M-Estimation

1.1 Motivation

Discussion 1.1.1. Consider an M-estimator $\hat{\boldsymbol{\theta}}_n$ obtained by solving:

$$\boldsymbol{\Psi}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\psi}(\boldsymbol{z}_i;\boldsymbol{\theta}) \overset{\text{Set}}{=} \mathbf{0}$$

Recall that the influence function expansion for $\hat{\boldsymbol{\theta}}_n$ is:

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\big) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{A}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{\psi}(\boldsymbol{z}_i;\boldsymbol{\theta}_0) + o_p(1)$$

where:

$$\boldsymbol{A}(\boldsymbol{\theta}_0) = -E\left\{\frac{\partial\boldsymbol{\psi}(\boldsymbol{z}_i;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},$$

and that the limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ is:

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\big) \xrightarrow{\mathcal{L}} N\big\{\mathbf{0},\ \boldsymbol{A}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{B}(\boldsymbol{\theta}_0)\boldsymbol{A}(\boldsymbol{\theta}_0)^{-T}\big\},$$

where:

$$\boldsymbol{B}(\boldsymbol{\theta}_0) = E\big\{\boldsymbol{\psi}(\boldsymbol{z}_i;\boldsymbol{\theta})\otimes\boldsymbol{\psi}(\boldsymbol{z}_i;\boldsymbol{\theta})\big\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

Let $\tilde{\boldsymbol{\theta}}_n$ denote the estimator obtained by solving the perturbed estimating equations:

$$\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\psi}(\boldsymbol{z}_i;\boldsymbol{\theta})\cdot\omega_i \overset{\text{Set}}{=} 0,$$

where (ω_i) are IID random weights, with $E(\omega_i) = 0$ and $E(\omega_i^2) = 1$, that are generated independent of the data.

The **score bootstrap** approximates the distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ by that of:

$$\sqrt{n}\big(\tilde{\boldsymbol{\theta}}_n - \hat{\boldsymbol{\theta}}_n\big) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{A}(\hat{\boldsymbol{\theta}}_n)^{-1}\boldsymbol{\psi}(\boldsymbol{z}_i;\hat{\boldsymbol{\theta}}_n)\cdot\omega_i$$

Conditional on the observed data, the perturbed influence function has expectation:

$$E\big\{\sqrt{n}\big(\tilde{\boldsymbol{\theta}}_n - \hat{\boldsymbol{\theta}}_n\big)\big\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{A}(\hat{\boldsymbol{\theta}}_n)^{-1}\boldsymbol{\psi}(\boldsymbol{z}_i;\hat{\boldsymbol{\theta}}_n)\cdot E(\omega_i) = \mathbf{0},$$

and variance:

$$\begin{aligned}\text{Var}\big\{\sqrt{n}\big(\tilde{\boldsymbol{\theta}}_n - \hat{\boldsymbol{\theta}}_n\big)\big\} &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}(\hat{\boldsymbol{\theta}}_n)^{-1}\boldsymbol{\psi}(\boldsymbol{z}_i;\hat{\boldsymbol{\theta}}_n)\otimes\boldsymbol{\psi}(\boldsymbol{z}_i;\hat{\boldsymbol{\theta}}_n)\boldsymbol{A}(\hat{\boldsymbol{\theta}}_n)^{-T}\cdot E(\omega_i^2) \\ &= \boldsymbol{A}(\hat{\boldsymbol{\theta}}_n)^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\psi}(\boldsymbol{z}_i;\hat{\boldsymbol{\theta}}_n)\otimes\boldsymbol{\psi}(\boldsymbol{z}_i;\hat{\boldsymbol{\theta}}_n)\right\}\boldsymbol{A}(\hat{\boldsymbol{\theta}}_n)^{-T} \xrightarrow{p} \boldsymbol{A}_0^{-1}\boldsymbol{B}_0\boldsymbol{A}_0^{-T}.\end{aligned}$$

