Setting and Model

Remark 1.1.1. Generalized estimating equations (GEEs) model dependence of the outcome mean on covariates, while treating correlation among clustered observations as a nuisance. GEEs are indicated when there are many clusters, and population averaged (marginal), rather than individual-level (conditional), inference is of interest. Inference on the parameters of the mean model remains valid even if the working covariance structure is misspecified. ♦

Discussion 1.1.1. Let $\boldsymbol{y}_k = (Y_{k1}, \cdots, Y_{kn_k})'$ denote the outcome for the kth cluster. Rather than fully specifying the distribution of Y_{ki}, GEE models specify the first two moments of Y_{ki}:

$$E(Y_{ki}|\boldsymbol{x}_{ki}) = \mu_{ki}, \qquad \text{Var}(Y_{ki}|\boldsymbol{x}_{ki}) = \phi\nu(\mu_{ki}). \tag{1.1.1}$$

Here ϕ is the dispersion parameter, and $\nu(\cdot)$ is the variance function. Within the linear exponential family, $\nu(\mu_{ki}) = \ddot{b} \circ \dot{b}^{-1}(\mu_{ki})$, where $b(\cdot)$ is the cumulant function. ♠

Definition 1.1.1. Consider the following mean model:

$$g\big\{E(Y_{ki}|\boldsymbol{x}_{ki})\big\} = g(\mu_{ki}) = \boldsymbol{x}_{ki}'\boldsymbol{\beta},$$

where $\boldsymbol{x}_{ki}$ is a $p \times 1$ covariate. The **generalized estimating equations** for $\boldsymbol{\beta}$ are:

$$\mathcal{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}) = \sum_{k=1}^{K} \boldsymbol{D}_k' \boldsymbol{V}_k^{-1}(\boldsymbol{y}_k - \boldsymbol{\mu}_k),$$

where $\boldsymbol{D}_k$ is the $n_k \times p$ Jacobian defined by:

$$\boldsymbol{D}_k = \frac{\partial \boldsymbol{\mu}_k}{\partial \boldsymbol{\beta}'},$$

$\boldsymbol{V}_k$ is the $n_k \times n_k$ working covariance structure, and $(\boldsymbol{y}_k - \boldsymbol{\mu}_k)$ is the $n_k \times 1$ residual. ■

Definition 1.1.2. Define the *marginal covariance structure* for cluster k as:

$$\boldsymbol{\Sigma}_k = \text{diag}\big\{\text{Var}(Y_{ki})\big\} = \text{diag}\big\{\phi\nu(\mu_{ki})\big\}.$$

Let $\boldsymbol{R}_k(\boldsymbol{\alpha})$ denote the cluster-specific working correlation matrix, parameterized by $\boldsymbol{\alpha}$. The **working covariance matrix** for cluster k is:

$$\boldsymbol{V}_k = \boldsymbol{\Sigma}_k^{1/2}(\boldsymbol{\beta})\boldsymbol{R}_k(\boldsymbol{\alpha})\boldsymbol{\Sigma}_k^{1/2}(\boldsymbol{\beta}).$$

■

Theorem 1.1.1. Suppose that 1. $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\phi}}$ are $\sqrt{K}$-consistent estimators for the correlation and dispersion parameters, and 2. $\partial_{\boldsymbol{\beta}'}\mathcal{U}_{\boldsymbol{\beta}}$ converges uniformly in probability to an invertible matrix $\boldsymbol{A}$ in an open neighborhood of the true $\boldsymbol{\beta}_0$:

$$\boldsymbol{A} \equiv \operatorname*{plim}_{K\to\infty} \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{D}_k'\boldsymbol{V}_k^{-1}\boldsymbol{D}_k.$$

