

Extremum Estimators

Introduction

Definition (Extremum Estimator): An estimator $\hat{\theta}$ is called an extremum estimator if there is a scalar objective function $Q_n(\mathbf{w}; \theta)$ such that

$$\hat{\theta} \in \arg \max Q_n(\mathbf{w}; \theta)$$

subject to $\theta \in \Theta \subset \mathbb{R}^p$, where

- n is the number of observations in the data
- $\mathbf{w} \equiv (\mathbf{w}_1, \dots, \mathbf{w}_n)$ is the sample or the data, and
- Θ is the set of possible parameter values

This maximization problem may not necessarily have a solution. The following lemma shows that $\hat{\theta}$ is measurable if $Q_n(\theta)$ is

Lemma (Existence of Extremum Estimators): Suppose that

1. the parameter space Θ is a compact subset of \mathbb{R}^p
2. $Q_n(\theta)$ is continuous in θ for any data \mathbf{w} , and
3. $Q_n(\theta)$ is a measurable function of \mathbf{w} for all $\theta \in \Theta$.

Then there exists $\hat{\theta}$ such that $\arg \max Q_n(\mathbf{w}; \theta)$ subject to $\theta \in \Theta$

Two Classes of Extremum Estimators

1. M-Estimators: $Q_n(\theta)$ is a simple average

$$Q_n(\theta) = \frac{1}{n} \sum_1^n m(\mathbf{w}_i; \theta)$$

- Examples: maximum likelihood (ML) and nonlinear least squares (NLS)
2. Generalized Method of Moments (GMM)

$$Q_n(\theta) = -g_n(\theta)' \hat{\mathbf{W}} g_n(\theta)$$

where

- $\hat{\mathbf{W}}$ is a $K \times K$ symmetric and positive definite matrix that defines the distance of $g_n(\theta)$ from zero.
- $g_n(\theta) = \frac{1}{n} \sum_1^n g(\mathbf{w}_i; \theta)$

M-Estimator Example: Maximum Likelihood

- \mathbf{w}_i is i.i.d.
- θ is a finite-dimensional vector

- a functional form of $f(\mathbf{w}_i; \theta)$ is known
- θ_0 is the true parameter value

The joint density of data $(\mathbf{w}_1, \dots, \mathbf{w}_n)$ is

$$f(\mathbf{w}_1, \dots, \mathbf{w}_n; \theta_0) = \prod_1^n f(\mathbf{w}_i; \theta_0)$$

The $Q_n(\theta)$ can either be the likelihood and the log-likelihood function:

$$f(\mathbf{w}_1, \dots, \mathbf{w}_n; \theta) = \prod_1^n f(\mathbf{w}_i; \theta)$$

$$\log f(\mathbf{w}_1, \dots, \mathbf{w}_n; \theta) = \log \left[\prod_1^n f(\mathbf{w}_i; \theta) \right] = \sum_1^n \log f(\mathbf{w}_i; \theta)$$

M-Estimator Example: Conditional Maximum Likelihood

- \mathbf{w}_i is partitioned into two groups, y_i and \mathbf{x}_i , and the interest is to examine how \mathbf{x}_i influences the conditional distribution of y_i
- $f(y_i|\mathbf{x}_i; \psi_0)$ be the conditional density of y_i given \mathbf{x}_i
- $f(\mathbf{x}_i; \psi_0)$ be the marginal density of \mathbf{x}_i

The joint density of data $(\mathbf{w}_1, \dots, \mathbf{w}_n) = (\mathbf{y}_t, \mathbf{x}_i)'$ is

$$f(y_t, \mathbf{x}_i; \theta_0, \psi_0) = f(y_i|\mathbf{x}_i; \theta_0)f(\mathbf{x}_i; \psi_0)$$

The $Q_n(\theta)$ can either be the likelihood and the log-likelihood function:

$$f(\mathbf{w}_i; \theta, \psi) = \prod_1^n f(y_i|\mathbf{x}_i; \theta) + \prod_1^n f(\mathbf{x}_i; \psi)$$

$$\sum_1^n \log f(\mathbf{w}_i; \theta, \psi) = \sum_1^n \log f(y_i|\mathbf{x}_i; \theta) + \sum_1^n \log f(\mathbf{x}_i; \psi)$$

M-Estimator Example: Nonlinear least square

- $y_i = \varphi_i(\mathbf{x}_i; \psi_0) + \epsilon_i$
- $\mathbb{E}(\epsilon_i|\mathbf{x}_i)$
- The functional form of φ is known

