Online Supplement of '*An empirical likelihood approach under cluster sampling with missing observations*'

Yves G. Berger

Appendix B

In order to simplify the notation, $\widehat{\boldsymbol{c}}_k^\star(\boldsymbol{\psi}_0, \boldsymbol{r})$ and $\widehat{\boldsymbol{a}}_k(\boldsymbol{\psi}_0, \boldsymbol{r})$ are respectively replaced by $\widehat{\boldsymbol{c}}_{0k}^\star$ and $\widehat{\boldsymbol{a}}_{0k}$. Hence, $\widehat{\boldsymbol{c}}_{0k}^\star = (\widehat{\boldsymbol{a}}_{0k}^\top, \widehat{\boldsymbol{c}}_k^\top)^\top$.

Lemma B1 *Let $\widehat{\boldsymbol{S}}$ be a random $r \times r$ matrix, where $r < \infty$. If there exists a population matrix $\boldsymbol{S}_U$ and a positive definite symmetric matrix $\boldsymbol{S}$ of constants such that $\widehat{\boldsymbol{S}} - \boldsymbol{S}_U = \boldsymbol{o}_p(1)$, $\boldsymbol{S}_U - \boldsymbol{S} = \boldsymbol{o}(1)$ and $\|\boldsymbol{S}\| < \infty$, we have that $\widehat{\boldsymbol{S}} = \boldsymbol{\mathcal{O}}_p(1)$, $\widehat{\boldsymbol{S}}^{-1} = \boldsymbol{\mathcal{O}}_p(1)$ and $\|\boldsymbol{S}_U^{-1}\| \asymp 1$.*

The proof of Lemma B1 can be found in Berger & Kabzinska (2016). Lemma B1 and [C11] imply

$$\widehat{\boldsymbol{S}}_0 = \boldsymbol{\mathcal{O}}_p(1), \tag{B.1}$$

$$\widehat{\boldsymbol{S}}_0^{-1} = \boldsymbol{\mathcal{O}}_p(1)\cdot \tag{B.2}$$

Proof (Lemma 3)

We have

$$n\widetilde{\boldsymbol{V}}_0 = \frac{n}{N^2}\sum_{h=1}^{H}\sum_{k\in\widetilde{\boldsymbol{S}}_h}\pi_k^{-2}\widehat{\boldsymbol{\epsilon}}_k\widehat{\boldsymbol{\epsilon}}_k^\top - \frac{n}{N^2}\sum_{h=1}^{H}\Big(n_h^{-1}\sum_{k\in\widetilde{\boldsymbol{S}}_h}\pi_k^{-1}\widehat{\boldsymbol{\epsilon}}_k\sum_{\ell\in\widetilde{\boldsymbol{S}}_h}\pi_\ell^{-1}\widehat{\boldsymbol{\epsilon}}_\ell^\top\Big)\cdot \tag{B.3}$$

It can be shown that

$$\widehat{\boldsymbol{\epsilon}}_k = \widetilde{\boldsymbol{\beta}}_0^\top\widehat{\boldsymbol{c}}_{0k}^\star, \tag{B.4}$$

Yves G. Berger
University of Southampton,
Southampton Statistical Sciences Research Institute,
Southampton, SO17 1BJ. United Kingdom
E-mail: Y.G.Berger@soton.ac.uk
URL: http://www.yvesberger.co.uk

where $\widetilde{\boldsymbol{\beta}}_0 := (\mathbf{1}_{d_\xi+p}^\top, \mathbf{0}_H^\top, \text{-}\boldsymbol{\beta}_0^\top)^\top$. Here, $\mathbf{1}_{d_\xi+p}$ denotes the $(d_\xi+p)\times 1$ unit vector and $\mathbf{0}_H$ the $H \times 1$ zero vector. Thus, using (B.4), expression (B.3) reduces to

$$n\widetilde{\boldsymbol{V}}_0 = -\widetilde{\boldsymbol{\beta}}_0^\top \left\{ \widehat{\boldsymbol{S}}_0 + \sum_{h=1}^{H} \psi_h^{-1} \varPsi_h^2 N_h^{-2} \widehat{\boldsymbol{C}}_{0h} \widehat{\boldsymbol{C}}_{0h}^\top \right\} \widetilde{\boldsymbol{\beta}}_0 \cdot \qquad \text{(B.5)}$$

where $\widehat{\boldsymbol{C}}_{0h}$, $\widehat{\boldsymbol{S}}_0$, ψ_h and $\varPsi_h$ are respectively defined by (20), (21), [C6] and [C7]. Consider the matrix of constants

$$\boldsymbol{\varSigma}_0 := -\widetilde{\boldsymbol{\beta}}_0^\top \left(\boldsymbol{S} + \sum_{h=1}^{H} \psi_h^{-1} \varPsi_h^2 \bar{\boldsymbol{\mathcal{C}}}_h \bar{\boldsymbol{\mathcal{C}}}_h^\top \right) \widetilde{\boldsymbol{\beta}}_0,$$

where $\bar{\boldsymbol{\mathcal{C}}}_h$ is defined in [C8]. We have that

$$n\widetilde{\boldsymbol{V}}_0 - \boldsymbol{\varSigma}_0 = -\widetilde{\boldsymbol{\beta}}_0^\top \left[\widehat{\boldsymbol{S}}_0 - \boldsymbol{S} + \sum_{h=1}^{H} \psi_h^{-1} \varPsi_h^2 \left\{ N_h^{-2} \widehat{\boldsymbol{C}}_{0h} \widehat{\boldsymbol{C}}_{0h}^\top - \bar{\boldsymbol{\mathcal{C}}}_h \bar{\boldsymbol{\mathcal{C}}}_h^\top \right\} \right] \widetilde{\boldsymbol{\beta}}_0 \cdot$$

Hence,

$$\| n\widetilde{\boldsymbol{V}}_0 \text{-} \boldsymbol{\varSigma}_0 \| \leqslant \| \widetilde{\boldsymbol{\beta}}_0 \|^2 \left\{ \| \widehat{\boldsymbol{S}}_0 \text{-} \boldsymbol{S} \| + \sum_{h=1}^{H} \psi_h^{-1} \varPsi_h^2 \left\| N_h^{-2} \widehat{\boldsymbol{C}}_{0h} \widehat{\boldsymbol{C}}_{0h}^\top \text{-} \bar{\boldsymbol{\mathcal{C}}}_h \bar{\boldsymbol{\mathcal{C}}}_h^\top \right\| \right\} \cdot \qquad \text{(B.6)}$$

The expression

$$N_h^{-2} \widehat{\boldsymbol{C}}_{0h} \widehat{\boldsymbol{C}}_{0h}^\top - \bar{\boldsymbol{\mathcal{C}}}_h \bar{\boldsymbol{\mathcal{C}}}_h^\top = \left\{ N_h^{-1} \widehat{\boldsymbol{C}}_{0h} - \bar{\boldsymbol{\mathcal{C}}}_h \right\} \bar{\boldsymbol{\mathcal{C}}}_h^\top + N_h^{-1} \widehat{\boldsymbol{C}}_{0h} \left\{ N_h^{-1} \widehat{\boldsymbol{C}}_{0h} - \bar{\boldsymbol{\mathcal{C}}}_h \right\}^\top$$

implies

$$\left\| N_h^{-2} \widehat{\boldsymbol{C}}_{0h} \widehat{\boldsymbol{C}}_{0h}^\top - \bar{\boldsymbol{\mathcal{C}}}_h \bar{\boldsymbol{\mathcal{C}}}_h^\top \right\| \leqslant \left\| N_h^{-1} \widehat{\boldsymbol{C}}_{0h} - \bar{\boldsymbol{\mathcal{C}}}_h \right\| \left\{ \| \bar{\boldsymbol{\mathcal{C}}}_h \| + N_h^{-1} \| \widehat{\boldsymbol{C}}_{0h} \| \right\} \cdot \qquad \text{(B.7)}$$

We have

$$N_h^{-1} \| \widehat{\boldsymbol{C}}_{0h} \| \leqslant \varPsi_h^{-1} N^{-1} \sum_{k \in \widetilde{\boldsymbol{S}}_h} \pi_k^{-1} \| \widehat{\boldsymbol{c}}_{0k}^\star \| \cdot$$