The estimate $\hat{\boldsymbol{\beta}}$ obtained by solving:

$$\mathcal{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta};\hat{\phi},\hat{\boldsymbol{\alpha}}) \overset{\text{Set}}{=} \mathbf{0}$$

is consistent and asymptotically normal:

$$\sqrt{K}\big(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\big) \xrightarrow{\mathcal{L}} N\big(\mathbf{0},\boldsymbol{\Omega}\big),$$

where $\boldsymbol{\Omega} = \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-T}$, with:

$$\boldsymbol{B} \equiv \operatorname*{plim}_{K\to\infty} \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{D}_k'\boldsymbol{V}_k^{-1}(\boldsymbol{y}_k-\boldsymbol{\mu}_k)(\boldsymbol{y}_k-\boldsymbol{\mu}_k)'\boldsymbol{V}_k^{-1}\boldsymbol{D}_k.$$

□

Proof. See section on Inference. ■

Remark 1.1.2. If the working covariance matrix is correctly specified, then:

$$\begin{aligned}
\boldsymbol{B} &= \lim_{n\to\infty}\frac{1}{K}\sum_{k=1}^{K} E\left\{\boldsymbol{D}_k'\boldsymbol{V}_k^{-1}(\boldsymbol{y}_k-\boldsymbol{\mu}_k)(\boldsymbol{y}_k-\boldsymbol{\mu}_k)'\boldsymbol{V}_k^{-1}\boldsymbol{D}_k\right\} \\
&= \lim_{n\to\infty}\frac{1}{K}\sum_{k=1}^{K} \boldsymbol{D}_k'\boldsymbol{V}_k^{-1}E\left\{(\boldsymbol{y}_k-\boldsymbol{\mu}_k)(\boldsymbol{y}_k-\boldsymbol{\mu}_k)'\right\}\boldsymbol{V}_k^{-1}\boldsymbol{D}_k \\
&= \lim_{n\to\infty}\frac{1}{K}\sum_{k=1}^{K} \boldsymbol{D}_k'\boldsymbol{V}_k^{-1}\boldsymbol{V}_k\boldsymbol{V}_k^{-1}\boldsymbol{D}_k = \lim_{n\to\infty}\frac{1}{K}\sum_{k=1}^{K} \boldsymbol{D}_k'\boldsymbol{V}_k^{-1}\boldsymbol{D}_k = \boldsymbol{A}.
\end{aligned}$$

Thus, under correct specification of the working covariance matrix, the limiting covariance $\boldsymbol{\Omega}$ reduces to $\boldsymbol{\Omega} = \boldsymbol{A}^{-1}$. ♦

Estimation

2.1 Fisher-Scoring Algorithm

Definition 1.2.1. Define the following $n_k \times n_k$ matrices:

$$\boldsymbol{\Delta}_k = \left\{\frac{\partial g(\mu_{ki})}{\partial \mu_{ki}}\right\} = \text{diag}\{\dot{g}(\mu_{ki})\}, \qquad \boldsymbol{W}_k^{-1} = \boldsymbol{\Delta}_k \boldsymbol{V}_k \boldsymbol{\Delta}_k.$$

■

Proposition 1.2.1. Let $g(\mu_{ki}) = \eta_{ki} = \boldsymbol{x}'_{ki}\boldsymbol{\beta}$ and $\mu_{ki} = h(\eta_{ki})$ s.t. $h = g^{-1}$. The first derivatives of $g(\mu_{ki})$ and $h(\eta_{ki})$ are related via:

$$\dot{h}(\eta_{ki}) = \frac{1}{\dot{g}(\mu_{ki})}.$$

♦

Proposition 1.2.2. The matrix $\boldsymbol{D}_k$ is expressible as:

$$\boldsymbol{D}_k = \frac{\partial \boldsymbol{\mu}_k}{\partial \boldsymbol{\beta}'_k} = \boldsymbol{\Delta}_k^{-1}\boldsymbol{X}_k$$

♦

Proof.

