Annals of the Institute of Statistical Mathematics (2023) 75:425–462
https://doi.org/10.1007/s10463-022-00850-6

Regression analysis for exponential family data in a finite population setup using two-stage cluster sample

Brajendra C. Sutradhar[1,2]

Received: 7 January 2022 / Revised: 22 July 2022 / Accepted: 2 August 2022 /
Published online: 14 September 2022

Abstract
Over the last four decades, the cluster regression analysis in a finite population (FP) setup for an exponential family such as linear or binary data was done by using a two-stage cluster sample chosen from the FP but by treating the sample as though it is a single-stage cluster sample from a super-population (SP) which contains the FP as a hypothetical sample. Because the responses within a cluster in the FP are correlated, the aforementioned sample mis-specification makes the sample-based so-called GLS (generalized least square) estimators design biased and inconsistent. In this paper, we demonstrate for the exponential family data how to avoid the sampling mis-specification and accommodate the cluster correlations to obtain unbiased and consistent estimates for the FP parameters. The asymptotic normality of the regression estimators is also given for the construction of confidence intervals when needed.

Keywords Clusters under a finite population · Clusters selected in first stage · Individuals selected in second stage from a selected cluster · Invalid inferences for regression effects using GLS estimates · Doubly weighted estimation · Unbiasedness · Consistency and asymptotic normality

1 Introduction

The inferences for regression parameters in a FP setup using two-stage cluster sample have been an important research topic over the last four decades. This type of regression analysis is encountered mainly by the national or provincial statistical agencies such as Statistics Canada, U.S. Bureau of Census, and similar organizations

✉ Brajendra C. Sutradhar
bsutradh@mun.ca

[1] Department of Mathematics and Statistics, Memorial University, St. John's, NL A1C 5S7, Canada

[2] Present Address: 28 Musgrave Street, St. John's, NL A1A 5E7, Canada

in other countries. For example, suppose that all hospitals in a country and the nurses working in these hospitals form a FP, and it is of interest to know the effects of certain covariates such as gender, age and education level, on the binary response, say job satisfaction status of the nurses in the country. Here, the regression effects/parameters for this country would be a FP regression parameter, whereas an infinite population regression parameter would refer to the effects of the covariates on the binary response of any nurse from any countries, for example. Thus, a FP regression parameter may be treated as an estimate of the infinite population-based regression parameter (e.g. Binder 1983; Godambe and Thompson 1986; Ghosh 1991, Section 14.2, p. 203), but in a FP study using two-stage cluster sample selected from this FP, we are interested to estimate the FP regression parameters. Furthermore, because there is also an invisible/random common cluster/hospital effect which is shared by all elements/nurses within the cluster, this would make the responses within a cluster correlated. Thus, it would be necessary to define a FP-based cluster correlation parameter and its estimation, which is, however, not yet addressed or addressed inadequately in the literature.

Over the last four decades, (a) some studies (e.g., Binder 1983) in a FP setup involving stratas, defined the FP regression parameters implicitly as the solution of a suitable such as the generalized linear model-based likelihood estimating equation constructed using the hypothetical responses of the finite population. Here, FP responses are hypothetical as they are not observed, until a sample is taken to observe a part of the FP. The FP regression parameters are estimated using a sample chosen from the FP. The possible cluster correlations are not accommodated in this approach in the estimating equations for the regression parameters. This approach may provide biased and hence inconsistent regression estimates, specially when the mean function of the responses involve the cluster variance/correlation parameters (e.g., Sutradhar 2020). (b) Some other studies suggested to sample the data from the finite population using the two-stage cluster sampling but estimated the regression parameters by using a generalized linear mixed model based (involving cluster correlations) such as generalized least square type estimating equation constructed by treating the second-stage data as a sample arising from an infinite population for the exponential family data [e.g., Valliant (1985) in a binary response setup, and Prasad and Rao (1990) in a linear data setup] . This second approach became popular but unfortunately as we show in this paper, it produces biased and hence inconsistent estimates for the regression parameters when these parameters are defined correctly under the finite population using the first (a) approach. Moreover, this second approach appears to be misleading as in a FP setup, it does not make sense to estimate the infinite population parameters using the sample from the FP. This paper provides a theoretical foundation, first, defines the regression parameters under the finite population but in the presence of cluster variance/correlation parameter. We then provide two-stage cluster sample-based estimation both for regression and cluster correlation parameters.

