Asymptotics for Nonlinear GMM

Eric Zivot

February 13, 2013

Asymptotic Properties of Nonlinear GMM

Under standard regularity conditions (to be discussed later), it can be shown that

$$\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}}) \xrightarrow{p} \boldsymbol{\theta}_0$$

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}}) - \boldsymbol{\theta}_0\right) \xrightarrow{d} N(\mathbf{0}, \text{avar}(\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}})))$$

where

$$\begin{aligned} \text{avar}(\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}})) &= \frac{1}{n}(\mathbf{G}'\mathbf{W}\mathbf{G})^{-1}\mathbf{G}'\mathbf{W}\mathbf{S}\mathbf{W}\mathbf{G}(\mathbf{G}'\mathbf{W}\mathbf{G})^{-1} \\ \mathbf{G} &= E\left[\frac{\partial \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right] \end{aligned}$$

For efficient GMM, set $\mathbf{W} = \mathbf{S}^{-1}$ and so

$$\text{avar}(\hat{\boldsymbol{\theta}}(\hat{\mathbf{S}}^{-1})) = \frac{1}{n}(\mathbf{G}'\mathbf{S}^{-1}\mathbf{G})^{-1}$$

Remark

Notice that with nonlinear GMM, the expression for avar$(\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}}))$ is of the same form as in linear GMM except that $\boldsymbol{\Sigma}_{xz} = E[\mathbf{x}_t\mathbf{z}_t']$ is replaced by the $K \times p$ matrix $\mathbf{G} = E\left[\frac{\partial \mathbf{g}(\mathbf{w}_t,\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right]$, where

$$
\begin{aligned}
\frac{\partial \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} &= \left[\frac{\partial \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)}{\partial \theta_1}, \ldots, \frac{\partial \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)}{\partial \theta_p}\right] \\
&= \begin{bmatrix} \frac{\partial g_1(\mathbf{w}_t,\boldsymbol{\theta}_0)}{\partial \theta_1} & \cdots & \frac{\partial g_1(\mathbf{w}_t,\boldsymbol{\theta}_0)}{\partial \theta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_K(\mathbf{w}_t,\boldsymbol{\theta}_0)}{\partial \theta_1} & \cdots & \frac{\partial g_K(\mathbf{w}_t,\boldsymbol{\theta}_0)}{\partial \theta_p} \end{bmatrix}
\end{aligned}
$$

is the $K \times p$ Jacobian matrix and

$$
\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0) = (g_1(\mathbf{w}_t, \boldsymbol{\theta}_0), \ldots, g_K(\mathbf{w}_t, \boldsymbol{\theta}_0))'
$$

Consistent estimate of avar$(\hat{\boldsymbol{\theta}}(\mathbf{W}))$

$$\widehat{\text{avar}}(\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}})) = \frac{1}{n}(\hat{\mathbf{G}}'\hat{\mathbf{W}}\hat{\mathbf{G}})^{-1}\hat{\mathbf{G}}\hat{\mathbf{W}}\hat{\mathbf{S}}\hat{\mathbf{W}}\hat{\mathbf{G}}(\hat{\mathbf{G}}'\hat{\mathbf{W}}\hat{\mathbf{G}})^{-1}$$

$$\hat{\mathbf{S}} \xrightarrow{p} \mathbf{S} = \text{avar}(\bar{\mathbf{g}})$$

where

$$\hat{\mathbf{G}} = \mathbf{G}_n(\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}}))$$

$$= n^{-1}\sum_{t=1}^{n}\frac{\partial \mathbf{g}(\mathbf{w}_t, \hat{\boldsymbol{\theta}}(\hat{\mathbf{W}}))}{\partial \boldsymbol{\theta}'} \xrightarrow{p} \mathbf{G}$$

$$\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}}) \xrightarrow{p} \boldsymbol{\theta}_0$$

For the efficient GMM estimator, $\hat{\mathbf{W}} = \hat{\mathbf{S}}^{-1}$, and

$$\widehat{\text{avar}}(\hat{\boldsymbol{\theta}}(\hat{\mathbf{S}}^{-1})) = \frac{1}{n}(\hat{\mathbf{G}}'\hat{\mathbf{S}}^{-1}\hat{\mathbf{G}})^{-1}$$

Estimation of $\mathbf{S}$

If $\{\mathbf{g}_t(\mathbf{w}_t, \boldsymbol{\theta}_0)\}$ is an ergodic stationary MDS then a consistent estimator of $\mathbf{S}$ takes the form

$$\begin{aligned}\hat{\mathbf{S}}_{\mathsf{HC}} &= n^{-1}\sum_{t=1}^{n}\mathbf{g}_t(\mathbf{w}_t,\hat{\boldsymbol{\theta}})\mathbf{g}_t(\mathbf{w}_t,\hat{\boldsymbol{\theta}})' \\ &\hat{\boldsymbol{\theta}}\xrightarrow{p}\boldsymbol{\theta}_0\end{aligned}$$