$$\frac{\partial \mu_{ki}}{\partial \boldsymbol{\beta}'} = \frac{\partial}{\partial \boldsymbol{\beta}'} h(\eta_{ki}) = \dot{h}(\eta_{ki})\boldsymbol{x}'_{ki} = \frac{1}{\dot{g}(\mu_{ki})}\boldsymbol{x}'_{ki}$$

■

Corollary 1.2.1. The GEEs for $\boldsymbol{\beta}$ are expressible as:

$$\mathcal{U}_\beta = \sum_{k=1}^K \boldsymbol{X}'_k \boldsymbol{\Delta}_k^{-1} \boldsymbol{V}_k^{-1}(\boldsymbol{y}_k - \boldsymbol{\mu}_k).$$

♣

Proposition 1.2.3. Define the $n_k \times 1$ working vector $\boldsymbol{z}_k$:

$$\boldsymbol{z}_k = \boldsymbol{\eta}_k + \boldsymbol{\Delta}_k(\boldsymbol{y}_k - \boldsymbol{\mu}_k).$$

The iteratively reweighted least squares (IRLS) update for $\boldsymbol{\beta}$ is:

$$\boldsymbol{\beta}^{(r+1)} \leftarrow \left(\sum_{k=1}^K \boldsymbol{X}'_k \boldsymbol{W}_k^{(r)} \boldsymbol{X}_k\right)^{-1} \sum_{k=1}^K \boldsymbol{X}'_k \boldsymbol{W}_k^{(r)} \boldsymbol{z}_k^{(r)}.$$

This is WLS of the rth working vector $\boldsymbol{z}^{(r)}$ on $\boldsymbol{X}$ using working weights $\boldsymbol{W}^{(r)}$. ♦

Proof. Let $\boldsymbol{\beta}^{(r)}$ denote the current state of $\boldsymbol{\beta}$. Take the first order Taylor expansion of the score $\mathcal{U}(\boldsymbol{\beta})$ about $\boldsymbol{\beta}^{(r)}$:

$$\mathcal{U}(\boldsymbol{\beta}) \doteq \mathcal{U}(\boldsymbol{\beta}^{(r)}) + \dot{\mathcal{U}}(\boldsymbol{\beta}^{(r)})\big(\boldsymbol{\beta} - \boldsymbol{\beta}^{(r)}\big).$$

At the solution $\mathcal{U}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$. Make the approximation:

$$\mathbf{0} = \mathcal{U}(\hat{\boldsymbol{\beta}}) \approx \mathcal{U}(\boldsymbol{\beta}^{(r+1)}) \doteq \mathcal{U}(\boldsymbol{\beta}^{(r)}) + \dot{\mathcal{U}}(\boldsymbol{\beta}^{(r)})\big(\boldsymbol{\beta}^{(r+1)} - \boldsymbol{\beta}^{(r)}\big).$$

Solving for $\boldsymbol{\beta}^{(r+1)}$:

$$-\dot{\mathcal{U}}(\boldsymbol{\beta}^{(r)})\boldsymbol{\beta}^{(r+1)} = -\dot{\mathcal{U}}(\boldsymbol{\beta}^{(r)})\boldsymbol{\beta}^{(r)} + \mathcal{U}(\boldsymbol{\beta}^{(r)}).$$

Consider the gradient of $\mathcal{U}$ w.r.t. $\boldsymbol{\beta}$:

$$\frac{\partial \mathcal{U}}{\partial \boldsymbol{\beta}'} = \sum_{k=1}^{K} \left\{ \frac{\partial}{\partial \boldsymbol{\beta}'} \boldsymbol{D}_k' \boldsymbol{V}_k^{-1} \right\} (\boldsymbol{y}_k - \boldsymbol{\mu}_k) - \boldsymbol{D}_k' \boldsymbol{V}_k^{-1} \boldsymbol{D}_k.$$