In notations, consider a finite population (FP:$\mathcal{F}$) consisting of K independent clusters with their sizes $N_1, \ldots, N_c, \ldots, N_K$, N_c being the size of the c-th cluster which is large but fixed. Suppose that $K \rightarrow \infty$. Here, $N = \text{limit}_{K \rightarrow \infty} \sum_{c=1}^{K} N_c \rightarrow \infty$ is the size of the FP. Let y_{ci} denotes a hypothetical response from the i-th ($i = 1, \ldots, N_c$)

individual of the c-th cluster under the finite population. It is hypothetical as it is unknown at the finite population level. Further let $\boldsymbol{x}_{ci}$ be a p-dimensional fixed covariate vector, and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_u, \dots, \beta_p)'$ be the regression effect of x_{ci} on y_{ci}, for all $c = 1, \dots, K$; $i = 1, \dots, N_c$. Notice that in this cluster setup, there is likely to be a cluster effect on the responses belonging to the same cluster. Let γ_c denote the random cluster effect of the c-th cluster which is shared by the responses belonging to this cluster. Thus, on top of $\boldsymbol{\beta}$, there is an influence of γ_c on the responses $(\{y_{ci}, i = 1, \dots, N_c\})$ belonging to the c-th cluster. To accommodate the influences of $\boldsymbol{x}_{ci}$ and γ_c on the response y_{ci}, or its mean, suppose that $\mathcal{F}$ follows an infinite/super-population (SP:$\mathcal{S}$)-based conditional mean model given by

$$\begin{aligned} E\left[Y_{ci}|\gamma_c\right] &= \mu^*_{ci}(\boldsymbol{\beta}, \gamma_c) \\ &\equiv \begin{cases} \tilde{m}_{ci}(\boldsymbol{\beta}, \gamma_c) = \boldsymbol{x}^\top_{ci}\boldsymbol{\beta} + \gamma_c & \text{for linear data} \\ m^*_{ci}(\boldsymbol{\beta}, \gamma_c) = \exp(\boldsymbol{x}^\top_{ci}\boldsymbol{\beta} + \gamma_c) & \text{for count data} \\ p^*_{ci}(\boldsymbol{\beta}, \gamma_c) = \exp(\boldsymbol{x}^\top_{ci}\boldsymbol{\beta} + \gamma_c)/\left[1 + \exp(\boldsymbol{x}^\top_{ci}\boldsymbol{\beta} + \gamma_c)\right] & \text{for binary data .} \end{cases} \end{aligned} \tag{1}$$

Suppose that $\gamma_c \overset{iid}{\sim} N(0, \sigma^2_\gamma)$, where σ^2_γ may be referred to as the cluster variance or correlation parameter. After some straightforward algebras, one may then obtain the unconditional means, variances, and pairwise covariances for the responses $\{y_{ci}, i = 1, \dots, N_c; c = 1, \dots, K\} \in \mathcal{F}$. More specifically, for the linear case, by writing

$$\begin{aligned} y_{ci} = \boldsymbol{x}^\top_{ci}\boldsymbol{\beta} + \gamma_c + \epsilon_{ci},\ c = 1, \dots, K \to \infty; i = 1, \dots, N_c, \\ \gamma_c \overset{iid}{\sim} (0, \sigma^2_\gamma)\ \epsilon_{ci} \overset{iid}{\sim} (0, \sigma^2_\epsilon)\ \gamma_c \text{ and } \epsilon_{ci} \text{ are independent;} \end{aligned} \tag{2}$$

one obtains the basic first- and second-order moment properties, as

$$\begin{aligned} E\left[Y_{ci}\right] &= \boldsymbol{x}^\top_{ci}\boldsymbol{\beta} = \mu_{ci}(\boldsymbol{\beta}) \\ \text{var}\left[Y_{ci}\right] &= \sigma_{c,ii} = (\sigma^2_\gamma + \sigma^2_\epsilon) = \sigma^2, \text{ for all } c, \text{ and } i \\ \text{cov}\left[Y_{ci}, Y_{cj}\right] &= \sigma_{c,ij} = \sigma^2_\gamma \Rightarrow \text{corr}\left[Y_{ci}, Y_{cj}\right] = \rho = \left[\sigma^2_\gamma/\sigma^2\right], \text{ for all } i \neq j. \end{aligned} \tag{3}$$