Thus, [C7] and [C10] imply

$$N_h^{-1} \| \widehat{\boldsymbol{C}}_{0h} \| = O_p(1) \cdot \qquad \text{(B.8)}$$

Thus, [C8], (B.7) and (B.8) imply

$$\left\| N_h^{-2} \widehat{\boldsymbol{C}}_{0h} \widehat{\boldsymbol{C}}_{0h}^\top - \bar{\boldsymbol{\mathcal{C}}}_h \bar{\boldsymbol{\mathcal{C}}}_h^\top \right\| = n_h^{-\frac{1}{2}} O_p(1) \cdot \qquad \text{(B.9)}$$

By using [C11] and substituting (B.9) into (B.6), we have

$$n\widetilde{\boldsymbol{V}}_0 - \boldsymbol{\varSigma}_0 = \boldsymbol{o}_p(1), \qquad \text{(B.10)}$$

which implies

$$n\widetilde{V}_{ij} - \Sigma_{ij} = o_p(1), \tag{B.11}$$

where $\widetilde{V}_{ij}$ and Σ_{ij} denote respectively the (i,j) component of $\widetilde{\boldsymbol{V}}_0$ and $\boldsymbol{\Sigma}_0$; that is,

$$\widetilde{V}_{ij} := N^{-2}\sum_{h=1}^{H}\Big(\sum_{k\in\widetilde{\boldsymbol{S}}_h}\pi_k^{-2}\widehat{\epsilon}_{ki}\widehat{\epsilon}_{kj} - \frac{1}{n_h}\sum_{k\in\widetilde{\boldsymbol{S}}_h}\pi_k^{-1}\widehat{\epsilon}_{ki}\sum_{\ell\in\widetilde{\boldsymbol{S}}_h}\pi_\ell^{-1}\widehat{\epsilon}_{\ell j}\Big).$$

where $\widehat{\epsilon}_{ki}$ denotes the i-th component of $\widehat{\boldsymbol{\epsilon}}_k$; that is,

$$\widehat{\epsilon}_{ki} = \widehat{a}_{0ki} - \boldsymbol{\beta}_{0i}^\top\widehat{\mathbf{f}}_k, \tag{B.12}$$

where $\widehat{a}_{0ki}$ is the i-th component of $\widehat{\boldsymbol{a}}_{0k}$ and $\boldsymbol{\beta}_{0i}$ is the i-th column of $\boldsymbol{\beta}_0$. Thus,

$$|\widetilde{V}_{ij}| \leqslant N^{-2}\sum_{h=1}^{H}\Big(\sum_{k\in\widetilde{\boldsymbol{S}}_h}\pi_k^{-2}|\widehat{\epsilon}_{ki}\widehat{\epsilon}_{kj}| + \frac{1}{n_h}\sum_{k\in\widetilde{\boldsymbol{S}}_h}\pi_k^{-1}|\widehat{\epsilon}_{ki}|\sum_{\ell\in\widetilde{\boldsymbol{S}}_h}\pi_\ell^{-1}|\widehat{\epsilon}_{\ell j}|\Big). \tag{B.13}$$

Using Cauchy's inequality, we have

$$\sum_{k\in\widetilde{\boldsymbol{S}}_h}\pi_k^{-2}|\widehat{\epsilon}_{ki}\widehat{\epsilon}_{kj}| \leqslant \Big(\sum_{k\in\widetilde{\boldsymbol{S}}_h}\pi_k^{-2}\ \widehat{\epsilon}_{ki}^2\sum_{\ell\in\widetilde{\boldsymbol{S}}_h}\pi_\ell^{-2}\ \widehat{\epsilon}_{\ell j}^2\Big)^{\frac{1}{2}}. \tag{B.14}$$

We also have

$$\sum_{k\in\widetilde{\boldsymbol{S}}}\pi_k^{-1}\ |\widehat{\epsilon}_{ki}| \leqslant \Big(n_h\sum_{k\in\widetilde{\boldsymbol{S}}}\pi_k^{-2}\ \widehat{\epsilon}_{ki}^2\Big)^{\frac{1}{2}}. \tag{B.15}$$

By substituting (B.14) and (B.15) into (B.13), we obtain

$$n|\widetilde{V}_{ij}| \leqslant 2\sum_{h=1}^{H}\Big(nN^{-2}\sum_{k\in\widetilde{\boldsymbol{S}}_h}\pi_k^{-2}\ \widehat{\epsilon}_{ki}^2\Big)^{\frac{1}{2}}\Big(nN^{-2}\sum_{\ell\in\widetilde{\boldsymbol{S}}_h}\pi_\ell^{-2}\ \widehat{\epsilon}_{\ell j}^2\Big)^{\frac{1}{2}}, \tag{B.16}$$

where $\widehat{\epsilon}_{ki}$ is defined by (B.12). Using (B.12), Minkowski inequality and $(\boldsymbol{\beta}_{0i}^\top\widehat{\mathbf{f}}_k)^2 \leqslant q\sum_{\ell=1}^{q}\beta_{0i\ell}^2\widehat{\mathrm{f}}_{k\ell}^2$, we obtain

$$\begin{aligned}\Big(\sum_{k\in\widetilde{\boldsymbol{S}}_h}\pi_k^{-2}\ \widehat{\epsilon}_{ki}^2\Big)^{\frac{1}{2}} &\leqslant \Big(\sum_{k\in\widetilde{\boldsymbol{S}}_h}\pi_k^{-2}\ \widehat{a}_{0ki}^2\Big)^{\frac{1}{2}} + \Big\{\sum_{k\in\widetilde{\boldsymbol{S}}_h}\pi_k^{-2}\ (\boldsymbol{\beta}_{0i}^\top\widehat{\mathbf{f}}_k)^2\Big\}^{\frac{1}{2}} \\ &\leqslant \Big(\sum_{k\in\widetilde{\boldsymbol{S}}_h}\pi_k^{-2}\ \widehat{a}_{0ki}^2\Big)^{\frac{1}{2}} + \sqrt{q}\sum_{\ell=1}^{q}|\beta_{0i\ell}|\Big\{\sum_{k\in\widetilde{\boldsymbol{S}}_h}\pi_k^{-2}\ \widehat{\mathrm{f}}_{k\ell}^2\Big\}^{\frac{1}{2}},\end{aligned} \tag{B.17}$$

where $\beta_{0i\ell}$ and $\widehat{\mathrm{f}}_{k\ell}$ are respectively the ℓ-th component of $\boldsymbol{\beta}_{0i}$ and $\widehat{\mathbf{f}}_k$.

Using conditions [C19] and [C20] and inequalities (B.16) and (B.17), we have that there exists a positive random variable v_{ij} such that

$$\mathbb{E}(v_{ij}) < \infty, \qquad n|\widetilde{V}_{ij}| \leqslant v_{ij} \quad \text{for all } n\cdot \tag{B.18}$$

The Dominated Convergence Theorem, (B.11) and (B.18) imply $n\mathbb{E}(\widetilde{V}_{ij}) - \Sigma_{ij} = o(1)$ or equivalently

$$n\mathbb{E}(\widetilde{\boldsymbol{V}}_0) - \boldsymbol{\Sigma}_0 = \boldsymbol{o}(1)\cdot \tag{B.19}$$

Minkowski inequality gives

$$\|\boldsymbol{\Sigma}_0 - n\boldsymbol{V}_0\| \leqslant \|n\mathbb{E}(\widetilde{\boldsymbol{V}}_0) - n\boldsymbol{V}_0\| + \|n\mathbb{E}(\widetilde{\boldsymbol{V}}_0) - \boldsymbol{\Sigma}_0\|,$$

which implies

$$\boldsymbol{\Sigma}_0 - n\boldsymbol{V}_0 = \boldsymbol{o}(1), \tag{B.20}$$

by using (31) and (B.19). Now, (B.10) and (B.20) implies

$$n(\widetilde{\boldsymbol{V}}_0 - \boldsymbol{V}_0) = \boldsymbol{o}(1), \tag{B.21}$$

Expression (B.5) can be re-written as

$$n\widetilde{\boldsymbol{V}}_0 = -\widetilde{\boldsymbol{\beta}}_0^\top \widehat{\boldsymbol{\Lambda}}_0 \; \widetilde{\boldsymbol{\beta}}_0, \tag{B.22}$$

where

$$\widehat{\boldsymbol{\Lambda}}_0 = \widehat{\boldsymbol{S}}_0 + \sum_{h=1}^{H} \psi_h^{-1} \Psi_h^2 N_h^{-2} \widehat{\boldsymbol{C}}_{0h} \widehat{\boldsymbol{C}}_{0h}^\top\cdot \tag{B.23}$$

Conditions [C7] and [C10] imply

$$N_h^{-1} \|\widehat{\boldsymbol{C}}_{0h}\| \leqslant \Psi_h^{-1} N^{-1} \sum_{k \in \widetilde{\boldsymbol{S}}_h} \pi_k^{-1} \|\widehat{\boldsymbol{c}}_{0k}^\star\| = O_p(1)\cdot$$

Using the last inequality and (B.1), expression (B.23) implies

$$\widehat{\boldsymbol{\Lambda}}_0 = \boldsymbol{O}_p(1), \tag{B.24}$$

because ψ_h, Ψ_h and H are finite.