The first term is eliminated upon taking the expectation:

$$E\left(\frac{\partial \mathcal{U}}{\partial \boldsymbol{\beta}'}\right) = E(\dot{\mathcal{U}}_\beta) = -\sum_{k=1}^{K} \boldsymbol{D}_k' \boldsymbol{V}_k^{-1} \boldsymbol{D}_k$$

As in Fisher scoring, adopt the following update equation for $\boldsymbol{\beta}^{(r)}$:

$$-E(\dot{\mathcal{U}}_\beta)\boldsymbol{\beta}^{(r+1)} = -E(\dot{\mathcal{U}}_\beta)\boldsymbol{\beta}^{(r)} + \mathcal{U}(\boldsymbol{\beta}^{(r)}),$$

$$\sum_{k=1}^{K} \boldsymbol{D}_k' \boldsymbol{V}_k^{-1} \boldsymbol{D}_k \boldsymbol{\beta}^{(r+1)} = \sum_{k=1}^{K} \boldsymbol{D}_k' \boldsymbol{V}_k^{-1} \boldsymbol{D}_k \boldsymbol{\beta}^{(r)} + \sum_{k=1}^{K} \boldsymbol{D}_k' \boldsymbol{V}_k^{-1} (\boldsymbol{y}_k - \boldsymbol{\mu}_k).$$

Re-expressing $\boldsymbol{D}_k' \boldsymbol{V}_k^{-1} \boldsymbol{D}_k$:

$$\boldsymbol{D}_k' \boldsymbol{V}_k^{-1} \boldsymbol{D}_k = \boldsymbol{X}_k' \boldsymbol{\Delta}_k^{-1} \boldsymbol{V}_k^{-1} \boldsymbol{\Delta}_k^{-1} \boldsymbol{X}_k = \boldsymbol{X}_k' \boldsymbol{W}_k \boldsymbol{X}_k$$

The system of equations becomes:

$$\left(\sum_{k=1}^{K} \boldsymbol{X}_k' \boldsymbol{W}_k \boldsymbol{X}_k\right) \boldsymbol{\beta}^{(r+1)} = \left(\sum_{k=1}^{K} \boldsymbol{X}_k' \boldsymbol{W}_k \boldsymbol{X}_k\right) \boldsymbol{\beta}^{(r)} + \sum_{k=1}^{K} \boldsymbol{X}_k' \boldsymbol{W}_k \boldsymbol{\Delta}_k (\boldsymbol{y}_k - \boldsymbol{\mu}_k).$$

Combining terms on the right:

$$\left(\sum_{k=1}^{K} \boldsymbol{X}_k' \boldsymbol{W}_k \boldsymbol{X}_k\right) \boldsymbol{\beta}^{(r+1)} = \sum_{k=1}^{K} \boldsymbol{X}_k' \boldsymbol{W}_k \big\{ \boldsymbol{X}_k \boldsymbol{\beta}^{(r)} + \boldsymbol{\Delta}_k (\boldsymbol{y}_k - \boldsymbol{\mu}_k) \big\}.$$

Identify $\boldsymbol{z}_k = \boldsymbol{X}_k \boldsymbol{\beta}^{(r)} + \boldsymbol{\Delta}_k (\boldsymbol{y}_k - \boldsymbol{\mu}_k)$. Rearranging gives the result. ■

2.2 Missingness

Remark 1.2.1. The GEE estimator $\boldsymbol{\beta}$ is consistent if data are missing completely at random. However, if the data are missing at random, consistency is not guaranteed. In this case, weighted estimating equations are needed, where the weighting is inversely proportional to the probability of missingness. ♦

