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
- $\mathbf{w} \equiv (\mathbf{w_1}, \ldots, \mathbf{w_n})$ 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 averate

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

   - 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 $(\mathbf{w_1},\ldots,\mathbf{w_n})$ is

$$f(\mathbf{w_1},\ldots,\mathbf{w_n};\theta_0)=\prod_{\mathbf{1}}^{\mathbf{n}}\mathbf{f}(\mathbf{w_i};\theta_\mathbf{0})$$

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

$$f(\mathbf{w_1},\ldots,\mathbf{w_n};\theta)=\prod_{\mathbf{1}}^{\mathbf{n}}\mathbf{f}(\mathbf{w_i};\theta)$$

$$\log f(\mathbf{w_1},\ldots,\mathbf{w_n};\theta)=\log\left[\prod_{\mathbf{1}}^{\mathbf{n}}\mathbf{f}(\mathbf{w_i};\theta)\right]=\sum_{\mathbf{1}}^{\mathbf{n}}\log\mathbf{f}(\mathbf{w_i};\theta)$$

M-Estimator Example: Conditional Maximum Likelihood

- $\mathbf{w_i}$ is partitioned into two groups, y_i an $\mathbf{x_i}$, and the interest is to examine how $\mathbf{x_i}$ influences the conditional distribution of y_i
- $f(y_i|\mathbf{x_i};\psi_\mathbf{0})$ be the conditional density of y_i given $\mathbf{x_i}$
- $f(\mathbf{x_i};\psi_\mathbf{0})$ be the marginal density of $\mathbf{x_i}$

The joint density of data $(\mathbf{w_1},\ldots,\mathbf{w_n})=(\mathbf{y_t},\mathbf{x_i'})'$ is

$$f(y_t,\mathbf{x_i};\theta_\mathbf{0},\psi_\mathbf{0})=\mathbf{f}(\mathbf{y_i}|\mathbf{x_i};\theta_\mathbf{0})\mathbf{f}(\mathbf{x_i};\psi_\mathbf{0})$$

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

$$f(\mathbf{w_i};\theta,\psi)=\prod_{\mathbf{1}}^{\mathbf{n}}\mathbf{f}(\mathbf{y_i}|\mathbf{x_i};\theta)+\prod_{\mathbf{1}}^{\mathbf{n}}\mathbf{f}(\mathbf{x_i};\psi)$$

$$\sum_{1}^{n}\log f(\mathbf{w_i};\theta,\psi)=\sum_{\mathbf{1}}^{\mathbf{n}}\log\mathbf{f}(\mathbf{y_i}|\mathbf{x_i};\theta)+\sum_{\mathbf{1}}^{\mathbf{n}}\log\mathbf{f}(\mathbf{x_i};\psi)$$

M-Estimator Example: Nonlinear least square

- $y_i=\varphi_i(\mathbf{x_i};\psi_\mathbf{0})+\epsilon_\mathbf{i}$
- $\mathbb{E}(\epsilon_i|\mathbf{x_i})$
- The functional form of φ is known

The $Q_n(\theta)$ is

$$-\frac{1}{n}\sum_{1}^{n}\left[y_i-\varphi_i(\mathbf{x_i};\psi)\right]^2$$

M-Estimator Example: Nonlinear GMM

- $y_i=\varphi_i(\mathbf{x_i};\psi_\mathbf{0})+\epsilon_\mathbf{i}$

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

Moment condition:

$$\mathbb{E}(\epsilon_i|\mathbf{x_i}) = \mathbf{0} \to \mathbb{E}(\epsilon_\mathbf{i} \cdot \mathbf{x_i}) = \mathbf{0} \to \mathbb{E}\Bigg(\Big[\mathbf{y_i} - \varphi_\mathbf{i}(\mathbf{x_i}; \psi)\Big] \cdot \mathbf{x_i}\Bigg) = \mathbf{0}$$

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

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

where

$$g_n(\theta) = \frac{1}{n}\sum_1^n \Big[y_i - \varphi_i(\mathbf{x_i}; \psi)\Big] \cdot \mathbf{x_i}$$

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) $\sup_{\theta\in\Theta} |Q_n(\theta) - Q_0(\theta)| \to_p 0$

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

If the parameter space is not compact,

Proposition (Consistency without Compact Parameter Space): Suppose that

1. $\theta_0 \in \textbf{interior}\Theta$ 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) $|Q_n(\theta) - Q_0(\theta)| \to_p 0$ for all $\theta \in \Theta$

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

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:

