Power and Sample Size Calculations for Generalized Estimating Equations via Local Asymptotics

Zhigang Li[1] and Ian W. McKeague[2]

[1]*Dartmouth College and* [2]*Columbia University*

Supplementary Material

S1 Regularity conditions

(C1) The parameter space $\mathbb{B}$ is compact and β_0 belongs to its interior.
(C2) There exists $\delta > 1$ such that

$$c_0 \equiv \sup_{m\geq 1} \max_{i=1,\ldots,m} E_{\beta_m}|h_i(\mathbf{y}_i)|^{1+\delta} < \infty \ \text{ and } \ c_1 \equiv \sup_{m\geq 1} \max_{i=1,\ldots,m} E_{\beta_m}|\mathbf{y}_i|^{\delta} < \infty,$$

where $h_i(\mathbf{y}_i) \equiv \sup_{b\in\mathbb{B}} |\Psi_i(\mathbf{y}_i, b)|$.
(C3) For any bounded sequence $\{y_i\}$ with $y_i \in \mathbb{R}^{n_i}$, the functions $b \mapsto \Psi_i(y_i, b)$ are equicontinuous and uniformly bounded on $\mathbb{B}$.
(C4) $\sup_{i\geq 1} |E_{b_1}\{\Psi_i(\mathbf{y}_i, b)\} - E_{b_1^*}\{\Psi_i(\mathbf{y}_i, b^*)\}| \lesssim |b_1 - b_1^*| + |b - b^*|$ for all $b_1, b_1^*, b, b^* \in \mathbb{B}$, where "$\lesssim$" means "smaller than up to a constant".
(C5) For all $\epsilon > 0$, $\inf_{m\geq 1, |b-\beta_0|>\epsilon} |f_{m,0}(b)| > 0 = |f_{m,0}(\beta_0)|$, where $f_{m,0}(b) = E_{\beta_0}\{m^{-1}s_m(b)\}$.
(C6) $\varphi_i(y, b) = \nabla_b \Psi_i(y, b)$ exists for all $y \in \mathbb{R}^{n_i}, b \in \mathbb{B}$, and its jth row (denoted by $\varphi_{ij}(y, b)$ for later use) satisfies conditions (C1)–(C4) in place of $\Psi_i(y, b)$, for each $j = 1, \ldots, k$.
(C7) There exists a neighborhood N of β_0 such that $\sup_{i\geq 1, j=1,\ldots,k} |V_{ij}(b) - V_{ij}(\beta_0)| \lesssim |b - \beta_0|$ for all $b \in N$, where $V_{ij}(b)$ is the jth column of $\mathrm{Var}_b(\Psi_i(\mathbf{y}_i, \beta_0))$.
(C8) The elements of $\frac{1}{m}M_m(\beta_0)$ and $\frac{1}{m}\sum_{i=1}^m \mathrm{Var}_{\beta_0}(\Psi_i(\mathbf{y}_i, \beta_0))$ converge to finite limits, where $M_m(\beta) = -E_{\beta_0}\{\nabla_\beta s_m(\beta)\}$.
(C9) $M(\beta_0) = \lim_{m\to\infty}\{m^{-1}M_m(\beta_0)\}$ is non-singular.

S2 Detailed formulae under marginal models

To obtain $\tilde{\nu}_m$, it is straightforward that we only need to give a derivation of (3.4) and (3.5) from (3.2) and (3.3), for which we need to calculate $M_m(\beta_0)$, $E_{\beta_m}(s_m(\beta_0))$ and $\mathrm{var}_{\beta_m}(s_m(\beta_0))$, with each mean and covariance understood to be conditioned on the covariates (which are suppressed in the notation) and cluster size. Under this setting, we have $\beta_m = (\phi_0, \alpha_0^T, \kappa_0^T, \psi_{1m}^T)^T$ and $\beta_0 = (\phi_0, \alpha_0^T, \kappa_0^T, \psi_0^T)^T$. As discussed at the end of Section 2.1, a combined estimating equation can be used to estimate β (see Fitzmaurice et al., 2009, Chap 3):

$$s_m(\beta) = \sum_{i=1}^{m} (\tilde{U}_i^T, u_i^T, U_i^T)^T = 0,$$

with $\tilde{U}_i$ and u_i defined as follows:

$$\tilde{U}_i(\theta, \phi) = \mathbf{1}_{n_i \times 1}^T \{\tilde{\mathtt{W}}_i(\theta) - \mathbf{1}_{n_i \times 1}\phi\}$$

and

$$u_i(\theta, \alpha, \phi) = E_i^T \{\mathtt{W}_i(\theta, \phi) - \rho_i(\alpha)\},$$

where $\tilde{\mathtt{W}}_i$ is the n_i-dimensional vector with jth element $\tilde{\mathtt{W}}_{ij} = (\mathbf{y}_{ij} - \mu_{ij})^2/\upsilon(\mu_{ij})$, $E_i = \partial\rho_i(\alpha)/\partial\alpha$, and $\rho_i(\alpha)$ is the vector [of dimension $n_i(n_i-1)/2$] consisting of the upper-triangular entries of $R_i(\alpha)$ in lexicographic order and $\mathtt{W}_i$ is defined in the same way as $\rho_i(\alpha)$ except from the matrix with jkth entry

$$\mathtt{W}_{ijk} = \frac{(\mathbf{y}_{ij} - \mu_{ij})(\mathbf{y}_{ik} - \mu_{ik})}{\phi\{\upsilon(\mu_{ij})\upsilon(\mu_{ik})\}^{1/2}}.$$

Thus, in the general setting described in Section 2.2, we have $\Psi_i(\mathbf{y}_i) = (\tilde{U}_i^T, u_i^T, U_i^T)^T$, where $U_i = D_i^T V_i(\mathbf{y}_i - \mu_i(\theta))$, $D_i = \partial\mu_i/\partial\theta$, and $\theta = (\kappa^T, \psi^T)^T$. It can now be shown that $M_m(\beta_0)$ has the form

$$M_m(\beta_0) = -\sum_{i=1}^{m} E_{\beta_0} \begin{pmatrix} \frac{\partial \tilde{U}_i}{\partial\phi} & \frac{\partial \tilde{U}_i}{\partial\alpha} & \frac{\partial \tilde{U}_i}{\partial\theta} \\ \frac{\partial u_i}{\partial\phi} & \frac{\partial u_i}{\partial\alpha} & \frac{\partial u_i}{\partial\theta} \\ \frac{\partial U_i}{\partial\phi} & \frac{\partial U_i}{\partial\alpha} & \frac{\partial U_i}{\partial\theta} \end{pmatrix} = \begin{pmatrix} F & G \\ 0 & H \end{pmatrix},$$

where, writing $W_i = D_i^T V_i^{-1}$ and denoting the jth row of W_i by W_{ij},

$$\begin{aligned} H &= -\sum_{i=1}^{m} E_{\beta_0}(\partial U_i/\partial\theta) \\ &= -\sum_{i=1}^{m} E_{\beta_0}[\partial\{W_i(\mathbf{y}_i - \mu_i)\}/\partial\theta] \\ &= -\sum_{i=1}^{m} \begin{pmatrix} E_{\beta_0}[\partial\{W_{i1}(\mathbf{y}_i - \mu_i)\}/\partial\theta] \\ \vdots \\ E_{\beta_0}[\partial\{W_{i,p+q}(\mathbf{y}_i - \mu_i)\}/\partial\theta] \end{pmatrix} \\ &= -\sum_{i=1}^{m} \begin{pmatrix} E_{\beta_0}\left\{(\mathbf{y}_i - \mu_i)^T \frac{\partial W_{i1}^T}{\partial\theta}\right\} - W_{i1}\frac{\partial\mu_i}{\partial\theta} \\ \vdots \\ E_{\beta_0}\left\{(\mathbf{y}_i - \mu_i)^T \frac{\partial W_{i,p+q}^T}{\partial\theta}\right\} - W_{i,p+q}\frac{\partial\mu_i}{\partial\theta} \end{pmatrix} \\ &= \sum_{i=1}^{m} D_i^T V_i^{-1} D_i. \end{aligned}$$

The zero submatrix at the bottom left-hand corner of $M_m(\beta_0)$ is due to the fact that $\partial U_i/\partial\phi$ and $\partial U_i/\partial\alpha$ are linear functions of $\mathbf{y}_i - \mu_i(\theta_0)$, which has zero expectation when $\beta = \beta_0$. Then

$$\{M_m(\beta_0)\}^{-1} = \begin{pmatrix} F^{-1} & -F^{-1}GH^{-1} \\ 0 & H^{-1} \end{pmatrix}$$

and substituting into (2), we obtain

$$\begin{aligned} \xi_{m\psi} &= B\{M_m(\beta_0)\}^{-1}E_{\beta_m}\{s_m(\beta_0)\} \\ &= \begin{pmatrix} 0_{p\times(k-q-p)} & \bar{B} \end{pmatrix} \begin{pmatrix} F^{-1} & -F^{-1}GH^{-1} \\ 0 & H^{-1} \end{pmatrix} E_{\beta_m}\{s_m(\beta_0)\} \\ &= \bar{B}H^{-1}\sum_{i=1}^{m} E_{\beta_m}(U_i) \\ &= \bar{B}A_m^{-1}\left\{\frac{1}{m}\sum_{i=1}^{m} D_i^T V_i^{-1}(\mu_i(\theta_m) - \mu_i(\theta_0))\right\}, \end{aligned}$$

where $A_m = \frac{1}{m}\sum_{i=1}^{m} D_i^T V_i^{-1} D_i$, $\bar{B} = (0_{p\times q}, I_p)$ and we used $E_{\beta_m}(U_i) = D_i^T V_i^{-1}(\mu_i(\theta_m) - \mu_i(\theta_0))$ in the last step. Here A_m, D_i and V_i are evaluated under H_0. Similarly, we have

$$\Sigma_{m\psi} = \bar{B}A_m^{-1}\left\{\frac{1}{m}\sum_{i=1}^{m} D_i^T V_i^{-1}\mathrm{Var}_{\beta_m}(\mathbf{y}_i|\mathbf{z}_i,\mathbf{x}_i)V_i^{-1}D_i\right\}A_m^{-1}\bar{B}^T$$

by noticing that $\mathrm{var}_{\beta_m}(U_i) = D_i^T V_i^{-1}\mathrm{var}_{\beta_m}(\mathbf{y}_i|\mathbf{z}_i,\mathbf{x}_i)V_i^{-1}D_i$. Again, A_m, D_i and V_i are evaluated under H_0. Now we replace h in the expressions of $\xi_{m\psi}$ and $\Sigma_{m\psi}$ by $\sqrt{m}(\psi_A - \psi_0)$. They become

$$\xi_{m\psi} = \bar{B}A_m^{-1}\left\{\frac{1}{m}\sum_{i=1}^{m} D_i^T V_i^{-1}(\mu_i(\theta_A) - \mu_i(\theta_0))\right\}$$

and

$$\Sigma_{m\psi} = \bar{B}A_m^{-1}\left\{\frac{1}{m}\sum_{i=1}^{m} D_i^T V_i^{-1}\mathrm{Var}_{\beta_A}(\mathbf{y}_i|\mathbf{z}_i,\mathbf{x}_i)V_i^{-1}D_i\right\}A_m^{-1}\bar{B}^T.$$

Then (3.4) and (3.5) are obtained by replacing A_m and the terms in curly brackets by their expectations under the joint distribution of the covariates.

