Classical Non-Linear Methods: Inference with M-Estimators

Bertel Schjerning

University of Copenhagen, Department of Economics

February 10, 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)

Recap: Setting

$$\boldsymbol{\theta}_o \in \operatorname*{argmin}_{\boldsymbol{\theta} \in \Theta} \mathrm{E}\left[q\left(\mathbf{w}, \boldsymbol{\theta}\right)\right], \qquad \text{(M-estimand)}$$

$$\widehat{\boldsymbol{\theta}} \in \operatorname*{argmin}_{\boldsymbol{\theta} \in \Theta} \frac{1}{N} \sum_{i=1}^{N} q\left(\mathbf{w}_i, \boldsymbol{\theta}\right). \qquad \text{(M-estimator)}$$

Know sufficient conditions for $\widehat{\boldsymbol{\theta}}_N$ $\sqrt{N}$-asymptotically normal,

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_o) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \mathbf{V}_o\right), \quad \mathbf{V}_o := \mathbf{A}_o^{-1}\mathbf{B}_o\mathbf{A}_o^{-1}.$$

To do inference (CIs, hypothesis testing, etc.)...

Q: How to estimate $\mathrm{Avar}(\widehat{\boldsymbol{\theta}}_N) = \mathbf{V}_o/N$?

Outline

Variance Estimation

- Example: Non-Linear Regression
- Multivariate Nonlinear Least Squares

Non-Linear Hypothesis Testing

Variance Estimation

How to Estimate Asymptotic Variance?

$$\mathrm{Avar}(\widehat{\boldsymbol{\theta}}) = \mathbf{A}_o^{-1}\mathbf{B}_o\mathbf{A}_o^{-1}/N.$$

Recall:

$$\begin{aligned}
\mathbf{A}_o &= \mathrm{E}\left[\mathbf{H}\left(\mathbf{w}, \boldsymbol{\theta}_o\right)\right], \\
\mathbf{B}_o &= \mathrm{E}\left[\mathbf{s}\left(\mathbf{w}, \boldsymbol{\theta}_o\right)\mathbf{s}\left(\mathbf{w}, \boldsymbol{\theta}_o\right)'\right], \\
\mathbf{s}\left(\mathbf{w}, \boldsymbol{\theta}\right) &= \frac{\partial}{\partial \boldsymbol{\theta}} q\left(\mathbf{w}, \boldsymbol{\theta}\right), & (P \times 1) \\
\mathbf{H}\left(\mathbf{w}, \boldsymbol{\theta}\right) &= \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} q\left(\mathbf{w}, \boldsymbol{\theta}\right). & (P \times P)
\end{aligned}$$

Q: How to consistently estimate $\mathbf{A}_o$ and $\mathbf{B}_o$?

Estimator 1: Most Structural

One (naï¿œeve) **idea**:

1. Analytically solve for expectations

$$\mathbf{A}(\boldsymbol{\theta}_o) := \mathrm{E}\left[\mathbf{H}(\mathbf{w}, \boldsymbol{\theta}_o)\right],$$
$$\mathbf{B}(\boldsymbol{\theta}_o) := \mathrm{E}\left[\mathbf{s}(\mathbf{w}, \boldsymbol{\theta}_o)\,\mathbf{s}(\mathbf{w}, \boldsymbol{\theta}_o)'\right].$$

2. Substitute $\boldsymbol{\theta}_o$ for $\widehat{\boldsymbol{\theta}}$.

Drawbacks:

- Requires complete specification of $\mathbf{w}$ distribution.
- Obtaining closed-form expression difficult.

Rarely an option...