The $Q_n(\theta)$ is

$$-\frac{1}{n} \sum_1^n [y_i - \varphi_i(\mathbf{x}_i; \psi)]^2$$

M-Estimator Example: Nonlinear GMM

- $y_i = \varphi_i(\mathbf{x}_i; \psi_0) + \epsilon_i$

- $\mathbb{E}(\epsilon_i | \mathbf{x}_i)$
- The functional form of φ is known

Moment condition:

$$\mathbb{E}(\epsilon_i | \mathbf{x}_i) = \mathbf{0} \rightarrow \mathbb{E}(\epsilon_i \cdot \mathbf{x}_i) = \mathbf{0} \rightarrow \mathbb{E}\left([y_i - \varphi_i(\mathbf{x}_i; \psi)] \cdot \mathbf{x}_i\right) = \mathbf{0}$$

Using the moment condition, the $Q_n(\theta)$ is

$$Q_n(\theta) = -g_n(\theta)' \hat{\mathbf{W}} g_n(\theta)$$

where

$$g_n(\theta) = \frac{1}{n} \sum_{i=1}^n [y_i - \varphi_i(\mathbf{x}_i; \psi)] \cdot \mathbf{x}_i$$

Consistency

Consistency

If the parameter space is compact,

Proposition (Consistency with Compact Parameter Space): Suppose that

1. Θ is a compact subset of \mathbb{R}^p
2. $Q_n(\mathbf{w}; \theta)$ is a continuous function of θ for any data \mathbf{w}
3. $Q_n(\mathbf{w}; \theta)$ is a measurable function of \mathbf{w} for all $\theta \in \Theta$
4. If there is a function $Q_0(\theta)$ such that
 - (identification) $Q_0(\theta)$ is uniquely maximized at $\theta_0 \in \Theta$
 - (uniform convergence) $\sup_{\theta \in \Theta} |Q_n(\theta) - Q_0(\theta)| \rightarrow_p 0$

Then, $\hat{\theta} \rightarrow_p \theta_0$

If the parameter space is not compact,

Proposition (Consistency without Compact Parameter Space): Suppose that

1. $\theta_0 \in \text{interior}\Theta$ and Θ is a convex subset of \mathbb{R}^p
2. $Q_n(\mathbf{w}; \theta)$ is a concave over Θ of for any data \mathbf{w}
3. $Q_n(\mathbf{w}; \theta)$ is a measurable function of \mathbf{w} for all $\theta \in \Theta$
4. If there is a function $Q_0(\theta)$ such that
 - (identification) $Q_0(\theta)$ is uniquely maximized at $\theta_0 \in \Theta$
 - (point-wise convergence) $|Q_n(\theta) - Q_0(\theta)| \rightarrow_p 0$ for all $\theta \in \Theta$

Then, $\hat{\theta} \rightarrow_p \theta_0$

Above proposition presents the set of sufficient conditions under which an extremum estimator is consistent. Now, let's specialize these conditions to M-estimators and GMM estimators.

1. What is $Q_n(\theta)$ for M-Estimators and GMM?
2. What are the conditions for an M-estimator $\hat{\theta}$ to be well-defined?
3. What is the identification condition for an M-estimator?
4. What is the uniform/point-wise convergence condition and the point-wise convergence condition?

Consistency of M-Estimators

(Q1) What is $Q_0(\theta)$ in the previous consistency propositions?

For M-estimator, the objective function is:

$$Q_n(\theta) = \frac{1}{n} \sum_{i=1}^n m(\mathbf{w}_i; \theta)$$

If $\mathbb{E}[m(\mathbf{w}_i; \theta)]$ exists and is finite,

$$Q_n(\theta) = \frac{1}{n} \sum_{i=1}^n m(\mathbf{w}_i; \theta) \rightarrow_p \mathbb{E}[m(\mathbf{w}_i; \theta)]$$

Therefore,

$$Q_0(\theta) = \mathbb{E}[m(\mathbf{w}_i; \theta)]$$

Consistency of M-Estimators

(Q2) What are the conditions for an M-estimator $\hat{\theta}$ to be well-defined?