Consequently, $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ and $\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \hat{\boldsymbol{\theta}}_n)$ converge to the same limiting distribution (see Kline & Santos 2012). ♠

1.2 Procedure

1. Estimate $\hat{\boldsymbol{\theta}}_n$ satisfying $\boldsymbol{\Psi}_n(\hat{\boldsymbol{\theta}}_n) = \mathbf{0}$.

   i. Calculate the individual score contributions evaluated at $\hat{\boldsymbol{\theta}}_n$:

   $$\hat{\boldsymbol{\psi}}_i = \boldsymbol{\psi}(\boldsymbol{z}_i; \hat{\boldsymbol{\theta}}_n).$$

   ii. Estimate the information matrix $\hat{\boldsymbol{\theta}}$:

   $$\hat{\boldsymbol{A}}_n = -\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\partial \boldsymbol{\psi}(\boldsymbol{z}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right\}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

2. For $b = 1, \cdots, B$:

   i. Generate the perturbation weights $\big(\omega_i^{(b)}\big)$.

   ii. Calculate the perturbed score:

   $$\mathcal{U}_n^{(b)} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\boldsymbol{A}}_n^{-1}\hat{\boldsymbol{\psi}}_i \cdot \omega_i^{(b)}.$$

   Note that for hypothesis testing perturbations should be performed under H_0.

3. Approximate the distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$ by that of $\mathcal{U}_n^{(b)}$.

Examples

2.1 Huber Estimator

Example 1.2.1 (Confidence Intervals)**.** Suppose (Y_i) are a random sample from a symmetric distribution with location θ_0. Recall that the Huber estimator $\hat{\theta}_n$ is a solution to the estimating equation:

$$\Psi_n(\theta) = \sum_{i=1}^{n}\psi_\tau(Y_i - \theta) \overset{\text{Set}}{=} 0,$$

where $\psi_\tau(u)$ is the Huber function with threshold τ:

$$\psi_\tau(u) = \begin{cases} -\tau & u < -\tau, \\ u & -\tau \leq u \leq \tau, \\ \tau & u > \tau. \end{cases}$$

The influence function expansion for $\hat{\theta}_n$ is:

$$\sqrt{n}\big(\hat{\theta}_n - \theta_0\big) = \frac{1}{\sqrt{n}} A_0^{-1} \sum_{i=1}^{n} \psi_\tau(Y_i - \theta_0) + o_p(1),$$

where A_0 is the information matrix:

$$A_0 = -E\left\{\frac{\partial \psi(Y_i - \theta)}{\partial \theta}\right\} = -E\big\{\dot{\psi}_\tau(Y_i - \theta_0)\big\}, \qquad \dot{\psi}_\tau(u) = \begin{cases} 1, & |u| \leq \tau, \\ 0, & |u| > \tau. \end{cases}$$

Asymptotically:

$$\sqrt{n}\big(\hat{\theta}_n - \theta_0\big) \xrightarrow{\mathcal{L}} N\big(0, A_0^{-1} B_0 A_0^{-1}\big),$$

where:

$$B_0 = E\big\{\psi_\tau^2(Y_i - \theta_0)\big\}.$$

The goal of perturbation is to approximate the distribution of the pivotal quantity:

$$Z_n = \sqrt{n}\frac{\big(\hat{\theta}_n - \theta_0\big)}{\sigma_n} = \frac{1}{\sqrt{n}} \sigma_n^{-1} A(\theta_0)^{-1} \sum_{i=1}^{n} \psi_\tau(Y_i - \theta_0) + o_p(1).$$

Towards this, obtain the individual score contributions $\hat{\psi}_i = \psi_\tau(Y_i - \hat{\theta}_n)$, and the empirical information:

$$\hat{A}_n = \frac{1}{n} \sum_{i=1}^{n} \dot{\psi}_\tau(Y_i - \hat{\theta}_n).$$