It can be shown that

$$n\widehat{\boldsymbol{V}}_0 = -\widetilde{\boldsymbol{b}}_0^\top \widehat{\boldsymbol{\Lambda}}_0 \; \widetilde{\boldsymbol{b}}_0\cdot \tag{B.25}$$

where $\widetilde{\boldsymbol{b}}_0 := (\mathbf{1}^\top, \mathbf{0}_H^\top, -\widehat{\boldsymbol{b}}_0^\top)^\top$, $\mathbf{1}$ is the unit vector and $\widehat{\boldsymbol{V}}_0$ is defined by (26). Equations (B.22) and (B.25) imply

$$n\widehat{\boldsymbol{V}}_0 - n\widetilde{\boldsymbol{V}}_0 = \{\widetilde{\boldsymbol{\beta}}_0 - \widetilde{\boldsymbol{b}}_0\}^\top \widehat{\boldsymbol{\Lambda}}_0 \; \widetilde{\boldsymbol{b}}_0 + \widetilde{\boldsymbol{\beta}}_0^\top \widehat{\boldsymbol{\Lambda}}_0 \; \{\widetilde{\boldsymbol{\beta}}_0 - \widetilde{\boldsymbol{b}}_0\}\cdot \tag{B.26}$$

By using (B.24) and Lemma 1, expression (B.26) gives

$$n\widehat{\boldsymbol{V}}_0 - n\widetilde{\boldsymbol{V}}_0 = \boldsymbol{o}_p(1), \tag{B.27}$$

which implies (32), because of equation (B.21). □

Proof (Theorem 1) Condition [C8] implies

$$n^{\frac{1}{2}} N^{-1}\big(\widehat{\boldsymbol{C}}_0 \text{-} \boldsymbol{C}^{\star}\big) = \sum_{h=1}^{H} \Psi_h \psi_h^{-1} n_h^{\frac{1}{2}} \big(N_h^{-1}\widehat{\boldsymbol{C}}_{0h} \text{-} \bar{\boldsymbol{\mathcal{C}}}_h\big) = \boldsymbol{\mathcal{O}}_p(1), \tag{B.28}$$

where

$$\widehat{\boldsymbol{C}}_0 := \sum_{k\in\widetilde{\boldsymbol{S}}} \pi_k^{-1} \widehat{\boldsymbol{c}}_{0k}^{\star}\cdot$$

Oğuz-Alper & Berger (2016, Th.2) showed that [C4], [C9], [C10], (B.1), (B.2) and (B.28) imply (33). □

Proof (Theorem 2)
Berger & Torres (2016) showed that [C4], [C3], [C9], [C10], (B.28), (B.1) and (B.2) imply

$$N^{-1} n^{\frac{1}{2}} \widehat{\boldsymbol{A}}(\boldsymbol{\psi}_0) = N^{-1} n^{\frac{1}{2}} \ddot{\boldsymbol{A}}(\boldsymbol{\psi}_0) + \boldsymbol{o}_p(1), \tag{B.29}$$

where $\ddot{\boldsymbol{A}}(\boldsymbol{\psi}_0)$ is defined by (22).

Using the fact that $\widehat{\boldsymbol{c}}_k$ is a sub-vector of $\widehat{\boldsymbol{c}}_{0k}^{\star}$ (see (8)), condition [C11] and Lemma B1 imply $nN^{-2}\|(\sum_{k\in\widetilde{\boldsymbol{S}}} \pi_k^{-2}\, \widehat{\boldsymbol{c}}_k\, \widehat{\boldsymbol{c}}_k^{\top})^{-1}\| = O_p(1)$. Cauchy's inequality and [C10], we have $nN^{-2}\| \sum_{k\in\widetilde{\boldsymbol{S}}} \pi_k^{-2}\, \widehat{\boldsymbol{c}}_k\, \widehat{\boldsymbol{a}}_{0k}^{\top}\| = O_p(1)$. Hence

$$\widehat{\boldsymbol{B}}_0 = \boldsymbol{\mathcal{O}}_p(1)\cdot \tag{B.30}$$

Expressions(B.28), (22) and (B.30) imply

$$N^{-1} n^{\frac{1}{2}} \ddot{\boldsymbol{A}}(\boldsymbol{\psi}_0) = \boldsymbol{\mathcal{O}}_p(1)\cdot \tag{B.31}$$

Hence, (B.29) implies

$$N^{-1} n^{\frac{1}{2}} \widehat{\boldsymbol{A}}(\boldsymbol{\psi}_0) = \boldsymbol{\mathcal{O}}_p(1)\cdot \tag{B.32}$$

Under conditions [C15] and [C16], we have the Taylor expansion of $\widehat{\boldsymbol{A}}(\boldsymbol{\psi})$ around $\boldsymbol{\psi}_0$,

$$N^{-1}\widehat{\boldsymbol{A}}(\widehat{\boldsymbol{\psi}}) = N^{-1}\widehat{\boldsymbol{A}}(\boldsymbol{\psi}_0) + N^{-1}\frac{\partial \widehat{\boldsymbol{A}}(\boldsymbol{\psi}_0)}{\partial \boldsymbol{\psi}_0}(\widehat{\boldsymbol{\psi}}\text{-}\boldsymbol{\psi}_0) + \|\widehat{\boldsymbol{\psi}}\text{-}\boldsymbol{\psi}_0\|^2 \boldsymbol{\mathcal{O}}_p(1)\cdot \tag{B.33}$$

We have $\widehat{\boldsymbol{A}}(\widehat{\boldsymbol{\psi}}) = \boldsymbol{0}$, because $\widehat{\boldsymbol{\psi}}$ is the solution to (14). Hence, (B.32), (B.33) and [C17] imply $n^{\frac{1}{2}}(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0) = \boldsymbol{\mathcal{O}}_p(1)$. Theorem 2 follows. □

Appendix C

In order to simplify the notation, $\widehat{\boldsymbol{c}}_k^{\star}(\boldsymbol{\psi}_0,\boldsymbol{r})$, $\boldsymbol{g}_i(\boldsymbol{\tau}_0)$, $\boldsymbol{a}_i(\boldsymbol{\psi}_0,\boldsymbol{r})$ and $\widehat{\boldsymbol{a}}_k(\boldsymbol{\psi}_0,\boldsymbol{r})$ are respectively replaced by $\widehat{\boldsymbol{c}}_{0k}^{\star}$, $\boldsymbol{g}_{0i}$, $\boldsymbol{a}_{0i}$ and $\widehat{\boldsymbol{a}}_{0k}$. Hence, $\widehat{\boldsymbol{c}}_{0k}^{\star}=(\widehat{\boldsymbol{a}}_{0k}^{\top},\widehat{\boldsymbol{c}}_k^{\top})^{\top}$.