Estimator 2: Least Structural

- $\mathbf{A}_o$ and $\mathbf{B}_o$ expectations of functions of $\boldsymbol{\theta}_o$.
- Invoke analogy principle:
  1. Replace expectations with averages.
  2. Insert $\widehat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}_o$.

$$\widehat{\mathbf{A}} := \frac{1}{N}\sum_{i=1}^{N} \widehat{\mathbf{H}}_i, \quad \widehat{\mathbf{H}}_i := \mathbf{H}(\mathbf{w}_i, \widehat{\boldsymbol{\theta}}),$$
$$\widehat{\mathbf{B}} := \frac{1}{N}\sum_{i=1}^{N} \widehat{\mathbf{s}}_i \widehat{\mathbf{s}}_i', \quad \widehat{\mathbf{s}}_i := \mathbf{s}(\mathbf{w}_i, \widehat{\boldsymbol{\theta}}).$$

\+ $\widehat{\mathbf{A}} \to_p \mathbf{A}_o$ and $\widehat{\mathbf{B}} \to_p \mathbf{B}_o$ under mild (add'l) cond's.

Estimator 2: Least Structural

$$\widehat{\mathbf{A}} := \frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{H}}_i, \quad \widehat{\mathbf{H}}_i := \mathbf{H}(\mathbf{w}_i, \widehat{\boldsymbol{\theta}}),$$

$$\widehat{\mathbf{B}} := \frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{s}}_i\widehat{\mathbf{s}}_i', \quad \widehat{\mathbf{s}}_i := \mathbf{s}(\mathbf{w}_i, \widehat{\boldsymbol{\theta}}).$$

+ With twice cont diff q, always available.

+ If $\widehat{\boldsymbol{\theta}}$ interior, $\widehat{\mathbf{A}}$ at least positive semi-definite.

÷ Requires calculation of second-order derivatives.

- If $q(\mathbf{w}, \cdot)$ strictly convex, $N^{-1}\sum_i \mathbf{H}(\mathbf{w}_i, \boldsymbol{\theta})$ p.d., *all* $\boldsymbol{\theta}$.

Estimator 3: Semistructural

Let $\mathbf{w} = (\mathbf{y}, \mathbf{x})$.

Let $\boldsymbol{\theta}_o$ index feature of $D(\mathbf{y}|\mathbf{x})$.

- E.g. cond'l mean, median, whole distribution.

Define conditional Hessian

$$\mathbf{A}(\mathbf{x}, \boldsymbol{\theta}_o) := \mathrm{E}\left[\mathbf{H}(\mathbf{w}, \boldsymbol{\theta}_o) \middle| \mathbf{x}\right].$$

Then estimator given by

$$\widetilde{\mathbf{A}} := \frac{1}{N} \sum_{i=1}^{N} \widehat{\mathbf{A}}_i, \quad \widehat{\mathbf{A}}_i := \mathbf{A}(\mathbf{x}_i, \widehat{\boldsymbol{\theta}}).$$

Estimator 3: Semistructural

Recall:

$$\widetilde{\mathbf{A}} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{A}}_i, \quad \widehat{\mathbf{A}}_i = \mathbf{A}(\mathbf{x}_i, \widehat{\boldsymbol{\theta}}).$$

+ Usually positive definite in sample.
- Useful if $\mathrm{E}\left[\mathbf{H}\left(\mathbf{w}, \boldsymbol{\theta}_o\right) \middle| \mathbf{x}\right]$ available in closed form
  - ... or easily approximated.

÷ Relies on more structure. Could be wrong.
  - Even more important for fully structural approach.

Asymptotic Variance Estimators

Least structural approach $\Rightarrow$

$$\widehat{\text{Avar}(\widehat{\boldsymbol{\theta}})} := \widehat{\mathbf{A}}^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{A}}^{-1}/N$$
$$= \left(\sum_{i=1}^{N}\widehat{\mathbf{H}}_i\right)^{-1}\left(\sum_{i=1}^{N}\widehat{\mathbf{s}}_i\widehat{\mathbf{s}}_i'\right)\left(\sum_{i=1}^{N}\widehat{\mathbf{H}}_i\right)^{-1}.$$

Semistructural approach $\Rightarrow$

$$\widetilde{\text{Avar}(\widehat{\boldsymbol{\theta}})} := \widetilde{\mathbf{A}}^{-1}\widehat{\mathbf{B}}\widetilde{\mathbf{A}}^{-1}/N$$
$$= \left(\sum_{i=1}^{N}\widehat{\mathbf{A}}_i\right)^{-1}\left(\sum_{i=1}^{N}\widehat{\mathbf{s}}_i\widehat{\mathbf{s}}_i'\right)\left(\sum_{i=1}^{N}\widehat{\mathbf{A}}_i\right)^{-1}.$$

Robust/semi-robust variance estimators.