For $b \in \{1, \cdots, B\}$, generate the perturbation weights $(\omega_i^{(b)})$ and calculate:

$$Z_n^{(b)} = \frac{1}{\sqrt{n}} \big(\sigma_n^{(b)}\big)^{-1} \hat{A}_n^{-1} \sum_{i=1}^{n} \hat{\psi}_i \cdot \omega_i^{(b)},$$

where the standard error is:

$$\big(\sigma_n^{(b)}\big)^2 = \frac{\hat{B}_n^{(b)}}{\hat{A}_n^2}, \qquad \hat{B}_n^{(b)} = \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_i^2 \cdot \big(\omega_i^{(b)}\big)^2.$$

Obtain critical values (ζ_*, ζ^*) from the quantiles of $\big(Z_n^{(b)}\big)$ such that:

$$P\left(\zeta_* \leq \sqrt{n}\frac{\big(\hat{\theta}_n - \theta\big)}{\sigma_n} \leq \zeta^*\right) = 1 - \alpha.$$

Rearranging provides the confidence interval for θ:

$$\text{CI} = \left(\theta : \hat{\theta}_n - \zeta^* \frac{\sigma_n}{\sqrt{n}} \leq \theta \leq \hat{\theta}_n - \zeta_* \frac{\sigma_n}{\sqrt{n}}\right).$$

♠

Example 1.2.2 (Testing)**.** The Wald statistic for evaluating the $H_0 : \theta = \theta_0$ is:

$$T_W = n\frac{\big(\hat{\theta} - \theta_0\big)^2}{\sigma_n^2}. \tag{1.2.1}$$

The goal of perturbation is to approximate the distribution of T_W under H_0. Towards this, for $b \in \{1, \cdots, B\}$, generate the perturbation weights $(\omega_i^{(b)})$ and calculate:

$$\mathcal{U}_n^{(b)} = \frac{1}{\sqrt{n}}\tilde{A}_n^{-1}\sum_{i=1}^{n}\tilde{\psi}_i \cdot \omega_i^{(b)}, \qquad \big(\tilde{\sigma}_n^{(b)}\big)^2 = \frac{\tilde{B}_n^{(b)}}{\tilde{A}_n^2}, \qquad \tilde{B}_n^{(b)} = \frac{1}{n}\sum_{i=1}^{n}\tilde{\psi}_i^2 \cdot \big(\omega_i^{(b)}\big)^2.$$

Construct the perturbation test statistics:

$$T_W^{(b)} = \frac{\big(\mathcal{U}_n^{(b)}\big)^2}{\big(\tilde{\sigma}_n^{(b)}\big)^2}.$$

A p-value assessing $H_0 : \theta = \theta_0$ is given by:

$$p = \frac{1}{B+1}\left\{1 + \sum_{b=1}^{B} I\big(T^{(b)} \geq T_{\text{obs}}\big)\right\}.$$

♠

2.2 Linear Models

Example 1.2.3 (Confidence Intervals)**.** Consider the model:

$$Y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i,$$

where $E(\epsilon_i|\boldsymbol{x}_i) = 0$ and $\text{Var}(\epsilon_i|\boldsymbol{x}_i) = \sigma_i^2$.

The OLS estimator $\hat{\boldsymbol{\beta}}_n$ is a solution to the estimating equation:

$$\boldsymbol{\Psi}_n(\boldsymbol{\beta}) = \boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) = \sum_{i=1}^{n}\boldsymbol{x}_i(Y_i - \boldsymbol{x}_i'\boldsymbol{\beta}) \overset{\text{Set}}{=} \mathbf{0}.$$

Identify the estimating function as:

$$\boldsymbol{\psi}_i = \boldsymbol{x}_i(Y_i - \boldsymbol{x}_i'\boldsymbol{\beta}).$$