The operators $\mathbb{E}_r(\cdot)$ and $\mathbb{V}_r(\cdot)$ denote the expectation and variance with respect to the response mechanism. The operators $\mathbb{V}_d(\cdot\,|\,\boldsymbol{r})$ and $\mathbb{E}_d(\cdot\,|\,\boldsymbol{r})$ denote the conditional expectation and variance with respect to the sampling design, given $\boldsymbol{r}$. The operators $\mathbb{E}(\cdot)$ and $\mathbb{V}(\cdot)$ denote respectively the expectation and variance operator with respect to the response mechanism and the sampling design.

Proof (Lemma 2) Using (8), (9) and (B.28), we have that

$$N^{-1}\widehat{\mathbf{f}}_{\pi}=n^{-\frac{1}{2}}\boldsymbol{\mathcal{O}}_p(1),\quad \text{where } \widehat{\mathbf{f}}_{\pi}:=\sum_{k\in\widetilde{\boldsymbol{S}}}\pi_k^{-1}\widehat{\mathbf{f}}_k\cdot \tag{C.1}$$

Expression (22) can be re-written as

$$\ddot{\boldsymbol{A}}(\boldsymbol{\psi}_0)=\widehat{\boldsymbol{A}}_{0\pi}-\sum_{\ell\in\widetilde{\boldsymbol{S}}}\pi_\ell^{-2}\,\widehat{\boldsymbol{a}}_{0\ell}\,\widehat{\boldsymbol{c}}_\ell^{\top}\Bigg(\sum_{k\in\widetilde{\boldsymbol{S}}}\pi_k^{-2}\,\widehat{\boldsymbol{c}}_k\,\widehat{\boldsymbol{c}}_k^{\top}\Bigg)^{-1}\begin{pmatrix}\mathbf{0}_H\\ \widehat{\mathbf{f}}_\pi\end{pmatrix}\cdot \tag{C.2}$$

We have that

$$\sum_{k\in\widetilde{\boldsymbol{S}}}\pi_k^{-2}\,\widehat{\boldsymbol{a}}_{0k}\,\widehat{\boldsymbol{c}}_k^{\top}=(\widehat{\boldsymbol{S}}_{za}^{\top},\widehat{\boldsymbol{S}}_{\mathrm{f}a}^{\top}), \tag{C.3}$$

$$\Bigg(\sum_{k\in\widetilde{\boldsymbol{S}}}\pi_k^{-2}\,\widehat{\boldsymbol{c}}_k\,\widehat{\boldsymbol{c}}_k^{\top}\Bigg)^{-1}=\begin{Bmatrix}\bullet & -\widehat{\boldsymbol{S}}_{zz}^{-1}\widehat{\boldsymbol{S}}_{z\mathrm{f}}(\widehat{\boldsymbol{S}}_{\mathrm{ff}}-\widehat{\boldsymbol{S}}_{z\mathrm{f}}^{\top}\widehat{\boldsymbol{S}}_{zz}^{-1}\widehat{\boldsymbol{S}}_{z\mathrm{f}})^{-1}\\ \bullet & (\widehat{\boldsymbol{S}}_{\mathrm{ff}}-\widehat{\boldsymbol{S}}_{z\mathrm{f}}^{\top}\widehat{\boldsymbol{S}}_{zz}^{-1}\widehat{\boldsymbol{S}}_{z\mathrm{f}})^{-1}\end{Bmatrix}, \tag{C.4}$$

where $\bullet$ represents sub-matrices for which expressions are not needed. By substituting (C.3) and (C.4) into (C.2), we obtain

$$\ddot{\boldsymbol{A}}(\boldsymbol{\psi}_0)=\widehat{\boldsymbol{A}}_{0\pi}-\widehat{\boldsymbol{b}}_0^{\top}\,\widehat{\mathbf{f}}_\pi\cdot$$

where $\widehat{\boldsymbol{A}}_{0\pi}$ and $\widehat{\boldsymbol{b}}_0$ are respectively defined by (23) and (27).

Lemma 1 and (C.1) imply

$$N^{-1}\ddot{\boldsymbol{A}}(\boldsymbol{\psi}_0)-\bar{\boldsymbol{\epsilon}}_\pi=(\widehat{\boldsymbol{b}}_0-\boldsymbol{\beta}_0)^{\top}N^{-1}\widehat{\mathbf{f}}_\pi=n^{-\frac{1}{2}}\boldsymbol{o}_p(1), \tag{C.5}$$

where $\bar{\boldsymbol{\epsilon}}_\pi$ is defined by (46).

Using (B.4), we have that $\bar{\boldsymbol{\epsilon}}_\pi=\widetilde{\boldsymbol{\beta}}_0^{\top}N^{-1}\sum_{k\in\widetilde{\boldsymbol{S}}}\pi_k^{-1}\widehat{\boldsymbol{c}}_{0k}^{\star}$. Thus, [C10] imply

$$\bar{\boldsymbol{\epsilon}}_\pi=\boldsymbol{\mathcal{O}}_p(1)\cdot \tag{C.6}$$

We have

$$N^{-2}\ddot{\boldsymbol{A}}(\boldsymbol{\psi}_0)\ddot{\boldsymbol{A}}(\boldsymbol{\psi}_0)^{\top}-\bar{\boldsymbol{\epsilon}}_\pi\bar{\boldsymbol{\epsilon}}_\pi^{\top}=\Big\{N^{-1}\ddot{\boldsymbol{A}}(\boldsymbol{\psi}_0)-\bar{\boldsymbol{\epsilon}}_\pi\Big\}N^{-1}\ddot{\boldsymbol{A}}(\boldsymbol{\psi}_0)^{\top}$$

$$+\,\bar{\boldsymbol{\epsilon}}_\pi\Big\{N^{-1}\ddot{\boldsymbol{A}}(\boldsymbol{\psi}_0)-\bar{\boldsymbol{\epsilon}}_\pi\Big\}^\top\cdot$$

By combining the last expressions with (B.31), (C.5) and (C.6), we obtain

$$N^{-2}\ddot{\boldsymbol{A}}(\boldsymbol{\psi}_0)\ddot{\boldsymbol{A}}(\boldsymbol{\psi}_0)^\top-\bar{\boldsymbol{\epsilon}}_\pi\bar{\boldsymbol{\epsilon}}_\pi^\top=n^{-\frac{1}{2}}\boldsymbol{\mathcal{O}}_p(1)\cdot \tag{C.7}$$

Let $\widehat{A}_{0\pi}^{(i)}$ and $\widehat{a}_{0ki}$ denote respectively the i-th component of $\widehat{\boldsymbol{A}}_{0\pi}$ and $\widehat{\boldsymbol{a}}_{0k}$. We have

$$N^{-1}|\widehat{A}_{0\pi}^{(i)}|\leqslant N^{-1}\sum_{k\in\widetilde{\boldsymbol{S}}}\pi_k^{-1}|\widehat{a}_{0ki}|\leqslant\Big(nN^{-2}\sum_{k\in\widetilde{\boldsymbol{S}}}\pi_k^{-2}\ \widehat{a}_{0ki}^2\Big)^{\frac{1}{2}}\cdot$$

Condition [C19] implies that for all i

$$N^{-1}|\widehat{A}_{0\pi}^{(i)}|\leqslant\mathcal{H}_i^{\frac{1}{2}},\quad\text{with}\quad\mathbb{E}(\mathcal{H}_i^{\frac{1}{2}})<\infty\cdot \tag{C.8}$$

Let C_i and $\widehat{C}_i$ denote respectively the i-th component of $\boldsymbol{C}$, $\widehat{\boldsymbol{C}}$. Note that $\widehat{C}_i-C_i=0$ for $i\leqslant H$. For $i>H$, we have

$$N^{-1}|\widehat{C}_i-C_i|\leqslant N^{-1}\sum_{k\in\widetilde{\boldsymbol{S}}}\pi_k^{-1}\ |\widehat{\mathrm{f}}_{ki}|\leqslant\Big(nN^{-2}\sum_{k\in\widetilde{\boldsymbol{S}}}\pi_k^{-2}\ \widehat{\mathrm{f}}_{ki}^2\Big)^{\frac{1}{2}}\cdot$$

Condition [C20] implies that for all i

$$N^{-1}|\widehat{C}_i-C_i|\leqslant\mathcal{F}_i^{\frac{1}{2}},\quad\text{with}\quad\mathbb{E}(\mathcal{F}_i^{\frac{1}{2}})<\infty\cdot \tag{C.9}$$