S3 Approach of Liu and Liang (1997)

LL's approach involes a multivariate extension of results of Self and Mauritsen (1988), who derived sample size and power formulae for generalized linear models based on the score statistic. They developed their results in the marginal model setting (as in Section 2.2), assuming discrete covariates distributed as

$$P(\mathbf{x}_i = \mathbf{u}_l, \mathbf{z}_i = \mathbf{w}_l) = \omega_l, \quad l = 1, \ldots, L, \tag{S3.1}$$

where $\{(\mathbf{u}_l, \mathbf{w}_l),\ l = 1, \dots, L\}$ are the L different possible values of covariates. LL also assumed that the structure of the true conditional correlation matrix of the outcome is known. The quasi-score test statistic T_m is given by

$$T_m = S_{m\psi}^T(\tilde{\kappa}_0, \psi_0)\hat{\Sigma}_m^{-1} S_{m\psi}(\tilde{\kappa}_0, \psi_0),$$

where $S_{m\psi}(\tilde{\kappa}_0, \psi_0) = \sum_{i=1}^m (\frac{\partial \mu_i}{\partial \psi})^T V_i^{-1}(\mathbf{y}_i - \mu_i)|_{\kappa=\hat{\kappa}_0, \psi=\psi_0}$, $\hat{\Sigma}_m = \mathrm{cov}_{\mathrm{H}_0}\{S_{m\psi}(\tilde{\kappa}_0, \psi_0)\}$ and $\tilde{\kappa}_0$ is an estimator of κ obtained by solving

$$S_{m\kappa}(\kappa, \psi_0) = \sum_{i=1}^m \left(\frac{\partial \mu_i}{\partial \kappa}\right)^T V_i^{-1}(\mathbf{y}_i - \mu_i)|_{\psi=\psi_0} = 0. \tag{S3.2}$$

Under $\psi = \psi_A$ and $\kappa = \kappa_0$, note that $\tilde{\kappa}_0$ is generally not a consistent estimator of κ_0 and it will converge to some value κ_0^*, namely the solution of

$$\mathrm{lim}_{m\to\infty} m^{-1}\mathrm{E}\{S_{m\kappa}(\kappa_0^*, \psi_0); \kappa_0, \psi_A\} = 0. \tag{S3.3}$$

LL used standard Taylor series arguments to obtain an approximation to $S_{m\psi}(\tilde{\kappa}_0, \psi_0)$. This approximation is $G(\kappa_0^*, \psi_0) = S_{m\psi}(\kappa_0^*, \psi_0) - I_{\psi\kappa}^*(I_{\lambda\kappa}^*)^{-1} S_{m\kappa}(\kappa_0^*, \psi_0)$, where

$$I_{\psi\kappa}^* = \sum_i \left(\frac{\partial \mu_i}{\partial \psi}\right)^T V_i^{-1} \left(\frac{\partial \mu_i}{\partial \kappa}\right)|_{\kappa=\kappa_0^*, \psi=\psi_0}, \tag{S3.4}$$

and

$$I_{\kappa\kappa}^* = \sum_i \left(\frac{\partial \mu_i}{\partial \kappa}\right)^T V_i^{-1} \left(\frac{\partial \mu_i}{\partial \kappa}\right)|_{\kappa=\kappa_0^*, \psi=\psi_0}. \tag{S3.5}$$

Let

$$\mu_{ij}^1 = g^{-1}(\kappa_0 + x_{ij}\psi_A), \tag{S3.6}$$

$$\mu_{ij}^* = g^{-1}(\kappa_0^* + x_{ij}\psi_0), \tag{S3.7}$$

$$\mu_i^1 = (\mu_{i1}^1, \mu_{i1}^1, \dots, \mu_{in}^1)^T, \tag{S3.8}$$

$$\mu_i^* = (\mu_{i1}^*, \mu_{i1}^*, \dots, \mu_{in}^*)^T, \tag{S3.9}$$

$$\mathbf{P}_i^* = \left\{ \left(\frac{\partial \mu_i}{\partial \psi}\right)^T - I_{\psi\kappa}^* I_{\lambda\kappa}^{*-1} \left(\frac{\partial \mu_i}{\partial \kappa}\right)^T \right\}|_{\kappa=\kappa_0^*, \psi=\psi_0}, \tag{S3.10}$$

and $V_i^* = V_i|_{\kappa=\kappa_0^*, \psi=\psi_0}$.

Notice that $G(\kappa_0^*, \psi_0) = \sum_i \mathbf{P}_i^*(V_i^*)^{-1}(\mathbf{y}_i - \mu_i^*)$. Under the allocation scheme specified in (S3.1), and $\psi = \psi_A$ and $\kappa = \kappa_0$, LL claim that $G(\kappa_0^*, \psi_0)$ is approximately normal with mean and variance given by $m\tilde{\xi}$ and $m\tilde{\Sigma}$, respectively, where

$$\tilde{\xi} = \sum_{l=1}^{L} \omega_l \mathbf{P}_l^*(V_l^*)^{-1}(\mu_l^1 - \mu_l^*) \tag{S3.11}$$

and

$$\tilde{\Sigma} = \sum_{l=1}^{L} \omega_l \mathbf{P}_l^*(V_l^*)^{-1}\text{cov}(\mathbf{y}_l; \kappa_0, \psi_A)(V_l^*)^{-1}\mathbf{P}_l^{*T}. \tag{S3.12}$$

Here μ_l^1, μ_l^*, $\mathbf{P}_l^*$ and V_l^* are defined as in (S3.8), (S3.9), (S3.10) and V_i^* with $(\mathbf{x}_i, \mathbf{z}_i) = (\mathbf{u}_l, \mathbf{w}_l)$, and $\text{cov}(\mathbf{y}_l; \kappa_0, \psi_A)$ equals $\text{cov}(\mathbf{y}_i; \kappa_0, \psi_A)$ with $(\mathbf{x}_i, \mathbf{z}_i) = (\mathbf{u}_l, \mathbf{w}_l)$. So the distribution of $S_{m\psi}(\tilde{\kappa}_0, \psi_0)$ is approximated by $N(m\tilde{\xi}, m\tilde{\Sigma})$, which implies that the distribution of T_m is approximated by a noncentral chi-square distribution with p degrees of freedom since T_m has p dimensions. The non-centrality parameter is given by

$$\tilde{\nu}_m = (m\tilde{\xi})^T(m\tilde{\Sigma})^{-1}(m\tilde{\xi}). \tag{S3.13}$$

The sample size m for achieving nominal power $1 - \eta$ at significance level ζ is obtained by solving the equation $\tilde{\nu}_m = \tilde{\nu}$, which implies that

$$m = \frac{\tilde{\nu}}{\tilde{\xi}^T\tilde{\Sigma}^{-1}\tilde{\xi}}. \tag{S3.14}$$

Under the allocation scheme (S3.1), equation (S3.3) becomes

$$\sum_{l=1}^{L} \omega_l \left(\frac{\partial \mu_l^*}{\partial \kappa}\right)^T (V_l^*)^{-1}(\mu_l^1 - \mu_l^*) = 0, \tag{S3.15}$$

where $\frac{\partial \mu_l^*}{\partial \kappa} = \frac{\partial \mu_l}{\partial \kappa}|_{\kappa=\kappa_0^*, \psi=\psi_0}$, and μ_l is defined to be the mean vector μ_i with $(\mathbf{x}_i, \mathbf{z}_i) = (\mathbf{u}_l, \mathbf{w}_l)$. This equation corresponds to equation (12) in LL, which needs to be solved to derive explicit formulae for $\tilde{\xi}$ and $\tilde{\Sigma}$.

Two potential problems of this method were pointed out by Self and Mauritsen (1988):

1. Since $\tilde{\kappa}_0$ is not a consistent estimator of κ under the alternative hypothesis, $\tilde{\Sigma}$ is not the asymptotic variance of $\frac{1}{\sqrt{m}} S_{m\psi}(\tilde{\kappa}_0, \psi_0)$. Thus the condition needed for the distributional result mentioned above does not hold, even asymptotically.

2. Even though the distribution of $\frac{1}{\sqrt{m}} S_{m\psi}(\tilde{\kappa}_0, \psi_0)$ approaches multivariate normality, the expected value of $S_{m\psi}(\tilde{\kappa}_0, \psi_0)$ is simultaneously going to infinity. Therefore, the result relies on the quality of the chi-square approximation at a sequence of points that move progressively farther out in the tail of the distribution as m becomes large.

The first problem is caused by inconsistency of the estimator under fixed alternatives. The second problem is caused by the test statistic converging in probability to a degenerate distribution (infinity) under fixed alternatives. Our approach using local asymptotic theory succeeds in overcoming both of these problems.

S4 Approach of Shih (1997)

Shih considered the case $p = 1$, with the working covariance identical to the true covariance, and approximated the distribution of W_m under the fixed alternative $\psi = \psi_A$ by a noncentral χ_1^2 with non-centrality parameter $\tilde{\nu}_m = m\psi_A^2/\mathrm{v}$, where v is the asymptotic variance of $\hat{\psi}$ (cf. Remark 2 in Section A) when $\psi = \psi_A$. This is similar to the approach discussed in Remark 4 of Section 3.2, where the asymptotic power function is used, except that the variance now depends on the value of parameter under a fixed alternative. In this approach, the sample size for achieving nominal power $1 - \eta$ at significance level ζ based on a two-sided test is simply given by $m = \mathrm{v}(z_{1-\zeta/2} + z_{1-\eta})^2/(\psi_A - \psi_0)^2$.

