Econ 582
Nonlinear Regression

Eric Zivot

June 3, 2013

Nonlinear Regression

In linear regression models

$$\begin{aligned} y_i &= \underset{(1\times K)}{\mathbf{x}_i'}\underset{(K\times 1)}{\boldsymbol{\beta}} + \varepsilon_i,\ E[\varepsilon_i|\mathbf{x}_i] = 0 \\ E[y_i|\mathbf{x}_i &= \mathbf{x}] = \mathbf{x}'\boldsymbol{\beta},\ \varepsilon_i = y_i - E[y_i|\mathbf{x}_i = \mathbf{x}] \\ \frac{\partial E[y_i|\mathbf{x}_i = \mathbf{x}]}{\partial \mathbf{x}} &= \boldsymbol{\beta} \end{aligned}$$

it is assumed that the regression function $m(\mathbf{x}) = E[y_i|\mathbf{x}_i = \mathbf{x}] = \mathbf{x}'\boldsymbol{\beta}$ is a linear (in $\mathbf{x}$) function of the $K \times 1$ vector $\boldsymbol{\beta}$.

In parametric nonlinear regression, the regression function $m(\mathbf{x}, \boldsymbol{\theta})$ is a nonlinear function of $p \geq K$ parameters $\boldsymbol{\theta}$

$$\begin{aligned} y_i &= m(\mathbf{x}, \boldsymbol{\theta}) + \varepsilon_i \\ E[y_i|\mathbf{x}_i &= \mathbf{x}] = m(\mathbf{x}, \boldsymbol{\theta}),\ \varepsilon_i = y_i - E[y_i|\mathbf{x}_i = \mathbf{x}] \end{aligned}$$

Examples of nonlinear regression functions

$$
\begin{aligned}
m(x,\boldsymbol{\theta}) &= \theta_1 + \theta_2 \frac{x}{1+\theta_3 x},\ K=1,\ p=3 \\
m(x,\boldsymbol{\theta}) &= \theta_1 + \theta_2 x^{\theta_3},\ K=1, p=3 \\
m(x,\boldsymbol{\theta}) &= \theta_1 + \theta_2 \exp(\theta_3 x),\ K=1, p=3 \\
m(x,\boldsymbol{\theta}) &= \theta_1 + \theta_2 x + \theta_3 (x-\theta_4) 1\,(x>\theta_4) \\
m(\mathbf{x},\boldsymbol{\theta}) &= G(\mathbf{x}'\boldsymbol{\theta}),\ G \text{ known},\ K=p \\
m(\mathbf{x},\boldsymbol{\theta}) &= \boldsymbol{\theta}_1'\mathbf{x}_1 + (\boldsymbol{\theta}_2'\mathbf{x}_1)\Phi\left(\frac{x_2-\theta_3}{\theta_4}\right) \\
m(\mathbf{x},\boldsymbol{\theta}) &= \left(\boldsymbol{\theta}_1'\mathbf{x}_1\right) 1\,(x_2<\theta_3) + (\boldsymbol{\theta}_2'\mathbf{x}_1) 1\,(x_2>\theta_3)
\end{aligned}
$$

Remarks

- Typically $m(\mathbf{x}, \boldsymbol{\theta})$ is a continuous and differentiable function of $\boldsymbol{\theta}$

- In the switching examples with the indicator function, $m(\mathbf{x}, \boldsymbol{\theta})$ is not differentiable in $\boldsymbol{\theta}$

- The form of $m(\mathbf{x}, \boldsymbol{\theta})$ is sometimes motivated by economic theory (e.g. cost function estimation)

- Sometimes the form of $m(\mathbf{x}, \boldsymbol{\theta})$ is adopted as a flexible approximation to an unknown regression function (e.g. $m(\mathbf{x}, \boldsymbol{\theta}) =$ polynomial in $\mathbf{x}$)

Nonlinear Least Squares Estimation

$$y_i = m(\underset{(K\times 1)}{\mathbf{x}_i}, \underset{(p\times 1)}{\boldsymbol{\theta}}) + \varepsilon_i$$