Next, consider $\boldsymbol{y}_c = (y_{c1}, \dots, y_{ci}, \dots, y_{cN_c})'$ as the $N_c \times 1$ hypothetical response vector under c-th cluster, and $\boldsymbol{X}_c$ as the $N_c \times p$ covariates matrix for N_c individuals in the c-th cluster. Suppose that $\boldsymbol{\Sigma}_{(c,N_c)}(\rho)$ is the $N_c \times N_c$ model-based covariance matrix for $\boldsymbol{y}_c$, which by (3), has the form given by

$$\text{cov}\left[Y_c\right] = \boldsymbol{\Sigma}_{(c,N_c)}(\rho) = \sigma^2\left[(1-\rho)\boldsymbol{I}_{N_c} + \rho\boldsymbol{U}_{N_c}\right] = \sigma^2 R_{(c,N_c)}(\rho), \tag{4}$$

with $\boldsymbol{I}_{N_c}$ and $\boldsymbol{U}_{N_c}$ as the $N_c \times N_c$ identity and unit matrices, respectively, and $R_{(c,N_c)}(\rho)$ represents the correlation matrix.

Now, by treating the fixed FP ($\mathcal{F}$) as though it is available and arose randomly from the SP ($\mathcal{S}$) ($\mathcal{F} \subset \mathcal{S}$), one could consistently and efficiently estimate $\boldsymbol{\beta}$ by solving the well-known GLS (generalized least square), more specifically the FP-based GLS (FPGLS) [equivalently HGLS (hypothetical GLS)] estimating equation

$$
\begin{aligned}
T_{y\in\mathcal{F}}(\boldsymbol{\beta}) &= \sum_{c=1}^{K} \boldsymbol{X}_c^{\top}\boldsymbol{\Sigma}_{(c,N_c)}^{-1}(\rho)(\boldsymbol{y}_c - \boldsymbol{X}_c\boldsymbol{\beta}) \\
&= \frac{1}{\sigma^2}\sum_{c=1}^{K} \boldsymbol{X}_c^{\top}\boldsymbol{R}_{(c,N_c)}^{-1}(\rho)(\boldsymbol{y}_c - \boldsymbol{X}_c\boldsymbol{\beta}) = 0,
\end{aligned}
\tag{5}
$$

where $\boldsymbol{\Sigma}_{(c,N_c)}(\rho) = \sigma^2\boldsymbol{R}_{(c,N_c)}(\rho)$ is the covariance matrix from (4). Suppose that the solution of (5) is denoted by $\boldsymbol{\beta}_{N,\mathrm{GLS}}$.

Notice that as in practice $(\mathcal{F})$ is unobserved, $\boldsymbol{\beta}_{N,\mathrm{GLS}}$ from the hypothetical GLS equation (5) is referred to as the FP-based regression parameter [Godambe and Thompson (1986) [see also Ghosh (1991, Section 14.2, p. 203), Binder (1983)]. More specifically, as indicated above, under approach (a), in a complex survey such as stratified sampling setup, by treating the responses as independent, Binder (1983, Eqn. (2.4)) has defined the regression parameters implicitly, as the solution of a suitable such as the generalized linear model (GLM)-based likelihood estimating equation (Nelder and Wedderburn 1972) constructed using the hypothetical/unknown responses of the finite population. We remark that Godambe and Thompson (1986) [see also Ghosh (1991, Section 14.2, p. 203)] referred to the solution of the aforementioned hypothetical estimating equation, as the N-dependent FP parameters, N being the size of the FP. Thus, the $\mathcal{S}$-based regression parameter $\boldsymbol{\beta}$, satisfying a $\mathcal{F}$-based estimating equation, becomes a FP parameter, say $\boldsymbol{\beta}_N$. Under the general exponential family data setup, the proposed $\mathcal{F}$-based estimating equation for $\boldsymbol{\beta}$ is developed in Sect. 4.1 [see (33) under Lemma 3]. We further remark that unlike Binder (1983), in a cluster setup, the responses within a cluster under the $\mathcal{F}$, are correlated as they are supposed to share a common random cluster effect. Hence, similar to the definition for $\boldsymbol{\beta}_N$, we will also define a N-dependent cluster variance/correlation parameter, say $\sigma^2_{\gamma,N}$, corresponding to σ^2_γ in (1), for the general exponential family data. This will be given in Sect. 4.2 [see (51) under Lemma 6].