Inequalities (C.8), (C.9) and [C21] imply that the components of $N^{-1}\ddot{\boldsymbol{A}}(\boldsymbol{\psi}_0)$ are dominated by positive random variables with finite expectations. Conditions [C19] and [C20] imply that the components of $\bar{\boldsymbol{\epsilon}}_\pi$ are dominated by positive random variables with finite expectations. Hence, the matrices in the left hand side of (C.5) and (C.7) are dominated by positive random variables with finite expectations. Thus,

$$N^{-1}\mathbb{E}\big\{\ddot{\boldsymbol{A}}(\boldsymbol{\psi}_0)\big\}=\mathbb{E}\big\{\bar{\boldsymbol{\epsilon}}_\pi\big\}+\boldsymbol{\mathcal{O}}(1),$$
$$N^{-2}\mathbb{E}\big\{\ddot{\boldsymbol{A}}(\boldsymbol{\psi}_0)\ddot{\boldsymbol{A}}(\boldsymbol{\psi}_0)^\top\big\}=\mathbb{E}\big\{\bar{\boldsymbol{\epsilon}}_\pi\bar{\boldsymbol{\epsilon}}_\pi^\top\big\}+\boldsymbol{\mathcal{O}}(1),$$

using the Dominated Convergence Theorem. The last two expressions imply that the variance of $N^{-1}\ddot{\boldsymbol{A}}(\boldsymbol{\psi}_0)$ is given by

$$\boldsymbol{V}_0=\mathbb{V}(\bar{\boldsymbol{\epsilon}}_\pi)+\boldsymbol{\mathcal{O}}(1)\cdot \tag{C.10}$$

The independence between the response mechanism and the sampling design implies that we can consider that non-response occurs before sampling as in Fay (1991) and Shao & Steel (1999). Thus, the variance is

$$\mathbb{V}(\bar{\boldsymbol{\epsilon}}_\pi)=\mathbb{E}_r\{\mathbb{V}_d(\bar{\boldsymbol{\epsilon}}_\pi\mid\boldsymbol{r})\}+\mathbb{V}_r\{\mathbb{E}_d(\bar{\boldsymbol{\epsilon}}_\pi\mid\boldsymbol{r})\}\cdot \tag{C.11}$$

We have

$$\begin{aligned}\mathbb{E}_d(\bar{\boldsymbol{\epsilon}}_\pi \mid \boldsymbol{r}) &= N^{-1}\sum_{i\in\mathcal{U}}\boldsymbol{a}_{0i} - \boldsymbol{\beta}_0^\top N^{-1}\sum_{i\in\mathcal{U}}\mathbf{f}_i \\ &= N^{-1}\sum_{i\in\mathcal{U}}\left\{\boldsymbol{g}_{0i}^\top r_i\,\rho_i^{-1}, \boldsymbol{\xi}_i^\top(r_i-\rho_i)\right\}^\top - \boldsymbol{\beta}_0^\top N^{-1}\sum_{i\in\mathcal{U}}\mathbf{f}_i \\ &= N^{-1}\sum_{i\in\mathcal{U}}\boldsymbol{\kappa}_{0i}\,r_i - N^{-1}\sum_{i\in\mathcal{U}}\left\{\mathbf{0}^\top, \boldsymbol{\xi}_i^\top\rho_i\right\}^\top - \boldsymbol{\beta}_0^\top N^{-1}\sum_{i\in\mathcal{U}}\mathbf{f}_i,\end{aligned}$$

where ρ_i is defined by (4) and

$$\boldsymbol{\kappa}_{0i} := \left\{\boldsymbol{g}_{0i}^\top\rho_i^{-1}, \boldsymbol{\xi}_i^\top\right\}^\top\cdot$$

Thus, as $r_i \perp\!\!\!\perp r_j$ for $i \neq j$ and $\mathbb{V}_r(r_i) = \rho_i(1-\rho_i)$, we have

$$N^2\mathbb{V}_r\{\mathbb{E}_d(\bar{\boldsymbol{\epsilon}}_\pi \mid \boldsymbol{r})\} = \sum_{i\in\mathcal{U}}\boldsymbol{\kappa}_{0i}^\top\boldsymbol{\kappa}_{0i}\;\rho_i(1-\rho_i)\cdot \tag{C.12}$$

The last expression implies

$$\begin{aligned}\left\|N^2\mathbb{V}_r\{\mathbb{E}_d(\bar{\boldsymbol{\epsilon}}_\pi \mid \boldsymbol{r})\}\right\| &\leqslant \sum_{i\in\mathcal{U}}\|\boldsymbol{\kappa}_{0i}\|^2\;\rho_i \\ &\leqslant \sum_{i\in\mathcal{U}}\|\boldsymbol{g}_{0i}\|^2\rho_i^{-1} + \sum_{i\in\mathcal{U}}\|\boldsymbol{\xi}_i\|^2\cdot\end{aligned}$$

The last inequality, $\boldsymbol{\beta}_0 = \boldsymbol{\mathcal{O}}(1)$, [C2], [C12] and [C13] imply

$$\mathbb{V}_r\{\mathbb{E}_d(\bar{\boldsymbol{\epsilon}}_\pi \mid \boldsymbol{r})\} = N^{-1}\boldsymbol{\mathcal{O}}(1)\cdot \tag{C.13}$$

Under the two-stage design, we have

$$\mathbb{E}_r\{\mathbb{V}_d(\bar{\boldsymbol{\epsilon}}_\pi \mid \boldsymbol{r})\} = \mathbb{E}_r(\boldsymbol{V}_{d;a}) + \mathbb{E}_r(\boldsymbol{V}_{d;b}), \tag{C.14}$$

where

$$\begin{aligned}\boldsymbol{V}_{d;a} &:= \mathbb{V}_d^{(1)}\{\mathbb{E}_d^{(2)}(\bar{\boldsymbol{\epsilon}}_\pi \mid \widetilde{\boldsymbol{S}}, \boldsymbol{r}) \mid \boldsymbol{r}\}, \\ \boldsymbol{V}_{d;b} &:= \mathbb{E}_d^{(1)}\{\mathbb{V}_d^{(2)}(\bar{\boldsymbol{\epsilon}}_\pi \mid \widetilde{\boldsymbol{S}}, \boldsymbol{r}) \mid \boldsymbol{r}\}\cdot\end{aligned} \tag{C.15}$$

Here, $\mathbb{E}_d^{(1)}$ and $\mathbb{E}_d^{(2)}$ respectively denote the first and second stage design expectation operator. The operators $\mathbb{V}_d^{(1)}$ and $\mathbb{V}_d^{(2)}$ denote the first and second stage design variance.

As the first stage is a stratified with-replacement design, we have the following Hansen & Hurwitz's (1943) variance

$$\boldsymbol{V}_{d;a} = N^{-2}\sum_{h=1}^{H}\left(\sum_{k\in\widetilde{U}_h}\pi_k^{-1}\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^\top - \frac{1}{n_h}\sum_{k\in\widetilde{U}_h}\boldsymbol{\epsilon}_k\sum_{\ell\in\widetilde{U}_h}\boldsymbol{\epsilon}_\ell^\top\right), \tag{C.16}$$

where

$$\boldsymbol{\epsilon}_k := \mathbb{E}_d^{(2)}(\widehat{\boldsymbol{\epsilon}}_k \mid \widetilde{\boldsymbol{S}}, \boldsymbol{r}) = \sum_{i\in\widetilde{U}_k} \boldsymbol{\epsilon}_{i|k}, \tag{C.17}$$

$$\boldsymbol{\epsilon}_{i|k} := \boldsymbol{a}_{0i} - \boldsymbol{\beta}_0^\top \mathbf{f}_i\cdot \tag{C.18}$$

We have

$$\mathbb{V}_r(\boldsymbol{\epsilon}_{i|k}) = (1-\rho_i)\big(\boldsymbol{g}_{0i}^\top, \rho_i\boldsymbol{\xi}_i^\top\big)^\top\big(\boldsymbol{g}_{0i}^\top, \rho_i\boldsymbol{\xi}_i^\top\big) \tag{C.19}$$

$$\mathbb{E}_r(\boldsymbol{\epsilon}_{i|k}) = \big(\boldsymbol{g}_{0i}^\top, \boldsymbol{0}^\top\big)^\top - \boldsymbol{\beta}_0^\top\mathbf{f}_i\cdot \tag{C.20}$$