The influence function expansion for $\hat{\boldsymbol{\beta}}_n$ is:

$$\sqrt{n}\big(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\big) = \frac{1}{\sqrt{n}}\boldsymbol{A}_0^{-1}\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + o_p(1),$$

where $\boldsymbol{A}_0$ is the unit information for $\boldsymbol{\beta}$:

$$\boldsymbol{A}_0 = \underset{n\to\infty}{\text{plim}}\frac{1}{n}\mathcal{I}_{\beta\beta'} = \underset{n\to\infty}{\text{plim}}\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}.$$

Asymptotically:

$$\sqrt{n}\big(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\big) \xrightarrow{\mathcal{L}} N\big(\mathbf{0}, \boldsymbol{A}_0^{-1}\boldsymbol{B}_0\boldsymbol{A}_0^{-T}\big),$$

where:

$$\boldsymbol{B}_0 = E\big\{\boldsymbol{\psi}(\boldsymbol{x}_i;\boldsymbol{\beta}_0) \otimes \boldsymbol{\psi}(\boldsymbol{x}_i;\boldsymbol{\beta}_0)\big\} = E\big\{(Y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2\boldsymbol{x}_i \otimes \boldsymbol{x}_i\big\}.$$

The goal of perturbation is to approximate the distribution of the pivotal quantity:

$$\boldsymbol{Z}_n = \sqrt{n}\boldsymbol{\Sigma}_n^{-1/2}\big(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\big) = \frac{1}{\sqrt{n}}\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{A}_0^{-1}\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_0) + o_p(1),$$

where $\boldsymbol{\Sigma}_n$ is a consistent estimator for the asymptotic variance:

$$\boldsymbol{\Sigma}_n = \hat{\boldsymbol{A}}_n^{-1}\hat{\boldsymbol{B}}_n\hat{\boldsymbol{A}}_n^{-T}.$$

The empirical information:

$$\hat{\boldsymbol{A}}_n = \frac{1}{n}\boldsymbol{X}'\boldsymbol{X}.$$

For $b \in \{1, \cdots, B\}$, generate the perturbation weights and calculate:

$$\boldsymbol{Z}_n^{(b)} = \frac{1}{\sqrt{n}}\big(\boldsymbol{\Sigma}_n^{(b)}\big)^{-1/2}\hat{\boldsymbol{A}}_n^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{(b)}\hat{\boldsymbol{e}}_n,$$

where $\hat{\boldsymbol{e}}_n = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_n)$, $\boldsymbol{\Omega}^{(b)} = \mathrm{diag}(\omega_i^{(b)})$, $\boldsymbol{\Sigma}_n^{(b)} = \hat{\boldsymbol{A}}_n^{-1}\hat{\boldsymbol{B}}_n^{(b)}\hat{\boldsymbol{A}}_n^{-T}$, and:

$$\hat{\boldsymbol{B}}_n^{(b)} = \frac{1}{n}\boldsymbol{X}'\boldsymbol{\Omega}^{(b)}\mathrm{diag}(\hat{\boldsymbol{e}}_n^2)\boldsymbol{\Omega}^{(b)}\boldsymbol{X}.$$

Obtain critical values $(\boldsymbol{\zeta}_*, \boldsymbol{\zeta}^*)$ from the quantiles of $\big(\boldsymbol{Z}_n^{(b)}\big)$ such that:

$$P\left(\boldsymbol{\zeta}_* \leq \sqrt{n}\boldsymbol{\Sigma}_n^{-1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \leq \boldsymbol{\zeta}^*\right) = 1 - \alpha.$$

Rearranging provides the confidence interval for θ:

$$\mathrm{CI} = \left(\boldsymbol{\beta} : \hat{\boldsymbol{\beta}}_n - n^{-1/2}\boldsymbol{\Sigma}_n^{1/2}\boldsymbol{\zeta}^* \leq \theta \leq \hat{\theta}_n - n^{-1/2}\boldsymbol{\Sigma}_n^{1/2}\boldsymbol{\zeta}_*\right).$$