S5 Two lemmas

The following lemmas are crucial for proving our main results.

Lemma 1 *Let* $H_m(b) = m^{-1}\sum_{i=1}^m \{\Psi_i(\mathbf{y}_i, b) - E_{\beta_m}(\Psi_i(\mathbf{y}_i, b))\}$. *Under conditions* (C1)–(C3)*,*

$$\sup_{b\in\mathbb{B}} |H_m(b)| \overset{P_m}{\to} 0.$$

This lemma is a routine extension of a uniform law of large numbers of Shao (Lemma 5.3) to a triangular array. The result is similar to a Glivenko–Cantelli theorem in that the convergence holds uniformly over a class of functions. The relative compactness condition (C3) plays a crucial role in the proof. The proof is similar to that for Lemma 5.3 in Shao.

Proof. Since we only need to consider components of Ψ_i's, without loss of generality we can assume that Ψ_i's are functions from $\mathbb{R}^{n_i} \times \mathbb{B}$ to $\mathbb{R}$. For any fixed $\epsilon > 0$ and any fixed subset $O \subset \mathbb{B}$,

$$P_m\left(\sup_{b\in O} |H_m(b)| > \epsilon\right) \leq P_m\left(\sup_{b\in O} H_m(b) > \epsilon\right) + P_m\left(\inf_{b\in O} H_m(b) < -\epsilon\right). \tag{S5.1}$$

We will show the first term on the right hand side converges to zero; the second term converges to zero by a similar argument. Clearly,

$$\begin{aligned} P_m\left(\sup_{b\in O} H_m(b) > \epsilon\right) &\leq P_m\left[m^{-1}\sum_{i=1}^m \left\{\sup_{b\in O} \Psi_i(\mathbf{y}_i, b) - E_{\beta_m}(\inf_{b\in O} \Psi_i(\mathbf{y}_i, b))\right\} > \epsilon\right] \\ &= P_m\left[m^{-1}\sum_{i=1}^m \Psi_i^{(m)} + m^{-1}\sum_{i=1}^m E_{\beta_m}\left\{\sup_{b\in O} \Psi_i(\mathbf{y}_i, b) - \inf_{b\in O} \Psi_i(\mathbf{y}_i, b)\right\} > \epsilon\right], \end{aligned} \tag{S5.2}$$

where $\Psi_i^{(m)} = \sup_{b\in O} \Psi_i(\mathbf{y}_i, b) - E_{\beta_m}(\sup_{b\in O} \Psi_i(\mathbf{y}_i, b))$. Since

$$\sup_{m\geq 1} \max_{i=1,\ldots,m} E_{\beta_m} |\sup_{b\in O} \Psi_i(\mathbf{y}_i, b)|^{1+\delta} \leq \sup_{m\geq 1} \max_{i=1,\ldots,m} E_{\beta_m} |h_i(\mathbf{y}_i)|^{1+\delta} < c_0,$$

where $h_i(\mathbf{y}_i)$ is defined in condition (C2), then $m^{-1}\sum_{i=1}^m \Psi_i^{(m)} = o_{P_m}(1)$ by Lemma 2. If we show that

$$m^{-1}\sum_{i=1}^m E_{\beta_m}\left\{\sup_{b\in O}\Psi_i(\mathbf{y}_i,b) - \inf_{b\in O}\Psi_i(\mathbf{y}_i,b)\right\} \quad < \quad \tilde{\epsilon},$$

for all $m \geq 1$ where $0 < \tilde{\epsilon} < \epsilon$, then (S5.2) converges to zero. Next we show that the above equation holds when the subset O is sufficiently small.

Using the Hölder and Markov inequalities, and condition (C2), for any $c > 0$

$$\begin{aligned}
& E_{\beta_m}\left\{m^{-1}\sum_{i=1}^m h_i(\mathbf{y}_i) I_{(c,\infty)}(|\mathbf{y}_i|)\right\} \\
& \leq \max_{i=1,\ldots,m} E_{\beta_m}\left\{h_i(\mathbf{y}_i) I_{(c,\infty)}(|\mathbf{y}_i|)\right\} \\
& \leq \max_{i=1,\ldots,m} \left\{E_{\beta_m}|h_i(\mathbf{y}_i)|^{1+\delta}\right\}^{1/(1+\delta)} \left\{P_m(|\mathbf{y}_i| > c)\right\}^{\delta/(1+\delta)} \\
& \leq \max_{i=1,\ldots,m} \left\{E_{\beta_m}|h_i(\mathbf{y}_i)|^{1+\delta}\right\}^{1/(1+\delta)} \left\{\frac{E_{\beta_m}|\mathbf{y}_i|^{\delta}}{c^{\delta}}\right\}^{\delta/(1+\delta)} \\
& = \left\{\max_{i=1,\ldots,m} E_{\beta_m}|h_i(\mathbf{y}_i)|^{1+\delta}\right\}^{1/(1+\delta)} \left\{\max_{i=1,\ldots,m} E_{\beta_m}|\mathbf{y}_i|^{\delta}\right\}^{\delta/(1+\delta)} c^{-\delta^2/(1+\delta)} \\
& \leq c_0^{1/(1+\delta)} c_1^{\delta/(1+\delta)} c^{-\delta^2/(1+\delta)}
\end{aligned}$$

for all $m \geq 1$. Thus for any $\epsilon > \tilde{\epsilon} > \epsilon/2$, there exists $c > 0$ such that

$$E_{\beta_m}\left\{\frac{1}{m}\sum_{i=1}^m h_i(\mathbf{y}_i) I_{(c,\infty)}(|\mathbf{y}_i|)\right\} < \tilde{\epsilon}/2 - \epsilon/4$$

for all $m \geq 1$. For this value of c,

$$\begin{aligned}
& E_{\beta_m}\left[m^{-1}\sum_{i=1}^m \left\{\sup_{b\in O}\Psi_i(\mathbf{y}_i,b) - \inf_{b\in O}\Psi_i(\mathbf{y}_i,b)\right\} I_{(c,\infty)}(|\mathbf{y}_i|)\right] \\
& \leq E_{\beta_m}\left[m^{-1}\sum_{i=1}^m \left\{\sup_{b\in\mathbb{B}}\Psi_i(\mathbf{y}_i,b) - \inf_{b\in\mathbb{B}}\Psi_i(\mathbf{y}_i,b)\right\} I_{(c,\infty)}(|\mathbf{y}_i|)\right] \\
& \leq E_{\beta_m}\left\{m^{-1}\sum_{i=1}^m |\sup_{b\in\mathbb{B}}\Psi_i(\mathbf{y}_i,b)| I_{(c,\infty)}(|\mathbf{y}_i|) + m^{-1}\sum_{i=1}^m |\inf_{b\in\mathbb{B}}\Psi_i(\mathbf{y}_i,b)| I_{(c,\infty)}(|\mathbf{y}_i|)\right\} \\
& \leq E_{\beta_m}\left\{m^{-1}\sum_{i=1}^m h_i(\mathbf{y}_i) I_{(c,\infty)}(|\mathbf{y}_i|) + m^{-1}\sum_{i=1}^m h_i(\mathbf{y}_i) I_{(c,\infty)}(|\mathbf{y}_i|)\right\} \\
& = 2E_{\beta_m}\left\{m^{-1}\sum_{i=1}^m h_i(\mathbf{y}_i) I_{(c,\infty)}(|\mathbf{y}_i|)\right\} \\
& < \tilde{\epsilon} - \epsilon/2
\end{aligned} \qquad \text{(S5.3)}$$

for all $m \geq 1$. By the equicontinuity of $\{\Psi_i(\mathbf{y}_i, b)\}$ in condition (C3), there exists a $\delta_\epsilon > 0$ such that

$$m^{-1}\sum_{i=1}^{m}\left\{\sup_{b\in O_\epsilon}\Psi_i(\mathbf{y}_i,b) - \inf_{b\in O_\epsilon}\Psi_i(\mathbf{y}_i,b)\right\} I_{[0,c]}(|\mathbf{y}_i|) < \epsilon/2$$

for all $m \geq 1$, where O_ϵ is any subset of $\mathbb{B}$ with $\mathrm{diam}(O_\epsilon) < \delta_\epsilon$. Here $\mathrm{diam}(O_\epsilon)$ is defined as the supremum of the distances between pairs of points in O_ϵ. The inequality (S5.3) holds with O replaced by O_ϵ which together with the above inequality implies

$$\begin{aligned}
& m^{-1}\sum_{i=1}^{m} E_{\beta_m}\left\{\sup_{b\in O_\epsilon}\Psi_i(\mathbf{y}_i,b) - \inf_{b\in O_\epsilon}\Psi_i(\mathbf{y}_i,b)\right\} \\
&= E_{\beta_m}\left[m^{-1}\sum_{i=1}^{m}\left\{\sup_{b\in O}\Psi_i(\mathbf{y}_i,b) - \inf_{b\in O}\Psi_i(\mathbf{y}_i,b)\right\} I_{(c,\infty)}(|\mathbf{y}_i|)\right] \\
&\quad + E_{\beta_m}\left[m^{-1}\sum_{i=1}^{m}\left\{\sup_{b\in O_\epsilon}\Psi_i(\mathbf{y}_i,b) - \inf_{b\in O_\epsilon}\Psi_i(\mathbf{y}_i,b)\right\} I_{[0,c]}(|\mathbf{y}_i|)\right] \\
&< \tilde{\epsilon} - \epsilon/2 + \epsilon/2 = \tilde{\epsilon} < \epsilon
\end{aligned}$$

for all $m \geq 1$. The right hand side of (S5.2) with O replaced by O_ϵ converges to zero since it is bounded above by

$$P_m\left(m^{-1}\sum_{i=1}^{m}\Psi_i^m > \epsilon - \tilde{\epsilon}\right) \to 0,$$

and we conclude that $P_m\left(\sup_{b\in O_\epsilon} H_m(b) > \epsilon\right) \to 0$. By a similar argument, $P_m\left(\inf_{b\in O_\epsilon} H_m(b) < -\epsilon\right) \to 0$. Thus by (S5.1), we have

$$P_m\left(\sup_{b\in O_\epsilon}|H_m(b)| > \epsilon\right) \to 0. \tag{S5.4}$$

Due to the compactness of $\mathbb{B}$, there exist finitely many open balls $\{O_\epsilon^j\}_{j=1,\dots,n_\epsilon}$ with $\mathrm{diam}(O_\epsilon^j) < \delta_\epsilon$ in $\mathbb{R}^k$ to cover $\mathbb{B}$. That implies

$$\left\{\sup_{b\in\mathbb{B}}|H_m(b)| > \epsilon\right\} \subset \cup_{j=1}^{n_\epsilon}\left\{\sup_{b\in O_\epsilon^j\cap\mathbb{B}}|H_m(b)| > \epsilon\right\},$$

which together with (S5.4) indicates

$$P_m\left(\sup_{b\in\mathbb{B}}|H_m(b)| > \epsilon\right) \leq \sum_{j=1}^{n_\epsilon} P_m\left(\sup_{b\in O_\epsilon^j\cap\mathbb{B}}|H_m(b)| > \epsilon\right) \to 0,$$

concluding the proof.