Inference

3.1 Asymptotics

Result 1.3.1. Under the regularity conditions for M-estimation, $\hat{\boldsymbol{\beta}}$ solving $\mathcal{U}_{\beta} \overset{\text{Set}}{=} \mathbf{0}$ converges in distribution as:

$$\sqrt{K}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \boldsymbol{\Omega}), \tag{1.3.2}$$

where $\boldsymbol{\Omega} = \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-T}$, and:

$$\boldsymbol{A} = \underset{K\to\infty}{\text{plim}} \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{D}_k'\boldsymbol{V}_k^{-1}\boldsymbol{D}_k,$$

$$\boldsymbol{B} = \underset{K\to\infty}{\text{plim}} \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{D}_k'\boldsymbol{V}_k^{-1}(\boldsymbol{y}_k - \boldsymbol{\mu}_k)(\boldsymbol{y}_k - \boldsymbol{\mu}_k)'\boldsymbol{V}_k^{-1}\boldsymbol{D}_k.$$

♣

Proof. Take the Taylor expansion of $\boldsymbol{U}(\boldsymbol{\beta})$ about $\boldsymbol{\beta}_0$:

$$\mathcal{U}(\boldsymbol{\beta}) = \mathcal{U}(\boldsymbol{\beta}_0) + \dot{\mathcal{U}}(\boldsymbol{\beta}_0)(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \mathcal{O}_p(1),$$

where the remainder is bounded in probability by assumption.

Evaluating the Taylor expansion at $\hat{\boldsymbol{\beta}}$:

$$\mathbf{0} = \mathcal{U}(\hat{\boldsymbol{\beta}}) = \mathcal{U}(\boldsymbol{\beta}_0) + \dot{\mathcal{U}}(\boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \mathcal{O}_p(1).$$

Solving for $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$:

$$\sqrt{K}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \left\{ -K^{-1}\dot{\mathcal{U}}(\boldsymbol{\beta}_0)\right\}^{-1} K^{-1/2}\mathcal{U}(\boldsymbol{\beta}_0) + \mathcal{O}_p(K^{-1/2}).$$

The term under the inverse converges in probability as:

$$\begin{aligned} -K^{-1}\dot{\mathcal{U}}(\boldsymbol{\beta}_0) &= -\frac{1}{K}\frac{\partial \mathcal{U}}{\partial \boldsymbol{\beta}'}(\boldsymbol{\beta}_0) = -\frac{1}{K}\sum_{k=1}^{K}\left\{\frac{\partial}{\partial \boldsymbol{\beta}'}\boldsymbol{D}_k'\boldsymbol{V}_k^{-1}\right\}(\boldsymbol{y}_k - \boldsymbol{\mu}_k) + \boldsymbol{D}_k'\boldsymbol{V}_k^{-1}\boldsymbol{D}_k \\ &\xrightarrow{p} -\underset{K\to\infty}{\text{plim}} \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{D}_k'\boldsymbol{V}_k^{-1}\boldsymbol{D}_k = -\boldsymbol{A}. \end{aligned}$$

Thus, the influence function expansion for $\hat{\boldsymbol{\beta}}$ is:

$$\sqrt{K}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = -\frac{1}{\sqrt{K}}\boldsymbol{A}^{-1}\mathcal{U}(\boldsymbol{\beta}_0) + o_p(1).$$

By the central limit theorem:

$$K^{-1/2}\mathcal{U}(\boldsymbol{\beta}_0) = \frac{1}{\sqrt{K}}\sum_{k=1}^{K}\boldsymbol{D}_k'\boldsymbol{V}_k^{-1}(\boldsymbol{y}_k - \boldsymbol{\mu}_k) \xrightarrow{\mathcal{L}} N\big(\boldsymbol{0}, \boldsymbol{B}\big),$$

where:

$$\boldsymbol{B} \equiv \lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{D}_k'\boldsymbol{V}_k^{-1}E\big\{(\boldsymbol{y}_k - \boldsymbol{\mu}_k)(\boldsymbol{y}_k - \boldsymbol{\mu}_k)'\big\}\boldsymbol{V}_k^{-1}\boldsymbol{D}_k.$$