- If Θ is compact,
 - $m(\mathbf{w}_i; \theta)$ is a continuous function of θ for any data \mathbf{w}
 - $m(\mathbf{w}_i; \theta)$ is a measurable function of \mathbf{w} for all $\theta \in \Theta$
- If Θ is not compact, but is convex and $\theta \in \text{interior}\Theta$:
 - $m(\mathbf{w}_i; \theta)$ is concave over Θ for any data \mathbf{w}
 - $m(\mathbf{w}_i; \theta)$ is a measurable function of \mathbf{w} for all $\theta \in \Theta$

Consistency of M-Estimators

(Q3) What is the identification condition for an M-estimator?

Identification condition for M-estimator is $\mathbb{E}[m(\mathbf{w}_i; \theta)]$ is uniquely identified at $\theta_0 \in \Theta$

- For ML, where $m(\mathbf{w}_i; \theta) = \log f(\mathbf{y}_i | \mathbf{x}_i; \theta)$, for all $\theta \neq \theta_0$,

$$\log f(y_i | \mathbf{x}_i; \theta) \neq \log f(y_i | \mathbf{x}_i; \theta_0)$$

- For NLS, where $m(\mathbf{w}_i; \theta) = -[y_i - \varphi_i(\mathbf{x}_i; \psi)]^2$, for all $\theta \neq \theta_0$,

$$\varphi(\mathbf{x}_i; \theta) \neq \varphi(\mathbf{x}_i; \theta_0)$$

Consistency of M-Estimators

(Q4) What is the uniform and point-wise convergence conditions?

- Uniform convergence condition: by the Law of the Large Numbers, the condition becomes

$$\mathbb{E} \left[\sup_{\theta \in \Theta} |m(\mathbf{w}_i; \theta)| \right] < \infty$$

- Point-wise convergence condition: by the Ergodic Theorem, the condition becomes

$$\mathbb{E} [|m(\mathbf{w}_i; \theta)|] < \infty$$

for all $\theta \in \Theta$, (i.e., $\mathbb{E} [m(\mathbf{w}_i; \theta)]$ exists and is finite)

Consistency of GMM Estimator

(Q1) What is $Q_0(\theta)$ in the previous consistency propositions?

For GMM estimator, the objective function is:

$$Q_n(\theta) = - \left[\frac{1}{n} \sum_1^n g_n(\mathbf{w}_i; \theta) \right]' \hat{\mathbf{W}} \left[\frac{1}{n} \sum_1^n \mathbf{g}_n(\mathbf{w}_i; \theta) \right]$$

$$Q_0(\theta) = -\mathbb{E}[g(\mathbf{w}_i; \theta)]' \hat{\mathbf{W}} \mathbb{E}[\mathbf{g}(\mathbf{w}_i; \theta)]$$

(Q2) What are the conditions for an M-estimator $\hat{\theta}$ to be well-defined?

1. $g(\mathbf{w}_i; \theta)$ is a continuous function of θ for any data \mathbf{w}
2. $g(\mathbf{w}_i; \theta)$ is a measurable function of \mathbf{w} for all $\theta \in \Theta$

Consistency of GMM Estimator

(Q3) What is the identification condition for an GMM estimator?

- Notice that the maximum is zero at θ_0 , because of the orthogonality conditions, $\mathbb{E}[g(\mathbf{w}_i; \theta)] = \mathbf{0}$.
- Therefore, the identification is satisfied if for all $\theta \in \Theta$,

$$\mathbb{E}[g(\mathbf{w}_i; \theta)] \neq \mathbb{E}[\mathbf{g}(\mathbf{w}_i; \theta_0)]$$

(Q4) What is the uniform convergence condition?

$$\mathbb{E} \left[\sup_{\theta \in \Theta} ||g(\mathbf{w}_i; \theta)|| \right] < \infty$$

Asymptotic Normality

Asymptotic Normality

The General Framework

- $\hat{\theta} = \arg \max Q_n(\theta)$
- If $\bar{\theta} \in [\theta_0, \hat{\theta}]$, [Mean Value Theorem](#) or first order Taylor Expansion:

$$0 = \frac{\partial Q_n(\hat{\theta})}{\partial \theta} = \frac{\partial Q_n(\theta_0)}{\partial \theta} + \frac{\partial Q_n^2(\bar{\theta})}{\partial \theta \partial \theta'} (\hat{\theta} - \theta_0)$$