Clearly, for the estimation of $\boldsymbol{\beta} \equiv \boldsymbol{\beta}_N$, as well as $\sigma^2_\gamma \equiv \sigma^2_{\gamma,N}$ (or $\rho \equiv \rho_N$), a suitable sample is needed, which would be a two-stage cluster sample in the present setup. This sample may be constructed [see e.g., Särndal et al. (1992, Section 4.2, p. 134)] as follows.

First stage A sample of, say k clusters $s_1^* \equiv \{(y_{ci}, x_{ci}), i = 1, \ldots, N_c;\ c = 1, \ldots, k\}$ is drawn from $(\mathcal{F}) \equiv \left\{(y_{ci}, x_{ci}), i = 1, \ldots, N_c;\ c = 1, \ldots, K\right\}$, according to a suitable design $p_1(\cdot)$. For simplicity, we will consider $p_1(\cdot)$ as an equal probability-based SRS (simple random sampling) without replacement.

Second stage For every cluster/family $c \in s_1^*$, a sample of, say s_{2c}^*, with n_c elements/individuals, is drawn from its parental cluster consisting of N_c elements/individuals, according to a suitable design $p_{2c}(\cdot)$. Once again, for simplicity, we will consider every $p_{2c}(\cdot)$ as an equal probability-based SRS (simple random sampling) without replacement.

We denote the resulting sample of individuals along with their responses and covariates, by s^*, which may be expressed as

$$\begin{aligned} s^* = \bigcup_{c \in s_1^*} s_{2c}^* &\equiv \left\{(y_{ci}, x_{ci}), i = 1, \dots, n_c;\ c = 1, \dots, k\right\} \\ &\subset \mathcal{F} \equiv \left\{(y_{ci}, x_{ci}), i = 1, \dots, N_c;\ c = 1, \dots, K\right\}. \end{aligned} \tag{6}$$

Next, based on the approach (b) indicated above, many authors estimated $\boldsymbol{\beta}$, by treating s^* as taken from the $\mathcal{S}$ directly, even though it is chosen from the $\mathcal{F}$. For example, Burdick and Sielken (1979), and Christensen (1984, 1987) have used the well-known OLS (ordinary least square) and Scott and Holt (1982), Valliant (1985, Eqn. (2); 1987), Prasad and Rao (1990, Eqn. (3.2)), Lehtonen and Veijanen (1998), and Fuller (2009, Section 2.6, Eqns. (2.6.6) and (2.6.13)) have used the well-known GLS (generalized least square) estimation approaches, where the OLS and GLS equations were written using s^* assuming it is taken from the $\mathcal{S}$ directly. Thus, in this approach (2), $\mathcal{F}$ has nothing to do for inferences, which is a major mistake, as s^* is truly chosen from $\mathcal{F}$ under the present two-stage cluster sampling setup. In this paper, we remove this error and provide a foundation for inferences using s^*, which has arisen from the $\mathcal{F}$, shown in (6).

In notations, using the s^* from (6) and utilizing the sample-based response vector $\boldsymbol{y}_{c\in s^*} = \left(y_{c1}, \dots, y_{ci}, \dots, y_{cn_c}\right)^\top : n_c \times 1$, and its corresponding covariate matrix $\boldsymbol{x}_c^\top = \left[\boldsymbol{x}_{c1}, \dots, \boldsymbol{x}_{ci}, \dots, \boldsymbol{x}_{cn_c}\right] : p \times n_c$, the aforementioned studies have constructed and solved the SS (survey sample)-based GLS (SSGLS) estimating equation