Example: Non-Linear Regression

Recap: Non-Linear Mean Regression Model

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

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].

Recap: Non-Linear Least Squares

Correct specification $\Rightarrow$ true theta solves population problem

$$\boldsymbol{\theta}_o \in \operatorname*{argmin}_{\boldsymbol{\theta}\in\Theta} \mathrm{E}\left[q\left(\mathbf{w},\boldsymbol{\theta}\right)\right], \tag{PP}$$

with $\mathbf{w} = (y, \mathbf{x})$ and the square loss

$$q\left(\mathbf{w},\boldsymbol{\theta}\right) = \left[y - m\left(\mathbf{x},\boldsymbol{\theta}\right)\right]^2 .$$

If given i.i.d. observations $\{(y_i, \mathbf{x}_i)\}_{i=1}^N$.

Suggests non-linear least squares (NLS) estimator

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

NLS Score and Hessian

In NLS,

$$q(\mathbf{w}, \boldsymbol{\theta}) = [y - m(\mathbf{x}, \boldsymbol{\theta})]^2 .$$

Chain and product rules $\Rightarrow$

$$\begin{aligned}
\mathbf{s}(\mathbf{w}, \boldsymbol{\theta}) &= -2[y - m(\mathbf{x}, \boldsymbol{\theta})] \frac{\partial}{\partial \boldsymbol{\theta}} m(\mathbf{x}, \boldsymbol{\theta}), \\
\mathbf{H}(\mathbf{w}, \boldsymbol{\theta}) &= 2 \frac{\partial}{\partial \boldsymbol{\theta}} m(\mathbf{x}, \boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}'} m(\mathbf{x}, \boldsymbol{\theta}) \\
&\quad - 2[y - m(\mathbf{x}, \boldsymbol{\theta})] \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} m(\mathbf{x}, \boldsymbol{\theta}).
\end{aligned}$$

Only one $D(y \mid \mathbf{x})$ feature specified. $(\mathrm{E}[y \mid \mathbf{x}])$

$\Rightarrow$ fully structural impossible.

NLS Variance Estimator: $\mathbf{A}_o^{-1}\mathbf{B}_o\mathbf{A}_o^{-1}/N$

Evaluate $\boldsymbol{\theta} = \boldsymbol{\theta}_o$,

$$\begin{aligned} \mathbf{s}(\mathbf{w}, \boldsymbol{\theta}_o) &= -2u\frac{\partial}{\partial\boldsymbol{\theta}} m(\mathbf{x}, \boldsymbol{\theta}_o), \\ \Rightarrow \mathbf{B}_o &= 4\mathrm{E}\left[u^2 \frac{\partial}{\partial\boldsymbol{\theta}} m(\mathbf{x}, \boldsymbol{\theta}_o) \frac{\partial}{\partial\boldsymbol{\theta}'} m(\mathbf{x}, \boldsymbol{\theta}_o)\right] \end{aligned}$$

Abbreviating

$$\widehat{u}_i := y_i - m(\mathbf{x}_i, \widehat{\boldsymbol{\theta}}), \qquad \text{(NLS residuals)}$$

$$\widehat{\nabla_\theta m_i} := \frac{\partial}{\partial\boldsymbol{\theta}'} m(\mathbf{x}_i, \widehat{\boldsymbol{\theta}}), \qquad (1 \times P)$$

$$\widehat{\mathbf{B}} = \frac{4}{N}\sum_{i=1}^{N} \widehat{u}_i^2 \widehat{\nabla_\theta m_i}' \widehat{\nabla_\theta m_i}.$$