Lemma 2 *Let $\{X_{mi}\}_{i=1,\dots,m}$ be independent random variables. If there is a constant $r > 1$ such that*

$$L = \sup_{m\geq 1}\max_{i=1,\dots,m} E|X_{mi}|^r < \infty,$$

then

$$\frac{1}{m}\sum_{i=1}^{m}\{X_{mi}-E(X_{mi})\}\xrightarrow{P}0.$$

This lemma is a version of the WLLN (see, e.g., Theorem 1.14 (ii) in Shao) in the setting of a triangular array. The proof is straightforward by Lyapounov's inequality, Jensen's inequality and Theorem 2 of von Bahr and Esseen (1965).

Proof. By Liapounov's inequality, it suffices to consider $r\in(1,2]$. For any $\epsilon>0$, by Markov's inequality

$$\begin{aligned} P\left[\frac{1}{m}\left|\sum_{i=1}^{m}\{X_{mi}-E(X_{mi})\}\right|>\epsilon\right] &= P\left[\frac{1}{m^r}\left|\sum_{i=1}^{m}\{X_{mi}-E(X_{mi})\}\right|^r>\epsilon^r\right] \\ &\leq \frac{1}{\epsilon^r m^r}E\left(\left|\sum_{i=1}^{m}\{X_{mi}-E(X_{mi})\}\right|^r\right). \end{aligned} \tag{S5.5}$$

By the inequality $|a+b|^r\leq 2^{r-1}(|a|^r+|b|^r)$ (Jensen's inequality) and Lyapounov's inequality,

$$\begin{aligned} \max_{i=1,\dots,m}E|X_{mi}-E(X_{mi})|^r &\leq \max_{i=1,\dots,m}2^{r-1}\{E|X_{mi}|^r+|E(X_{mi})|^r\} \\ &\leq \max_{i=1,\dots,m}2^{r-1}\{E|X_{mi}|^r+E|X_{mi}|^r\} \\ &= 2^r\max_{i=1,\dots,m}E|X_{mi}|^r \\ &\leq 2^rL. \end{aligned}$$

Then, according to Theorem 2 of von Bahr and Esseen (1965), we have

$$\begin{aligned} E\left[\left|\sum_{i=1}^{m}\{\mathbf{X}_{mi}-E(X_{mi})\}\right|^r\right] &\leq 2\sum_{i=1}^{m}E|X_{mi}-E(X_{mi})|^r \\ &\leq 2^{r+1}mL. \end{aligned}$$

Combining the above inequality and (S5.5),

$$P\left[\frac{1}{m}\left\{\sum_{i=1}^{m}(X_{mi}-E(X_{mi}))\right\}>\epsilon\right]\leq\frac{2^{r+1}L}{\epsilon^r m^{r-1}}\to 0,$$

concluding the proof.

S6 Proof of Theorem 1

We first show that $\hat{\psi}\xrightarrow{P_m}\psi_0$ under conditions (C1)–(C5) by adapting the proof of Theorem 5.7 of van der Vaart (1998). Let $f_{m,m}(b)=E_{\beta_m}\{m^{-1}s_m(b)\}$. By Lemma 1,

$$\left|m^{-1}s_m(\hat{\beta})-f_{m,m}(\hat{\beta})\right|\leq\sup_{b\in\mathbb{B}}|H_m(b)|\xrightarrow{P_m}0.$$

Since $s_m(\hat{\beta}) = 0$, it follows that $f_{m,m}(\hat{\beta}) \stackrel{P_m}{\to} 0$, which together with condition (C4) shows that there exists $L > 0$ such that

$$|f_{m,0}(\hat{\beta})| \leq |f_{m,0}(\hat{\beta}) - f_{m,m}(\hat{\beta})| + |f_{m,m}(\hat{\beta})| \leq L|\beta_m - \beta_0| + o_{P_m}(1) = o_{P_m}(1), \quad \text{(S6.1)}$$

where $f_{m,0}$ is defined in (C5). According to (C5), which indicates that the solution β_0 of $f_m(b) = 0$ is well-separated from other points in $\mathbb{B}$, for all $\epsilon > 0$ there exists $\delta > 0$ such that

$$\{|\hat{\beta} - \beta_0| > \epsilon\} \subset \{|f_{m,0}(\hat{\beta})| > \delta\}$$

for all $m \geq 1$, which together with (S6.1) implies that $|\hat{\beta} - \beta_0| \stackrel{P_m}{\to} 0$. Therefore, $\hat{\beta} \stackrel{P_m}{\to} \beta_0$ and $\hat{\psi} \stackrel{P_m}{\to} \psi_0$.

Next, assuming conditions (C1)–(C9), we will show $\sqrt{m}(\hat{\psi} - \psi_0)$ converges to $N_p(h, B\Sigma_{\beta_0}B^T)$ in distribution under P_m. It suffices to consider $\sqrt{m}(\hat{\beta} - \beta_0)$, since $\hat{\psi}$ is a subvector of $\hat{\beta}$. By the fundamental theorem of calculus, the chain rule, and the fact that $s_m(\hat{\beta}) = 0$, in terms of $g(t) = s_m(\beta_0 + t(\hat{\beta} - \beta_0))$, $0 \leq t \leq 1$, we have

$$\begin{aligned} -s_m(\beta_0) &= s_m(\hat{\beta}) - s_m(\beta_0) = g(1) - g(0) = \int_0^1 g'(t)\, dt \\ &= \left\{ \int_0^1 \nabla s_m(\beta_0 + t(\hat{\beta} - \beta_0))\, dt \right\} (\hat{\beta} - \beta_0). \end{aligned} \quad \text{(S6.2)}$$

We also have

$$\begin{aligned} &\left\| m^{-1} \left[\int_0^1 \nabla s_m(\beta_0 + t(\hat{\beta} - \beta_0)) dt - E_{\beta_m}\{\nabla s_m(\beta_0)\} \right] \right\| \\ &\qquad \leq \int_0^1 \left\| m^{-1} \left[\nabla s_m(\beta_0 + t(\hat{\beta} - \beta_0)) - E_{\beta_m}\{\nabla s_m(\beta_0)\} \right] \right\| dt, \end{aligned} \quad \text{(S6.3)}$$

where $\|\cdot\|$ denotes the Frobenius matrix norm, that is $\|A\| = \sqrt{\mathrm{tr}(AA^T)}$ for any matrix A. We will show that (S6.3) has order $o_{P_m}(1)$, for which it suffices to show that

$$\int_0^1 m^{-1} \left| \sum_{i=1}^m \varphi_{ij}(\mathbf{y}_i, \beta_0 + t(\hat{\beta} - \beta_0)) - \sum_{i=1}^m E_{\beta_m}\{\varphi_{ij}(\mathbf{y}_i, \beta_0)\} \right| dt = o_{P_m}(1), \quad \text{(S6.4)}$$

for $j = 1, \ldots, k$. The integrand in the above expression is (uniformly) bounded by

$$\sup_{b \in \mathbb{B}} m^{-1} \left| \sum_{i=1}^m \varphi_{ij}(\mathbf{y}_i, b) - \sum_{i=1}^m E_{\beta_m}\{\varphi_{ij}(\mathbf{y}_i, b)\} \right| \quad \text{(S6.5)}$$

$$+ \sup_{0 \leq t \leq 1} m^{-1} \left| \sum_{i=1}^m E_{\beta_m}\{\varphi_{ij}(\mathbf{y}_i, \beta_0 + t(\hat{\beta} - \beta_0))\} - \sum_{i=1}^m E_{\beta_m}\{\varphi_{ij}(\mathbf{y}_i, \beta_0)\} \right|. \quad \text{(S6.6)}$$

Using an argument similar to the proof of Lemma 1, we can show that under conditions (C1)–(C3) and (C6), (S6.5) $= o_{P_m}(1)$. By conditions (C4) and (C6), there exists a $L > 0$ such that

(S6.6) $\leq \sup_{0\leq t\leq 1} Lt|\hat{\beta}-\beta_0| = L|\hat{\beta}-\beta_0| = o_{P_m}(1)$ by the first part of the proof. Thus (S6.3) $= o_{P_m}(1)$, and since under conditions (C4) and (C6) we have $m^{-1}\|E_{\beta_m}\{\nabla s_m(\beta_0)\} - M_m(\beta_0)\| = o(1)$, it follows that

$$m^{-1}\left\|\left\{\int_0^1 \nabla s_m(\beta_0 + t(\hat{\beta}-\beta_0))dt\right\} - M_m(\beta_0)\right\| = o_{P_m}(1).$$

The above display together with (S6.2) and conditions (C8) and (C9) give

$$\sqrt{m}\{M_m(\beta_0)\}^{-1}s_m(\beta_0) = (1+o_{P_m}(1))\sqrt{m}(\hat{\beta}-\beta_0). \tag{S6.7}$$