♠

Example 1.2.4 (Testing)**.** Consider the model:

$$Y_i = \boldsymbol{x}_i'\boldsymbol{\alpha} + \boldsymbol{z}_i'\boldsymbol{\beta} + \epsilon_i,$$

where $E(\epsilon_i|\boldsymbol{x}_i, \boldsymbol{z}_i) = 0$ and $\mathrm{Var}(\epsilon_i|\boldsymbol{x}_i, \boldsymbol{z}_i) = \sigma_i^2$. Suppose $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ is of interest. Let $\tilde{\boldsymbol{\alpha}}_n$ denote a solution to the OLS estimating equation:

$$\mathcal{U}_\alpha(\boldsymbol{\alpha}, \boldsymbol{\beta}_0) = \boldsymbol{Z}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha} - \boldsymbol{Z}\boldsymbol{\beta}_0) \overset{\text{Set}}{=} \mathbf{0}.$$

Consider the distribution of the score statistic:

$$\mathcal{U}_\beta(\tilde{\boldsymbol{\alpha}}_n, \boldsymbol{\beta}_0) = \boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\alpha}} - \boldsymbol{Z}\boldsymbol{\beta}_0).$$

Take the Taylor expansions of $\mathcal{U}_\alpha$ and $\mathcal{U}_\beta$ about the true $\boldsymbol{\alpha}_0$:

$$\begin{aligned}
\mathbf{0} = \mathcal{U}_\alpha(\tilde{\boldsymbol{\alpha}}_n, \boldsymbol{\beta}_0) &= \mathcal{U}_\alpha(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) - \mathcal{I}_{\alpha\alpha'}(\tilde{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0) + O_p(1), \\
\mathcal{U}_\beta(\tilde{\boldsymbol{\alpha}}_n, \boldsymbol{\beta}_0) &= \mathcal{U}_\beta(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) - \mathcal{I}_{\beta\alpha'}(\tilde{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0) + \mathcal{O}_p(1),
\end{aligned}$$

where $\mathcal{I}$ denotes the information matrix, and the remainder is bounded in probability by assumption. Substitute the first expansion into the second to obtain:

$$\begin{aligned}
\mathcal{U}_\beta(\tilde{\boldsymbol{\alpha}}_n, \boldsymbol{\beta}_0) &= \mathcal{U}_\beta(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) - \mathcal{I}_{\beta\alpha'}\mathcal{I}_{\alpha\alpha'}^{-1}\mathcal{U}_\alpha(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) + \mathcal{O}_p(1) \\
&= \big(-\mathcal{I}_{\beta\alpha'}\mathcal{I}_{\alpha\alpha'}^{-1}, \boldsymbol{I}\big)\begin{pmatrix}\mathcal{U}_\alpha(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) \\ \mathcal{U}_\beta(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\end{pmatrix} + \mathcal{O}_p(1).
\end{aligned}$$

Define the following structures:

$$\boldsymbol{C} = \big(-\mathcal{I}_{\beta\alpha'}\mathcal{I}_{\alpha\alpha'}^{-1}, \boldsymbol{I}\big), \qquad \mathcal{U}_\theta(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) = \begin{pmatrix}\mathcal{U}_\alpha(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) \\ \mathcal{U}_\beta(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\end{pmatrix}, \qquad \boldsymbol{\Xi} = (\boldsymbol{X}, \boldsymbol{Z}).$$

The score statistic for $\boldsymbol{\beta}$ is expressible as:

$$\begin{aligned}
\frac{1}{\sqrt{n}}\mathcal{U}_\beta(\tilde{\boldsymbol{\alpha}}_n, \boldsymbol{\beta}_0) &= \frac{1}{\sqrt{n}}\boldsymbol{C}\mathcal{U}_\theta(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\boldsymbol{C}\boldsymbol{\Xi}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}_0 - \boldsymbol{Z}\boldsymbol{\beta}_0) + o_p(1).
\end{aligned}$$