- If $\frac{\partial Q_n^2(\bar{\theta})}{\partial \theta \partial \theta'}$ is [nonsingular](#) and $\frac{\partial Q_n(\hat{\theta})}{\partial \theta} = 0$, then

$$\sqrt{n}(\hat{\theta} - \theta_0) = - \left[\frac{\partial Q_n^2(\bar{\theta})}{\partial \theta \partial \theta'} \right]^{-1} \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta}$$

$$\sqrt{n}(\hat{\theta} - \theta_0) = - \left[\frac{\partial Q_n^2(\theta_0)}{\partial \theta \partial \theta'} \right]^{-1} \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} \quad (1)$$

$$\xrightarrow{d} N(0, A^{-1} B A^{-1}) \quad (2)$$

where

$$A = \frac{\partial Q_n^2(\theta_0)}{\partial \theta \partial \theta'}$$

$$B = \text{Var} \left(\sqrt{n} \frac{\partial Q_n^2(\theta_0)}{\partial \theta \partial \theta'} \right)$$

Asymptotic Normality for M-Estimators

Let's denote

- **Score vector** as

$$\mathbf{s}(\mathbf{w}_i; \theta) = \frac{\partial \mathbf{Q}_n(\theta)}{\partial \theta} = \frac{\partial \mathbf{m}(\mathbf{w}_i; \theta)}{\partial \theta}$$

- **Hessian** as

$$\mathbf{H}(\mathbf{w}_i; \theta) = \frac{\partial^2 \mathbf{Q}_n(\theta)}{\partial \theta \partial \theta'} = \frac{\partial^2 \mathbf{m}(\mathbf{w}_i; \theta)}{\partial \theta \partial \theta'}$$

$$\frac{1}{n} \sum_1^n \mathbf{H}(\mathbf{w}_i; \bar{\theta}) \xrightarrow{p} \mathbb{E} [\mathbf{H}(\mathbf{w}_i; \theta_0)]$$

$$\frac{1}{\sqrt{n}} \sum_1^n \mathbf{s}(\mathbf{w}_i; \theta_0) \xrightarrow{d} N(\mathbf{0}, \Sigma)$$

Then by [Slutsky's theorem](#),

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d N\left(0, \mathbb{E}[\mathbf{H}(\mathbf{w}_i; \theta_0)]^{-1} \Sigma \mathbb{E}[\mathbf{H}(\mathbf{w}_i; \theta_0)]^{-1}\right)$$

Asymptotic Normality for GMM-Estimators

$$Q_n(\theta) = g_n(\theta)' W g_n(\theta)$$

where

$$g_n(\theta) = \frac{1}{n} \sum_1^n g(w_i; \theta)$$

Let $G_n(\theta)$ is the Jacobian of $g_n(\theta)$

$$\mathbf{G}_n(\theta) = \frac{\partial \mathbf{g}_n(\theta)}{\partial \theta}$$

- If $\bar{\theta} \in [\theta_0, \hat{\theta}]$,

$$0 = \mathbf{G}_n(\hat{\theta})' \mathbf{W} \mathbf{g}_n(\hat{\theta}) = \mathbf{G}_n(\hat{\theta})' \mathbf{W} \left(\mathbf{g}_n(\theta_0) + \mathbf{G}_n(\bar{\theta})(\hat{\theta} - \theta_0) \right) \quad (3)$$

$$= \mathbf{G}_n(\hat{\theta})' \mathbf{W} \mathbf{g}_n(\theta_0) + \mathbf{G}_n(\hat{\theta})' \mathbf{W} \mathbf{G}_n(\bar{\theta})(\hat{\theta} - \theta_0) \quad (4)$$

because $Q_n(\theta)$ is already a quadratic form in $g_n(\theta)$

- If $\mathbf{G}_n(\hat{\theta})' \mathbf{W} \mathbf{G}_n(\bar{\theta})$ is nonsingular, then

$$\sqrt{n}(\hat{\theta} - \theta_0) = -[\mathbf{G}_n(\hat{\theta})' \mathbf{W} \mathbf{G}_n(\bar{\theta})]^{-1} \mathbf{G}_n(\hat{\theta})' \mathbf{W} \sqrt{n} \mathbf{g}_n(\theta_0)$$

Let $G = \mathbb{E}[G_n(\theta_0)]$ and $\Omega = \mathbb{E}[g(\mathbf{w}; \theta_0)g(\mathbf{w}; \theta_0)']$

$$\sqrt{n}(\hat{\theta} - \theta_0) = (G' W G)^{-1} G' W \sqrt{n} g_n(\theta_0) \quad (5)$$

$$= (G' W G)^{-1} G' W N(0, \Omega) \quad (6)$$

$$= N\left(0, (G' W G)^{-1} G' W \Omega W (G' W G)^{-1}\right) \quad (7)$$

What is the optimal choice of the weighting matrix W ?