$$\begin{aligned} T^*_{y\in s^*}(\boldsymbol{\beta}) &= \sum_{c=1}^{k} \boldsymbol{x}_c^\top \boldsymbol{\Sigma}^{-1}_{(c,n_c)}(\rho)\left(\boldsymbol{y}_{c\in s^*} - \boldsymbol{x}_c\boldsymbol{\beta}\right) \\ &= \frac{1}{\sigma^2} \sum_{c=1}^{k} \boldsymbol{x}_c^\top \boldsymbol{R}^{-1}_{(c,n_c)}(\rho)(\boldsymbol{y}_{c\in s^*} - \boldsymbol{x}_c\boldsymbol{\beta}) = 0, \end{aligned} \tag{7}$$

which, with $\boldsymbol{\Sigma}_{(c,n_c)}(\rho) = \sigma^2 \boldsymbol{R}_{(c,n_c)}(\rho) = \sigma^2[(1-\rho)\boldsymbol{I}_{n_c} + \rho \boldsymbol{U}_{n_c}]$, is quite similar to that of the FP-based GLS equation $T_{y\in\mathcal{F}}(\boldsymbol{\beta}) = 0$, defined in (5). Here similar to (4), $\boldsymbol{I}_{n_c}$ and $\boldsymbol{U}_{n_c}$ represent the $n_c \times n_c$ identity and unit matrix, respectively. Notice that even though the SSGLS estimating equation $T^*_{y\in s^*}(\boldsymbol{\beta}) = 0$ in (7) apparently uses the s^*-based data, unfortunately its construction fails, as mentioned above, to accommodate the fact that s^* is chosen from the FP ($\mathcal{F} \equiv \left\{(y_{ci}, x_{ci}), i = 1, \dots, N_c;\ c = 1, \dots, K\right\}$). More clearly, under the present two-stage cluster setup, the SSGLS estimating equation $T^*_{y\in s^*}(\boldsymbol{\beta}) = 0$ would have been valid provided the SSGLS estimating function $T^*_{y\in s^*}(\boldsymbol{\beta})$ in (7), were sampling design (D_{s^*}) unbiased for the HGLS estimation function $T_{y\in\mathcal{F}}(\boldsymbol{\beta})$ in (5), i.e., if $E_{D_{s^*}}\left[T^*_{y\in s^*}(\boldsymbol{\beta})\right] = T_{y\in\mathcal{F}}(\boldsymbol{\beta})$, which is, as shown in Sect. 2, however, not the case in the present setup. Thus, the existing SSGLS estimates become biased and inconsistent, and subsequently provide invalid inferences, which must be remedied.

For the purpose, we turn back to the general exponential family-based response model indicated by (1) which includes the linear, count and binary responses as special cases. In Sect. 3, we provide the necessary moments including the unconditional means, variances and correlations for the $\mathcal{F}$-based exponential responses. In Sect. 4, we develop the $\mathcal{F}$-based HGQL (hypothetical GQL) estimating functions for both $\boldsymbol{\beta}$ and σ_γ^2. The corresponding SS (s^*)-based doubly weighted (SSDW) estimating equations are developed in Sect. 5. In Sect. 6, we show that the proposed estimators are consistent for the respective parameters. In the same section, the asymptotic normality property for the estimators of the main regression effects is also given for the convenience of confidence interval construction when needed.