We have that $r_i \perp\!\!\!\perp r_j$ for $i \neq j$ implies $\boldsymbol{\epsilon}_{i|k} \perp\!\!\!\perp \boldsymbol{\epsilon}_{j|k}$ for $i \neq j$. Thus,

$$\mathbb{E}_r(\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_\ell^\top) = \mathbb{E}_r(\boldsymbol{\epsilon}_k)\ \mathbb{E}_r(\boldsymbol{\epsilon}_\ell)^\top, \quad \text{for } k \neq \ell\cdot$$

Now, (C.16) implies

$$\mathbb{E}_r(\boldsymbol{V}_{d;a}) = N^{-2}\sum_{h=1}^{H}\sum_{k\in\widetilde{U}_h}(\pi_k^{-1} - n_h^{-1})\mathbb{V}_r(\boldsymbol{\epsilon}_k) + \boldsymbol{V}_{d,r}, \tag{C.21}$$

where

$$\boldsymbol{V}_{d,r} := N^{-2}\sum_{h=1}^{H}\Big\{\sum_{k\in\widetilde{U}_h}\pi_k^{-1}\mathbb{E}_r(\boldsymbol{\epsilon}_k)\mathbb{E}_r(\boldsymbol{\epsilon}_k)^\top - n_h^{-1}\sum_{k\in\widetilde{U}_h}\mathbb{E}_r(\boldsymbol{\epsilon}_k)\sum_{\ell\in\widetilde{U}_h}\mathbb{E}_r(\boldsymbol{\epsilon}_\ell)^\top\Big\}\cdot \tag{C.22}$$

Using (C.17) and (C.19), we have

$$\|\mathbb{V}_r(\boldsymbol{\epsilon}_k)\| \leqslant \sum_{i\in\widetilde{U}_k}\|\mathbb{V}_r(\boldsymbol{\epsilon}_{i|k})\| \leqslant O(1)\sum_{i\in\widetilde{U}_k}\|\boldsymbol{g}_{0i}\|^2 + \sum_{i\in\widetilde{U}_k}\|\boldsymbol{\xi}_i\|^2\cdot$$

The last inequality combined with [C4], [C6], [C12], and [C13] implies

$$\begin{aligned}
N^{-2}\Big\|\sum_{h=1}^{H}\sum_{k\in\widetilde{U}_h}(\pi_k^{-1} - n_h^{-1})\mathbb{V}_r(\boldsymbol{\epsilon}_k)\Big\| &\leqslant N^{-1}n^{-1}\sum_{h=1}^{H}\sum_{k\in\widetilde{U}_h}|nN^{-1}\pi_k^{-1} - nN^{-1}n_h^{-1}|\|\mathbb{V}_r(\boldsymbol{\epsilon}_k)\| \\
&\leqslant N^{-1}n^{-1}O_p(1)\sum_{h=1}^{H}\sum_{k\in\widetilde{U}_h}\|\mathbb{V}_r(\boldsymbol{\epsilon}_k)\| \\
&\leqslant N^{-1}n^{-1}\Big\{O(1)\sum_{i\in\mathcal{U}}\|\boldsymbol{g}_{0i}\|^2 + O(1)\sum_{i\in\mathcal{U}}\|\boldsymbol{\xi}_i\|^2\Big\} \\
&= n^{-1}O(1)\cdot
\end{aligned} \tag{C.23}$$

We have that $\mathbb{E}_r(\boldsymbol{\epsilon}_k) = \sum_{i\in\widetilde{U}_k}(\boldsymbol{g}_{0i}^\top, \boldsymbol{0}^\top)^\top - \boldsymbol{\beta}_0^\top\sum_{i\in\widetilde{U}_k}\mathbf{f}_i$ implies

$$\|\mathbb{E}_r(\boldsymbol{\epsilon}_k)\| \leqslant \sum_{i\in\widetilde{U}_k}\|\boldsymbol{g}_{0i}\| + \|\boldsymbol{\beta}_0\|\sum_{i\in\widetilde{U}_k}\|\mathbf{f}_i\|,$$

Minkowski inequality implies

$$\Big(\sum_{h=1}^{H}\sum_{k\in\widetilde{U}_h}\|\mathbb{E}_r(\boldsymbol{\epsilon}_k)\|^2\Big)^{\frac{1}{2}} \leqslant \Big\{\sum_{h=1}^{H}\sum_{k\in\widetilde{U}_h}\Big(\sum_{i\in\widetilde{U}_k}\|\boldsymbol{g}_{0i}\|\Big)^2\Big\}^{\frac{1}{2}} + \|\boldsymbol{\beta}_0\|\Big\{\sum_{h=1}^{H}\sum_{k\in\widetilde{U}_h}\Big(\sum_{i\in\widetilde{U}_k}\|\mathbf{f}_i\|\Big)^2\Big\}^{\frac{1}{2}}$$
$$\leqslant \Big\{\sum_{h=1}^{H}\sum_{k\in\widetilde{U}_h}\widetilde{N}_k\sum_{i\in\widetilde{U}_k}\|\boldsymbol{g}_{0i}\|^2\Big\}^{\frac{1}{2}} + \|\boldsymbol{\beta}_0\|\Big\{\sum_{h=1}^{H}\sum_{k\in\widetilde{U}_h}\widetilde{N}_k\sum_{i\in\widetilde{U}_k}\|\mathbf{f}_i\|^2\Big\}^{\frac{1}{2}}$$
$$\leqslant O(1)\Big\{\sum_{i\in\mathcal{U}}\|\boldsymbol{g}_{0i}\|^2\Big\}^{\frac{1}{2}} + O(1)\|\boldsymbol{\beta}_0\|\Big\{\sum_{i\in\mathcal{U}}\|\mathbf{f}_i\|^2\Big\}^{\frac{1}{2}}, \tag{C.24}$$

because of condition [C1]. Now (C.24), [C13] and [C14] imply

$$N^{-1}\sum_{h=1}^{H}\sum_{k\in\widetilde{U}_h}\|\mathbb{E}_r(\boldsymbol{\epsilon}_k)\|^2 = O(1)\cdot \tag{C.25}$$

Using [C4] and [C6], definition (C.22) implies

$$\|\boldsymbol{V}_{d,r}\| \leqslant n^{-1}N^{-1}\sum_{h=1}^{H}\sum_{k\in\widetilde{U}_h}\|\mathbb{E}_r(\boldsymbol{\epsilon}_k)\|^2 + N^{-2}\sum_{h=1}^{H}n_h^{-1}\Big(\sum_{k\in\widetilde{U}_h}\|\mathbb{E}_r(\boldsymbol{\epsilon}_k)\|\Big)^2$$
$$\leqslant n^{-1}N^{-1}\sum_{h=1}^{H}\sum_{k\in\widetilde{U}_h}\|\mathbb{E}_r(\boldsymbol{\epsilon}_k)\|^2 + N^{-2}\sum_{h=1}^{H}n_h^{-1}N_h\sum_{k\in\widetilde{U}_h}\|\mathbb{E}_r(\boldsymbol{\epsilon}_k)\|^2$$
$$\leqslant n^{-1}N^{-1}\sum_{h=1}^{H}\sum_{k\in\widetilde{U}_h}\|\mathbb{E}_r(\boldsymbol{\epsilon}_k)\|^2 + n^{-1}N^{-1}\sum_{h=1}^{H}\Psi_h\psi_h^{-1}\sum_{k\in\widetilde{U}_h}\|\mathbb{E}_r(\boldsymbol{\epsilon}_k)\|^2$$
$$\leqslant n^{-1}N^{-1}\sum_{h=1}^{H}\sum_{k\in\widetilde{U}_h}\|\mathbb{E}_r(\boldsymbol{\epsilon}_k)\|^2 + O(1)n^{-1}N^{-1}\sum_{h=1}^{H}\sum_{k\in\widetilde{U}_h}\|\mathbb{E}_r(\boldsymbol{\epsilon}_k)\|^2$$
$$= O(1)n^{-1}N^{-1}\sum_{h=1}^{H}\sum_{k\in\widetilde{U}_h}\|\mathbb{E}_r(\boldsymbol{\epsilon}_k)\|^2$$
$$= n^{-1}O(1), \tag{C.26}$$

using (C.25).