The conclusion follows from Slutsky's theorem. ■

Corollary 1.3.1. The influence function expansion for $\boldsymbol{\beta}$ is:

$$\sqrt{K}\big(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\big) = -\boldsymbol{A}^{-1}\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\boldsymbol{D}_k'\boldsymbol{V}_k^{-1}(\boldsymbol{y}_k - \boldsymbol{\mu}_k) + o_p(1).$$

♣

Remark 1.3.1. The empirical estimators of $\boldsymbol{A}$ and $\boldsymbol{B}$ are:

$$\hat{\boldsymbol{A}} = \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{D}_k'\boldsymbol{V}_k^{-1}\boldsymbol{D}_k$$

$$\hat{\boldsymbol{B}} = \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{D}_k'\boldsymbol{V}_k^{-1}(\boldsymbol{y}_k - \boldsymbol{\mu}_k)(\boldsymbol{y}_k - \boldsymbol{\mu}_k)'\boldsymbol{V}_k^{-1}\boldsymbol{D}_k.$$

♦

3.2 Hypothesis Testing

Proposition 1.3.1. Partition the regression coefficient $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)'$, where $\boldsymbol{\beta}_1$ is the parameter of interest, and $\boldsymbol{\beta}_2$ is a nusiance parameter. Let

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{12}' & \boldsymbol{\Omega}_{22} \end{pmatrix},$$

denote the corresponding partition of the asymptotic covariance matrix.

The Wald statistic for assessing $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{0}$ is:

$$T_W = K \cdot \hat{\boldsymbol{\beta}}_1'\boldsymbol{\Omega}_{11}^{-1}\hat{\boldsymbol{\beta}}_1 \xrightarrow{\mathcal{L}} \chi^2_{\dim(\boldsymbol{\beta}_1)}$$

♦

Proof. Follows from (1.3.2). ■

Proposition 1.3.2. Let $\tilde{\boldsymbol{\beta}}_2$ denote the estimate of the nuisance parameter $\boldsymbol{\beta}_2$ under $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$, i.e. a solution to the estimating equation:

$$\mathcal{U}_2(\boldsymbol{\beta}_1 = \mathbf{0}, \boldsymbol{\beta}_2) \overset{\text{Set}}{=} \mathbf{0}.$$

Define the score for $\boldsymbol{\beta}_1$:

$$\tilde{\mathcal{U}_1} = \mathcal{U}_1\big(\boldsymbol{\beta}_1 = \mathbf{0}, \boldsymbol{\beta}_2 = \tilde{\boldsymbol{\beta}}_2\big).$$

The score test of $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ is:

$$T_S = \frac{1}{K}\tilde{\mathcal{U}_1}'\boldsymbol{\Omega}_{11|2}^{-1}\tilde{\mathcal{U}_1},$$

where:

$$\boldsymbol{\Omega}_{11|2} = \boldsymbol{B}_{11} - \boldsymbol{B}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{B}_{21} + \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{B}_{22}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}.$$

♦

Proof. Partition the estimating equations $\mathcal{U} = (\mathcal{U}_1, \mathcal{U}_2)'$, where:

$$\mathcal{U}_l = \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{D}_{k,l}'\boldsymbol{V}_k^{-1}(\boldsymbol{y}_k - \boldsymbol{\mu}_k), \qquad \boldsymbol{D}_{l,k} = \frac{\partial\boldsymbol{\mu}_k}{\partial\boldsymbol{\beta}_l'}.$$

Similarly, partition $\boldsymbol{A}$ and $\boldsymbol{B}$ as:

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}, \qquad \boldsymbol{B} = \begin{pmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{pmatrix}.$$

Here, the component matrices are defined as:

$$\boldsymbol{A}_{l,l'} = \operatorname*{plim}_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{D}_{l,k}'\boldsymbol{V}_k^{-1}\boldsymbol{D}_{l',k},$$

$$\boldsymbol{B}_{l,l'} = \operatorname*{plim}_{K\to\infty}\frac{1}{K}\sum_{k=1}^{m}\boldsymbol{D}_{l,k}'\boldsymbol{V}_k^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_k)(\boldsymbol{y} - \boldsymbol{\mu}_k)'\boldsymbol{V}_k^{-1}\boldsymbol{D}_{l',k}.$$