The result then follows if the left hand side above converges in distribution under P_m to $N(h^*, \Sigma_{\beta_0})$, where $h^* = (\mathbf{0}_{k-p}^T, h^T)^T$ and $\mathbf{0}_{k-p}$ is the $(k-p)$-dimensional zero vector. We will establish this using the Lindeberg–Feller theorem and the Cramér–Wold device. Fix a nonzero k-vector l and $\epsilon > 0$. The Lindeberg condition is checked using condition (C2) and the Hölder and Markov inequalities:

$$\begin{aligned}
&\sum_{i=1}^m E_{\beta_m}|\sqrt{m}l^T\{M_m(\beta_0)\}^{-1}\Psi_i(\mathbf{y}_i,\beta_0)|^2 1\{|\sqrt{m}l^T\{M_m(\beta_0)\}^{-1}\Psi_i(\mathbf{y}_i,\beta_0)| > \epsilon\} \\
&\quad \leq \sum_{i=1}^m \left[E_{\beta_m}|\sqrt{m}l^T\{M_m(\beta_0)\}^{-1}\Psi_i(\mathbf{y}_i,\beta_0)|^{(2+\tilde{\delta})}\right]^{\frac{2}{2+\tilde{\delta}}} \\
&\qquad\quad \times \left[P_m\{|\sqrt{m}l^T\{M_m(\beta_0)\}^{-1}\Psi_i(\mathbf{y}_i,\beta_0)| > \epsilon\}\right]^{\frac{\tilde{\delta}}{2+\tilde{\delta}}} \\
&\quad \leq m|\{M_m(\beta_0)\}^{-1}l|^2 \sum_{i=1}^m \left\{E_{\beta_m}|\Psi_i(\mathbf{y}_i,\beta_0)|^{(2+\tilde{\delta})}\right\}^{\frac{2}{2+\tilde{\delta}}} \\
&\qquad\quad \times \left[\frac{(\sqrt{m}|\{M_m(\beta_0)\}^{-1}l|)^{2+\tilde{\delta}} E_{\beta_m}|\Psi_i(\mathbf{y}_i,\beta_0)|^{(2+\tilde{\delta})}}{\epsilon^{2+\tilde{\delta}}}\right]^{\frac{\tilde{\delta}}{2+\tilde{\delta}}} \\
&\quad = m^{1+\tilde{\delta}/2}|\{M_m(\beta_0)\}^{-1}l|^{2+\tilde{\delta}} \sum_{i=1}^m \frac{E_{\beta_m}|\Psi_i(\mathbf{y}_i,\beta_0)|^{2+\tilde{\delta}}}{\epsilon^{\tilde{\delta}}} \\
&\quad \leq \frac{m^{2+\tilde{\delta}/2}|\{M_m(\beta_0)\}^{-1}l|^{2+\tilde{\delta}}c_0}{\epsilon^{\tilde{\delta}}} = \frac{|\{m^{-1}M_m(\beta_0)\}^{-1}l|^{2+\tilde{\delta}}c_0}{m^{\tilde{\delta}/2}\epsilon^{\tilde{\delta}}} \to 0,
\end{aligned}$$

where $1\{\cdot\}$ is an indicator function, $\tilde{\delta} = \delta - 1$ and the last step is from condition (C9). Also under condition (C7),

$$\begin{aligned}
&\sum_{i=1}^m \mathrm{Var}_{\beta_m}(\sqrt{m}l^T\{M_m(\beta_0)\}^{-1}\Psi_i(\mathbf{y}_i,\beta_0)) \\
&\quad = \sum_{i=1}^m ml^T\{M_m(\beta_0)\}^{-1}\mathrm{Var}_{\beta_m}(\Psi_i(\mathbf{y}_i,\beta_0))\{M_m(\beta_0)\}^{-1}l \\
&\quad = l^T\{m^{-1}M_m(\beta_0)\}^{-1}\{m^{-1}\mathrm{Var}_{\beta_m}(s_m(\beta_0))\}\{m^{-1}M_m(\beta_0)\}^{-1}l \to l^T\Sigma_{\beta_0}l.
\end{aligned}$$

Thus $\sqrt{m}l^T\{M_m(\beta_0)\}^{-1}\{s_m(\beta_0) - E_{\beta_m}(s_m(\beta_0))\}$ converges to $N(0, l^T\Sigma_{\beta_0}l)$ in distribution under P_m. Now we show that $\sqrt{m}l^T\{M_m(\beta_0)\}^{-1}E_{\beta_m}(s_m(\beta_0)) \to l^T h^*$. From condition

(C6),

$$\sup_{m\geq 1}\max_{i=1,\ldots,m} E_{\beta_m}\left\{\sup_{b\in\mathbb{B}}|\varphi_{ij}(\mathbf{y}_i,b)|\right\}<\infty, \tag{S6.8}$$

$$\sup_{0\leq t\leq 1}\left|m^{-1}E_{\beta_m}\left\{\sum_{i=1}^m \varphi_{ij}(\mathbf{y}_i,\beta_0+th^*/\sqrt{m})\right\}-m^{-1}E_{\beta_0}\left\{\sum_{i=1}^m\varphi_{ij}(\mathbf{y}_i,\beta_0)\right\}\right|$$
$$\lesssim \sup_{0\leq t\leq 1}(|\beta_m-\beta_0|+|th^*/\sqrt{m}|)\to 0 \tag{S6.9}$$

for $j=1,\ldots,k$. Since $m^{-1}\sum_{i=1}^m\varphi_{ij}(\mathbf{y}_i,b)$ is the jth row of $\nabla_b s(b)$, by (S6.8) and (S6.9)

$$\begin{aligned}
&l^T\sqrt{m}\{M_m(\beta_0)\}^{-1}E_{\beta_m}(s_m(\beta_0))\\
&\quad= l^T\sqrt{m}\{M_m(\beta_0)\}^{-1}E_{\beta_m}(s_m(\beta_0)-s_m(\beta_m))\\
&\quad= l^T\sqrt{m}\{M_m(\beta_0)\}^{-1}E_{\beta_m}\left\{\int_0^1\nabla_\beta s_m(\beta_0+th^*/\sqrt{m})\,dt\right\}\frac{-h^*}{\sqrt{m}}\\
&\quad= l^T\{M_m(\beta_0)\}^{-1}\int_0^1 E_{\beta_m}\left\{-\nabla_\beta s_m(\beta_0+th^*/\sqrt{m})\right\}\,dt\,h^* &\text{(S6.10)}\\
&\quad= l^T\{m^{-1}M_m(\beta_0)\}^{-1}\int_0^1 m^{-1}E_{\beta_m}\left\{-\nabla_\beta s_m(\beta_0+th^*/\sqrt{m})\right\}\,dt\,h^*\\
&\quad\to l^T\{M(\beta_0)\}^{-1}\int_0^1 M(\beta_0)\,dt\,h^*=l^Th^*, &\text{(S6.11)}
\end{aligned}$$

where (S6.10) and (S6.11) use Fubini's theorem and the dominated convergence theorem, respectively, and $\nabla_\beta s_m(\beta_0+th^*/\sqrt{m})=\partial s_m(b)/\partial b|_{b=\beta_0+th^*/\sqrt{m}}$. Combining the above result and that $\sqrt{m}l^T\{M_m(\beta_0)\}^{-1}\{s_m(\beta_0)-E_{\beta_m}(s_m(\beta_0))\}$ converges in distribution to $N(0,l^T\Sigma_{\beta_0}l)$ under P_m, we have that $\sqrt{m}l^T\{M_m(\beta_0)\}^{-1}s_m(\beta_0)$ converges under P_m in distribution to $N(l^Th^*,l^T\Sigma_{\beta_0}l)$ for any non-zero k-vector l. Thus $\sqrt{m}\{M_m(\beta_0)\}^{-1}s_m(\beta_0)$ converges in distribution to $N(h^*,\Sigma_{\beta_0})$ under P_m, and from (S6.7) and using Slutsky's lemma, $\sqrt{m}(\hat\beta-\beta_0)$ converges in distribution under P_m to $N(h^*,\Sigma_{\beta_0})$. The proof is completed by noticing that $\hat\psi=B\hat\beta$.

S7 Proof of Theorem 2

Asymptotic distribution of W_m

To prove W_m converges in distribution under H_{1m} to $\chi^2_p(\nu)$, the key step is to show that $\hat\Sigma$ converges in probability under H_{1m} to Σ_{β_0}, which implies that $B\hat\Sigma B^T$ converges in probability under H_{1m} to $B\Sigma_{\beta_0}B^T$. Note that, under conditions (C7)–(C9),

$$\Sigma_{\beta_0}=\{M(\beta_0)\}^{-1}\left[\lim_{m\to\infty}m^{-1}\sum_{i=1}^m\mathrm{Var}_{\beta_0}\{\Psi_i(\mathbf{y}_i,\beta_0)\}\right]\{M(\beta_0)\}^{-1}.$$

Write

$$\hat{\Sigma} = \left\{m^{-1}M_m(\hat{\beta})\right\}^{-1}\left\{m^{-1}\sum_{i=1}^{m}\Psi_i(\mathbf{y}_i,\hat{\beta})\Psi_i(\mathbf{y}_i,\hat{\beta})^T\right\}\left\{m^{-1}M_m(\hat{\beta})\right\}^{-1}.$$

We will show that

$$m^{-1}M_m(\hat{\beta}) = M(\beta_0) + o_{P_m}(1) \tag{S7.1}$$

and

$$m^{-1}\sum_{i=1}^{m}\Psi_i(\mathbf{y}_i,\hat{\beta})\Psi_i(\mathbf{y}_i,\hat{\beta})^T - m^{-1}\sum_{i=1}^{m}\mathrm{Var}_{\beta_0}[\Psi_i(\mathbf{y}_i,\beta_0)] = o_{P_m}(1). \tag{S7.2}$$

The proof of (S7.1) is straightforward using conditions (C4), (C6) and (C9) and the consistency of $\hat{\beta}$ shown in Section B:

$$\begin{aligned}\left\|m^{-1}M_m(\hat{\beta}) - M(\beta_0)\right\| &= \left\|m^{-1}M_m(\hat{\beta}) - m^{-1}M_m(\beta_0)\right\| \\ &\quad + \left\|m^{-1}M_m(\beta_0) - M(\beta_0)\right\| \\ &\lesssim |\hat{\beta}-\beta_0| + o(1) = o_{P_m}(1).\end{aligned}$$