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

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

$$Q_n(\theta) = \frac{1}{n}\sum_{1}^{n} m(\mathbf{w_i};\theta) \rightarrow_{\mathbf{p}} \mathbb{E}\left[\mathbf{m}(\mathbf{w_i};\theta)\right]$$

Therefore,

$$Q_0(\theta) = \mathbb{E}\left[m(\mathbf{w_i};\theta)\right]$$

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 $\mathbb{E}\left[m(\mathbf{w_i};\theta)\right]$ is uniquelyidentified at $\theta_0 \in \Theta$

- For ML, where $m(\mathbf{w_i};\theta) = \log \mathbf{f}(\mathbf{y_i}|\mathbf{x_i};\theta_\mathbf{0})$, for all $\theta \neq \theta_0$,

$$\log f(y_i|\mathbf{x_i};\theta) \neq \log \mathbf{f}(\mathbf{y_i}|\mathbf{x_i};\theta_\mathbf{0})$$

- For NLS, where $m(\mathbf{w_i};\theta) = -\left[\mathbf{y_i} - \varphi_\mathbf{i}(\mathbf{x_i};\psi)\right]^\mathbf{2}$, for all $\theta \neq \theta_0$,

$$\varphi(\mathbf{x_i};\theta) \neq \varphi(\mathbf{x_i};\theta_\mathbf{0})$$

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

$$\mathbb{E}\left[|m(\mathbf{w_i};\theta)|\right] < \infty$$

  for all $\theta \in \Theta$, (i.e., $\mathbb{E}\left[m(\mathbf{w_i};\theta)\right]$ 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:

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

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

(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, $\mathbb{E}\big[g(\mathbf{w_i};\theta)\big] = \mathbf{0}$.
- Therefore, the identification is satisfied if for all $\theta \in \Theta$,

$$\mathbb{E}\big[g(\mathbf{w_i};\theta)\big] \neq \mathbb{E}\big[\mathbf{g}(\mathbf{w_i};\theta_\mathbf{0})\big]$$

(Q4) What is the uniform convergence condition?

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

Aymptotic 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

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

$$\begin{aligned}
\sqrt{n}(\hat{\theta} - \theta_0) &= -\left[\frac{\partial Q_n^2(\theta_0)}{\partial \theta \partial \theta'}\right]^{-1} \sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial \theta} && (1) \\
&\overset{d}{\to} N(0, A^{-1}BA^{-1}) && (2)
\end{aligned}$$

where

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

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

Asymptotic Normality for M-Estimators

Let's denote

- **Score vector** as

$$\mathbf{s(w_i;\theta) = \frac{\partial Q_n(\theta)}{\partial \theta} = \frac{\partial m(w_i;\theta)}{\partial \theta}}$$

- **Hessian** as

$$\mathbf{H(w_i;\theta) = \frac{\partial Q_n^2(\theta)}{\partial \theta \partial \theta'} = \frac{\partial^2 m(w_i;\theta)}{\partial \theta \partial \theta'}}$$

$$\frac{1}{n}\sum_1^n \mathbf{H(w_i;\bar{\theta})} \overset{\mathbf{p}}{\to} \mathbb{E}\left[\mathbf{H(w_i;\theta_0)}\right]$$

$$\frac{1}{\sqrt{n}}\sum_1^n \mathbf{s(w_i;\theta_0)} \overset{\mathbf{d}}{\to} \mathbf{N(0,\Sigma)}$$

Then by Slutzky's theorem,

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

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

$$\begin{aligned} 0 = \mathbf{G_n}(\hat{\theta})'\mathbf{W}\mathbf{g_n}(\hat{\theta}) &= \mathbf{G_n}(\hat{\theta})'\mathbf{W}\Big(\mathbf{g_n}(\theta_\mathbf{0}) + \mathbf{G_n}(\bar{\theta})(\hat{\theta} - \theta_\mathbf{0})\Big) &(3)\\ &= \mathbf{G_n}(\hat{\theta})'\mathbf{W}\mathbf{g_n}(\theta_\mathbf{0}) + \mathbf{G_n}(\hat{\theta})'\mathbf{W}\mathbf{G_n}(\bar{\theta})(\hat{\theta} - \theta_\mathbf{0}) &(4)\end{aligned}$$

  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

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\big[\mathbf{G_n}(\hat{\theta})'\mathbf{W}\mathbf{G_n}(\bar{\theta})\big]^{\mathbf{-1}}\mathbf{G_n}(\hat{\theta})'\mathbf{W}\sqrt{\mathbf{n}}\mathbf{g_n}(\theta_\mathbf{0})$$