By the standard central limit theorem:

$$\frac{1}{\sqrt{n}}\mathcal{U}_\beta(\tilde{\boldsymbol{\alpha}}_n, \boldsymbol{\beta}_0) \xrightarrow{\mathcal{L}} N\big(\mathbf{0}, \boldsymbol{\Sigma}_0\big),$$

where:

$$\boldsymbol{\Sigma}_0 = \underset{n\to\infty}{\mathrm{plim}}\frac{1}{n}\boldsymbol{C}\boldsymbol{\Xi}'\big\{\mathrm{diag}(\sigma_i^2)\big\}\boldsymbol{\Xi}\boldsymbol{C}'.$$

A score statistic for evaluating $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ is:

$$T_S = \frac{1}{n}\mathcal{U}_\beta(\tilde{\boldsymbol{\alpha}}_n, \boldsymbol{\beta}_0)'\big\{\boldsymbol{C}\Xi'\text{diag}(\sigma_i^2)\Xi\boldsymbol{C}'\big\}^{-1}\mathcal{U}_\beta(\tilde{\boldsymbol{\alpha}}_n, \boldsymbol{\beta}_0).$$

The goal of perturbation is to calculate the distribution of T_S under H_0. Towards, this $b \in \{1, \cdots, B\}$, generate the perturbation weights and calculate:

$$\mathcal{U}_\beta^{(b)} = \boldsymbol{C}\Xi'\boldsymbol{\Omega}^{(b)}\tilde{\boldsymbol{e}}_n,$$
$$T_S^{(b)} = \frac{1}{n}\mathcal{U}_\beta^{(b)'}\big\{\boldsymbol{C}\Xi'\boldsymbol{\Omega}^{(b)}\text{diag}(\tilde{\boldsymbol{e}}_n^2)\boldsymbol{\Omega}^{(b)}\Xi\boldsymbol{C}'\big\}^{-1}\mathcal{U}_\beta^{(b)},$$

where $\tilde{\boldsymbol{e}}_n = (\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\alpha}}_n - \boldsymbol{Z}\boldsymbol{\beta}_0)$ and $\boldsymbol{\Omega}^{(b)} = \text{diag}(\omega_i^{(b)})$.

A p-value assessing $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ is given by:

$$p = \frac{1}{B+1}\left\{1 + \sum_{b=1}^{B} I\big(T_S^{(b)} \geq T_{\text{obs}}\big)\right\}.$$

♠

2.3 Generalized Linear Models

Example 1.2.5 (Confidence Intervals)**.** Consider a GLM with linear predictor:

$$\eta_i = \boldsymbol{x}_i'\boldsymbol{\beta}.$$

The maximum likelihood estimator $\hat{\boldsymbol{\beta}}_n$ is a solution to the score equations:

$$\mathcal{U}_\beta = \boldsymbol{X}'\boldsymbol{W}\boldsymbol{\Delta}\big(\boldsymbol{y} - \boldsymbol{\mu}\big), \qquad \boldsymbol{\Delta} = \text{diag}\big\{\dot{g}(\mu_i)\big\}, \qquad \boldsymbol{W} = \text{diag}\left\{\frac{1}{\phi\nu(\mu_i)\dot{g}^2(\mu_i)}\right\}.$$

Identify the estimating function as:

$$\boldsymbol{\psi}_i = \boldsymbol{x}_i\frac{\big(Y_i - \boldsymbol{x}_i'\boldsymbol{\beta}\big)}{\phi\nu(\mu_i)\dot{g}(\mu_i)}.$$

The influence function expansion for $\hat{\boldsymbol{\beta}}_n$ is:

$$\sqrt{n}\big(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\big) = \frac{1}{\sqrt{n}}\boldsymbol{A}_0^{-1}\mathcal{U}_\beta(\boldsymbol{\beta}_0) + o_p(1),$$

where $\boldsymbol{A}_0$ is the unit information for $\boldsymbol{\beta}$:

$$\boldsymbol{A}_0 = \operatorname*{plim}_{n\to\infty}\frac{1}{n}\mathcal{I}_{\beta\beta'} = \operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}.$$