If $\{\mathbf{g}_t(\mathbf{w}_t, \boldsymbol{\theta}_0)\}$ is a mean-zero serially correlated ergodic-stationary process then

$$\begin{aligned}\mathbf{S} &= \mathsf{LRV} = \boldsymbol{\Gamma}_0 + 2\sum_{j=1}^{\infty}(\boldsymbol{\Gamma}_j + \boldsymbol{\Gamma}_j') \\ \boldsymbol{\Gamma}_j &= E[\mathbf{g}_t(\mathbf{w}_t,\boldsymbol{\theta}_0)\mathbf{g}_{t-j}(\mathbf{w}_{t-j},\boldsymbol{\theta}_0)']\end{aligned}$$

Consistent estimators of LRV and Γ_J have the form

$$\begin{aligned}
\hat{\mathbf{S}}_{\mathsf{HAC}} &= \hat{\boldsymbol{\Gamma}}_0(\hat{\boldsymbol{\theta}}) + \sum_{j=1}^{n-1} k\left(\frac{j}{q(n)}\right)(\hat{\boldsymbol{\Gamma}}_j(\hat{\boldsymbol{\theta}}) + \hat{\boldsymbol{\Gamma}}'_j(\hat{\boldsymbol{\theta}})) \\
\hat{\boldsymbol{\Gamma}}_j(\hat{\boldsymbol{\theta}}) &= \frac{1}{n-j}\sum_{t=j+1}^{n} \mathbf{g}_t(\mathbf{w}_t, \hat{\boldsymbol{\theta}})\mathbf{g}_{t-j}(\mathbf{w}_{t-j}\hat{\boldsymbol{\theta}})' \\
& \hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0
\end{aligned}$$

Proving Consistency and Asymptotic Normality for Nonlinear GMM

Hayashi's Chapter 7 (see also Hall Chapter 3 and the handbook chapter by Newey and McFadden) discusses the techniques for proving consistency and asymptotic normality for a general class of nonlinear extremum estimators.

Extremum estimators have the form

$$\begin{aligned}
\hat{\boldsymbol{\theta}} &= \arg \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset R^p} Q_n(\boldsymbol{\theta}) \\
Q_n(\boldsymbol{\theta}) &= \text{objective function} \\
\boldsymbol{\Theta} &= \text{parameter space}
\end{aligned}$$

GMM as an Extremum Estimator

For GMM, define the objective function to be maximized as

$$\begin{aligned} Q_n(\boldsymbol{\theta}) &= -\frac{1}{2n} J(\boldsymbol{\theta}, \hat{\mathbf{W}}) = -\frac{1}{2} \mathbf{g}_n(\boldsymbol{\theta})' \hat{\mathbf{W}} \mathbf{g}_n(\boldsymbol{\theta}) \\ \mathbf{g}_n(\boldsymbol{\theta}) &= \frac{1}{n} \sum_{t=1}^{n} \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0) \end{aligned}$$

If $\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}})$ minimizes $J(\boldsymbol{\theta}, \hat{\mathbf{W}})$, then it maximizes $Q_n(\boldsymbol{\theta})$

Remark: $\frac{1}{2n}$ is used in the definition of $Q_n(\boldsymbol{\theta})$ to simplify derivations (e.g. remove unnecessary constants)

M-Estimators

For the class of M-estimators, the objective function to be maximized has the form

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} m(\mathbf{w}_t, \boldsymbol{\theta})$$

$$m(\mathbf{w}_t, \boldsymbol{\theta}) = \text{real valued function}$$

Three common M-estimators are:

(1) maximum likelihood (ML) estimators;

(2) nonlinear least squares (NLS) estimators;

(3) bounded influence function robust estimators

MLE as an M-Estimator

Let $y_t|\mathbf{x}_t \sim f(y_t|\mathbf{x}_t;\boldsymbol{\theta})$, for $t = 1, \ldots, n$, with log-likelihood function

$$\ln L(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X}) = \sum_{t=1}^{n} \ln f(y_t|\mathbf{x}_t;\boldsymbol{\theta}).$$

The MLE solves

$$\hat{\boldsymbol{\theta}}_{MLE} = \arg\max_{\boldsymbol{\theta}} \sum_{t=1}^{n} \ln f(y_t|\mathbf{x}_t;\boldsymbol{\theta})$$

Then the MLE is an M-estimator with

$$\begin{aligned} m(\mathbf{w}_t,\boldsymbol{\theta}) &= \ln f(y_t|\mathbf{x}_t;\boldsymbol{\theta}) \\ Q_n(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{t=1}^{n} \ln f(y_t|\mathbf{x}_t;\boldsymbol{\theta}) \end{aligned}$$

NLS Estimator as an M-Estimator

Consider the nonlinear regression

$$y_t = \varphi(\mathbf{x}_t, \boldsymbol{\theta}) + \varepsilon_t \; t = 1, \ldots, n$$

The NLS estimator solves

$$\hat{\boldsymbol{\theta}}_{NLS} = \arg\min_{\boldsymbol{\theta}} \text{RSS}(\boldsymbol{\theta}) = \sum_{t=1}^{n} (y_t - \varphi(\mathbf{x}_t, \boldsymbol{\theta}))^2$$