- The most efficient choice of $W = \Omega^{-1}$

$$\sqrt{n}(\hat{\theta} - \theta_0) = N\left(0, (G' \Omega^{-1} G)^{-1} G' \Omega^{-1} \Omega \Omega^{-1} (G' \Omega^{-1} G)^{-1}\right) \quad (8)$$

$$\xrightarrow{d} N\left(0, (G' \Omega^{-1} G)^{-1}\right) \quad (9)$$

- When G is invertible, W is irrelevant

$$\sqrt{n}(\hat{\theta} - \theta_0) = N\left(0, G^{-1}\Omega G'^{-1}\right) \quad (10)$$

$$\xrightarrow{d} N\left(0, (G'\Omega^{-1}G)^{-1}\right) \quad (11)$$

Restrictions and Hypothesis Testing

Restrictions

Let $\hat{\theta}$ be the extremum estimator in either ML or GMM. The constrained estimator, denoted $\tilde{\theta}$, solves

$$\max_{\theta \in \Theta} Q_n(\theta) \quad s. t. \quad \mathbf{a}(\theta) = \mathbf{0}$$

In many cases, economic theory suggests restrictions on the parameters of a model. For example, a demand function is supposed to be homogeneous of degree zero in prices and income.

If we have a Cobb-Douglas (log-linear) model,

$$\ln q = \beta_0 + \beta_1 \ln p_1 + \beta_2 \ln p_2 + \beta_3 \ln m + \varepsilon$$

, then we need that

$$k \ln q = \beta_0 + \beta_1 \ln kp_1 + \beta_2 \ln kp_2 + \beta_3 \ln km + \varepsilon$$

, so

$$\begin{aligned} \beta_1 \ln p_1 + \beta_2 \ln p_2 + \beta_3 \ln m &= \beta_1 \ln kp_1 + \beta_2 \ln kp_2 + \beta_3 \ln km \\ &= (\ln k) (\beta_1 + \beta_2 + \beta_3) + \beta_1 \ln p_1 + \beta_2 \ln p_2 + \beta_3 \ln m. \end{aligned}$$

The only way to guarantee this for arbitrary k is to set

$$\beta_1 + \beta_2 + \beta_3 = 0,$$

which is a **parameter restriction**. In particular, this is a linear equality restriction, which is probably the most commonly encountered case.

The general formulation of linear equality restrictions is the model

$$y = X\beta + \epsilon \quad (12)$$

$$R\beta = r \quad (13)$$

- We assume R is of rank Q , so that there are no redundant restrictions
- We also assume that $\exists \beta$ that satisfies the restrictions: they aren't infeasible Taking Lagrangean,

$$\min_{\beta, \lambda} Q_n(\beta, \lambda) = \frac{1}{n} (y - X\beta)' + 2\lambda' (R\beta - r)$$

$$H_0 : R\beta_0 = r$$

Hypothesis Testing

In many cases, one wishes to test economic theories. If theory suggests parameter restrictions, as in the above homogeneity example, one can test theory by testing parameter restrictions. A number of tests are available.

- Wald
- Lagrange multiplier (LM) - for constrained estimator
- Likelihood ratio (LR)

There is a trio of statistics called **the trinity**:

1. Wald - for unconstrained estimator
2. Lagrange multiplier (LM) - for constrained estimator
3. Likelihood ratio (LR)

that can be used for testing the null hypothesis.

- The three statistics share the same asymptotic distribution (of χ^2)
- Applicable for both ML and GMM

Null Hypothesis

Consider the problem of testing a set of r possibly nonlinear restrictions and p -dimensional model parameter:

$$H_0 : \mathbf{a}(\theta_0) = \mathbf{0}$$

- $\mathbf{a}(\theta_0)$ has dimension $(r \times 1)$
- $\mathbf{A}(\theta)$ has dimension $(r \times p)$

Assume

- $\mathbf{a}(\cdot)$ is continuously differentiable
- $\mathbf{A}(\theta)$ is the Jacobian of $\mathbf{a}(\theta)$

$$\mathbf{A}(\theta) = \frac{\partial \mathbf{a}(\theta)}{\partial \theta'}$$

- $\mathbf{A}(\theta)$ is of full (row) rank (i.e. r restrictions are not redundant)

Assumptions for the Trinity

1. Taylor expression for the sampling error:

$$\sqrt{n}(\hat{\theta} - \theta_0) = \Psi^{-1} \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} + o_p$$

where the term o_p means some random variable that converges to zero in probability, which will depend on the context.