Asymptotically:

$$\sqrt{n}\big(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\big) \xrightarrow{\mathcal{L}} N\big(\boldsymbol{0}, \boldsymbol{A}_0^{-1}\boldsymbol{B}_0\boldsymbol{A}_0^{-T}\big).$$

Recall that:

$$\mathrm{Var}(Y_i - \boldsymbol{x}_i'\boldsymbol{\beta}) = \phi\nu(\mu_i),$$

therefore:

$$\boldsymbol{B}_0 = E\big\{\boldsymbol{\psi}(\boldsymbol{x}_i;\boldsymbol{\beta}_0) \otimes \boldsymbol{\psi}(\boldsymbol{x}_i;\boldsymbol{\beta}_0)\big\} = E\left\{\frac{\boldsymbol{x}_i \otimes \boldsymbol{x}_i}{\phi\nu(\mu_i)\dot{g}^2(\mu_i)}\right\} = \operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}.$$

The goal of perturbation is to approximate the distribution of the pivotal quantity:

$$\boldsymbol{Z}_n = \sqrt{n}\boldsymbol{\Sigma}_n^{-1/2}\big(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\big) = \frac{1}{\sqrt{n}}\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{A}_0^{-1}\boldsymbol{X}'\boldsymbol{W}\boldsymbol{\Delta}\big(\boldsymbol{y} - \boldsymbol{\mu}_0\big) + o_p(1),$$

where $\boldsymbol{\Sigma}_n$ is a consistent estimator for the asymptotic variance:

$$\boldsymbol{\Sigma}_n = \hat{\boldsymbol{A}}_n^{-1}\hat{\boldsymbol{B}}_n\hat{\boldsymbol{A}}_n^{-T}.$$

The empirical information is:

$$\hat{\boldsymbol{A}}_n = \frac{1}{n}\boldsymbol{X}'\hat{\boldsymbol{W}}\boldsymbol{X}.$$

For $b \in \{1, \cdots, B\}$, generate the perturbation weights and calculate:

$$\boldsymbol{Z}_n^{(b)} = \frac{1}{\sqrt{n}}\big(\boldsymbol{\Sigma}_n^{(b)}\big)^{-1/2}\boldsymbol{A}_0^{-1}\boldsymbol{X}'\boldsymbol{W}\boldsymbol{\Delta}\boldsymbol{\Omega}^{(b)}\hat{\boldsymbol{e}},$$

where $\hat{\boldsymbol{e}}_n = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_n)$, $\boldsymbol{\Omega}^{(b)} = \mathrm{diag}(\omega_i^{(b)})$, $\boldsymbol{\Sigma}_n^{(b)} = \hat{\boldsymbol{A}}_n^{-1}\hat{\boldsymbol{B}}_n^{(b)}\hat{\boldsymbol{A}}_n^{-T}$, and:

$$\hat{\boldsymbol{B}}_n^{(b)} = \frac{1}{n}\boldsymbol{X}'\boldsymbol{W}\boldsymbol{\Delta}\boldsymbol{\Omega}^{(b)}\mathrm{diag}(\hat{\boldsymbol{e}}_n^2)\boldsymbol{\Omega}^{(b)}\boldsymbol{\Delta}\boldsymbol{W}\boldsymbol{X}.$$

Obtain critical values $(\boldsymbol{\zeta}_*, \boldsymbol{\zeta}^*)$ from the quantiles of $\big(\boldsymbol{Z}_n^{(b)}\big)$ such that:

$$P\left(\boldsymbol{\zeta}_* \le \sqrt{n}\boldsymbol{\Sigma}_n^{-1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \le \boldsymbol{\zeta}^*\right) = 1 - \alpha.$$

Rearranging provides the confidence interval for θ:

$$\mathrm{CI} = \left(\boldsymbol{\beta} : \hat{\boldsymbol{\beta}}_n - n^{-1/2}\boldsymbol{\Sigma}_n^{1/2}\boldsymbol{\zeta}^* \le \theta \le \hat{\theta}_n - n^{-1/2}\boldsymbol{\Sigma}_n^{1/2}\boldsymbol{\zeta}_*\right).$$