Example (Cobb-Douglas production function)

$$\begin{aligned} y_i &= \theta_1 x_{1i}^{\theta_2} x_{2i}^{\theta_3} + \varepsilon_i \\ K &= 2,\ p = 3 \end{aligned}$$

The nonlinear least squares estimator solves

$$\min_{\theta} S(\boldsymbol{\theta}) = \sum_{i=1}^{n} (y_i - m(\mathbf{x}_i, \boldsymbol{\theta}))^2 = \sum_{i=1}^{n} \varepsilon_i^2$$

Assume $m(\mathbf{x}, \boldsymbol{\theta})$ is a continuous and differentiable function of $\boldsymbol{\theta}$. The FOCs for a minimum are

$$\begin{aligned} \frac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= -2 \sum_{i=1}^{n} (y_i - m(\mathbf{x}_i, \hat{\boldsymbol{\theta}})) \frac{\partial m(\mathbf{x}_i, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} \\ &= -2 \sum_{i=1}^{n} (y_i - m(\mathbf{x}_i, \hat{\boldsymbol{\theta}})) \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{x}_i, \hat{\boldsymbol{\theta}}) = \mathbf{0} \end{aligned}$$

where

$$\underset{p \times 1}{\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta})} = \frac{\partial m(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \frac{\partial m(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \theta_1} \\ \vdots \\ \frac{\partial m(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \theta_p} \end{pmatrix}$$

Note

$$-2 \sum_{i=1}^{n} (y_i - m(\mathbf{x}_i, \hat{\boldsymbol{\theta}})) \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{x}_i, \hat{\boldsymbol{\theta}}) = -2 \sum_{i=1}^{n} \hat{\varepsilon}_i \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{x}_i, \hat{\boldsymbol{\theta}})$$

Matrix Notation

$$\underset{n\times 1}{\mathbf{y}} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \ \underset{n\times 1}{\mathbf{m}(\mathbf{X},\boldsymbol{\theta})} = \begin{pmatrix} m(\mathbf{x}_1,\boldsymbol{\theta}) \\ \vdots \\ m(\mathbf{x}_n,\boldsymbol{\theta}) \end{pmatrix},$$

$$\underset{n\times K}{\mathbf{X}} = \begin{pmatrix} \mathbf{x}_1' \\ \\ \mathbf{x}_n' \end{pmatrix}, \ \underset{n\times 1}{\boldsymbol{\varepsilon}} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$S(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{m}(\mathbf{X},\boldsymbol{\theta}))'(\mathbf{y} - \mathbf{m}(\mathbf{X},\boldsymbol{\theta}))$$

$$\underset{n\times p}{\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\boldsymbol{\theta})} = \frac{\partial \mathbf{m}(\mathbf{X},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \begin{pmatrix} \frac{\partial m(\mathbf{x}_1,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1} & \cdots & \frac{\partial m(\mathbf{x}_1,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial m(\mathbf{x}_n,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1} & \cdots & \frac{\partial m(\mathbf{x}_n,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_p} \end{pmatrix}$$

FOCs

$$
\begin{aligned}
\frac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= -2\left(\frac{\partial \mathbf{m}(\mathbf{X},\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}\right)'(\mathbf{y}-\mathbf{m}(\mathbf{X},\hat{\boldsymbol{\theta}})) \\
&= -2\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\hat{\boldsymbol{\theta}})'(\mathbf{y}-\mathbf{m}(\mathbf{X},\hat{\boldsymbol{\theta}})) \\
&= -2\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\hat{\boldsymbol{\theta}})'\hat{\boldsymbol{\varepsilon}} = \mathbf{0}
\end{aligned}
$$

Note: In general we have p nonlinear equations in p unknown and there is no analytical solution. Hence, $\hat{\boldsymbol{\theta}}$ must be found numerically using an iterative algorithm. The most commonly used algorithm is Gauss-Newton iteration.