Write the left hand side of (S7.2) as

$$m^{-1}\sum_{i=1}^{m}\Psi_i(\mathbf{y}_i,\hat{\beta})\Psi_i(\mathbf{y}_i,\hat{\beta})^T - m^{-1}\sum_{i=1}^{m}\Psi_i\left(\mathbf{y}_i,\beta_0\right)\Psi_i\left(\mathbf{y}_i,\beta_0\right)^T \tag{S7.3}$$

$$+\, m^{-1}\sum_{i=1}^{m}\left[\Psi_i\left(\mathbf{y}_i,\beta_0\right)\Psi_i\left(\mathbf{y}_i,\beta_0\right)^T - E_{\beta_m}\left\{\Psi_i\left(\mathbf{y}_i,\beta_0\right)\Psi_i\left(\mathbf{y}_i,\beta_0\right)^T\right\}\right] \tag{S7.4}$$

$$+\, m^{-1}\sum_{i=1}^{m}(E_{\beta_m}-E_{\beta_0})\left\{\Psi_i\left(\mathbf{y}_i,\beta_0\right)\Psi_i\left(\mathbf{y}_i,\beta_0\right)^T\right\}. \tag{S7.5}$$

Let $\Psi_{ij}(\mathbf{y}_i,\beta_0)$ be the jth element of the vector $\Psi_i\left(\mathbf{y}_i,\beta_0\right)$ for $j=1,\ldots,k$. The term (S7.4) converges to zero in probability under H_{1m} by Lemma 2; the condition needed to apply that lemma can be shown to hold using the Cauchy–Schwarz inequality and condition (C2). It can be shown that (S7.5) $= o(1)$ since under conditions (C2), (C4) and (C7)

$$\begin{aligned}&|(E_{\beta_m}-E_{\beta_0})\{\Psi_{ij}(\mathbf{y}_i,\beta_0)\Psi_{il}(\mathbf{y}_i,\beta_0)\}| \\ &\quad\le |\mathrm{cov}_{\beta_m}(\Psi_{ij}(\mathbf{y}_i,\beta_0),\Psi_{il}(\mathbf{y}_i,\beta_0)) - \mathrm{cov}_{\beta_0}(\Psi_{ij}(\mathbf{y}_i,\beta_0),\Psi_{il}(\mathbf{y}_i,\beta_0))| \\ &\qquad + |E_{\beta_m}(\Psi_{ij}(\mathbf{y}_i,\beta_0))||\{E_{\beta_m}(\Psi_{il}(\mathbf{y}_i,\beta_0)) - E_{\beta_0}(\Psi_{il}(\mathbf{y}_i,\beta_0))\}| \\ &\qquad + |E_{\beta_0}(\Psi_{il}(\mathbf{y}_i,\beta_0))||\{E_{\beta_m}(\Psi_{ij}(\mathbf{y}_i,\beta_0)) - E_{\beta_0}(\Psi_{ij}(\mathbf{y}_i,\beta_0))\}| \\ &\quad\lesssim |\beta_m-\beta_0| = o(1).\end{aligned}$$

Next consider the matrix (S7.3). If each component can be shown to converge in probability under H_{1m} to zero, then we have completed the proof of (S7.2). Let

$$\begin{aligned}g_{ijl}(\mathbf{y}_i,b) &= \partial\{\Psi_{ij}(\mathbf{y}_i,b)\Psi_{il}(\mathbf{y}_i,b)\}/\partial b \\ &= \varphi_{ij}(\mathbf{y}_i,b)\Psi_{il}(\mathbf{y}_i,b) + \varphi_{il}(\mathbf{y}_i,b)\Psi_{ij}(\mathbf{y}_i,b) \equiv g^1_{ijl}(\mathbf{y}_i,b) + g^2_{ijl}(\mathbf{y}_i,b).\end{aligned}$$

As in (S6.2), but with the role of $g(t)$ now played by $m^{-1}\sum_{i=1}^m \Psi_{ij}(\mathbf{y}_i,\beta_0+t(\hat\beta-\beta_0))\Psi_{il}(\mathbf{y}_i,\beta_0+t(\hat\beta-\beta_0))$, $0\le t\le 1$, for any fixed $\epsilon>0$, the jlth entry of (S7.3) is bounded above by

$$\begin{aligned}
&\left|\left\{\int_0^1 m^{-1}\sum_{i=1}^m g_{ijl}(\mathbf{y}_i,\beta_0+t(\hat\beta-\beta_0))\,dt\right\}(\hat\beta-\beta_0)\right|\\
&\qquad\le \sup_{b\in\mathbb{B}}\left|m^{-1}\sum_{i=1}^m g^1_{ijl}(\mathbf{y}_i,b)\right| o_{P_m}(1)+\sup_{b\in\mathbb{B}}\left|m^{-1}\sum_{i=1}^m g^2_{ijl}(\mathbf{y}_i,b)\right| o_{P_m}(1). \qquad \text{(S7.6)}
\end{aligned}$$

Next we show that the supremum terms above are of order $O_{P_m}(1)$. The first of these terms

$$\begin{aligned}
&\sup_{b\in\mathbb{B}}\left|\frac{1}{m}\sum_{i=1}^m g^1_{ijl}(\mathbf{y}_i,b)\right|\\
&\qquad\le \sup_{b\in\mathbb{B}}\left|m^{-1}\sum_{i=1}^m g^1_{ijl}(\mathbf{y}_i,b)-E_{\beta_m}\left\{m^{-1}\sum_{i=1}^m g^1_{ijl}(\mathbf{y}_i,b)\right\}\right| \qquad \text{(S7.7)}\\
&\qquad+\sup_{b\in\mathbb{B}}\left|E_{\beta_m}\left\{m^{-1}\sum_{i=1}^m g^1_{ijl}(\mathbf{y}_i,b)\right\}\right|. \qquad \text{(S7.8)}
\end{aligned}$$

For $\delta^*=(\delta-1)/2>0$, we have by the Cauchy–Schwarz inequality

$$\begin{aligned}
&E_{\beta_m}\left\{\sup_{b\in\mathbb{B}}|g^1_{ijl}(\mathbf{y}_i,b)|\right\}^{1+\delta^*}\\
&\quad\le E_{\beta_m}\left\{\sup_{b\in\mathbb{B}}|\varphi_{ij}(\mathbf{y}_i,b)|\sup_{b\in\mathbb{B}}|\Psi_{il}(\mathbf{y}_i,b)|\right\}^{1+\delta^*}\\
&\quad\le \left\{E_{\beta_m}\left\{\sup_{b\in\mathbb{B}}|\varphi_{ij}(\mathbf{y}_i,b)|\right\}^{1+\delta}\right\}^{\frac12}\left\{E_{\beta_m}\left\{\sup_{b\in\mathbb{B}}|\Psi_{il}(\mathbf{y}_i,b)|\right\}^{1+\delta}\right\}^{\frac12},
\end{aligned}$$

so using conditions (C2) and (C6) we have (S7.8) $=O(1)$. The term (S7.7) can be shown to be of order $o_{P_m}(1)$ using Lemma 1 with g^1_{ijl} playing the role of Ψ_i. To that end we need to check that condition (C3) holds for g^1_{ijl}, i.e., that for any $c>0$ and sequence $\{y_i\}$ satisfying $|y_i|\le c$, the sequence of functions $\{g^1_{ijl}(y_i,b)\}_{i=1,2,\dots}$ is equicontinuous on $\mathbb{B}$. This follows from (C3) and (C6) using the inequality

$$\begin{aligned}
&|g^1_{ijl}(y_i,t)-g^1_{ijl}(y_i,s)|\\
&\quad\le|\varphi_{ij}(y_i,t)||\Psi_{il}(y_i,t)-\Psi_{il}(y_i,s)|+|\varphi_{ij}(y_i,t)-\varphi_{ij}(y_i,s)||\Psi_{il}(y_i,s)|.
\end{aligned}$$

We have now shown that (S7.3) $=o_{P_m}(1)$, so (S7.2) holds. Then using the second part of condition (C8) combined with (S7.2) and (S7.1), we have $\hat\Sigma$ converges in probability under H_{1m} to Σ_{β_0}. Thus $B\hat\Sigma B^T$ converges in probability under H_{1m} to $B\Sigma_{\beta_0}B^T$. From Theorem 1, $\sqrt{m}(\hat\psi-\psi_0)$ converges in distribution under H_{1m} to $N(h,B\Sigma_{\beta_0}B^T)$. Therefore, by Slutsky's lemma and the continuous mapping theorem, W_m converges in distribution under H_{1m} to non-central χ^2_p with non-centrality parameter $\nu=h^T(B\Sigma_{\beta_0}B^T)^{-1}h$. Next we derive the asymptotic distribution for T_m.

Asymptotic distribution of T_m

An estimate $\tilde{\lambda}$ of the nuisance parameter vector λ under H_0 is needed to calculate the quasi-score statistic. For this purpose it suffices to use the first $k-p$ estimating equations, so $\tilde{\lambda}$ can be taken as a solution of $Cs_m(\lambda, \psi_0) = 0$, where $C = (I_{(k-p)}, 0_{(k-p)\times p})$. Write the quasi-score statistic as

$$T_m = \left\{m^{-1/2} B s_m(\tilde{\beta})\right\}^T \left(m^{-1} V_T\right)^{-1} \left\{m^{-1/2} B s_m(\tilde{\beta})\right\},$$

where

$$\begin{aligned} V_T = \; & \{BM_m(\tilde{\beta})^{-1}B^T\}^{-1} \\ & \times \left[BM_m(\tilde{\beta})^{-1} \left\{ \sum_{i=1}^m \Psi_i(\mathbf{y}_i, \tilde{\beta}) \Psi_i(\mathbf{y}_i, \tilde{\beta})^T \right\} M_m(\tilde{\beta})^{-1} B^T \right] \\ & \times \{BM_m(\tilde{\beta})^{-1}B^T\}^{-1}, \end{aligned}$$

and $\tilde{\beta} = (\tilde{\lambda}^T, \psi_0^T)^T$. We first establish a connection between $\hat{\psi} - \psi_0$ and $s_m(\tilde{\beta})$:

$$\sqrt{m}(\hat{\psi} - \psi_0) = \sqrt{m} B \{M_m(\beta_0)\}^{-1} B^T B s_m(\tilde{\beta}) + o_{P_m}(1). \tag{S7.9}$$