Then NLS estimator is an M-estimator with

$$\begin{aligned} m(\mathbf{w}_t, \boldsymbol{\theta}) &= -(y_t - \varphi(\mathbf{x}_t, \boldsymbol{\theta}))^2 \\ Q_n(\boldsymbol{\theta}) &= -\frac{1}{n} \sum_{t=1}^{n} (y_t - \varphi(\mathbf{x}_t, \boldsymbol{\theta}))^2 \end{aligned}$$

Robust Regression as an M-estimator

Consider the linear regression

$$y_t = \mathbf{x}_t'\boldsymbol{\theta} + \varepsilon_t,\ t = 1, \ldots, n$$

To be robust to potential outliers and/or fat-tail error distributions, Huber defined the robust regression estimator as

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{RR} &= \operatorname*{arg\,min}_{\boldsymbol{\theta}} \sum_{t=1}^{n} \rho\left(y_t - \mathbf{x}_t'\boldsymbol{\theta}\right) \\ & \rho(\cdot) \text{ is a convex bounded function} \end{aligned}$$

Then Huber's robust regression estimator is an M-estimator with

$$\begin{aligned} m(\mathbf{w}_t, \boldsymbol{\theta}) &= \rho\left(y_t - \mathbf{x}_t'\boldsymbol{\theta}\right) \\ Q_n(\boldsymbol{\theta}) &= -\frac{1}{n} \sum_{t=1}^{n} \rho\left(y_t - \mathbf{x}_t'\boldsymbol{\theta}\right) \end{aligned}$$

M-estimators as Method of Moments Estimators

The first order conditions (FOCs) for an M-estimator are

$$\begin{aligned} \mathbf{0} &= \frac{\partial Q_n(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial m(\mathbf{w}_t, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{\psi}(\mathbf{w}_t, \hat{\boldsymbol{\theta}}), \\ \boldsymbol{\psi}(\mathbf{w}_t, \boldsymbol{\theta}) &= \frac{\partial m(\mathbf{w}_t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \end{aligned}$$

Therefore, M-estimators can be thought of as method of moment estimators based on the population moment

$$E[\boldsymbol{\psi}(\mathbf{w}_t, \boldsymbol{\theta}_0)] = \mathbf{0}$$

Note: this is useful for deriving the asymptotic distribution of M-estimators.

Example: MLE as Method of Moments Estimator

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{MLE} &= \arg\max_{\boldsymbol{\theta}} Q_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} \ln f(y_t|\mathbf{x}_t;\boldsymbol{\theta}) \Rightarrow \\
\mathbf{0} &= \frac{\partial Q_n(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial \ln f(y_t|\mathbf{x}_t,\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} \\
&= \frac{1}{n}\sum_{t=1}^{n} \hat{\boldsymbol{\psi}}(\mathbf{w}_t,\hat{\boldsymbol{\theta}}) \\
\hat{\boldsymbol{\psi}}(\mathbf{w}_t,\hat{\boldsymbol{\theta}}) &= \frac{\partial \ln f(y_t|\mathbf{x}_t,\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \text{likelihood score}
\end{aligned}$$

The population moment is

$$E[\boldsymbol{\psi}(\mathbf{w}_t,\boldsymbol{\theta}_0)] = E\left[\frac{\partial \ln f(y_t|\mathbf{x}_t,\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right] = \mathbf{0}$$

GMM as an M-estimator

GMM can be thought of as an M-estimator. The first order condtions for GMM are

$$\begin{aligned}\mathbf{0} &= \frac{\partial Q_n(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = -\mathbf{G}_n(\hat{\boldsymbol{\theta}})'\hat{\mathbf{W}}\mathbf{g}_n(\hat{\boldsymbol{\theta}}) = -\mathbf{G}_n(\hat{\boldsymbol{\theta}})'\hat{\mathbf{W}}\frac{1}{n}\sum_{t=1}^{n}\mathbf{g}(\mathbf{w}_t,\hat{\boldsymbol{\theta}}) \\ &= \frac{1}{n}\sum_{t=1}^{n}\hat{\boldsymbol{\psi}}(\mathbf{w}_t,\hat{\boldsymbol{\theta}}), \\ \hat{\boldsymbol{\psi}}(\mathbf{w}_t,\boldsymbol{\theta}) &= -\mathbf{G}_n(\hat{\boldsymbol{\theta}})'\hat{\mathbf{W}}\mathbf{g}(\mathbf{w}_t,\boldsymbol{\theta})\end{aligned}$$

The population moment is

$$E[\boldsymbol{\psi}(\mathbf{w}_t,\boldsymbol{\theta}_0)] = E\left[\mathbf{G}'\mathbf{W}\mathbf{g}(\mathbf{w}_t,\boldsymbol{\theta}_0)\right] = \mathbf{G}'\mathbf{W}E[\mathbf{g}(\mathbf{w}_t,\boldsymbol{\theta}_0)] = \mathbf{0}$$

Note: Defined this way, there are $K = L$ population moments defined by $\mathbf{G}'\mathbf{W}E[\mathbf{g}(\mathbf{w}_t,\boldsymbol{\theta}_0)] = \mathbf{0}$.

