Classical Non-Linear Methods: Introduction to M-Estimation

Bertel Schjerning

University of Copenhagen, Department of Economics

January 29, 2026

Plan for Classical Non-Linear Methods

Outline

Introduction

Non-Linear Estimation Chapters

W. Chapters 12–13: Abstract and technical.

- ▶ I'll try to limit technicalities.

But generality can be useful! Unified framework.

- ▶ **Ex:** OLS, Non-linear LS, MLE, Least absolute deviations...

There will be <u>no</u> exam questions in Ch. 12–13 *specifically*.

But important—and required—background knowledge.

Big Picture: Steps in Econometric Analysis

1. **Identification:** Given distribution of observables, can we recover parameters? If so, how?

2. **Estimation:** Given finite sample of observations, how to construct parameter estimates?

3. **Inference:** Confidence intervals, prediction intervals, hypothesis testing, etc.

Steps in Econometrics Analysis

- ▶ Identification: As if sample infinitely large.

- ▶ Estimation: Finite sample. Which formula(e)/algorithm?

- ▶ Inference: Builds on (asymptotic) distribution theory.

Steps highly interdependent.

- ▶ Identification method may suggest estimator.

- ▶ Inference method hinges on estimation method.

Non-Linear Regression

Non-Linear Regression Model

y: scalar outcome variable.

$\mathbf{x}$: K-vector of explanatory variables.

Parametric model for $\mathrm{E}\left[y \mid \mathbf{x}\right]$:

$$\mathcal{M} := \{\mathbf{x} \mapsto m\left(\mathbf{x}, \boldsymbol{\theta}\right); \boldsymbol{\theta} \in \Theta\}$$

- ▶ $m\left(\mathbf{x}, \boldsymbol{\theta}\right)$: Candidate for $\mathrm{E}\left[y \mid \mathbf{x}\right]$.

- ▶ $\Theta \subseteq \mathbb{R}^P$ parameter space.   (P fixed!)

Mean model correctly specified if fctn $\mathrm{E}\left[y \mid \mathbf{x}\right] \in \mathcal{M}$, i.e.

for some $\boldsymbol{\theta}_o \in \Theta$, $\mathrm{E}\left[y \mid \mathbf{x}\right] = m\left(\mathbf{x}, \boldsymbol{\theta}_o\right)$ for all $\mathbf{x}$.

$\boldsymbol{\theta}_o$ often called "true value of theta."

Examples of Functional Form

Ex 1. If y unrestricted, may take

$$m(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{x}\boldsymbol{\theta}.$$ (linear (mean) regression)

Ex 2. If $y \geqslant 0$, may take

$$m(\mathbf{x}, \boldsymbol{\theta}) = \exp(\mathbf{x}\boldsymbol{\theta}).$$ (exponential regression)

Ex 3. If $y \in [0, 1]$, may take

$$m(\mathbf{x}, \boldsymbol{\theta}) = \frac{1}{1 + \exp(-\mathbf{x}\boldsymbol{\theta})}.$$ (logistic regression)

Here: $K = P$. But $K \gtrless P$ allowed.

Error Formulation

▶ Assume correct specification (NLS.1).

▶ Defining $u := y - m(\mathbf{x}, \boldsymbol{\theta}_o)$, may write

$$y = m(\mathbf{x}, \boldsymbol{\theta}_o) + u, \quad \mathrm{E}[u \mid \mathbf{x}] = 0.$$

▶ $\mathrm{E}[u \mid \mathbf{x}] = 0$ a consequence of correct specification.

▶ Error formulation here only useful for abbreviations.

Discussion

$E[u \mid \mathbf{x}] = 0$ does *not* imply u and $\mathbf{x}$ independent.

... only cond'l *mean* independence.

May have $\text{var}(u \mid \mathbf{x})$ nonconstant (in $\mathbf{x}$).

▶ If $y \geqslant 0$, must have $u \geqslant -m(\mathbf{x}, \boldsymbol{\theta}_o) \ldots$

Dealing with *semi*parametric model for $D(y|\mathbf{x})$.

▶ Parametric model for mean $E[y \mid \mathbf{x}]$.

▶ Other characteristics (var, skewness, etc.) unrestricted.

Identification

Identification

We'll show: $\boldsymbol{\theta_o}$ solves population problem (PP)

$$\min_{\boldsymbol{\theta} \in \Theta} \mathrm{E} \left[\left| y - m\left(\mathbf{x}, \boldsymbol{\theta}\right) \right|^2 \right]. \qquad \text{(PP)}$$

▶ Function $m(\cdot, \cdot)$ + parameter space Θ known quantities.

▶ Hence, **IF** given $D\left(y, \mathbf{x}\right)$, PP problem known.

▶ $\boldsymbol{\theta_o}$ identified if PP solution *unique*.

