrightarrow{p} A(\theta_0), \\
&V_n \xrightarrow{p} V, \\
&R(\widetilde{\theta}_n) \xrightarrow{p} R(\theta_0), \\
&R(\overline{\theta}_n) \xrightarrow{p} R(\theta_0), \quad \text{and} \\
&\sqrt{n} m_n(\theta_0) \xrightarrow{D} N\left(0, W(\theta_0)\right).
\end{aligned} \tag{11.28}$$

Define

$$B(\theta_0) = A(\theta_0)V^{-1}A(\theta_0)',$$

and substitute $B(\theta_0)$ and the probability limits from (11.28) into (11.26) and (11.27) gives

$$2A(\theta_0)V^{-1}\sqrt{n}m_n(\theta_0) + 2B(\theta_0)\sqrt{n}(\widetilde{\theta}_n - \theta_0) + R(\theta_0)\sqrt{n}\lambda_n = 0, \tag{11.29}$$

$$R(\theta_0)'\sqrt{n}(\widetilde{\theta}_n - \theta_0) = 0. \tag{11.30}$$

Now, pre-multiply (11.29) by $R(\theta_0)'B^{-1}(\theta_0)$ we get

$$\begin{aligned}
0 &= 2R(\theta_0)'B^{-1}(\theta_0)A(\theta_0)V^{-1}\sqrt{n}m_n(\theta_0) + 2R(\theta_0)'\sqrt{n}(\widetilde{\theta}_n - \theta_0) + R(\theta_0)'B^{-1}(\theta_0)R(\theta_0)\sqrt{n}\lambda_n, \\
&= 2R(\theta_0)'B^{-1}(\theta_0)A(\theta_0)V^{-1}\sqrt{n}m_n(\theta_0) + R(\theta_0)'B^{-1}(\theta_0)R(\theta_0)\sqrt{n}\lambda_n.
\end{aligned}$$

where the second equality follows from (11.30).

Hence,

$$\begin{aligned}
\sqrt{n}\lambda_n &= -2\left(R(\theta_0)'B^{-1}(\theta_0)R(\theta_0)\right)^{-1} R(\theta_0)'B^{-1}(\theta_0)A(\theta_0)V^{-1}\sqrt{n}m_n(\theta_0) \\
&\xrightarrow{D} N\left(0, C(\theta_0)\right),
\end{aligned}$$

where

$$\begin{aligned}
C(\theta_0) = 4&\left(R(\theta_0)'B^{-1}(\theta_0)R(\theta_0)\right)^{-1} R(\theta_0)'B^{-1}(\theta_0)A(\theta_0)V^{-1} \\
&\times W(\theta_0) \times V^{-1}A(\theta_0)B^{-1}(\theta_0)R(\theta_0)\left(R(\theta_0)'B^{-1}(\theta_0)R(\theta_0)\right)^{-1}.
\end{aligned}$$

Note that if we choose V optimally, so that $V = W(\theta_0)$,then $C(\theta_0)$ simplifies considerably to

$$C(\theta_0) = 4\left(R(\theta_0)'B^{-1}(\theta_0)R(\theta_0)\right)^{-1}.$$

Hence, the GMM LM statistic is

$$\begin{aligned} LM_n &= \frac{n}{4}\lambda_n' R(\widetilde{\theta}_n)' \left(A(\widetilde{\theta}_n)W^{-1}(\widetilde{\theta}_n)A(\widetilde{\theta}_n)'\right)^{-1} R(\widetilde{\theta}_n)\lambda_n, \\ &\xrightarrow{D} \chi^2(q), \end{aligned} \tag{11.31}$$

with q being the number of restrictions imposed by the function $r(\cdot)$, under the null hypothesis.

Now, we use the first-order conditions (11.24) to obtain

$$\begin{aligned} LM_n &= n m_n(\widetilde{\theta}_n)'V_n^{-1}\frac{\partial}{\partial\theta'}m_n(\widetilde{\theta}_n)\left(A(\widetilde{\theta}_n)W^{-1}(\widetilde{\theta}_n)A(\widetilde{\theta}_n)'\right)^{-1}\frac{\partial}{\partial\theta}m_n(\widetilde{\theta}_n)V_n^{-1}m_n(\widetilde{\theta}_n), \\ &\xrightarrow{D} \chi^2(q). \end{aligned} \tag{11.32}$$

Remark: Note that the LM test statistic developed here is for the *optimal* weight matrix.

Likelihood Ratio Test:

Using analogous arguments as for the LR statistic developed above for the MLE, one can show that the *Distance Difference* (DD) test

$$LR_n = n\left(m_n(\widetilde{\theta}_n)'W(\widetilde{\theta}_n)m_n(\widetilde{\theta}_n) - m_n(\widehat{\theta}_n)'W(\widehat{\theta}_n)m_n(\widehat{\theta}_n)\right),$$

is asymptotically equivalent to the LM statistic defined for the GMM estimator, that is

$$LR_n \xrightarrow{D} \chi^2(q).$$

Again, note that we used here the optimal weight matrix. For other choices of the weight matrix the LR statistic has a different limiting distribution.