NLS Variance Estimator: $\mathbf{A}_o^{-1}\mathbf{B}_o\mathbf{A}_o^{-1}/N$

Evaluate $\boldsymbol{\theta} = \boldsymbol{\theta}_o$,

$$\mathbf{H}\left(\mathbf{w}, \boldsymbol{\theta}_o\right) = 2\frac{\partial}{\partial \boldsymbol{\theta}} m\left(\mathbf{x}, \boldsymbol{\theta}_o\right) \frac{\partial}{\partial \boldsymbol{\theta}'} m\left(\mathbf{x}, \boldsymbol{\theta}_o\right) - 2u \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} m\left(\mathbf{x}, \boldsymbol{\theta}_o\right).$$

Abbreviate:

$$\widehat{\nabla_\theta^2 m_i} := \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} m(\mathbf{x}_i, \widehat{\boldsymbol{\theta}}).$$

Least structural approach $\Rightarrow$

$$\widehat{\mathbf{A}} = \frac{1}{N}\sum_{i=1}^{N} \widehat{\mathbf{H}}_i, \quad \widehat{\mathbf{H}}_i = 2\widehat{\nabla_\theta m_i}' \widehat{\nabla_\theta m_i} - 2\widehat{u}_i \widehat{\nabla_\theta^2 m_i}.$$

$\Rightarrow$ Fully robust estimator: $\widehat{\mathbf{A}}^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{A}}^{-1}/N$.

NLS Variance Estimator: $\mathbf{A}_o^{-1}\mathbf{B}_o\mathbf{A}_o^{-1}/N$

$$\mathbf{H}\left(\mathbf{w},\boldsymbol{\theta}_o\right) = 2\frac{\partial}{\partial\boldsymbol{\theta}} m\left(\mathbf{x},\boldsymbol{\theta}_o\right)\frac{\partial}{\partial\boldsymbol{\theta}'} m\left(\mathbf{x},\boldsymbol{\theta}_o\right) - 2u\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} m\left(\mathbf{x},\boldsymbol{\theta}_o\right).$$

Model well-specified $\Rightarrow \mathrm{E}\left[u \mid \mathbf{x}\right] = 0$.

$$\Rightarrow \mathbf{A}\left(\mathbf{x},\boldsymbol{\theta}_o\right) = \mathrm{E}\left[\mathbf{H}\left(\mathbf{w},\boldsymbol{\theta}_o\right)\middle|\,\mathbf{x}\right] = 2\frac{\partial}{\partial\boldsymbol{\theta}} m\left(\mathbf{x},\boldsymbol{\theta}_o\right)\frac{\partial}{\partial\boldsymbol{\theta}'} m\left(\mathbf{x},\boldsymbol{\theta}_o\right).$$

Semistructural approach $\Rightarrow$

$$\widetilde{\mathbf{A}} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{A}}_i, \quad \widehat{\mathbf{A}}_i = 2\widehat{\nabla_\theta m_i}'\widehat{\nabla_\theta m_i}.$$

$\Rightarrow$ Semirobust estimator: $\widetilde{\mathbf{A}}^{-1}\widehat{\mathbf{B}}\widetilde{\mathbf{A}}^{-1}/N$.

NLS Variance Estimator

Semirobust estimator $\widetilde{\text{Avar}(\widehat{\boldsymbol{\theta}})} =$

$$\left(\sum_{i=1}^{N} \widehat{\nabla_\theta m_i}' \widehat{\nabla_\theta m_i}\right)^{-1} \left(\sum_{i=1}^{N} \hat{u}_i^2 \widehat{\nabla_\theta m_i}' \widehat{\nabla_\theta m_i}\right) \left(\sum_{i=1}^{N} \widehat{\nabla_\theta m_i}' \widehat{\nabla_\theta m_i}\right)^{-1}.$$

No restrictions on var $(y|\mathbf{x})$.

[⇒] Heteroskedasticity-robust variance estimator for NLS.

- Output of canned software packages.

Asymptotic standard errors = square root of diagonal.