Let $\tilde{\boldsymbol{\beta}}_2$ denote the solution to $\mathcal{U}_2 \overset{\text{Set}}{=} \mathbf{0}$ under $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$:

$$\mathbf{0} = \mathcal{U}_2(\boldsymbol{\beta}_1 = \mathbf{0}, \boldsymbol{\beta}_2 = \tilde{\boldsymbol{\beta}}_2).$$

Take the Taylor expansion of $\mathcal{U}_2$ about $(\boldsymbol{\beta}_1 = \mathbf{0}, \boldsymbol{\beta}_{2,0})$:

$$\mathcal{U}_2(\boldsymbol{\beta}_2) = \mathcal{U}_2(\boldsymbol{\beta}_{2,0}) + \left\{\frac{\partial\mathcal{U}_2}{\partial\boldsymbol{\beta}_2'}(\boldsymbol{\beta}_{2,0})\right\}(\boldsymbol{\beta}_2 - \boldsymbol{\beta}_{2,0}) + \mathcal{O}_p(1).$$

Evaluating at $\tilde{\boldsymbol{\beta}}_2$:

$$\mathbf{0} = \mathcal{U}_2(\tilde{\boldsymbol{\beta}}_{2,0}) = \mathcal{U}_2(\boldsymbol{\beta}_{2,0}) + \big\{\dot{\mathcal{U}}_2(\boldsymbol{\beta}_{2,0})\big\}(\tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{2,0}) + \mathcal{O}_p(1).$$

Re-arranging:

$$\sqrt{K}(\tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{2,0}) = \big\{-K^{-1}\dot{\mathcal{U}}_2(\boldsymbol{\beta}_{2,0})\big\}^{-1}K^{-1/2}\mathcal{U}_2(\boldsymbol{\beta}_{2,0}) + \mathcal{O}_p(K^{-1/2}). \tag{1.3.3}$$

The term under the inverse converges as:

$$\begin{aligned} -K^{-1}\dot{\mathcal{U}}_2(\boldsymbol{\beta}_{2,0}) &= -\frac{1}{K}\sum_{k=1}^{K}\left\{\frac{\partial}{\partial\boldsymbol{\beta}'}\boldsymbol{D}_{2,k}'\boldsymbol{V}_k^{-1}\right\}(\boldsymbol{y}_k - \boldsymbol{\mu}_k) + \boldsymbol{D}_{2,k}'\boldsymbol{V}_k^{-1}\boldsymbol{D}_{2,k} \\ &\xrightarrow{p} -\operatorname*{plim}_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{D}_{2,k}'\boldsymbol{V}_k^{-1}\boldsymbol{D}_{2,k} = \boldsymbol{A}_{22}. \end{aligned}$$

Re-expressing (1.3.3):

$$\sqrt{K}(\tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) = K^{-1/2}\boldsymbol{A}_{22}^{-1}\mathcal{U}_2(\boldsymbol{\beta}_2 = \boldsymbol{\beta}_{2,0}) + o_p(1) \tag{1.3.4}$$

Next, take the Taylor expansion of $\boldsymbol{U}_1$ about $(\boldsymbol{\beta}_1 = \mathbf{0}, \boldsymbol{\beta}_{2,0})$:

$$\mathcal{U}_1(\boldsymbol{\beta}_2) = \mathcal{U}_1(\boldsymbol{\beta}_{2,0}) + \left\{\frac{\partial\mathcal{U}_1}{\partial\boldsymbol{\beta}_2'}(\boldsymbol{\beta}_{2,0})\right\}(\boldsymbol{\beta}_2 - \boldsymbol{\beta}_{2,0}) + \mathcal{O}_p(1).$$