♠

Example 1.2.6 (Testing)**.** Consider a GLM with linear predictor:

$$\eta_i = \boldsymbol{x}_i'\boldsymbol{\alpha} + \boldsymbol{z}_i'\boldsymbol{\beta}.$$

Suppose $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ is of interest. Let $\tilde{\boldsymbol{\alpha}}$ denote a solution to the estimating equation:

$$\mathcal{U}_\alpha(\boldsymbol{\alpha}, \boldsymbol{\beta}_0) = \boldsymbol{Z}'\boldsymbol{W}\boldsymbol{\Delta}\big\{\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\alpha}, \boldsymbol{\beta}_0)\big\}.$$

Following the same derivation used for linear models:

$$\begin{aligned}\frac{1}{\sqrt{n}}\mathcal{U}_\beta(\tilde{\boldsymbol{\alpha}}_n, \boldsymbol{\beta}_0) &= \frac{1}{\sqrt{n}}\boldsymbol{C}\mathcal{U}_\theta(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) + o_p(1) \\ &= \frac{1}{\sqrt{n}}\boldsymbol{C}\boldsymbol{\Xi}'\boldsymbol{W}\boldsymbol{\Delta}(\boldsymbol{y} - \boldsymbol{\mu}_0) + o_p(1).\end{aligned}$$

A score statistic for evaluating $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ is:

$$T_S = \frac{1}{n}\mathcal{U}_\beta(\tilde{\boldsymbol{\alpha}}_n, \boldsymbol{\beta}_0)'\big\{\boldsymbol{\Xi}'\boldsymbol{W}\boldsymbol{\Delta}\mathrm{Var}(\boldsymbol{y})\boldsymbol{\Delta}\boldsymbol{W}\boldsymbol{\Xi}\big\}^{-1}\mathcal{U}_\beta(\tilde{\boldsymbol{\alpha}}_n, \boldsymbol{\beta}_0).$$

The goal of perturbation is to calculate the distribution of T_S under H_0. Towards, this $b \in \{1, \cdots, B\}$, generate the perturbation weights and calculate:

$$\begin{gathered}\mathcal{U}_\beta^{(b)} = \boldsymbol{C}\boldsymbol{\Xi}'\boldsymbol{W}\boldsymbol{\Delta}\boldsymbol{\Omega}^{(b)}\tilde{\boldsymbol{e}}_n, \\ T_S^{(b)} = \frac{1}{n}\mathcal{U}_\beta^{(b)\prime}\big\{\boldsymbol{C}\boldsymbol{\Xi}'\boldsymbol{W}\boldsymbol{\Delta}\boldsymbol{\Omega}^{(b)}\mathrm{diag}(\tilde{\boldsymbol{e}}_n^2)\boldsymbol{\Omega}^{(b)}\boldsymbol{\Delta}\boldsymbol{W}\boldsymbol{\Xi}\boldsymbol{C}'\big\}^{-1}\mathcal{U}_\beta^{(b)},\end{gathered}$$

where $\tilde{\boldsymbol{e}}_n = (\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\alpha}}_n - \boldsymbol{Z}\boldsymbol{\beta}_0)$ and $\boldsymbol{\Omega}^{(b)} = \mathrm{diag}(\omega_i^{(b)})$.

A p-value assessing $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ is given by:

$$p = \frac{1}{B+1}\left\{1 + \sum_{b=1}^{B} I\big(T_S^{(b)} \geq T_{\mathrm{obs}}\big)\right\}.$$

Under the canonical link, the preceding are simplified by the identity $\boldsymbol{W}\boldsymbol{\Delta} = \boldsymbol{I}$. ♠

References

- Kline & Santos. *A Score Based Approach to Wild Bootstrap Inference*. Journal of Econometric Methods (2012).