NLS Variance Estimator: Special Cases

$$\widetilde{\text{Avar}}(\widehat{\boldsymbol{\theta}}) = \left(\sum_{i=1}^{N} \widehat{\nabla_\theta m_i}' \widehat{\nabla_\theta m_i}\right)^{-1} \left(\sum_{i=1}^{N} \widehat{u}_i^2 \widehat{\nabla_\theta m_i}' \widehat{\nabla_\theta m_i}\right) \left(\sum_{i=1}^{N} \widehat{\nabla_\theta m_i}' \widehat{\nabla_\theta m_i}\right)^{-1}.$$

Exponential regression:

$$\begin{aligned} m(\mathbf{x}, \boldsymbol{\theta}) &= \exp(\mathbf{x}\boldsymbol{\theta}), \\ \Rightarrow \nabla_\theta m(\mathbf{x}, \boldsymbol{\theta}_o) &= \exp(\mathbf{x}\boldsymbol{\theta}_o)\,\mathbf{x}, \\ \Rightarrow \widehat{\nabla_\theta m_i}' \widehat{\nabla_\theta m_i} &= \exp(2\mathbf{x}_i\widehat{\boldsymbol{\theta}})\mathbf{x}_i'\mathbf{x}_i, \\ \text{and} \quad \widehat{u}_i &= y_i - \exp(\mathbf{x}_i\widehat{\boldsymbol{\theta}}). \end{aligned}$$

NLS Variance Estimator: Special Cases

$$\widetilde{\text{Avar}}(\widehat{\boldsymbol{\theta}}) = \left(\sum_{i=1}^{N} \widehat{\nabla_\theta m_i}' \widehat{\nabla_\theta m_i}\right)^{-1} \left(\sum_{i=1}^{N} \widehat{u}_i^2 \widehat{\nabla_\theta m_i}' \widehat{\nabla_\theta m_i}\right) \left(\sum_{i=1}^{N} \widehat{\nabla_\theta m_i}' \widehat{\nabla_\theta m_i}\right)^{-1}.$$

Linear regression:

$$\begin{aligned} m(\mathbf{x}, \boldsymbol{\theta}) &= \mathbf{x}\boldsymbol{\theta}, \\ \Rightarrow \nabla_\theta m(\mathbf{x}, \boldsymbol{\theta}_o) &= \mathbf{x}, \\ \Rightarrow \widehat{\nabla_\theta m_i}' \widehat{\nabla_\theta m_i} &= \mathbf{x}_i' \mathbf{x}_i, \\ \text{and} \quad \widehat{u}_i &= y_i - \mathbf{x}_i \widehat{\boldsymbol{\theta}}. \qquad \text{(OLS residuals)} \end{aligned}$$

...usual heteroskedasticity-robust variance estimator for OLS.

Multivariate Nonlinear Least Squares

Nonlinear Vector Regression

- Now: outcome $\mathbf{y}$ $(G \times 1)$.
- Parametric model $\mathbf{m}(\mathbf{x}, \theta)$ for $E(\mathbf{y} \mid \mathbf{x})$.
- Multivariate NLS estimator = M-estimator with

$$q(\mathbf{w}, \theta) = \|\mathbf{y} - \mathbf{m}(\mathbf{x}, \theta)\|^2 = \sum_{g=1}^{G} \left[y_g - m_g(\mathbf{x}, \theta)\right]^2.$$

$\Rightarrow$ general theorems apply.

Nonlinear Vector Regression

- Now $\mathbf{u} := \mathbf{y} - \mathbf{m}(\mathbf{x}, \boldsymbol{\theta}_{\mathbf{o}})$.
- Asymptotic variance of sandwich form,

$$
\begin{aligned}
A_0 &:= E[\nabla_\theta \mathbf{m}(\mathbf{x}, \boldsymbol{\theta}_{\mathbf{o}})' \, \nabla_\theta \mathbf{m}(\mathbf{x}, \boldsymbol{\theta}_{\mathbf{o}})] \,, \\
B_0 &:= E[\nabla_\theta \mathbf{m}(\mathbf{x}, \boldsymbol{\theta}_{\mathbf{o}})' \, \mathbf{u}\mathbf{u}' \, \nabla_\theta \mathbf{m}(\mathbf{x}, \boldsymbol{\theta}_{\mathbf{o}})] \,.
\end{aligned}
$$

- Robust estimation analogous to scalar case.
- Also robust to cross-equation correlation.