Consistency of Extremum Estimators

Basic Idea: Let $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset R^p} Q_n(\boldsymbol{\theta})$. If

- $Q_n(\boldsymbol{\theta}) \xrightarrow{p} Q_0(\boldsymbol{\theta})$ uniformly in $\boldsymbol{\theta}$,

- $\boldsymbol{\theta}_0$ uniquely maximizes $Q_0(\boldsymbol{\theta})$,

then $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}$.

Technical requirements

1. Continuity of $Q_n(\boldsymbol{\theta})$ and $Q_0(\boldsymbol{\theta})$

2. Uniform convergence of $Q_n(\boldsymbol{\theta})$ to $Q_0(\boldsymbol{\theta})$

$$\sup_{\theta \in \Theta} |Q_n(\boldsymbol{\theta}) - Q_0(\boldsymbol{\theta})| \xrightarrow{p} 0 \text{ as } n \to \infty$$

3. Compact parameter space Θ

4. $\boldsymbol{\theta}_0$ uniquely maximizes $Q_0(\boldsymbol{\theta})$

Note: See Hall Chapter 3 or Newey and McFadden (1994) for a formal proof based on ε and δ arguments

Application to GMM

Recall

$$\begin{aligned} Q_n(\boldsymbol{\theta}) &= -\frac{1}{2n} J(\boldsymbol{\theta}, \hat{\mathbf{W}}) = -\frac{1}{2}\mathbf{g}_n(\boldsymbol{\theta})'\hat{\mathbf{W}}\mathbf{g}_n(\boldsymbol{\theta}) \\ \hat{\boldsymbol{\theta}} &= \arg \max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}\subset R^p} Q_n(\boldsymbol{\theta}) \\ &= \arg \min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}\subset R^p} J(\boldsymbol{\theta}, \hat{\mathbf{W}}) \end{aligned}$$

Now

1. $Q_n(\boldsymbol{\theta})$ is continuous provided $\mathbf{g}_n(\boldsymbol{\theta})$ is continuous

2. If $\{\mathbf{w}_t\}$ is ergodic-stationary and $E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})]$ exists, then

$$\begin{aligned} \mathbf{g}_n(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{t=1}^{n} \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}) \xrightarrow{p} E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})] \text{ (for fixed } \theta) \\ Q_0(\boldsymbol{\theta}) &= -\frac{1}{2}E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})]'\mathbf{W}E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})] \end{aligned}$$

3. If $E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)] = \mathbf{0}$ and $E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})] \neq \mathbf{0}$ for any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ then

$$\begin{aligned} Q_0(\boldsymbol{\theta}_0) &= -\frac{1}{2}E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)]'\mathbf{W}E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)] = 0 \\ Q_0(\boldsymbol{\theta}) &< 0 \text{ for } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{aligned}$$

Therefore, $Q_0(\boldsymbol{\theta})$ is uniquely maximized at $\boldsymbol{\theta}_0$

4. Uniform convergence of $Q_n(\boldsymbol{\theta})$ to $Q_0(\boldsymbol{\theta})$ requires uniform convergence of $\mathbf{g}_n(\boldsymbol{\theta})$ to $E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})]$. Hayashi states that a sufficient condition for

$$\sup_{\theta \in \boldsymbol{\Theta}} |\mathbf{g}_n(\boldsymbol{\theta}) - E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})]| \xrightarrow{p} \mathbf{0} \text{ as } n \to \infty$$

is

$$E[\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} ||\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})||] < \infty$$

where

$$\begin{aligned} ||\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})|| &= \left(g_1(\mathbf{w}_t, \boldsymbol{\theta})^2 + \cdots + g_K(\mathbf{w}_t, \boldsymbol{\theta})^2\right)^{1/2} \\ &= \text{Euclidean norm} \end{aligned}$$

Asymptotic Normality of GMM

Recall

$$\begin{aligned} Q_n(\boldsymbol{\theta}) &= -\frac{1}{2n} J(\boldsymbol{\theta}, \hat{\mathbf{W}}) = -\frac{1}{2}\mathbf{g}_n(\boldsymbol{\theta})' \hat{\mathbf{W}} \mathbf{g}_n(\boldsymbol{\theta}) \\ \hat{\boldsymbol{\theta}} &= \arg \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset R^p} Q_n(\boldsymbol{\theta}) \end{aligned}$$

Assume

1. $\boldsymbol{\theta}_0 \subset \boldsymbol{\Theta}$ compact

2. $\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})$ is $K \times 1$ and continuously differentiable in $\boldsymbol{\theta}$ for any $\mathbf{w}_t$

3. $E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)] = \mathbf{0}$ and $E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)] \neq \mathbf{0}$ for $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$

4. $\mathbf{G} = E[\frac{\partial \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}]$ has full column rank P

5. $\sqrt{n}\mathbf{g}_n(\boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{S})$

