M-Estimation, II: Inference with M-Estimators

Jesper Riis–Vestergaard Sørensen

University of Copenhagen, Department of Economics

Recap: Setting

optimization

candidate parameters found in parameter space \theta

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

w=observables like outcome y and explanatory vars x

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

Under conditions, $\widehat{\boldsymbol{\theta}}$ $\sqrt{N}$-asymptotically normal,

As if $\hat{\theta} \approx N\left(\theta_0, \frac{A_0^{-1}B_0A_0^{-1}}{N}\right)$

$$\sqrt{N}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o) \xrightarrow{d} \mathbb{N}\left(\mathbf{0}, \mathbf{A}_o^{-1}\mathbf{B}_o\mathbf{A}_o^{-1}\right).$$

as $N \to \infty$ for fixed $P = \dim(\Theta)$ fixed

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

Q: How to estimate $\text{Avar}(\widehat{\boldsymbol{\theta}}) = \mathbf{A}_o^{-1}\mathbf{B}_o\mathbf{A}_o^{-1}/N$?

Outline

Variance Estimation

Example: Nonlinear Least Squares

Multivariate Nonlinear Least Squares

Nonlinear Hypothesis Testing

Variance Estimation

How to Estimate Asymptotic Variance?

\theta_0 the true parameter

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

$$\approx \frac{1}{N}\sum_i H(w_i, \theta_0)$$

Recall:

$$\mathbf{A}_o = E\left[\mathbf{H}\left(\mathbf{w}, \boldsymbol{\theta}_o\right)\right],$$

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

$$\mathbf{s}\left(\mathbf{w}, \boldsymbol{\theta}\right) = \frac{\partial}{\partial \boldsymbol{\theta}} q\left(\mathbf{w}, \boldsymbol{\theta}\right), \qquad (P \times 1)$$

$$\mathbf{H}\left(\mathbf{w}, \boldsymbol{\theta}\right) = \frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} q\left(\mathbf{w}, \boldsymbol{\theta}\right). \qquad (P \times P)$$

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

Estimator 1: Most Structural

one extreme

Most structural = the approach that requires the most assumptions

One (naïve) **idea**:

1. Analytically solve for expectations

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

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

Drawbacks:

(y,x)

Typically, only model Dist(y|x)

- Requires complete specification of $\mathbf{w}$ distribution.

  we have imposed a distribution of outcome(y)|regressors(x), not the distribution of the regressors(x) themselves
- Obtaining closed-form expression difficult.

Rarely an option...

Estimator 2: Least Structural

requires the fewest assumptions

- $\mathbf{A}_o$ and $\mathbf{B}_o$ expectations of functions of $\boldsymbol{\theta}_o$.

A good way to create estimators is the analogy principle

Goldberger explains the main idea of it:

the analogy principle of estimation...proposes that population parameters be estimated by sample statistics which have the same property in the sample as the parameters do in the population (Goldberger, 1968, as cited in Manski, 1988)

- Invoke analogy principle:

  1. Replace expectations with averages.
     replacing the population expectations with sample averages
  2. Insert $\widehat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}_o$.
     estimator for true estimate

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

average Hessian contributes evaluated at the i'th observation and M-estimator

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

average of the outer product of the estimate of the score contributions, evaluated at the i'th observation and M-estimator

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

Estimator 2: Least Structural

come for 'free' from optimization algorithm

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

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

+ 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.

p.d. positive definite
full rank

- 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

$$A_o = E[H(w, \theta_o)]$$
$$= E[E[H(w, \theta_o) | x]]$$

- Let $\mathbf{w} = (\mathbf{y}, \mathbf{x})$.
- Let $\boldsymbol{\theta}_o$ index feature of $\mathbf{y}|\mathbf{x}$ distribution.
  - E.g. mean, median, whole distribution.

wrt x

wrt. y|x

- Define

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

- Estimator:

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

Estimator 3: Semistructural

Recall:

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

+ Usually positive definite in sample.
  this means we can invert it

- Useful when $E\left[\mathbf{H}\left(\mathbf{w}, \boldsymbol{\theta}_o\right) \middle| \mathbf{x}\right]$ 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 $\Rightarrow$

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

(Semi)Robust variance estimators.

Example: Nonlinear Least Squares

NLS Score and Hessian

$E[y|x] = m(x, \theta_0)$
(correctly specified)

In NLS,

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

Chain and product rules $\Rightarrow$

$$\frac{\partial}{\partial \theta} q(w, \theta) = \mathbf{s}(\mathbf{w}, \boldsymbol{\theta}) = -2 [y - m(\mathbf{x}, \boldsymbol{\theta})] \frac{\partial}{\partial \boldsymbol{\theta}} m(\mathbf{x}, \boldsymbol{\theta}), \qquad p \times 1$$

$$\frac{\partial}{\partial \theta'} s(w, \theta) = \mathbf{H}(\mathbf{w}, \boldsymbol{\theta}) = 2 \frac{\partial}{\partial \boldsymbol{\theta}} m(\mathbf{x}, \boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}'} m(\mathbf{x}, \boldsymbol{\theta})$$

$$- 2 [y - m(\mathbf{x}, \boldsymbol{\theta})] \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} m(\mathbf{x}, \boldsymbol{\theta}) . \qquad p \times p$$

Only $E(y|\mathbf{x})$ specified $\Rightarrow$ fully structural impossible.