2 Sampling mis-specification effects on regression parameters estimation using GLM-based GLS approach

Even though $s^* \equiv \{(y_{ci}, x_{ci}), i = 1, \ldots, n_c;\ c = 1, \ldots, k\}$ in (6) is a two-stage sample collected from the FP $\mathcal{F} \equiv \{(y_{ci}, x_{ci}), i = 1, \ldots, N_c;\ c = 1, \ldots, K\}$, the GLS estimating equation $T^*_{y\in s^*}(\boldsymbol{\beta}) = \sum_{c=1}^{k} \boldsymbol{x}_c^\top \boldsymbol{\Sigma}^{-1}_{(c,n_c)}(\rho)\big(\boldsymbol{y}_{c\in s^*} - \boldsymbol{x}_c\boldsymbol{\beta}\big) = 0$ for $\boldsymbol{\beta}$ in (7) used by the existing studies [e.g., Valliant 1985, Eqn. (2) for binary data, and Prasad and Rao [1990, Eqn. (3.2)] and Fuller (2009, Section 2.6, Eqns. (2.6.6) and (2.6.13)) for linear data] ignores the fact that $\boldsymbol{y}_{c\in s^*}$ is sampled from $\boldsymbol{y}_{c\in\mathcal{F}}$ of size N_c. More clearly, $T^*_{y\in s^*}(\boldsymbol{\beta}) = \sum_{c=1}^{k} \boldsymbol{x}_c^\top \boldsymbol{\Sigma}^{-1}_{(c,n_c)}(\rho)(\boldsymbol{y}_{c\in s^*} - \boldsymbol{x}_c\boldsymbol{\beta})$ is constructed by pretending that k clusters, are chosen from an infinite population ($\mathcal{S}$) consisting of a large number of independent clusters. Suppose that the c-th selected cluster has size n_c. Thus, the existing studies used $s^* \equiv \big\{(\boldsymbol{y}_1, \boldsymbol{x}_1), \ldots, (\boldsymbol{y}_c, \boldsymbol{x}_c), \ldots, (\boldsymbol{y}_k, \boldsymbol{x}_k)\big\}$ as a single-stage cluster sample from the $\mathcal{S}$. Here $\boldsymbol{y}_c$ has the same dimension $n_c \times 1$, as for the c-th selected cluster. Notice that this s^* chosen from the $\mathcal{S}$, appears to be the same as the s^* in (6). But they are completely different. This anomaly would naturally make the SSGLS estimating function $T^*_{y\in s^*}(\boldsymbol{\beta}) = \sum_{c=1}^{k} \boldsymbol{x}_c^\top \boldsymbol{\Sigma}^{-1}_{(c,n_c)}(\rho)(\boldsymbol{y}_{c\in s^*} - \boldsymbol{x}_c\boldsymbol{\beta})$ biased for the $\mathcal{F}$-based GLS estimating function $T_{y\in\mathcal{F}}(\boldsymbol{\beta}) = \frac{1}{\sigma^2}\sum_{c=1}^{K} \boldsymbol{X}_c^\top \boldsymbol{R}^{-1}_{(c,N_c)}(\rho)(\boldsymbol{y}_c - \boldsymbol{X}_c\boldsymbol{\beta})$, defined in (5). This sample selection bias would subsequently produce an invalid estimate for $\boldsymbol{\beta}$.

We now examine the bias performance of the SSGLS estimating function $T^*_{y\in s^*}(\boldsymbol{\beta})$ as follows. For the purpose, we need to compute the design expectation of the SSGLS estimating function, $T^*_{y\in s^*}(\boldsymbol{\beta})$, under the true two-stage cluster sampling scheme (6). That is, we compute

$$E_{D_{s^*}}[T^*_{y\in s^*}(\boldsymbol{\beta})] = E_{p_1}\left[E_{p_{2c}}\left(T^*_{y\in s^*}(\boldsymbol{\beta})\right)\Big|p_1\right], \tag{8}$$

where, as explained in Sect. 1 [see (6)], p_1 is a suitable sampling design for the selection of the first-stage sample s_1^*, and p_{2c} is the sampling design for the selection of the second stage sample s_{2c}^*. For simplicity, we consider them as equal probability based without replacement designs so that

$$Pr(s_1^*) = 1/\binom{K}{k}, \text{ and } Pr(s_{2c}^*) = 1/\binom{N_c}{n_c}, \tag{9}$$

or, equivalently, (i) $\Pr\left[(\text{c-th cluster}) \in s_1^*\right] = k/K$, and
(ii) $\Pr\left[(\text{i-th individual from the c-th selected cluster}) \in s_{2c}^*\right] = n_c/N_c$, respectively.