6. $E[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} ||\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})||] < \infty$

Basic trick: Apply exact TSE (i.e., Mean Value Theorem) to nonlinear sample moment $\mathbf{g}_n(\boldsymbol{\theta})$ to get a linear function of $\boldsymbol{\theta}$

Since $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}\subset R^p} Q_n(\boldsymbol{\theta})$, the first order condtions (FOC) are

$$\underset{p\times 1}{\mathbf{0}} = \frac{\partial Q_n(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = -\mathbf{G}_n(\hat{\boldsymbol{\theta}})'\hat{\mathbf{W}}\mathbf{g}_n(\hat{\boldsymbol{\theta}})$$
$$\underset{K\times p}{\mathbf{G}_n(\boldsymbol{\theta})} = \frac{\partial \mathbf{g}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \left[\frac{\partial \mathbf{g}_n(\boldsymbol{\theta})}{\partial \theta_1}, \cdots, \frac{\partial \mathbf{g}_n(\boldsymbol{\theta})}{\partial \theta_p}\right]$$

Now apply exact TSE to $\mathbf{g}_n(\hat{\boldsymbol{\theta}})$ about $\boldsymbol{\theta}_0$

$$\mathbf{g}_n(\hat{\boldsymbol{\theta}}) = \mathbf{g}_n(\boldsymbol{\theta}_0) + \mathbf{G}_n(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$$
$$\tilde{\boldsymbol{\theta}}_i = \lambda_i\hat{\boldsymbol{\theta}}_i + (1-\lambda_i)\boldsymbol{\theta}_{i,0}, \; i = 1, \ldots, p$$
$$\lambda_i \in (0, 1)$$

Substitute TSE into FOC

$$\underset{p\times 1}{\mathbf{0}} = -\mathbf{G}_n(\hat{\boldsymbol{\theta}})'\hat{\mathbf{W}}\left[\mathbf{g}_n(\boldsymbol{\theta}_0) + \mathbf{G}_n(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\right]$$
$$= -\mathbf{G}_n(\hat{\boldsymbol{\theta}})'\hat{\mathbf{W}}\mathbf{g}_n(\boldsymbol{\theta}_0) - \mathbf{G}_n(\hat{\boldsymbol{\theta}})'\hat{\mathbf{W}}\mathbf{G}_n(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$$

Solve for $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\left[\mathbf{G}_n(\hat{\boldsymbol{\theta}})'\hat{\mathbf{W}}\mathbf{G}_n(\tilde{\boldsymbol{\theta}})\right]^{-1}\mathbf{G}_n(\hat{\boldsymbol{\theta}})'\hat{\mathbf{W}}\mathbf{g}_n(\boldsymbol{\theta}_0)$$

Multiply both sides by $\sqrt{n}$

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\left[\mathbf{G}_n(\hat{\boldsymbol{\theta}})'\hat{\mathbf{W}}\mathbf{G}_n(\tilde{\boldsymbol{\theta}})\right]^{-1}\mathbf{G}_n(\hat{\boldsymbol{\theta}})'\hat{\mathbf{W}}\sqrt{n}\mathbf{g}_n(\boldsymbol{\theta}_0)$$

Asymptotics

Now,

$$\underset{K\times p}{\mathbf{G}_n(\boldsymbol{\theta})}=\frac{\partial \mathbf{g}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$$

By the ergodic theorem

$$\mathbf{G}_n(\boldsymbol{\theta}_0) \xrightarrow{p} E\left[\frac{\partial \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right] = \mathbf{G}$$

Also, we know that

$$\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0 \Rightarrow \tilde{\boldsymbol{\theta}}_i = \lambda_i \hat{\boldsymbol{\theta}}_i + (1-\lambda_i)\boldsymbol{\theta}_{i,0} \xrightarrow{p} \boldsymbol{\theta}_{i,0}$$

If

$$\mathbf{G}_n(\boldsymbol{\theta}) \xrightarrow{p} E\left[\frac{\partial \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right] \text{ uniformly in } \boldsymbol{\theta}$$

then

$$\mathbf{G}_n(\hat{\boldsymbol{\theta}}), \mathbf{G}_n(\tilde{\boldsymbol{\theta}}) \xrightarrow{p} \mathbf{G}$$

A sufficient condition for uniform convergence of $\mathbf{G}_n(\boldsymbol{\theta})$ to $E\left[\frac{\partial \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right]$ is

$$E\left[\sup_{\theta}\left\|\frac{\partial \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right\|\right] < \infty$$

Finally, by assumption

$$\sqrt{n}\mathbf{g}_n(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{S})$$

Therefore,

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &\xrightarrow{d} (\mathbf{G}'\mathbf{W}\mathbf{G})^{-1}\mathbf{G}'\mathbf{W} \times N(\mathbf{0}, \mathbf{S}) \\ &\equiv N(0, \mathbf{V}) \\ \mathbf{V} &= (\mathbf{G}'\mathbf{W}\mathbf{G})^{-1}\mathbf{G}'\mathbf{W}\mathbf{S}\mathbf{W}\mathbf{G}(\mathbf{G}'\mathbf{W}\mathbf{G})^{-1} \end{aligned}$$