Let $G = \mathbb{E}\big[G_n(\theta_0)\big]$ and $\Omega = \mathbb{E} = \big[g(\mathbf{w};\theta_\mathbf{0})\mathbf{g}(\mathbf{w};\theta_\mathbf{0})'\big]$

$$\begin{aligned}\sqrt{n}(\hat{\theta} - \theta_0) &= (G'WG)^{-1}G'W\sqrt{n}g_n(\theta_0) &(5)\\ &= (G'WG)^{-1}G'WN(0,\Omega) &(6)\\ &= N\Big(0, (G'WG)^{-1}G'W\Omega W(G'WG)^{-1}\Big) &(7)\end{aligned}$$

What is the optimal choice of the weighting matrix W?

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

$$\begin{aligned}\sqrt{n}(\hat{\theta} - \theta_0) &= N\Big(0, (G'\Omega^{-1}G)^{-1}G'\Omega^{-1}\Omega\Omega^{-1}(G'\Omega^{-1}G)^{-1}\Big) &(8)\\ &\xrightarrow{d} N\Big(0, (G'\Omega^{-1}G)^{-1}\Big) &(9)\end{aligned}$$

- When G is invertible, W is irrelevant

$$\begin{aligned}\sqrt{n}(\hat{\theta} - \theta_0) &= N\Big(0, G^{-1}\Omega G'^{-1}\Big) &(10)\\ &\xrightarrow{d} N\Big(0, (G'\Omega^{-1}G)^{-1}\Big) &(11)\end{aligned}$$

GMM vs. ML

$$\mathrm{Avar}(\hat{\theta}) \geq \mathbb{E}\big[\mathbf{s}(\mathbf{w_i};\theta_\mathbf{0})\mathbf{s}(\mathbf{w_i};\theta_\mathbf{0})'\big]^{\mathbf{-1}}$$

where

$$\mathbf{s(w_i;\theta_0)} \equiv \frac{\partial \log \mathbf{f(w_i;\theta_0)}}{\partial\theta}$$

- The lower bound for the asymptotic variance of GMM estimators is asymptotic variance of the ML estimator.
- ML is more efficient than GMM in general
- GMM with the optimal orthogonal condition is numerically equivalent to ML
- ML exploits the knowledge of the parametric form of $f(\mathbf{w_i};\theta)$ while GMM doesn't
- GMM is more robust than ML to the specification error in $f(\mathbf{w_i};\theta)$

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

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

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

$$\begin{aligned} y &= X\beta + \epsilon && (12) \\ R\beta &= r && (13) \end{aligned}$$

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

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

$$H_0 : R\beta_0 = r$$

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:

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

- $\mathbf{a}(\theta_\mathbf{0})$ 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.

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

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

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

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

Recall

- For M-estimator:

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

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

- For GMM:

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\big[\mathbf{G_n}(\hat{\theta})'\hat{\mathbf{W}}\mathbf{G_n}(\bar{\theta})\big]^{-1}\mathbf{G_n}(\hat{\theta})'\hat{\mathbf{W}}\sqrt{\mathbf{n}}\mathbf{g_n}(\theta_0)$$

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

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:

$$\begin{aligned} \sqrt{n}\, \mathbf{a}(\hat{\theta}) &= \mathbf{A}(\theta_0)\sqrt{\mathbf{n}}\ (\hat{\theta} - \theta_0) + \mathbf{o_p} && (14) \\ &= -\mathbf{A}(\theta_0)\mathbf{\Psi^{-1}}\sqrt{\mathbf{n}}\ \mathbf{\Psi^{-1}}\sqrt{\mathbf{n}}\frac{\partial \mathbf{Q_n}(\theta_0)}{\partial \theta} + \mathbf{o_p} + \mathbf{o_p} && (15) \end{aligned}$$

and the asymptotic variance is:

$$\begin{aligned} \mathrm{AVar}\big(\mathbf{a}(\hat{\theta})\big) &= \mathbf{A}(\theta_0)\mathbf{\Psi^{-1}\Sigma\Psi^{-1}}\mathbf{A}(\theta_0) && (16) \\ &= \mathbf{A}(\theta_0)\mathbf{\Sigma^{-1}}\mathbf{A}'(\theta_0) && (17) \end{aligned}$$

Since the $\mathbf{A_0}$ and Σ is positive definite $\mathrm{AVar}\big(\mathbf{a}(\hat{\theta})\big)$ is positive definite. Therefore, the associated quadratic form

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

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

Lagrange Multiplier (LM) Statistic

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

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

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

Likelihood Ratio Multiplier (LR) Statistic

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

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

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