Example: Linear Panel Model

Consider linear panel model with strict exogeneity

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta}_{\mathbf{o}} + c_i + u_{it}, \qquad E(u_{it} \mid \mathbf{x}_i, c_i) = 0, \quad t = 1, 2, \ldots, T.$$

First-differencing and stacking:

$$\Delta\mathbf{y}_i = \Delta\mathbf{X}_i\boldsymbol{\beta}_{\mathbf{o}} + \Delta\mathbf{u}_i, \qquad ((T-1)\times 1).$$

Strict exogeneity $\Rightarrow E(\Delta\mathbf{u}_i \mid \mathbf{x}_i) = 0$, so

$$E(\Delta\mathbf{y}_i \mid \mathbf{x}_i) = \Delta\mathbf{X}_i\boldsymbol{\beta}_{\mathbf{o}}.$$

$\Rightarrow$ Suggests multivariate NLS!

Example: Linear Panel Model

- Here $\nabla_\theta \mathbf{m}(\mathbf{x}_i, \theta) = \Delta \mathbf{X}_i$.

$$\Rightarrow \text{ (semi)robust variance estimator:}$$

$$\left(\sum_{i=1}^{N} \Delta \mathbf{X}_i' \Delta \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \Delta \mathbf{X}_i' \widehat{\Delta \mathbf{u}}_i \widehat{\Delta \mathbf{u}}_i' \Delta \mathbf{X}_i \right) \left(\sum_{i=1}^{N} \Delta \mathbf{X}_i' \Delta \mathbf{X}_i \right)^{-1} .$$

- Just robust estimator of $\text{Avar}(\hat{\beta}_{FD})$!
- Robust to both heteroskedasticity and serial correlation.
- FE estimation similarly embedded. (Check!)

Non-Linear Hypothesis Testing

Non-linear Hypotheses

Want to test $Q\,(\leqslant P)$ nonlinear restrictions

$$\mathrm{H}_0 : \mathbf{c}\,(\boldsymbol{\theta}_o) = \mathbf{0}. \qquad (Q \times 1)$$

Ex. 1: $\theta_{o1} = \theta_{o2}^2$.

$\Rightarrow$ $\mathbf{c}\,(\boldsymbol{\theta}) = \theta_1 - \theta_2^2$.

Ex. 2: $\theta_{o1}\theta_{o2} = 1$.

$\Rightarrow$ $\mathbf{c}\,(\boldsymbol{\theta}) = \theta_1\theta_2 - 1$.

Wald Statistic

- Suppose $\mathbf{c}$ continuously differentiable.
- Let $\nabla \mathbf{c}$ denotes its Jacobian $(Q \times P)$
- Suppose $\widehat{\boldsymbol{\theta}}_N$ $\sqrt{N}$-asymptotically normal.
- Let $\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\theta}}_N)}$ be consistent for $\mathrm{Avar}(\widehat{\boldsymbol{\theta}}_N)$.
- Wald statistic:

$$W_N := \mathbf{c}(\widehat{\boldsymbol{\theta}}_N)' \left[\widehat{\mathbf{C}}_N \widehat{\mathrm{Avar}(\widehat{\boldsymbol{\theta}}_N)} \widehat{\mathbf{C}}_N'\right]^{-1} \mathbf{c}(\widehat{\boldsymbol{\theta}}_N), \quad \widehat{\mathbf{C}}_N := \nabla \mathbf{c}(\widehat{\boldsymbol{\theta}}_N).$$

Wald Test

$$\mathrm{H}_0 : \mathbf{c}(\boldsymbol{\theta}_o) = \mathbf{0}. \qquad (Q \times 1)$$

- Under $\mathrm{H}_0 : W_N \to_d \chi^2_Q$.
- Let $\alpha \in (0, 1)$ denote significance level.
- Wald test:

$$\text{Reject } \mathrm{H}_0 \Leftrightarrow W_N > (1 - \alpha)\text{-quantile of } \chi^2_Q.$$

Discussion

Above testing procedure presumes:

1. $\mathbf{c}$ continuously differentiable.
2. $\nabla \mathbf{c}(\boldsymbol{\theta}_o)$ full rank (Q).