Evaluating at $\tilde{\boldsymbol{\beta}}_2$:

$$\mathcal{U}_1(\tilde{\boldsymbol{\beta}}_2) = \mathcal{U}_1(\boldsymbol{\beta}_{2,0}) + \big\{\dot{\mathcal{U}}_1(\boldsymbol{\beta}_{2,0})\big\}(\tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{2,0}) + \mathcal{O}_p(1).$$

Re-scaling:

$$K^{-1/2}\mathcal{U}_1(\tilde{\boldsymbol{\beta}}_2) = K^{-1/2}\mathcal{U}_1(\boldsymbol{\beta}_{2,0}) + \big\{K^{-1}\dot{\mathcal{U}}_1(\boldsymbol{\beta}_{2,0})\big\}K^{1/2}(\tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{2,0}) + \mathcal{O}_p(K^{-1/2}).$$

By the law of large numbers $K^{-1}\dot{\mathcal{U}}_1(\boldsymbol{\beta}_{2,0}) = \boldsymbol{A}_{12} + o_p(1)$, thus:

$$K^{-1/2}\mathcal{U}_1(\tilde{\boldsymbol{\beta}}_2) = K^{-1/2}\mathcal{U}_1(\boldsymbol{\beta}_{2,0}) - \boldsymbol{A}_{12}K^{1/2}(\tilde{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{2,0}) + \mathcal{O}_p(K^{-1/2}). \tag{1.3.5}$$

Substituting (1.3.4) into (1.3.5):

$$K^{-1/2}\mathcal{U}_1(\tilde{\boldsymbol{\beta}}_2) = K^{-1/2}\mathcal{U}_1(\boldsymbol{\beta}_{2,0}) - K^{-1/2}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\mathcal{U}_2(\boldsymbol{\beta}_{2,0}) + o_p(1). \tag{1.3.6}$$

Re-expressing the Taylor expansion of $\mathcal{U}_1$ in terms of the total score $\mathcal{U} = (\mathcal{U}_1, \mathcal{U}_2)'$:

$$\frac{1}{\sqrt{K}}\mathcal{U}_1(\tilde{\boldsymbol{\beta}}_2) = \Big(\boldsymbol{I}, -\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\Big)\cdot\frac{1}{\sqrt{K}}\begin{pmatrix}\mathcal{U}_{1,0}\\ \mathcal{U}_{2,0}\end{pmatrix} + o_p(1).$$

Since by the central limit theorem:

$$\frac{1}{\sqrt{K}}\begin{pmatrix}\mathcal{U}_{1,0}\\ \mathcal{U}_{2,0}\end{pmatrix} \xrightarrow{\mathcal{L}} N\big(\mathbf{0}, \boldsymbol{B}\big),$$

by Slutsky's theorem:

$$\frac{1}{\sqrt{K}}\mathcal{U}_1(\tilde{\boldsymbol{\beta}}_2) = \xrightarrow{\mathcal{L}} N(\mathbf{0}, \boldsymbol{\Omega}_{11|2}),$$

where the limiting variance is:

$$\begin{aligned}\boldsymbol{\Omega}_{11|2} &\equiv \Big(\boldsymbol{I}, -\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\Big)\left(\begin{array}{cc}\boldsymbol{B}_{11} & \boldsymbol{B}_{12}\\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22}\end{array}\right)\begin{pmatrix}\boldsymbol{I}\\ -\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\end{pmatrix}\\ &= \boldsymbol{B}_{11} - \boldsymbol{B}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{B}_{21} + \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{B}_{22}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\end{aligned}$$

Observe that if $\boldsymbol{A} = \boldsymbol{B}$, the limiting variance reduces to:

$$\boldsymbol{\Omega}_{11|2} = \boldsymbol{B}_{11} - \boldsymbol{B}_{12}\boldsymbol{B}_{22}^{-1}\boldsymbol{B}_{21},$$

where is the standard form of the efficient information. ■