Identification

$\pm m\left(\mathbf{x}, \boldsymbol{\theta}_o\right)$ and expanding square,

$$
\begin{aligned}
\left|y - m\left(\mathbf{x}, \boldsymbol{\theta}\right)\right|^2 &= \left|\left[y - m\left(\mathbf{x}, \boldsymbol{\theta}_o\right)\right] - \left[m\left(\mathbf{x}, \boldsymbol{\theta}\right) - m\left(\mathbf{x}, \boldsymbol{\theta}_o\right)\right]\right|^2 \\
&= \left|y - m\left(\mathbf{x}, \boldsymbol{\theta}_o\right)\right|^2 + \left|m\left(\mathbf{x}, \boldsymbol{\theta}\right) - m\left(\mathbf{x}, \boldsymbol{\theta}_o\right)\right|^2 \\
&\quad - 2u\left[m\left(\mathbf{x}, \boldsymbol{\theta}\right) - m\left(\mathbf{x}, \boldsymbol{\theta}_o\right)\right].
\end{aligned}
$$

Taking expectations,

$$
\begin{aligned}
\mathrm{E}\left[\left|y - m\left(\mathbf{x}, \boldsymbol{\theta}\right)\right|^2\right] &= \mathrm{E}\left[\left|y - m\left(\mathbf{x}, \boldsymbol{\theta}_o\right)\right|^2\right] \\
&\quad + \mathrm{E}\left[\left|m\left(\mathbf{x}, \boldsymbol{\theta}\right) - m\left(\mathbf{x}, \boldsymbol{\theta}_o\right)\right|^2\right].
\end{aligned}
$$

Identification

Have shown

$$\mathrm{E}\left[|y - m(\mathbf{x}, \boldsymbol{\theta})|^2\right] = \mathrm{E}\left[|y - m(\mathbf{x}, \boldsymbol{\theta}_o)|^2\right]$$
$$+ \mathrm{E}\left[|m(\mathbf{x}, \boldsymbol{\theta}) - m(\mathbf{x}, \boldsymbol{\theta}_o)|^2\right].$$

It follows that

$$\mathrm{E}\left[|y - m(\mathbf{x}, \boldsymbol{\theta})|^2\right] \geqslant \mathrm{E}\left[|y - m(\mathbf{x}, \boldsymbol{\theta}_o)|^2\right] \text{ for all } \theta \in \Theta.$$

$\Rightarrow \boldsymbol{\theta}_o$ solves PP.

Q: When unique?

Identification Condition

Have shown

$$\mathrm{E}\left[\left|y - m\left(\mathbf{x}, \boldsymbol{\theta}\right)\right|^2\right] = \mathrm{E}\left[\left|y - m\left(\mathbf{x}, \boldsymbol{\theta}_o\right)\right|^2\right]$$
$$+ \mathrm{E}\left[\left|m\left(\mathbf{x}, \boldsymbol{\theta}\right) - m\left(\mathbf{x}, \boldsymbol{\theta}_o\right)\right|^2\right].$$

$\boldsymbol{\theta}_o$ uniquely solves PP if and only if

$$\mathrm{E}\left[\left|m\left(\mathbf{x}, \boldsymbol{\theta}\right) - m\left(\mathbf{x}, \boldsymbol{\theta}_o\right)\right|^2\right] > 0 \text{ for all } \boldsymbol{\theta} \in \Theta \backslash \{\boldsymbol{\theta}_o\}.$$

Q: When will identification fail?

Identification Failure: Linear Case

Ex: *Linear* regression, $m(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{x}\boldsymbol{\theta}$ with $\Theta = \mathbb{R}^K$.

Here

$$\begin{aligned}
\mathrm{E}\left[\left|m(\mathbf{x}, \boldsymbol{\theta}) - m(\mathbf{x}, \boldsymbol{\theta}_o)\right|^2\right] &= \mathrm{E}\left[\left|\mathbf{x}(\boldsymbol{\theta} - \boldsymbol{\theta}_o)\right|^2\right] \\
&= (\boldsymbol{\theta} - \boldsymbol{\theta}_o)' \, \mathrm{E}\left[\mathbf{x}'\mathbf{x}\right] (\boldsymbol{\theta} - \boldsymbol{\theta}_o).
\end{aligned}$$

▶ > 0 if $\mathrm{E}\left[\mathbf{x}'\mathbf{x}\right]$ positive definite.

▶ Just usual (population) rank condition (OLS.2).

If not full rank...

Identification Failure: Non-linear example

Ex: *Non*linear regression with $\Theta = \mathbb{R}^4$ and

$$m\left(\mathbf{x}, \boldsymbol{\theta}\right) = \theta_1 + \theta_2 x_2 + \theta_3 x_3^{\theta_4}.$$

▶ Suppose $\theta_{o,3} = 0.$ (Truth linear.)

▶ At $\boldsymbol{\theta}$ with $\theta_3 = 0 \left(= \theta_{o,3}\right)$...

▶ ... criterion function *independent of* θ_4.

▶ For this $\boldsymbol{\theta}_o$, identification fails.

▶ Example of poorly identified model.

Estimation

Estimation

$\boldsymbol{\theta_o}$ solves PP,

$$\boldsymbol{\theta_o} \in \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmin}} \, \mathrm{E}\left[\left|y - m\left(\mathbf{x}, \boldsymbol{\theta}\right)\right|^2\right].$$

Analogy principle suggests,

$$\widehat{\boldsymbol{\theta}} \in \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmin}} \, \frac{1}{N} \sum_{i=1}^{N} \left[y_i - m\left(\mathbf{x}_i, \boldsymbol{\theta}\right)\right]^2.$$

Non-linear least squares (NLS) estimator.

For now, assume existence (but not uniqueness) of solution.

Consistency?

Q: Does NLS consistently estimate $\boldsymbol{\theta}_o$?

It turns out answer is "yes," provided (roughly)

1. $\boldsymbol{\theta}_o$ is identified,

2. Criterion function convergence

$$\frac{1}{N}\sum_{i=1}^{N}\left[y_i - m\left(\mathbf{x}_i, \boldsymbol{\theta}\right)\right]^2 \overset{``\rightarrow"}{} \mathrm{E}\left[\left|y - m\left(\mathbf{x}, \boldsymbol{\theta}\right)\right|^2\right]$$

   is suitable strong (functional!) sense.

Next: More detail in general setting.