Remark: For efficient GMM, $\mathbf{W} = \mathbf{S}^{-1}$ and

$$\mathbf{V} = \text{avar}(\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)) = (\mathbf{G}'\mathbf{S}^{-1}\mathbf{G})^{-1}$$

Remark

The preceeding results imply the following simplifications provided $\hat{\boldsymbol{\theta}}$ is close to $\boldsymbol{\theta}_0$ so that $\mathbf{G}_n(\tilde{\boldsymbol{\theta}}) = \mathbf{G}_n(\boldsymbol{\theta}_0) + o_p(1)$:

$$\begin{aligned} \mathbf{g}_n(\hat{\boldsymbol{\theta}}) &= \mathbf{g}_n(\boldsymbol{\theta}_0) + \mathbf{G}_n(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ &= \mathbf{g}_n(\boldsymbol{\theta}_0) + \mathbf{G}_n(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o_p(1) \end{aligned}$$

and

$$\begin{aligned} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &= -\left[\mathbf{G}_n(\hat{\boldsymbol{\theta}})'\hat{\mathbf{W}}\mathbf{G}_n(\tilde{\boldsymbol{\theta}})\right]^{-1}\mathbf{G}_n(\hat{\boldsymbol{\theta}})'\hat{\mathbf{W}}\mathbf{g}_n(\boldsymbol{\theta}_0) \\ &= -\left[\mathbf{G}_n(\boldsymbol{\theta}_0)'\hat{\mathbf{W}}\mathbf{G}_n(\boldsymbol{\theta}_0)\right]^{-1}\mathbf{G}_n(\boldsymbol{\theta}_0)'\hat{\mathbf{W}}\mathbf{g}_n(\boldsymbol{\theta}_0) \\ &\quad + o_p(1) \end{aligned}$$

where $o_p(1)$ represents terms that converge in probability to zero.

These simplifications will be useful in the following derivations.

Remark

It is not wise to try to prove consistency by linearizing the FOCs

$$\underset{p\times 1}{\mathbf{0}} = \frac{\partial Q_n(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = -\mathbf{G}_n(\hat{\boldsymbol{\theta}})'\hat{\mathbf{W}}\mathbf{g}_n(\hat{\boldsymbol{\theta}})$$

using the Mean Value Theorem. This is because $\hat{\boldsymbol{\theta}}$ that solves the FOCs may be a local-maximum to $Q_n(\boldsymbol{\theta})$ and so $\hat{\boldsymbol{\theta}}$ may not converge to $\boldsymbol{\theta}_0$.

Note that the proof of asymptotic normality assumes that $\hat{\boldsymbol{\theta}}$ is consistent for $\boldsymbol{\theta}$ so we avoid complications associated with local maximum.

Hypothesis Testing in Nonlinear GMM Models

The main types of hypothesis tests

- Overidentification restrictions
- Coefficient restrictions (linear and nonlinear)
- Subsets of orthogonality restrictions

work the same way in nonlinear GMM as they do with linear GMM.

Remarks:

1. One should always first test the overidentifying restrictions before conducting the other tests. If the model specification is rejected, it does not make sense to do the remaining tests.

2. Testing instrument relevance is not straightforward in nonlinear models.

Asymptotic Distribution of Sample Moments and J-statistic

By assumption, the normalized sample moment evaluated at $\boldsymbol{\theta}_0$ is asymptotically normally distributed

$$\begin{aligned}\sqrt{n}\mathbf{g}_n(\boldsymbol{\theta}_0) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} g(\mathbf{w}_t, \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{S}) \\ \underset{K\times K}{\mathbf{S}} &= \text{avar}(\sqrt{n}\mathbf{g}_n(\boldsymbol{\theta}_0))\end{aligned}$$

As a result, the J-statistic evaluated at $\boldsymbol{\theta}_0$ is asymptotically chi-square distributed with K degrees of freedom

$$J = J(\boldsymbol{\theta}_0, \hat{\mathbf{S}}^{-1}) = n\mathbf{g}_n(\boldsymbol{\theta}_0)'\hat{\mathbf{S}}^{-1}\mathbf{g}_n(\boldsymbol{\theta}_0) \xrightarrow{d} \chi^2(K)$$

provided $\hat{\mathbf{S}}\xrightarrow{p}\mathbf{S}$.