According to (S6.7),

$$\begin{aligned} (1 + o_{P_m}(1))\sqrt{m}(\hat{\beta} - \beta_0) &= \sqrt{m}\{M_m(\beta_0)\}^{-1} s_m(\beta_0) \\ &= \sqrt{m}\{M_m(\beta_0)\}^{-1} s_m(\tilde{\beta}) \\ &\quad + \sqrt{m}\{M_m(\beta_0)\}^{-1} \{s_m(\beta_0) - s_m(\tilde{\beta})\}. \end{aligned}$$

Under conditions (C1)–(C9), it can be shown that $\sqrt{m}(\tilde{\lambda} - \lambda_0) = O_{P_m}(1)$ using a similar proof as Theorem 1. Following the steps between (S6.2) and (S6.7) with $s_m(\tilde{\beta})$ in place of $s_m(\hat{\beta})$, we have

$$\begin{aligned} \sqrt{m}\{M_m(\beta_0)\}^{-1}\{s_m(\beta_0) - s_m(\tilde{\beta})\} &= (1 + o_{P_m}(1))\sqrt{m}(\beta_0 - \tilde{\beta}) \\ &= (1 + o_{P_m}(1))\sqrt{m} C^T (\lambda_0 - \tilde{\lambda}) \\ &= O_{P_m}(1) C + o_{P_m}(1). \end{aligned}$$

Combining the results of the above two displays, we have

$$\begin{aligned} & (1 + o_{P_m}(1))\sqrt{m}(\hat{\psi} - \psi_0) \\ & \quad = \sqrt{m} B\{M_m(\beta_0)\}^{-1} s_m(\tilde{\beta}) + B\{C^T O_{P_m}(1) + o_{P_m}(1)\} \\ & \quad = \sqrt{m} B\{M_m(\beta_0)\}^{-1} s_m(\tilde{\beta}) + o_{P_m}(1) \\ & \quad = \sqrt{m} B\{M_m(\beta_0)\}^{-1} B^T B s_m(\tilde{\beta}) + o_{P_m}(1) \\ & \quad = B\{m^{-1} M_m(\beta_0)\}^{-1} B^T m^{-1/2} B s_m(\tilde{\beta}) + o_{P_m}(1). \end{aligned}$$

The above display implies equation (S7.9), which together with Theorem 1 and Slutsky's lemma shows that $m^{-1/2} B s_m(\tilde{\beta})$ converges under P_m in distribution to a normal distribution with mean $\{BM(\beta_0)^{-1}B^T\}^{-1} h$ and variance

$$\{BM(\beta_0)^{-1}B^T\}^{-1} (B \Sigma_{\beta_0} B^T) \{BM(\beta_0)^{-1}B^T\}^{-1}.$$

We have that $\left(m^{-1}\sum_{i=1}^m \Psi_i(\mathbf{y}_i,\tilde\beta)\Psi_i(\mathbf{y}_i,\tilde\beta)^T\right)$ and $m^{-1}M_m(\tilde\beta)$ converge in probability under P_m to $\lim_{m\to\infty}\left[m^{-1}\sum_{i=1}^m \mathrm{Var}_{\beta_0}\{\Psi_i(\mathbf{y}_i,\beta_0)\}\right]$ and $M(\beta_0)$ respectively using the same argument for (S7.1) and (S7.2) in Theorem 2. Therefore, $m^{-1}V_T$ converges in probability under P_m to the asymptotic variance of $m^{-1/2}Bs_m(\tilde\beta)$.

By Slutsky's lemma and the continuous mapping theorem, it follows that T_m converges in distribution under P_m to noncentral chi-squared with non-centrality parameter

$$\begin{aligned}&\left[\left\{BM(\beta_0)^{-1}B^T\right\}^{-1}h\right]^T\left[\left\{BM(\beta_0)^{-1}B^T\right\}^{-1}(B\Sigma_{\beta_0}B^T)\left\{BM(\beta_0)^{-1}B^T\right\}^{-1}\right]^{-1}\\&\left[\left\{BM(\beta_0)^{-1}B^T\right\}^{-1}h\right]\\&\quad=h^T(B\Sigma_{\beta_0}B^T)^{-1}h,\end{aligned}$$

concluding the proof.

S8 Derivation of (4.1)

In Example 4.1, the matrix $\bar B$ becomes the vector $(0,1)$ since κ and ψ are both univariate, $\theta_A=(\kappa_0,\psi_A)^T$, $\theta_0=(\kappa_0,\psi_0)^T$, $\beta_A=(\sigma,\alpha_0^T,\kappa_0,\psi_A)^T$ and $\mu_i=\mathbf{1}_n(\kappa+\psi x_i)$, where $\mathbf{1}_n$ is the $n\times 1$ vector with all elements being 1. Following the sample size calculation procedure given near the end of Section 3 in the manuscript, we first choose type I error rate (ζ) and desired power ($1-\eta$) in step 1. The cluster sizes are the same (n) and the covariate x_i's could have any arbitrary distribution in step 2. We give values of ψ_0, ψ_A, κ_0, α and σ and calculate D_1 and V_1 in step 3. The $n\times 2$ matrix $D_i=\partial\mu_i/\partial\theta$ evaluated under H_0 is $D_i=\mathbf{1}_n(1,x_i)$. The $n\times n$ variance matrices evaluated under H_0 are $V_i=\sigma^2R$, where R is a $n\times n$ correlation matrix. Then we calculate $\mathrm{Var}_{\beta_A}(\mathbf{y}_1|\mathbf{z}_1,\mathbf{x}_1,n_1)$ in step 4. In this example, that conditional variance is equal to V_i obtained in the previous step. We calculate the following quantities in step 5:

$$E(D_1^TV_1^{-1}D_1)\quad=\quad\frac{\mathbf{1}_n^TR^{-1}\mathbf{1}_n}{\sigma^2}\begin{pmatrix}1 & E(x_1)\\ E(x_1) & E(x_1^2)\end{pmatrix},$$

$$E(D_1^TV_1^{-1}\{\mu_1(\theta_A)-\mu_1(\theta_0)\})=\frac{(\mathbf{1}_n^TR^{-1}\mathbf{1}_n)(\psi_A-\psi_0)}{\sigma^2}\begin{pmatrix}E(x_1)\\ E(x_1^2)\end{pmatrix},$$

and

$$\begin{aligned}E[D_1^TV_1^{-1}\mathrm{cov}_{\beta_A}(\mathbf{y}_1|x_1)V_1^{-1}D_1]&=E(D_1^TV_1^{-1}D_1)\\&=\frac{\mathbf{1}_n^TR^{-1}\mathbf{1}_n}{\sigma^2}\begin{pmatrix}1 & E(x_1)\\ E(x_1) & E(x_1^2)\end{pmatrix}.\end{aligned}$$

In step 6, we calculate

$$\begin{aligned}\tilde\xi_\psi&=\bar B[E(D_1^TV_1^{-1}D_1)]^{-1}E\{D_1^TV_1^{-1}[\mu_1(\theta_A)-\mu_1(\theta_0)]\}\\&=\psi_A-\psi_0,\end{aligned}$$

$$\begin{aligned}\tilde{\Sigma}_\psi &= \bar{B}[E(D_1^T V_1^{-1} D_1)]^{-1} E[D_1^T V_1^{-1} \mathrm{cov}_{\beta_A}(\mathbf{y}_1|x_1) V_1^{-1} D_1][E(D_1^T V_1^{-1} D_1)]^{-1} \bar{B}^T \\ &= \frac{\sigma^2}{(\mathbf{1}_n^T R^{-1} \mathbf{1}_n)\mathrm{var}(x_1)},\end{aligned}$$

and $\tilde{\nu} = (z_{1-\zeta/2} + z_{1-\eta})^2$. Formula (4.1) is obtained by (3.7).

S9 Derivation of the sample size formula in Example 4.2

In this example, again the matrix $\bar{B}$ becomes the vector $(0, 1)$ since κ and ψ are both univariate, $\theta_A = (\kappa_0, \psi_A)^T$, $\theta_0 = (\kappa_0, \psi_0)^T$, $\beta_A = (\alpha_0^T, \kappa_0, \psi_A)^T$ and $\mu_i = \mathbf{1}_n \mathrm{expit}(\kappa + \psi x_i)$ where $\mathbf{1}_n$ is the $n \times 1$ vector with all elements being 1. As in the previous derivation, we similarly follow the steps in the sample size calculation procedure . The $2 \times n$ matrix $D_i = \partial\mu_i/\partial\theta$ evaluated under H_0 is $D_i = \mathbf{1}_n(1, x_i)\mathrm{v}_{0x_i}$ where $\mathrm{v}_{0x_i} = p_{0x_i}(1 - p_{0x_i})$ and $p_{0x_i} = \mathrm{expit}(\kappa_0 + \psi_0 x_i)$. The $n \times n$ variance matrices evaluated under H_0 are $V_i = \mathrm{v}_{0x_i} R$, where R is a $n \times n$ correlation matrix. Therefore,

$$E(D_1^T V_1^{-1} D_1) = (\mathbf{1}_n^T R^{-1} \mathbf{1}_n) \begin{pmatrix} E(\mathrm{v}_{0x_1}) & E(x_1 \mathrm{v}_{0x_1}) \\ E(x_1 \mathrm{v}_{0x_1}) & E(x_1^2 \mathrm{v}_{0x_1}) \end{pmatrix},$$

$$E(D_1^T V_1^{-1}[\mu_1(\theta_A) - \mu_1(\theta_0)]) = (\mathbf{1}_n^T R^{-1} \mathbf{1}_n) \begin{pmatrix} E(p_{1x_1}) - E(p_{0x_1}) \\ E(x_1 p_{1x_1}) - E(x_1 p_{0x_1}) \end{pmatrix},$$

and

$$E[D_1^T V_1^{-1} \mathrm{cov}_{\beta_A}(\mathbf{y}_1|x_1) V_1^{-1} D_1] = (\mathbf{1}_n^T R^{-1} \mathbf{1}_n) \begin{pmatrix} E(\mathrm{v}_{1x_1}) & E(x_1 \mathrm{v}_{1x_1}) \\ E(x_1 \mathrm{v}_{1x_1}) & E(x_1^2 \mathrm{v}_{1x_1}) \end{pmatrix},$$

where $\mathrm{v}_{1x_1} = p_{1x_1}(1 - p_{1x_1})$ and $p_{1x_1} = \mathrm{expit}(\kappa_0 + \psi_A x_1)$. Then