Notice from (7) that $\boldsymbol{\Sigma}_{(c,n_c)}(\rho) = \sigma^2 \boldsymbol{R}_{(c,n_c)}(\rho) = \sigma^2\left[(1-\rho)\boldsymbol{I}_{n_c} + \rho \boldsymbol{U}_{n_c}\right]$, yielding $\boldsymbol{\Sigma}^{-1}_{(c,n_c)}(\rho) = \frac{1}{\sigma^2}\boldsymbol{R}^{-1}_{(c,n_c)}(\rho)$, where

$$\boldsymbol{R}^{-1}_{(c,n_c)}(\rho) = \left[((a(n_c,\rho) - b(n_c,\rho))\boldsymbol{I}_{n_c} + b(n_c,\rho)\boldsymbol{U}_{n_c})\right] \tag{10}$$

(e.g., Seber 1984, p. 520), with $a(n_c,\rho)$ and $b(n_c,\rho)$ defined as

$$a(n_c,\rho) = \frac{1 + (n_c - 2)\rho}{(1-\rho)\{1 + (n_c - 1)\rho\}} \quad \text{and} \quad b(n_c,\rho) = \frac{-\rho}{(1-\rho)\{1 + (n_c - 1)\rho\}},$$

respectively. Using the inversion formula from (10), after an algebra, one may re-express the SSGLS estimating function as

$$\begin{aligned} T^*_{y\in s^*}(\boldsymbol{\beta}) &= \sum_{c=1}^{k} \boldsymbol{x}_c^{\top} \boldsymbol{\Sigma}^{-1}_{(c,n_c)}(\rho)(\boldsymbol{y}_{c\in s^*} - \boldsymbol{x}_c\boldsymbol{\beta}) \\ &= \frac{1}{\sigma^2}\sum_{c=1}^{k}\sum_{i=1}^{n_c} \boldsymbol{x}^*_{w,ci}(n_c,\rho)\left(y_{ci} - \boldsymbol{x}'_{ci}\beta\right), \end{aligned} \tag{11}$$

where the weighted covariate vector $\boldsymbol{x}^*_{w,ci}(n_c,\rho)$ is given by

$$\boldsymbol{x}^*_{w,ci}(n_c,\rho) = a(n_c,\rho)\boldsymbol{x}_{ci} + b(n_c,\rho)\sum_{j\neq i, j\in s^*}^{n_c} \boldsymbol{x}_{cj}. \tag{12}$$

Next, by applying (8)–(9), we can compute the two-stage sampling-based design expectation of the SSGLS function in (11) as

$$\begin{aligned} E_{D_{s^*}}\left[T^*_{y\in s^*}(\boldsymbol{\beta})\right] &= E_{p_1}\left[\sum_{c=1}^{k} E_{p_{2c}}\left\{\frac{1}{\sigma^2}\sum_{i=1}^{n_c}\boldsymbol{x}^*_{w,ci}(n_c,\rho)\left(y_{ci} - \boldsymbol{x}'_{ci}\beta\right)\right\}|p_1\right] \\ &= \frac{1}{\sigma^2}E_{p_1}\left[\sum_{c=1}^{k}\frac{n_c}{N_c}\sum_{i=1}^{N_c}\boldsymbol{x}^*_{w,ci}(n_c,\rho)\left(y_{ci} - \boldsymbol{x}'_{ci}\beta\right)\right] \\ &= \frac{1}{\sigma^2}\frac{k}{K}\sum_{c=1}^{K}\frac{n_c}{N_c}\left[\sum_{i=1}^{N_c}\boldsymbol{x}^*_{w,ci}(n_c,\rho)\left(y_{ci} - \boldsymbol{x}'_{ci}\beta\right)\right], \end{aligned} \tag{13}$$

which is, however, not equal to the $\mathcal{F}$-based GLS estimating function $T_{y\in\mathcal{F}}(\boldsymbol{\beta}) = \frac{1}{\sigma^2}\sum_{c=1}^{K}\boldsymbol{X}_c^{\top}\boldsymbol{R}^{-1}_{(c,N_c)}(\rho)(\boldsymbol{y}_c - \boldsymbol{X}_c\boldsymbol{\beta})$, defined in (5). This is because, by similar calculations as in (10), one obtains

$$\boldsymbol{R}^{-1}_{(c,N_c)}(\rho) = \left[((\tilde{a}(N_c,\rho) - \tilde{b}(N_c,\rho))\boldsymbol{I}_{N_c} + \tilde{b}(N_c,\rho)\boldsymbol{U}_{N_c})\right] \tag{14}$$