Expression (C.21) combined with (C.23) and (C.26) implies

$$\mathbb{E}_r(\boldsymbol{V}_{d;a}) = n^{-1}\boldsymbol{\mathcal{O}}(1)\cdot \tag{C.27}$$

We also need to derive the order of the second term of (C.14): $\mathbb{E}_r(\boldsymbol{V}_{d;b})$, where $\boldsymbol{V}_{d;b}$ is defined by (C.15). We have that (46) and (C.15) imply

$$\boldsymbol{V}_{d;b} = N^{-2}\sum_{k=1}^{N}\pi_k^{-1}\mathbb{V}_d^{(2)}(\widehat{\boldsymbol{\epsilon}}_k|\widetilde{\boldsymbol{S}},\boldsymbol{r}), \tag{C.28}$$

where $\widehat{\boldsymbol{\epsilon}}_k$ defined by (29) can be re-written as

$$\widehat{\boldsymbol{\epsilon}}_k = \sum_{i\in \boldsymbol{S}_k} \pi_{i|k}^{-1} \boldsymbol{\epsilon}_{i|k},$$

with $\boldsymbol{\epsilon}_{i|k}$ defined by (C.18) and

$$\mathbb{V}_d^{(2)}(\widehat{\boldsymbol{\epsilon}}_k|\widetilde{\boldsymbol{S}}, \boldsymbol{r}) = \sum_{i\in\widetilde{U}_k}\sum_{j\in\widetilde{U}_k} D_{ij|k}\ \boldsymbol{\epsilon}_{i|k}\ \boldsymbol{\epsilon}_{j|k}^{\top}.$$

Here, $D_{ij|k}$ is defined by

$$D_{ij|k} = \pi_{ij|k}\pi_{i|k}^{-1}\,\pi_{j|k}^{-1} - 1,$$

where $\pi_{ij|k}$ denotes the conditional joint-inclusion probability of units $i, j \in \widetilde{U}_k$.

Condition [C1] implies that there exists a constant γ_2 such that

$$\widetilde{n}_k^{-1}\widetilde{N}_k \leqslant \gamma_2, \tag{C.29}$$

where $\widetilde{n}_k$ and $\widetilde{N}_k$ be respectively the sample and population size of $\widetilde{U}_k$. Using [C5], (C.29) and the fact that $\widetilde{U}_k$ is a finite set, we have that there exists a finite constants γ_3 such that for all k

$$\widetilde{n}_k^2\widetilde{N}_k^{-2}\sum_{i\in\widetilde{U}_k}\sum_{\substack{j\in\widetilde{U}_k\\ j\neq i}} D_{ij|k}^2 \leqslant \gamma_3. \tag{C.30}$$

Furthermore,

$$\begin{aligned}\|\mathbb{E}_r(\boldsymbol{V}_{d;b})\| &\leqslant N^{-2}\sum_{k=1}^{N}\pi_k^{-1}\sum_{i\in\widetilde{U}_k}\sum_{j\in\widetilde{U}_k} D_{ij|k}\ \|\mathbb{E}_r(\boldsymbol{\epsilon}_{i|k}\ \boldsymbol{\epsilon}_{j|k}^{\top})\| \\ &= N^{-2}\sum_{k=1}^{N}\pi_k^{-1}\Big\{\sum_{i\in\widetilde{U}_k}\sum_{\substack{j\in\widetilde{U}_k\\ j\neq i}} D_{ij|k}\,\|\mathbb{E}_r(\boldsymbol{\epsilon}_{i|k}\ \boldsymbol{\epsilon}_{j|k}^{\top})\| \\ &\quad + \sum_{i\in\widetilde{U}_k}\pi_{i|k}^{-1}(1-\pi_{i|k})\,\|\mathbb{E}_r(\boldsymbol{\epsilon}_{i|k}\ \boldsymbol{\epsilon}_{i|k}^{\top})\|\Big\}.\end{aligned}$$

Cauchy's inequality, [C4], (C.29), [C5] and (C.30) imply

$$\begin{aligned}\|\mathbb{E}_r(\boldsymbol{V}_{d;b})\| \leqslant N^{-2}\sum_{k=1}^{N}\pi_k^{-1}\Big\{&\Big(\sum_{i\in\widetilde{U}_k}\sum_{\substack{j\in\widetilde{U}_k\\ j\neq i}} D_{ij|k}^2\Big)^{\frac{1}{2}}\Big(\sum_{i\in\widetilde{U}_k}\sum_{\substack{j\in\widetilde{U}_k\\ j\neq i}}\|\mathbb{E}_r(\boldsymbol{\epsilon}_{i|k}\ \boldsymbol{\epsilon}_{j|k}^{\top})\|^2\Big)^{\frac{1}{2}} \\ &+ \sum_{i\in\widetilde{U}_k}\pi_{i|k}^{-1}(1-\pi_{i|k})\,\|\mathbb{E}_r(\boldsymbol{\epsilon}_{i|k}\ \boldsymbol{\epsilon}_{i|k}^{\top})\|\Big\}\end{aligned}$$

$$
\begin{aligned}
&\leqslant O(1)\gamma_3^{\frac{1}{2}} n^{-1}N^{-1}\sum_{k=1}^{N}\Big\{\widetilde{n}_k^{-1}\widetilde{N}_k\Big(\sum_{i\in\widetilde{U}_k}\sum_{\substack{j\in\widetilde{U}_k\\ j\neq i}}\|\mathbb{E}_r(\boldsymbol{\epsilon}_{i|k}\ \boldsymbol{\epsilon}_{j|k}^\top)\|^2\Big)^{\frac{1}{2}}\\
&\quad+\sum_{i\in\widetilde{U}_k}\pi_{i|k}^{-1}(1-\pi_{i|k})\,\|\mathbb{E}_r(\boldsymbol{\epsilon}_{i|k}\ \boldsymbol{\epsilon}_{i|k}^\top)\|\Big\}\\
&\leqslant O(1)n^{-1}N^{-1}\sum_{k=1}^{N}\Big\{\Big(\sum_{i\in\widetilde{U}_k}\sum_{\substack{j\in\widetilde{U}_k\\ j\neq i}}\|\mathbb{E}_r(\boldsymbol{\epsilon}_{i|k}\ \boldsymbol{\epsilon}_{j|k}^\top)\|^2\Big)^{\frac{1}{2}}\\
&\quad+O(1)\sum_{i\in\widetilde{U}_k}\|\mathbb{E}_r(\boldsymbol{\epsilon}_{i|k}\ \boldsymbol{\epsilon}_{i|k}^\top)\|\Big\}\cdot
\end{aligned}
\tag{C.31}
$$

Using $\boldsymbol{\epsilon}_{i|k} \perp\!\!\!\perp \boldsymbol{\epsilon}_{j|k}$ for $i \neq j$, we have that

$$
\mathbb{E}_r(\boldsymbol{\epsilon}_{i|k}\ \boldsymbol{\epsilon}_{j|k}^\top) = \mathbb{E}_r(\boldsymbol{\epsilon}_{i|k})\mathbb{E}_r(\boldsymbol{\epsilon}_{j|k})^\top, \quad \text{when } i \neq j\cdot
$$

Thus, (C.31) implies

$$
\|\mathbb{E}_r(\boldsymbol{V}_{d;b})\| \leqslant O(1)n^{-1}N^{-1}\sum_{k=1}^{N}\Big\{\sum_{i\in\widetilde{U}_k}\|\mathbb{E}_r(\boldsymbol{\epsilon}_{i|k})\|^2 + O(1)\sum_{i\in\widetilde{U}_k}\|\mathbb{E}_r(\boldsymbol{\epsilon}_{i|k}\boldsymbol{\epsilon}_{i|k}^\top)\|\Big\}\cdot
\tag{C.32}
$$