$$\begin{aligned}\tilde{\xi}_\psi &= \bar{B}[E(D_1^T V_1^{-1} D_1)]^{-1} E\{D_1^T V_1^{-1}[\mu_1(\theta_A) - \mu_1(\theta_0)]\} \\ &= \frac{E(\mathrm{v}_{0x_1})(E(x_1 p_{1x_1}) - E(x_1 p_{0x_1})) - E(x_1 \mathrm{v}_{0x_1})(E(p_{1x_1}) - E(p_{0x_1}))}{E(\mathrm{v}_{0x_1})E(x_1^2 \mathrm{v}_{0x_1}) - [E(x_1 \mathrm{v}_{0x_1})]^2}\end{aligned}$$

and

$$\begin{aligned}\tilde{\Sigma}_\psi &= \bar{B}[E(D_1^T V_1^{-1} D_1)]^{-1} E[D_1^T V_1^{-1} \mathrm{cov}_{\beta_1}(\mathbf{y}_1|x_1) V_1^{-1} D_1][E(D_1^T V_1^{-1} D_1)]^{-1} \bar{B}^T \\ &= \frac{E(\mathrm{v}_{1x_1})[E(x_1 \mathrm{v}_{0x_1})]^2 + E(x_1^2 \mathrm{v}_{1x_1})[E(\mathrm{v}_{0x_1})]^2 - 2E(x_1 \mathrm{v}_{1x_1})E(x_1 \mathrm{v}_{0x_1})E(\mathrm{v}_{0x_1})}{[E(\mathrm{v}_{0x_1})E(x_1^2 \mathrm{v}_{0x_1}) - (E(x_1 \mathrm{v}_{0x_1}))^2]^2 (\mathbf{1}_n^T R^{-1} \mathbf{1}_n)}.\end{aligned}$$

By (3.7), we can obtain the formula.

S10 Derivation of the sample size formula in Example 4.3

In this example, again the matrix $\bar{B} = (0, 1)$, $\theta_A = (\kappa_0, \psi_A)^T$, $\theta_0 = (\kappa_0, \psi_0)^T$ and $\beta_A = (\rho, \kappa_0, \psi_A)^T$, where the correlation ρ is a scalar since the cluster size is 2. There is only one type of clusters with cluster sizes being 2. That is, $\mathbf{x}_i = (x_{i1}, x_{i2})^T$ follows a degenerate distribution $P(\mathbf{x}_i = (0, 1)^T) = 1$. The corresponding vector mean is $\mu_i = (\text{expit}(\kappa), \text{expit}(\kappa + \psi))^T$. As in the derivation of 4.1, we also follow the steps in the sample size calculation procedure. The matrix $D_i = \partial\mu_i/\partial\theta$ evaluated under H_0 equals

$$D_i = \begin{pmatrix} \text{v}_0 & 0 \\ \tilde{\text{v}}_0 & \tilde{\text{v}}_0 \end{pmatrix},$$

where $\tilde{\text{v}}_0 = \tilde{p}_0(1 - \tilde{p}_0)$, $\text{v}_0 = p_0(1 - p_0)$, $\tilde{p}_0 = \text{expit}(\kappa_0 + \psi_0)$ and $p_0 = \text{expit}(\kappa_0)$. Then the variance matrix V_i evaluated under H_0 is

$$V_i = \begin{pmatrix} \text{v}_0 & \rho\sqrt{\text{v}_0\tilde{\text{v}}_0} \\ \rho\sqrt{\text{v}_0\tilde{\text{v}}_0} & \tilde{\text{v}}_0 \end{pmatrix}.$$

Thus

$$[E(D_1^T V_1^{-1} D_1)]^{-1} = (D_1^T V_1^{-1} D_1)^{-1} = D_1^{-1} V_1 (D_1^T)^{-1}.$$

$$\begin{aligned} E\left\{D_1^T V_1^{-1}[\mu_1(\theta_A) - \mu_1(\theta_0)]\right\} &= D_1^T V_1^{-1}[\mu_1(\theta_A) - \mu_1(\theta_0)] \\ &= D_1^T V_1^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} (p_1 - \tilde{p}_0), \end{aligned}$$

where $p_1 = \text{expit}(\kappa_0 + \psi_A)$. Then

$$\begin{aligned} \tilde{\xi}_\psi &= \bar{B}[E(D_1^T V_1^{-1} D_1)]^{-1} E\left\{D_1^T V_1^{-1}[\mu_1(\theta_A) - \mu_1(\theta_0)]\right\} \\ &= \bar{B} D_1^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} (p_1 - \tilde{p}_0) \\ &= \frac{1}{\text{v}_0\tilde{\text{v}}_0} \bar{B} \begin{pmatrix} \tilde{\text{v}}_0 & 0 \\ -\tilde{\text{v}}_0 & \text{v}_0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} (p_1 - \tilde{p}_0) \\ &= \frac{p_1 - \tilde{p}_0}{\tilde{\text{v}}_0}. \end{aligned}$$

We have

$$\text{cov}_{\beta_A}(\mathbf{y}_i|\mathbf{x}_i) = \begin{pmatrix} \text{v}_0 & \rho\sqrt{\text{v}_0\text{v}_1} \\ \rho\sqrt{\text{v}_0\text{v}_1} & \text{v}_1 \end{pmatrix},$$

where $\mathrm{v}_1 = p_1(1-p_1)$. Thus

$$\begin{aligned}
\tilde{\Sigma}_\psi &= \bar{B}[E(D_1^T V_1^{-1} D_1)]^{-1} E\left(D_1^T V_1^{-1} \mathrm{cov}_{\beta_A}(\mathbf{y}_1|\mathbf{x}_1) V_1^{-1} D_1\right)[E(D_1^T V_1^{-1} D_1)]^{-1} \bar{B}^T \\
&= \bar{B} D_1^{-1} \mathrm{cov}_{\beta_A}(\mathbf{y}_1|\mathbf{x}_1)(D_1^T)^{-1} \bar{B}^T \\
&= \frac{1}{\mathrm{v}_0\tilde{\mathrm{v}}_0} \bar{B} \begin{pmatrix} \tilde{\mathrm{v}}_0 & 0 \\ -\tilde{\mathrm{v}}_0 & \mathrm{v}_0 \end{pmatrix} \begin{pmatrix} \mathrm{v}_0 & \rho\sqrt{\mathrm{v}_0\mathrm{v}_1} \\ \rho\sqrt{\mathrm{v}_0\mathrm{v}_1} & \mathrm{v}_1 \end{pmatrix} \frac{1}{\mathrm{v}_0\tilde{\mathrm{v}}_0} \begin{pmatrix} \tilde{\mathrm{v}}_0 & -\tilde{\mathrm{v}}_0 \\ 0 & \mathrm{v}_0 \end{pmatrix} \bar{B}^T \\
&= \frac{1}{(\tilde{\mathrm{v}}_0\mathrm{v}_0)^2} \bar{B} \begin{pmatrix} \mathrm{v}_0\tilde{\mathrm{v}}_0^2 & \rho\mathrm{v}_0\tilde{\mathrm{v}}_0\sqrt{\mathrm{v}_0\mathrm{v}_1} - \mathrm{v}_0\tilde{\mathrm{v}}_0^2 \\ \rho\mathrm{v}_0\tilde{\mathrm{v}}_0\sqrt{\mathrm{v}_0\mathrm{v}_1} - \mathrm{v}_0\tilde{\mathrm{v}}_0^2 & \mathrm{v}_0^2\mathrm{v}_1 - 2\rho\mathrm{v}_0\tilde{\mathrm{v}}_0\sqrt{\mathrm{v}_0\mathrm{v}_1} + \mathrm{v}_0\tilde{\mathrm{v}}_0^2 \end{pmatrix} \bar{B}^T \\
&= \frac{\mathrm{v}_0^2\mathrm{v}_1 - 2\rho\mathrm{v}_0\tilde{\mathrm{v}}_0\sqrt{\mathrm{v}_0\mathrm{v}_1} + \mathrm{v}_0\tilde{\mathrm{v}}_0^2}{(\mathrm{v}_0\tilde{\mathrm{v}}_0)^2}.
\end{aligned}$$

Then the formula is obtained by (3.7).

```
/****************************************************************;
SAS code of sample size calculation for the Arsenic study;

alpha: type I error rate;
eta: type II error rate;
lambda: intercept in the logistic regression;
psi: coefficient of the Arsenic exposure in the logistic regression;
rho: correlation in exchangeable correlation structure or correlation
between adjacent measurements in AR1 structure;
a1: mean of the natural log transformed exposure;
b1: standard deviation of the natural log transformed exposure;
n: number of iterations of Monte Carlo integral;
cluster: cluster size;
****************************************************************/;
%macro sample(alpha,eta,lambda,psi,rho,a1,b1,n,cluster);
proc iml;
************************************************read the design
parameters from prespecification or pilot data;
lambda=&lambda;psi=&psi;a1=&a1;b1=&b1;rho=&rho; n=&n;
********************************************create the correlation
structure;
R=J(&cluster,&cluster,.);
do t=1 to &cluster;
  do s=1 to &cluster;
    ***AR1 structure;
    *R[t,s]=rho**abs(t-s);
    ***Exchangeable structure;
    R[t,s]=rho;
      if t=s then R[t,s]=1;
  end;
end;

****************Monte Carlo integration for the expectations in the
sample size formulae in Example 4.2;
seed=100;
numex=0;
denoxpx=0;

mx=a1;***mean value of the log-arsenic;

do i=1 to n;
x=a1+b1*rannor(seed);
px=exp(lambda+psi*x)/(1+exp(lambda+psi*x));
vx=px*(1-px);
numex=numex+mx*mx*vx+x*x*vx-2*mx*x*vx;
denoxpx=denoxpx+(x-mx)*px;
end;

mnumex=numex/n;
mdenoxpx=denoxpx/n;

****************************final steps for sample size calculation;
NUTILDE=(probit(1-&alpha/2)+probit(1-&eta))**2;
numer=NUTILDE*mnumex;
deno=J(1,&cluster,1)*inv(R)*J(&cluster,1,1)*mdenoxpx*mdenoxpx;
m=numer/deno;
```

```
print m;
quit;

%mend;

%sample(alpha=0.05,eta=0.1,lambda=-
2.71662,psi=0.4055,rho=0.2,a1=0.653,b1=2.00,n=10000000,cluster=4);
```