Gauss-Newton (GM) Algorithm

The GN algorithm can be motivated as follows. Consider a 1st order Taylor series approximation to $\mathbf{m}(\mathbf{X},\boldsymbol{\theta})$ at $\boldsymbol{\theta}=\boldsymbol{\theta}_1$ (starting value)

$$\underset{(n\times 1)}{\mathbf{m}(\mathbf{X},\boldsymbol{\theta})} = \underset{(n\times 1)}{\mathbf{m}(\mathbf{X},\boldsymbol{\theta}_1)} + \underset{(n\times p)}{\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\boldsymbol{\theta}_1)}\underset{(p\times 1)}{(\boldsymbol{\theta}-\boldsymbol{\theta}_1)} + \text{ error}$$

Approximate the nonlinear regression using the TS approximation

$$\begin{aligned}\mathbf{y} &= \mathbf{m}(\mathbf{X},\boldsymbol{\theta}) + \boldsymbol{\varepsilon} \\ &\approx \mathbf{m}(\mathbf{X},\boldsymbol{\theta}_1) + \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\boldsymbol{\theta}_1)(\boldsymbol{\theta}-\boldsymbol{\theta}_1) + \boldsymbol{\varepsilon}\end{aligned}$$

Using

$$\begin{aligned} & \mathbf{m}(\mathbf{X}, \boldsymbol{\theta}_1) + \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_1)(\boldsymbol{\theta} - \boldsymbol{\theta}_1) \\ = \; & [\mathbf{m}(\mathbf{X}, \boldsymbol{\theta}_1) - \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_1)\boldsymbol{\theta}_1] + \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_1)\boldsymbol{\theta} \end{aligned}$$

rewrite the model as

$$\mathbf{y} - \mathbf{m}(\mathbf{X}, \boldsymbol{\theta}_1) + \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_1)\boldsymbol{\theta}_1 = \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_1)\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

or

$$\begin{aligned} \bar{\mathbf{y}}(\boldsymbol{\theta}_1) &= \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_1)\boldsymbol{\theta} + \boldsymbol{\varepsilon} \\ \bar{\mathbf{y}}(\boldsymbol{\theta}_1) &= \mathbf{y} - \mathbf{m}(\mathbf{X}, \boldsymbol{\theta}_1) + \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_1)\boldsymbol{\theta}_1 \end{aligned}$$

This approximate model is linear in $\boldsymbol{\theta}$

Estimate the approximate linear model by least squares

$$\begin{aligned}
\min_{\boldsymbol{\theta}} \bar{S}_1(\boldsymbol{\theta}) &= (\bar{\mathbf{y}}(\boldsymbol{\theta}_1) - \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_1)\boldsymbol{\theta})' (\bar{\mathbf{y}}(\boldsymbol{\theta}_1) - \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_1)\boldsymbol{\theta}) \\
&\Rightarrow \boldsymbol{\theta}_2 = \left(\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_1)' \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_1)\right)^{-1} \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_1)' \bar{\mathbf{y}}(\boldsymbol{\theta}_1)
\end{aligned}$$

Then repeat estimation of approximate linear model using updated estimate $\boldsymbol{\theta}_2$

$$\begin{aligned}
\min_{\boldsymbol{\theta}} \bar{S}_2(\boldsymbol{\theta}) &= (\bar{\mathbf{y}}(\boldsymbol{\theta}_2) - \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_2)\boldsymbol{\theta})' (\bar{\mathbf{y}}(\boldsymbol{\theta}_2) - \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_2)\boldsymbol{\theta}) \\
&\Rightarrow \boldsymbol{\theta}_3 = \left(\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_2)' \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_2)\right)^{-1} \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_2)' \bar{\mathbf{y}}(\boldsymbol{\theta}_2)
\end{aligned}$$

