Classical Non-Linear Methods: Introduction to M-Estimation

Bertel Schjerning

University of Copenhagen, Department of Economics

February 9, 2026

Plan for Classical Non-Linear Methods

Lecture 4: M-estimation, Intro, Non-linear LS	(W.12)
Lecture 5: Asymptotic properties of M-estimators	(W.12)
Lecture 6: M-estimator inference, Variance estimation	(W.12)
Lecture 7: Maximum likelihood estimation	(W.13)

Outline

Introduction

Non-Linear Regression

- Identification
- Estimation

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 no 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

$$E[y \mid \mathbf{x}] = m(\mathbf{x}, \boldsymbol{\theta}_o) \quad (1)$$

y : random scalar outcome variable

$\mathbf{x}$: random k-vector of explanatory variables

m : Nonlinear *parametric* model for $E(y \mid \mathbf{x})$, i.e. $m(\cdot, \cdot)$ is known up to a set of parameters $\boldsymbol{\theta}_o$

$\boldsymbol{\theta}_o$: $P \times 1$ vector of that index the model for $E[y \mid \mathbf{x}]$, $\boldsymbol{\theta}_o \in \Theta \subset R^P$

- $\Theta \subseteq \mathbb{R}^P$ parameter space. (P fixed!)
- $\boldsymbol{\theta}_o$ often called "true value of theta."
- If $E[y|\mathbf{x}] = m(\mathbf{x}, \boldsymbol{\theta}_o)$ holds for some $\theta_0 \in \Theta$, we say that the model is *correctly specified* [NLS.1].

Examples of Functional Form

Ex 1. If y unrestricted, may take

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

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

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

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

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

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

Model in Error Formulation

Equivalent way of writing $E[y \mid \mathbf{x}] = m(\mathbf{x}, \boldsymbol{\theta}_o)$ is

$$y = m(\mathbf{x}, \boldsymbol{\theta}_o) + u, \quad \text{with } E[u \mid \mathbf{x}] = 0 \tag{2}$$

$E[u \mid \mathbf{x}] = 0$ is not an additional assumption, it is simply a consequence of equation (1).

To see this

1. Take conditional expectations of (2)

$$\begin{aligned} E[y \mid \mathbf{x}] &= E[m(\mathbf{x}, \boldsymbol{\theta}_o) \mid \mathbf{x}] + E[u \mid \mathbf{x}] \\ &= m(\mathbf{x}, \boldsymbol{\theta}_o) \end{aligned}$$

2. Take equation (1) and define $u \equiv y - m(x, \boldsymbol{\theta}_o)$. Take conditional expectations

$$\begin{aligned} E[u \mid \mathbf{x}] &= E[y \mid \mathbf{x}] - E[m(\mathbf{x}, \theta_0) \mid \mathbf{x}] \\ &= E[y \mid \mathbf{x}] - m(\mathbf{x}, \boldsymbol{\theta}_o) = 0 \end{aligned}$$

Discussion

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

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

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

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

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

- Parametric model for mean $\mathrm{E}\left[y|\ \mathbf{x}\right]$.
- Other characteristics (var, skewness, etc.) unrestricted.

Identification

Identification

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

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

- 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

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

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

Taking expectations,

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

Identification

Have shown

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

It follows that

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

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

Q: When unique?

Identification Condition

Have shown

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

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

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

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

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

- > 0 if $\mathrm{E}[\mathbf{x}'\mathbf{x}]$ 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 $\theta_3 = 0\,(= \theta_{o,3})$...
- ... 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,

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

Analogy principle suggests,

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

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

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

is suitable strong (functional!) sense.

Next: More detail in general setting.