

Classical Non-Linear Methods: Inference with M-Estimators

Bertel Schjerning

University of Copenhagen, Department of Economics

January 29, 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

$$\boldsymbol{\theta}_o \in \operatorname{argmin}_{\boldsymbol{\theta} \in \Theta} \mathbb{E} [q(\mathbf{w}, \boldsymbol{\theta})], \quad (\text{M-estimand})$$

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

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

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

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

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

Outline

Variance Estimation

Example: Non-Linear Regression

Non-Linear Hypothesis Testing

Variance Estimation

How to Estimate Asymptotic Variance?

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

Recall:

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

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

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

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

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

Estimator 1: Most Structural

One (nāi;œve) **idea**:

1. Analytically solve for expectations

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

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

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

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 θ_o .
- ▶ Invoke analogy principle:
 1. Replace expectations with averages.
 2. Insert $\hat{\theta}$ for θ_o .

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

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

Estimator 2: Least Structural

$$\begin{aligned}\hat{\mathbf{A}} &:= \frac{1}{N} \sum_{i=1}^N \hat{\mathbf{H}}_i, & \hat{\mathbf{H}}_i &:= \mathbf{H}(\mathbf{w}_i, \hat{\boldsymbol{\theta}}), \\ \hat{\mathbf{B}} &:= \frac{1}{N} \sum_{i=1}^N \hat{\mathbf{s}}_i \hat{\mathbf{s}}_i', & \hat{\mathbf{s}}_i &:= \mathbf{s}(\mathbf{w}_i, \hat{\boldsymbol{\theta}}).\end{aligned}$$

- + With twice cont diff q , always available.
- + If $\hat{\boldsymbol{\theta}}$ interior, $\hat{\mathbf{A}}$ 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 θ_o index feature of $D(\mathbf{y}|\mathbf{x})$.

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

Define **conditional Hessian**

$$\mathbf{A}(\mathbf{x}, \theta_o) := \mathbb{E}[\mathbf{H}(\mathbf{w}, \theta_o) | \mathbf{x}].$$

Then estimator given by

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

Estimator 3: Semistructural

Recall:

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

+ Usually positive definite in sample.

► Useful if $E[\mathbf{H}(\mathbf{w}, \boldsymbol{\theta}_o) | \mathbf{x}]$ 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

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

Semistructural approach \Rightarrow

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

Robust/semi-robust variance estimators.

Example: Non-Linear Regression

Recap: Non-Linear Mean Regression Model

y : scalar outcome variable.

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

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

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

► $m(\mathbf{x}, \boldsymbol{\theta})$: Candidate for $E[y \mid \mathbf{x}]$.

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

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

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

$\boldsymbol{\theta}_o$ called “**true value of theta**.”

Recap: Non-Linear Least Squares

Correct specification \Rightarrow true θ solves population problem

$$\theta_o \in \operatorname{argmin}_{\theta \in \Theta} E[q(\mathbf{w}, \theta)], \quad (\text{PP})$$

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

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

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

Suggests **non-linear least squares (NLS)** estimator

$$\hat{\theta} \in \operatorname{argmin}_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^N [y_i - m(\mathbf{x}_i, \theta)]^2. \quad (\text{SP})$$

NLS Score and Hessian

In NLS,

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

Chain and product rules \Rightarrow

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

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

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

\Rightarrow fully structural impossible.

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

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

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

Abbreviating

$$\begin{aligned} \widehat{u}_i &:= y_i - m(\mathbf{x}_i, \widehat{\boldsymbol{\theta}}), & (\text{NLS residuals}) \\ \widehat{\nabla_{\boldsymbol{\theta}} m}_i &:= \frac{\partial}{\partial \boldsymbol{\theta}'} m(\mathbf{x}_i, \widehat{\boldsymbol{\theta}}), & (1 \times P) \\ \widehat{\mathbf{B}} &= \frac{4}{N} \sum_{i=1}^N \widehat{u}_i^2 \widehat{\nabla_{\boldsymbol{\theta}} m}_i' \widehat{\nabla_{\boldsymbol{\theta}} m}_i. \end{aligned}$$

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:

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

Least structural approach \Rightarrow

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

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

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

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

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

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

Semistructural approach \Rightarrow

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

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

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 canned software packages.

Asymptotic standard errors = square root of diagonal.

NLS Variance Estimator: Special Cases

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

Exponential regression:

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

NLS Variance Estimator: Special Cases

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

Linear regression:

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

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

Non-Linear Hypothesis Testing

Non-linear Hypotheses

Want to test $Q (\leq P)$ nonlinear restrictions

$$H_0 : \mathbf{c}(\boldsymbol{\theta}_o) = \mathbf{0}. \quad (Q \times 1)$$

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 $\hat{\boldsymbol{\theta}}_N$ \sqrt{N} -asymptotically normal.
- ▶ Let $\widehat{\text{Avar}}(\hat{\boldsymbol{\theta}}_N)$ be consistent for $\text{Avar}(\hat{\boldsymbol{\theta}}_N)$.
- ▶ Wald statistic:

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

Wald Test

$$H_0 : \mathbf{c}(\boldsymbol{\theta}_o) = \mathbf{0}. \quad (Q \times 1)$$

- ▶ Under $H_0 : W_N \rightarrow_d \chi_Q^2$.
- ▶ Let $\alpha \in (0, 1)$ denote significance level.
- ▶ Wald test:

Reject $H_0 \Leftrightarrow W_N > (1 - \alpha)$ -quantile of χ_Q^2 .

Discussion

Above testing procedure presumes:

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

Ex. 1: $\mathbf{c}(\boldsymbol{\theta}) = \theta_1 - \theta_2^2 \Rightarrow \nabla \mathbf{c}(\boldsymbol{\theta}_o) = \begin{bmatrix} 1 & -2\theta_{o2} \end{bmatrix}$.

► Rank? _____

Ex. 2: $\mathbf{c}(\boldsymbol{\theta}) = \theta_1\theta_2 - 1 \Rightarrow \nabla \mathbf{c}(\boldsymbol{\theta}_o) = \begin{bmatrix} \theta_{o2} & \theta_{o1} \end{bmatrix}$.

► Rank? _____

Why Chi Square?

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

Two ingredients

(1) Normal/Chi-Square relation:

- ▶ If $\mathbf{Z} \sim N(\mathbf{0}_{G \times 1}, \mathbf{V})$ then $\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z} \sim \chi_G^2$.
- ▶ Multivariate version of $Z \sim N(0, \sigma^2) \Rightarrow (Z/\sigma)^2 \sim \chi_1^2$.

(2) Delta Method...

Delta Method

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

- ▶ $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_o) \rightarrow_d \mathbf{N}(\mathbf{0}_{P \times 1}, \mathbf{V})$
- ▶ $\mathbf{c} : \mathbb{R}^P \rightarrow \mathbb{R}^Q$, continuously differentiable at $\boldsymbol{\theta}_o$.

Then

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

Why?

Our Case

Have assumed

- ▶ $\hat{\theta}_N$ \sqrt{N} -asymptotically normal
- ▶ \mathbf{c} cont diff at θ_o
- ▶ $\mathbf{C} = \nabla \mathbf{c}(\theta_o)$ full rank (Q)

Hence

$$\sqrt{N}[\mathbf{c}(\hat{\theta}_N) - \mathbf{c}(\theta_o)]'(\mathbf{CVC}')^{-1}\sqrt{N}[\mathbf{c}(\hat{\theta}_N) - \mathbf{c}(\theta_o)] \xrightarrow{d} \chi_Q^2.$$

Wald arises from _____