At iteration n we have

$$\boldsymbol{\theta}_{n+1} = \left(\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\boldsymbol{\theta}_n)'\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\boldsymbol{\theta}_n)\right)^{-1}\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\boldsymbol{\theta}_n)'\bar{\mathbf{y}}(\boldsymbol{\theta}_n)$$

Substituting in $\bar{\mathbf{y}}(\boldsymbol{\theta}_n) = \mathbf{y} - \mathbf{m}(\mathbf{X},\boldsymbol{\theta}_n) + \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\boldsymbol{\theta}_n)\boldsymbol{\theta}_n$ we have

$$\begin{aligned}
\boldsymbol{\theta}_{n+1} &= \left(\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\boldsymbol{\theta}_n)'\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\boldsymbol{\theta}_n)\right)^{-1}\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\boldsymbol{\theta}_n)' \\
&\quad \times \left[\mathbf{y} - \mathbf{m}(\mathbf{X},\boldsymbol{\theta}_n) + \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\boldsymbol{\theta}_n)\boldsymbol{\theta}_n\right] \\
&= \left(\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\boldsymbol{\theta}_n)'\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\boldsymbol{\theta}_n)\right)^{-1}\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\boldsymbol{\theta}_n)'\left[\mathbf{y} - \mathbf{m}(\mathbf{X},\boldsymbol{\theta}_n)\right] \\
&\quad + \left(\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\boldsymbol{\theta}_n)'\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\boldsymbol{\theta}_n)\right)^{-1}\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\boldsymbol{\theta}_n)'\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\boldsymbol{\theta}_n)\boldsymbol{\theta}_n \\
&= \boldsymbol{\theta}_n + \left(\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\boldsymbol{\theta}_n)'\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\boldsymbol{\theta}_n)\right)^{-1}\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X},\boldsymbol{\theta}_n)'\left[\mathbf{y} - \mathbf{m}(\mathbf{X},\boldsymbol{\theta}_n)\right]
\end{aligned}$$

Note: Using $\frac{\partial S(\boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}} = -2\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_n)' \left[\mathbf{y} - \mathbf{m}(\mathbf{X}, \boldsymbol{\theta}_n)\right]$ we have

$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n - \frac{1}{2}\left(\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_n)'\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_n)\right)^{-1} \frac{\partial S(\boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}}$$

Provided that

$$\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_n)'\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_n) \text{ is pd}$$
$$\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_n) \text{ is full rank } p$$

Then the FOCs are satisfied if $\boldsymbol{\theta}_{n+1} \approx \boldsymbol{\theta}_n$. That is,

$$\frac{\partial S(\boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}} = \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_n)' \left[\mathbf{y} - \mathbf{m}(\mathbf{X}, \boldsymbol{\theta}_n)\right] \approx \mathbf{0}$$

Common Convergence Criteria

- Stop when

$$
\begin{aligned}
||\boldsymbol{\theta}_{n+1} - \boldsymbol{\theta}_n|| &< \varepsilon \approx 10^{-6} \\
||\mathbf{x}|| &= \left(x_1^2 + \cdots + x_p^2\right)^{1/2}
\end{aligned}
$$

  To avoid issues with the units of $\boldsymbol{\theta}$, it is better to stop when

$$
\frac{||\boldsymbol{\theta}_{n+1} - \boldsymbol{\theta}_n||}{||\boldsymbol{\theta}_n + \tau||} < \eta \approx 10^{-5}
$$

- Stop when

$$
\left\| \frac{\partial S(\boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}} \right\| < \varepsilon
$$

- Stop when

$$\frac{||S(\boldsymbol{\theta}_{n+1}) - S(\boldsymbol{\theta}_n)||}{||S(\boldsymbol{\theta}_n) + \tau||} < \eta$$

Remarks

- The solution to the FOCs can be a local minimum, local maximum or a global minimum.