Since, $\mathbb{E}_r(\boldsymbol{\epsilon}_{i|k}\ \boldsymbol{\epsilon}_{i|k}^\top) = \mathbb{V}_r(\boldsymbol{\epsilon}_{i|k}) + \mathbb{E}_r(\boldsymbol{\epsilon}_{i|k})\mathbb{E}_r(\boldsymbol{\epsilon}_{i|k})^\top$, expression (C.19) imply

$$
\|\mathbb{E}_r(\boldsymbol{\epsilon}_{i|k}\ \boldsymbol{\epsilon}_{i|k}^\top)\| \leqslant \big\|\big(\boldsymbol{g}_{0i}^\top, \rho_i\boldsymbol{\xi}_i^\top\big)^\top\big\|^2 + \|\mathbb{E}_r(\boldsymbol{\epsilon}_{i|k})\|^2\cdot
\tag{C.33}
$$

Using [C12], [C13], [C14], (C.33) and Minkowski inequality, inequality (C.32) implies

$$
\begin{aligned}
\|\mathbb{E}_r(\boldsymbol{V}_{d;b})\| &\leqslant O(1)n^{-1}\Bigg[O(1)\Big\{\Big(N^{-1}\sum_{i\in\mathcal{U}}\|\boldsymbol{g}_{0i}\|^2 + N^{-1}\sum_{i\in\mathcal{U}}\|\boldsymbol{\xi}_i\|^2\Big)^{\frac{1}{2}}\Big\}^2\\
&\qquad + O(1)N^{-1}\sum_{k=1}^{N}\sum_{i\in\widetilde{U}_k}\|\mathbb{E}_r(\boldsymbol{\epsilon}_{i|k})\|^2\Bigg]\\
&\leqslant O(1)n^{-1} + O(1)n^{-1}N^{-1}\sum_{k=1}^{N}\sum_{i\in\widetilde{U}_k}\|\mathbb{E}_r(\boldsymbol{\epsilon}_{i|k})\|^2\cdot
\end{aligned}
\tag{C.34}
$$

Using [C13], [C14] and Minkowski inequality, we have that (C.20) implies

$$
\begin{aligned}
\Big(N^{-1}\sum_{k=1}^{N}\sum_{i\in\widetilde{U}_k}\|\mathbb{E}_r(\boldsymbol{\epsilon}_{i|k})\|^2\Big)^{\frac{1}{2}} &\leqslant \Big\{N^{-1}\sum_{i\in\mathcal{U}}\Big(\|\boldsymbol{g}_{0i}\| + \|\boldsymbol{\beta}_0\|\|\mathbf{f}_i\|\Big)^2\Big\}^{\frac{1}{2}}\\
&\leqslant \Big\{N^{-1}\sum_{i\in\mathcal{U}}\|\boldsymbol{g}_{0i}\|^2\Big\}^{\frac{1}{2}} + \|\boldsymbol{\beta}_0\|\Big\{N^{-1}\sum_{i\in\mathcal{U}}\|\mathbf{f}_i\|^2\Big\}^{\frac{1}{2}}
\end{aligned}
$$

$$= O(1)\cdot$$

The last inequality combined with (C.34) implies

$$\|\mathbb{E}_r(\boldsymbol{V}_{d;b})\| = n^{-1}O(1)\cdot \tag{C.35}$$

Hence, expressions (C.10), (C.11), (C.13), (C.14), (C.27) and (C.35) imply

$$n\boldsymbol{V}_0 = \boldsymbol{\mathcal{O}}(1) + nN^{-1}\boldsymbol{\mathcal{O}}(1) + \boldsymbol{o}(1) = \boldsymbol{\mathcal{O}}(1),$$

because $nN^{-1} = o(1)$. Thus, (30) holds.

Using (28) and (C.17), we have

$$\begin{aligned}
\mathbb{E}_d^{(2)}(\widetilde{\boldsymbol{V}}_0|\widetilde{\boldsymbol{S}},\boldsymbol{r}) &= N^{-2}\sum_{h=1}^{H}\Big\{\sum_{k\in\widetilde{\boldsymbol{S}}_h}\pi_k^{-2}\mathbb{E}_d^{(2)}(\widehat{\boldsymbol{\epsilon}}_k\widehat{\boldsymbol{\epsilon}}_k^\top|\widetilde{\boldsymbol{S}},\boldsymbol{r}) - n_h^{-1}\sum_{k\in\widetilde{\boldsymbol{S}}_h}\pi_k^{-1}\boldsymbol{\epsilon}_k\sum_{\ell\in\widetilde{\boldsymbol{S}}_h}\pi_\ell^{-1}\boldsymbol{\epsilon}_\ell\Big\} \\
&= N^{-2}\sum_{h=1}^{H}\Big\{\sum_{k\in\widetilde{\boldsymbol{S}}_h}\pi_k^{-2}\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^\top - n_h^{-1}\sum_{k\in\widetilde{\boldsymbol{S}}_h}\pi_k^{-1}\boldsymbol{\epsilon}_k\sum_{\ell\in\widetilde{\boldsymbol{S}}_h}\pi_\ell^{-1}\boldsymbol{\epsilon}_\ell^\top\Big\} \\
&\quad + N^{-2}\sum_{h=1}^{H}\sum_{k\in\widetilde{\boldsymbol{S}}_h}\pi_k^{-2}\mathbb{V}_d^{(2)}(\widehat{\boldsymbol{\epsilon}}_k|\widetilde{\boldsymbol{S}},\boldsymbol{r})\cdot
\end{aligned} \tag{C.36}$$

Hence, by combining (C.36) with (C.15), (C.16) and (C.28), we have

$$\mathbb{E}_d^{(1)}\big\{\mathbb{E}_d^{(2)}(\widetilde{\boldsymbol{V}}_0 \mid \widetilde{\boldsymbol{S}},\boldsymbol{r})\big\} = \boldsymbol{V}_{d;a} + \boldsymbol{V}_{d;b}\cdot \tag{C.37}$$

Thus, (C.14) and (C.37) imply

$$\begin{aligned}
\mathbb{E}(\widetilde{\boldsymbol{V}}_0) &= \mathbb{E}_r[\mathbb{E}_d^{(1)}\{\mathbb{E}_d^{(2)}(\widetilde{\boldsymbol{V}}_0 \mid \widetilde{\boldsymbol{S}},\boldsymbol{r})\}] \\
&= \mathbb{E}_r(\boldsymbol{V}_{d;a}) + \mathbb{E}_r(\boldsymbol{V}_{d;b}) \\
&= \mathbb{E}_r\{\mathbb{V}_d(\bar{\boldsymbol{\epsilon}}_\pi \mid \boldsymbol{r})\}\cdot
\end{aligned}$$

Expressions (C.11), (C.13) imply $\mathbb{E}_r\{\mathbb{V}_d(\bar{\boldsymbol{\epsilon}}_\pi \mid \boldsymbol{r})\} = \mathbb{V}(\bar{\boldsymbol{\epsilon}}_\pi) + N^{-1}\boldsymbol{\mathcal{O}}(1)$ and

$$\mathbb{E}(\widetilde{\boldsymbol{V}}_0) = \mathbb{V}(\bar{\boldsymbol{\epsilon}}_\pi) + N^{-1}\boldsymbol{\mathcal{O}}(1),$$

which implies (31), using (C.10) and $nN^{-1} = o(1)$. □

References

Berger, Y. G. & Kabzinska, E. (2016). Empirical likelihood approach for aligning information from multiple surveys. Southampton Statistical Sciences Research Institute.

Berger, Y. G. & Torres, O. D. L. R. (2016). An empirical likelihood approach for inference under complex sampling design. *Journal of the Royal Statistical Society Series B* 78, 319–341.

Fay, B. E. (1991). A design-based perspective on missing data variance. *Proceeding of the 1991 Annual Research Conference. U.S. Bureau of the Census* , 429–440.

Hansen, M. H. & Hurwitz, W. N. (1943). On the theory of sampling from finite populations. *The Annals of Mathematical Statistics* 14, pp. 333–362.

Oğuz-Alper, M. & Berger, Y. G. (2016). Empirical likelihood approach for modelling survey data. *Biometrika* 103, 447–459.

Shao, J. & Steel, P. (1999). Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *Journal of the American Statistical Association* 94, 254–265.