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

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

$s(w, \theta) = -2[y - m(x, \theta)]\frac{\partial}{\partial \theta} m(x, \theta)$

$u = y - m(x, \theta_o)$

$$\mathbf{s}(\mathbf{w}, \boldsymbol{\theta}_o) = -2u\frac{\partial}{\partial \boldsymbol{\theta}} m(\mathbf{x}, \boldsymbol{\theta}_o),$$

$$\Rightarrow \mathbf{B}_o = 4E\left[u^2 \frac{\partial}{\partial \boldsymbol{\theta}} m(\mathbf{x}, \boldsymbol{\theta}_o) \frac{\partial}{\partial \boldsymbol{\theta}'} m(\mathbf{x}, \boldsymbol{\theta}_o)\right]$$

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

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

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

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

Model well-specified $\Rightarrow E(u|\mathbf{x}) = 0.$

$y - m(x, \theta_0)$

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

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

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

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

$\Rightarrow$ Heteroskedasticity-robust variance estimator for NLS.

  - Output of standard 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: $y \geq 0$ (e.g. wage)

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

$$\Rightarrow \nabla_\theta m(\mathbf{x}, \boldsymbol{\theta}_o) = \exp(\mathbf{x}\boldsymbol{\theta}_o)\,\mathbf{x},$$

$$\Rightarrow \widehat{\nabla_\theta m_i}' \widehat{\nabla_\theta m_i} = \exp(2\mathbf{x}_i\widehat{\boldsymbol{\theta}})\mathbf{x}_i'\mathbf{x}_i,$$

$$\text{and} \quad \widehat{u}_i = y_i - \exp(\mathbf{x}_i\widehat{\boldsymbol{\theta}}).$$

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 $\mathbf{m}(\mathbf{x}, \boldsymbol{\theta})$ for $E(\mathbf{y}|\, \mathbf{x})$.
- Multivariate NLS estimator = M-estimator with

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

$\Rightarrow$ general theorems apply.

Nonlinear Vector Regression

- Now $\mathbf{u} := \mathbf{y} - \mathbf{m}(\mathbf{x}, \boldsymbol{\theta}_o)$.

  assuming correct specification
- Asymptotic variance of sandwich form,

Jacobian of m w.r.t \theta_0 (the true parameter)

$$\mathbf{A}_o := E\left[\nabla_\theta \mathbf{m}(\mathbf{x}, \boldsymbol{\theta}_o)' \nabla_\theta \mathbf{m}(\mathbf{x}, \boldsymbol{\theta}_o)\right],$$

$$\mathbf{B}_o := E\left[\nabla_\theta \mathbf{m}(\mathbf{x}, \boldsymbol{\theta}_o)' \mathbf{u}\mathbf{u}' \nabla_\theta \mathbf{m}(\mathbf{x}, \boldsymbol{\theta}_o)\right].$$

- Robust estimation analogous to scalar case.

  Robustness is not only to heteroskedasticity, but also to:
- Also robust to *cross-equation correlation*.

Example: Linear Panel Model

Consider linear panel model with strict exogeneity

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

First-differencing and stacking:

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

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

$$E\left(\Delta\mathbf{y}_i|\,\mathbf{x}_i\right) = \Delta\mathbf{X}_i\boldsymbol{\beta}_o.$$

Suggests multivariate NLS!

Example: Linear Panel Model

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

$\Rightarrow$ (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 $\mathrm{Avar}(\widehat{\boldsymbol{\beta}}_{FD})$!
- Robust to both heteroskedasticity and serial correlation.
- FE estimation similarly embedded. (Check!)

Nonlinear Hypothesis Testing

Nonlinear Hypotheses

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

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

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

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

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

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

Wald Statistic

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

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

Wald Test

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

- Under H_0 : $W \to_d \chi^2_Q$.
- Let $\alpha \in (0, 1)$ denote significance level.
- Wald test:

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

Discussion

Above testing procedure presumes:

1. $\mathbf{c}$ continuously differentiable.
2. $\nabla \mathbf{c}(\boldsymbol{\theta}_o)$ full rank (Q).
   Jacobian has dimensions Qx1

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? This is rank 1 (full rank). All good in the hood

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? It is possible that the two elements are zero, thus yielding a zero row -> rank zero
  not full rank!

Why Chi Square?

Q: Where does $W \to_d \chi^2_Q$ 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

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

Then

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

Why?

Our Case

Have assumed

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

Hence

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

Wald arises from ____