Ex. 1: $\mathbf{c}(\boldsymbol{\theta}) = \theta_1 - \theta_2^2 \Rightarrow \nabla \mathbf{c}(\boldsymbol{\theta}_o) = \left[\begin{array}{cc} 1 & , -2\theta_{o2} \end{array}\right]$.

- Rank? In this case, always full rank (1) since $\nabla \mathbf{c}(\boldsymbol{\theta}_o) \neq \mathbf{0}$.

Ex. 2: $\mathbf{c}(\boldsymbol{\theta}) = \theta_1\theta_2 - 1 \Rightarrow \nabla \mathbf{c}(\boldsymbol{\theta}_o) = \left[\begin{array}{cc} \theta_{o2}, & \theta_{o1} \end{array}\right]$.

- Rank? Can rank $< Q = 1$ if $\theta_{o1} = \theta_{o2} = 0$.

Why Chi Square?

Q: Where does $W_N \rightarrow_d \chi^2_Q$ under null come from?

Two ingredients

(1) Normal/Chi-Square relation:

- If $\mathbf{Z} \sim \mathrm{N}\left(\mathbf{0}_{G\times 1}, \mathbf{V}\right)$ then $\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z} \sim \chi^2_G$.
- Multivariate version of $Z \sim \mathrm{N}\left(0, \sigma^2\right) \Rightarrow \left(Z/\sigma\right)^2 \sim \chi^2_1$.

(2) Delta Method...

Delta Method

Suppose interest lies in $\mathbf{c}(\boldsymbol{\theta}_o)$, where

- $\sqrt{N}(\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_o) \to_d \mathbf{N}\left(\mathbf{0}_{P\times 1}, \mathbf{V}\right)$
- $\mathbf{c} : \mathbb{R}^P \to \mathbb{R}^Q$, continuously differentiable at $\boldsymbol{\theta}_o$.

Then

$$\sqrt{N}[\mathbf{c}(\widehat{\boldsymbol{\theta}}_N) - \mathbf{c}(\boldsymbol{\theta}_o)] \xrightarrow{d} \mathbf{N}\left(\mathbf{0}_{Q\times 1}, \mathbf{CVC}'\right), \quad \mathbf{C} := \nabla \mathbf{c}\left(\boldsymbol{\theta}_o\right).$$

Why?

- Taylor expansion of $\mathbf{c}(\widehat{\boldsymbol{\theta}}_N)$
- Slutsky's theorem.
- Multivariate CLT.
- ... and some algebra.
- Delta method is workhorse for inference in M-estimation.

Our Case

Have assumed

- $\widehat{\boldsymbol{\theta}}_N$ $\sqrt{N}$-asymptotically normal
- $\mathbf{c}$ cont diff at $\boldsymbol{\theta}_o$
- $\mathbf{C} = \nabla \mathbf{c}\,(\boldsymbol{\theta}_o)$ full rank (Q)

Hence Wald arises from the usual Normal/Chi-Square relation

$$\sqrt{N}[\mathbf{c}(\widehat{\boldsymbol{\theta}}_N) - \mathbf{c}(\boldsymbol{\theta}_o)]'(\mathbf{CVC}')^{-1}\sqrt{N}[\mathbf{c}(\widehat{\boldsymbol{\theta}}_N) - \mathbf{c}(\boldsymbol{\theta}_o)] \xrightarrow{d} \chi^2_Q.$$

where we estimate $\mathbf{CVC}'$ using $\widehat{\mathbf{C}}_N \widehat{\mathrm{Avar}(\widehat{\boldsymbol{\theta}}_N)} \widehat{\mathbf{C}}'_N$.