As with linear GMM, the normalized sample moment evaluated at the efficient GMM estimate $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\hat{\mathrm{S}}^{-1})$ is asymptotically normally distributed

$$\hat{\mathbf{S}}^{-1/2}\sqrt{n}\mathbf{g}_n(\hat{\boldsymbol{\theta}}) \xrightarrow{d} N(\mathbf{0}, (\mathbf{I}_K - \mathbf{P}_F))$$

$$\mathbf{P}_F = \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}',\ \mathbf{F} = \mathbf{S}^{-1/2}\mathbf{G},\ \mathbf{G} = E\left[\frac{\partial \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right]$$

$$\text{rank}\ (\mathbf{I}_K - \mathbf{P}_F) = K - P$$

To derive this result, combine the TSE of $\mathbf{g}_n(\hat{\boldsymbol{\theta}})$ about $\boldsymbol{\theta}_0$ with the the F.O.C for the efficient GMM optimization

$$\begin{aligned}
\mathbf{g}_n(\hat{\boldsymbol{\theta}}) &= \mathbf{g}_n(\boldsymbol{\theta}_0) + \mathbf{G}_n(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o_p(1) \\
(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &= -\left[\mathbf{G}_n(\boldsymbol{\theta}_0)'\hat{\mathbf{S}}^{-1}\mathbf{G}_n(\boldsymbol{\theta}_0)\right]^{-1}\mathbf{G}_n(\boldsymbol{\theta}_0)'\hat{\mathbf{S}}^{-1}\mathbf{g}_n(\boldsymbol{\theta}_0) + o_p(1)
\end{aligned}$$

to give

$$\begin{gathered}
\mathbf{g}_n(\hat{\boldsymbol{\theta}}) = \mathbf{g}_n(\boldsymbol{\theta}_0) \\
-\mathbf{G}_n(\boldsymbol{\theta}_0)\left[\mathbf{G}_n(\boldsymbol{\theta}_0)'\hat{\mathbf{S}}^{-1}\mathbf{G}_n(\boldsymbol{\theta}_0)\right]^{-1}\mathbf{G}_n(\boldsymbol{\theta}_0)'\hat{\mathbf{S}}^{-1}\mathbf{g}_n(\boldsymbol{\theta}_0) + o_p(1)
\end{gathered}$$

Let $\hat{\mathbf{S}}^{-1} = \hat{\mathbf{S}}^{-1/2\prime}\hat{\mathbf{S}}^{-1/2}$. Then write the normalized sample moment as

$$\hat{\mathbf{S}}^{-1/2}\sqrt{n}\mathbf{g}_n(\hat{\boldsymbol{\theta}}) = \left[\mathbf{I}_K - \mathbf{P}_{\hat{F}}\right]\hat{\mathbf{S}}^{-1/2}\sqrt{n}\mathbf{g}_n(\boldsymbol{\theta}_0) + o_p(1)$$

where

$$\hat{\mathbf{F}} = \hat{\mathbf{S}}^{-1/2}\mathbf{G}_n(\boldsymbol{\theta}_0),\ \mathbf{P}_{\hat{F}} = \hat{\mathbf{F}}(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}'$$

Then by the asymptotic normality of $\sqrt{n}\mathbf{g}_n(\boldsymbol{\theta}_0)$ and Slutsky's theorem

$$\begin{aligned}\hat{\mathbf{S}}^{-1/2}\sqrt{n}\mathbf{g}_n(\hat{\boldsymbol{\theta}}) &\xrightarrow{d} [\mathbf{I}_k - \mathbf{P}_F] \times N(\mathbf{0}, \mathbf{I}_K) \\ &\equiv N(\mathbf{0}, \mathbf{I}_k - \mathbf{P}_F)\end{aligned}$$

since $\mathbf{I}_k - \mathbf{P}_F$ is idempotent.

From the previous result, it follows that the asymptotic distribution of the J-statistic evaluated at the efficient GMM estimate $\hat{\boldsymbol{\theta}}(\hat{\mathrm{S}}^{-1})$ is chi-square with $K-P$ degrees of freedom

$$\begin{aligned} J(\hat{\boldsymbol{\theta}}(\hat{\mathrm{S}}^{-1}),\hat{\mathrm{S}}^{-1}) &= n\mathbf{g}_n(\hat{\boldsymbol{\theta}}(\hat{\mathrm{S}}^{-1}))'\hat{\mathrm{S}}^{-1}\mathbf{g}_n(\hat{\boldsymbol{\theta}}(\hat{\mathrm{S}}^{-1})) \\ &= \left(\hat{\mathrm{S}}^{-1/2}\sqrt{n}\mathbf{g}_n(\hat{\boldsymbol{\theta}}(\hat{\mathrm{S}}^{-1}))\right)' \\ &\quad \times \left(\hat{\mathrm{S}}^{-1/2}\sqrt{n}\mathbf{g}_n(\hat{\boldsymbol{\theta}}(\hat{\mathrm{S}}^{-1}))\right) \\ &\quad \xrightarrow{d} \chi^2(K-P) \end{aligned}$$

Therefore, we reject the overidentifying restrictions at the 5% level if

$$J(\hat{\boldsymbol{\theta}}(\hat{\mathrm{S}}^{-1}),\hat{\mathrm{S}}^{-1}) > \chi^2_{0.95}(K-P)$$

Testing Restrictions on Parameters: The Trilogy of Tests

For ease of exposition, consider testing the simple hypotheses

$$
\begin{aligned}
H_0 &: \underset{P\times 1}{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 \\
H_1 &: \boldsymbol{\theta} \neq \boldsymbol{\theta}_0
\end{aligned}
$$