- The GN iteration scheme always leads in the direction of a minimum instead of a maximum provided $\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_n)'\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_n)$ is pd

$$\begin{aligned} \boldsymbol{\theta}_{n+1} &= \boldsymbol{\theta}_n + \left(\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_n)'\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_n)\right)^{-1} \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_n)' \left[\mathbf{y} - \mathbf{m}(\mathbf{X}, \boldsymbol{\theta}_n)\right] \\ &= \boldsymbol{\theta}_n - \frac{1}{2}\left(\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_n)'\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_n)\right)^{-1} \frac{\partial S(\boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}} \end{aligned}$$

If $\frac{\partial S(\boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}} > 0$ then $\boldsymbol{\theta}_{n+1} < \boldsymbol{\theta}_n$; if $\frac{\partial S(\boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}} < 0$ then $\boldsymbol{\theta}_{n+1} > \boldsymbol{\theta}_n$

- The GN iteration scheme can overshoot the global minimum. To guard against this a step-length correction, λ_n, is often added to the algorithm

$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n - \lambda_n \left(\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_n)' \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{X}, \boldsymbol{\theta}_n) \right)^{-1} \frac{\partial S(\boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}}$$

  where λ_n is chosen such that $S(\boldsymbol{\theta}_{n+1}) > S(\boldsymbol{\theta}_n)$

- To guard against getting stuck at a local minimum it is often suggested that different starting values be used.

Asymptotic Distribution of NLS estimator (Homoskedastic Case)

$$y_i = m(\mathbf{x}_i, \boldsymbol{\theta}) + \varepsilon_i, \; E[\varepsilon_i^2] = \sigma^2$$

Consider the linear approximation evaluated at the true value of $\boldsymbol{\theta}$

$$\bar{y}_i(\boldsymbol{\theta}) = \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta})'\boldsymbol{\theta} + \varepsilon_i$$

Assuming

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta})\varepsilon_i \xrightarrow{p} 0$$

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta})\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta})' \xrightarrow{p} E\left[\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta})\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta})'\right] = \mathbf{M}_{\boldsymbol{\theta\theta}}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbf{m}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta})\varepsilon_i \xrightarrow{d} N(0, \sigma^2\mathbf{M}_{\boldsymbol{\theta\theta}})$$

Then

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \sigma^2\mathbf{M}_{\boldsymbol{\theta\theta}}^{-1})$$

Equivalently,

$$\hat{\boldsymbol{\theta}} \sim N\left(\theta, \frac{1}{n}\sigma^2 \mathbf{M}_{\boldsymbol{\theta\theta}}^{-1}\right)$$

The asymptotic variance $\sigma^2 \mathbf{M}_{\boldsymbol{\theta\theta}}^{-1}$ can be consistently estimated using

$$\hat{\sigma}^2 \widehat{\mathbf{M}}_{\boldsymbol{\theta\theta}}^{-1}$$

with

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n-p}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 \\ \widehat{\mathbf{M}}_{\boldsymbol{\theta\theta}} &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{x}_i, \hat{\boldsymbol{\theta}})\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{x}_i, \hat{\boldsymbol{\theta}})' \end{aligned}$$

Asymptotic Distribution of NLS estimator (Heteroskedastic Case)

$$y_i = m(\mathbf{x}_i, \boldsymbol{\theta}) + \varepsilon_i,\ \ E[\varepsilon_i^2|\mathbf{x}_i] = \sigma^2(\mathbf{x}_i)$$

Then

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{M}_{\boldsymbol{\theta\theta}}^{-1}\mathbf{V}\mathbf{M}_{\boldsymbol{\theta\theta}}^{-1})$$

$$\mathbf{V} = E\left[\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta})\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{x}_i, \boldsymbol{\theta})'\varepsilon_i^2\right]$$

The matrix $\mathbf{V}$ can be consistently estimated using the White-type HC estimator

$$\hat{\mathbf{V}}_{HC} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{x}_i, \hat{\boldsymbol{\theta}})\mathbf{m}_{\boldsymbol{\theta}}(\mathbf{x}_i, \hat{\boldsymbol{\theta}})'\hat{\varepsilon}_i^2$$