- 2.

$$\frac{\partial Q_n(\theta_0)}{\partial \theta} \xrightarrow{d} N(0, \Sigma)$$

3. $\sqrt{n}(\tilde{\theta} - \theta_0)$ converges in distribution to a random variable, where $\tilde{\theta}$ is the constrained estimator:

$$\tilde{\theta} \in \arg \max_{\theta \in \Theta} Q_n(\theta) \quad s.t. \quad \mathbf{a}(\theta) = \mathbf{0}$$

$$\sqrt{n}(\hat{\theta} - \theta_0) = \Psi^{-1} \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} + o_p$$

Recall

- For M-estimator:

$$\sqrt{n}(\hat{\theta} - \theta_0) = - \left[\frac{\partial Q_n^2(\bar{\theta})}{\partial \theta \partial \theta'} \right]^{-1} \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta}$$

$$\Psi = \frac{\partial Q_n^2(\bar{\theta})}{\partial \theta \partial \theta'} = \mathbb{E}[\mathbf{H}(\mathbf{w}_i; \bar{\theta})]$$

- For GMM:

$$\sqrt{n}(\hat{\theta} - \theta_0) = - [\mathbf{G}_n(\hat{\theta})' \hat{\mathbf{W}} \mathbf{G}_n(\bar{\theta})]^{-1} \mathbf{G}_n(\hat{\theta})' \hat{\mathbf{W}} \sqrt{n} \mathbf{g}_n(\theta_0)$$

$$\Psi = \mathbf{G}_n(\hat{\theta})' \hat{\mathbf{W}} \mathbf{G}_n(\bar{\theta})$$

Notice that for ML and efficient GMM ($\mathbf{W} = \mathbf{\Omega}^{-1}$), then

$$\Sigma = -\Psi$$

Wald Statistic

Based on the Mean Value Theorem and Taylor expansion, under the null:

$$\sqrt{n} \mathbf{a}(\hat{\theta}) = \mathbf{A}(\theta_0) \sqrt{n} (\hat{\theta} - \theta_0) + \mathbf{o}_p \quad (14)$$

$$= -\mathbf{A}(\theta_0) \Psi^{-1} \sqrt{n} \Psi^{-1} \sqrt{n} \frac{\partial \mathbf{Q}_n(\theta_0)}{\partial \theta} + \mathbf{o}_p + \mathbf{o}_p \quad (15)$$

and the asymptotic variance is:

$$\text{AVar}(\mathbf{a}(\hat{\theta})) = \mathbf{A}(\theta_0) \Psi^{-1} \Sigma \Psi^{-1} \mathbf{A}(\theta_0) \quad (16)$$

$$= \mathbf{A}(\theta_0) \Sigma^{-1} \mathbf{A}'(\theta_0) \quad (17)$$

Since the \mathbf{A}_0 and Σ is positive definite $\text{AVar}(\mathbf{a}(\hat{\theta}))$ is positive definite. Therefore, the associated quadratic form

$$W \equiv n \mathbf{a}(\hat{\theta})' [\mathbf{A}(\hat{\theta}) \hat{\Sigma}^{-1} \mathbf{A}(\hat{\theta})']^{-1} \mathbf{a}(\hat{\theta})$$

is asymptotically $\chi^2(r)$ under the null hypothesis.

Lagrange Multiplier (LM) Statistic

$$LM \equiv n \left(\frac{\partial Q_n(\tilde{\theta})}{\partial \theta} \right)' \tilde{\Sigma}^{-1} \left(\frac{\partial Q_n(\tilde{\theta})}{\partial \theta} \right) \quad (18)$$

$$= n \gamma_n' [\mathbf{A}(\hat{\theta}) \hat{\Sigma}^{-1} \mathbf{A}(\hat{\theta})'] \gamma_n \quad (19)$$

is asymptotically $\chi^2(r)$ under the null hypothesis.

Likelihood Ratio Multiplier (LR) Statistic

$$LR \equiv 2n [Q_n(\hat{\theta}) - Q_n(\tilde{\theta})] \quad (20)$$

$$= n \gamma_n' [\mathbf{A}(\hat{\theta}) \hat{\Sigma}^{-1} \mathbf{A}(\hat{\theta})'] \gamma_n \quad (21)$$

is asymptotically $\chi^2(r)$ under the null hypothesis.