We consider three types of tests

- Wald test
- GMM-LR test (difference in J-statistics)
- GMM-LM test (GMM score test)

Let $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\hat{\mathbf{S}}^{-1})$ denote the efficient GMM estimator. The three test statistics have the form

$$
\begin{aligned}
Wald_{\mathsf{GMM}} &= n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' \left[\mathbf{G}_n(\hat{\boldsymbol{\theta}})'\hat{\mathbf{S}}^{-1}\mathbf{G}_n(\hat{\boldsymbol{\theta}})\right] (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\
LR_{\mathsf{GMM}} &= J(\boldsymbol{\theta}_0, \hat{\mathbf{S}}^{-1}) - J(\hat{\boldsymbol{\theta}}, \hat{\mathbf{S}}^{-1}) \\
LM_{\mathsf{GMM}} &= n\mathbf{g}_n(\boldsymbol{\theta}_0)'\hat{\mathbf{S}}^{-1}\mathbf{G}_n(\boldsymbol{\theta}_0) \left[\mathbf{G}_n(\boldsymbol{\theta}_0)'\hat{\mathbf{S}}^{-1}\mathbf{G}_n(\boldsymbol{\theta}_0)\right] \\
&\quad \times \mathbf{G}_n(\boldsymbol{\theta}_0)'\hat{\mathbf{S}}^{-1}\mathbf{g}_n(\boldsymbol{\theta}_0)
\end{aligned}
$$

where the same value of $\hat{\mathbf{S}}^{-1}$ is used for all statistics.

Result: Under $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$

$$Wald_{\mathsf{GMM}}, LR_{\mathsf{GMM}}, LM_{\mathsf{GMM}} \xrightarrow{d} \chi^2(P)$$

Derivation of LM_{GMM} (score) statistic

The score statistic is based on the score of the GMM objective function

$$\underset{p\times 1}{\mathbf{0}} = \frac{\partial Q_n(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = -\mathbf{G}_n(\hat{\boldsymbol{\theta}})'\hat{\mathbf{S}}^{-1}\mathbf{g}_n(\hat{\boldsymbol{\theta}})$$

Intuition: if $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ is true then

$$\underset{p\times 1}{\mathbf{0}} \approx \frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = -\mathbf{G}_n(\boldsymbol{\theta}_0)'\hat{\mathbf{S}}^{-1}\mathbf{g}_n(\boldsymbol{\theta}_0)$$

whereas if $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ is not true then

$$\underset{p\times 1}{\mathbf{0}} \neq \frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = -\mathbf{G}_n(\boldsymbol{\theta}_0)'\hat{\mathbf{S}}^{-1}\mathbf{g}_n(\boldsymbol{\theta}_0)$$

Assuming $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ is true

$$\begin{aligned} \sqrt{n}\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} &= -\mathbf{G}_n(\boldsymbol{\theta}_0)'\hat{\mathbf{S}}^{-1}\sqrt{n}\mathbf{g}_n(\boldsymbol{\theta}_0) \\ &\xrightarrow{d} \mathbf{G}'\mathbf{S}^{-1} \times N(\mathbf{0}, \mathbf{S}) \\ &\equiv N(\mathbf{0}, \mathbf{G}'\mathbf{S}^{-1}\mathbf{G}) \end{aligned}$$

Notice that

$$\text{avar}\left(\sqrt{n}\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right) = \mathbf{G}'\mathbf{S}^{-1}\mathbf{G}$$

Then, the GMM score statistic is defined as

$$\begin{aligned} LM_{\mathsf{GMM}} &= \left(\sqrt{n}\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)' \left[\mathbf{G}_n'(\boldsymbol{\theta}_0)\hat{\mathbf{S}}^{-1}\mathbf{G}_n(\boldsymbol{\theta}_0)\right]^{-1} \\ &\quad \times \left(\sqrt{n}\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right) \\ &= n\mathbf{g}_n(\boldsymbol{\theta}_0)'\hat{\mathbf{S}}^{-1}\mathbf{G}_n(\boldsymbol{\theta}_0)\left[\mathbf{G}_n(\boldsymbol{\theta}_0)'\hat{\mathbf{S}}^{-1}\mathbf{G}_n(\boldsymbol{\theta}_0)\right] \\ &\quad \times \mathbf{G}_n(\boldsymbol{\theta}_0)'\hat{\mathbf{S}}^{-1}\mathbf{g}_n(\boldsymbol{\theta}_0) \end{aligned}$$

Then by the asymptotic normality of $\sqrt{n}\mathbf{g}_n(\boldsymbol{\theta}_0)$ and Slutsky's theorem

$$\begin{aligned} LM_{\mathsf{GMM}} &\xrightarrow{d} N(\mathbf{0}, \mathbf{G}'\mathbf{S}^{-1}\mathbf{G})' \left[\mathbf{G}'\mathbf{S}^{-1}\mathbf{G}\right]^{-1} \\ &\times N(\mathbf{0}, \mathbf{G}'\mathbf{S}^{-1}\mathbf{G}) \\ \equiv\ & \chi^2(P) \end{aligned}$$