with

$$\tilde{a}(N_c,\rho) = \frac{1+(N_c-2)\rho}{(1-\rho)\{1+(N_c-1)\rho\}} \quad \text{and} \quad \tilde{b}(N_c,\rho) = \frac{-\rho}{(1-\rho)\{1+(N_c-1)\rho\}},$$

yielding

$$T_{y\in\mathcal{F}}(\boldsymbol{\beta}) = \frac{1}{\sigma^2}\sum_{c=1}^{K}\sum_{i=1}^{N_c}\tilde{\boldsymbol{x}}_{w,ci}(N_c,\rho)\big(y_{ci} - \boldsymbol{x}'_{ci}\beta\big), \tag{15}$$

where

$$\tilde{\boldsymbol{x}}_{w,ci}(N_c,\rho) = \left[\tilde{a}(N_c,\rho)\boldsymbol{x}_{ci} + \tilde{b}(N_c,\rho)\sum_{j\neq i,j=1}^{N_c}\boldsymbol{x}_{cj}\right] : p\times 1. \tag{16}$$

Clearly, the $\mathcal{F}$-based population total function in (15) is quite different than the $\mathcal{F}$ -based population total function in the right hand side of (13). Hence, the SSGLS estimating function in (7) is a design biased estimating function for the $\mathcal{F}$-based GLS or HGLS estimating function in (5), implying that SSGLS estimate would be biased and inconsistent for $\boldsymbol{\beta}$.

We further remark that even if we consider a sampling weight-based weighted SSGLS (SSWGLS) estimating function

$$\tilde{T}_{y\in s^*}(\boldsymbol{\beta}) = \frac{K}{k}\sum_{c=1}^{k}\frac{N_c}{n_c}\boldsymbol{x}_c^\top\boldsymbol{\Sigma}^{-1}_{(c,n_c)}(\rho)(\boldsymbol{y}_{c\in s^*} - \boldsymbol{x}_c\boldsymbol{\beta}),$$

we would have obtained

$$\begin{aligned} E_{D_{s^*}}\left[\tilde{T}_{y\in s^*}(\boldsymbol{\beta})\right] &= E_{p_1}\left[\sum_{c=1}^{k}E_{p_{2c}}\frac{K}{k}\frac{N_c}{n_c}\left\{\frac{1}{\sigma^2}\sum_{i=1}^{n_c}\boldsymbol{x}^*_{w,ci}(n_c,\rho)\big(y_{ci} - \boldsymbol{x}'_{ci}\beta\big)\right\}|p_1\right] \\ &= \frac{1}{\sigma^2}\sum_{c=1}^{K}\left[\sum_{i=1}^{N_c}\boldsymbol{x}^*_{w,ci}(n_c,\rho)\big(y_{ci} - \boldsymbol{x}'_{ci}\beta\big)\right], \end{aligned} \tag{17}$$

which is still quite different than the $\mathcal{F}$-based population total function in (15) because of the significant difference between $\boldsymbol{x}^*_{w,ci}(n_c,\rho)$ in (17) and $\tilde{\boldsymbol{x}}_{w,ci}(N_c,\rho)$ in (15). They become the same only when $\rho = 0$, which, however, does not hold for the clustered correlated data.

In the next section, we consider more general clustered exponential data and provide their correlation model. This correlation structure will be exploited in Sect. 4 to develop suitable HGQL estimating functions for both regression parameters $\boldsymbol{\beta}$ and cluster variance/correlation parameter σ^2_γ (1). Note that as the HGQL estimation of σ^2_γ would also require additional higher order moments up to order four (Sutradhar 2004), these moments for the exponential data are provided first before constructing the HGQL estimating equation. The corresponding SS-based estimating functions for $\boldsymbol{\beta}$ and σ^2_γ will then be developed in Sect. 5.

3 Cluster correlation model for exponential family data

In this section, as a summary of the models in (1), we follow the exponential family-based $\mathcal{S}$ (super-population) model with its probability and moment properties up to order four as given in Lemma 1 below.

Lemma 1 [Exponential family-based $\mathcal{S}$ model for the $\mathcal{F}$ in (6)] *For the $\mathcal{F}$ in (6), let $r_{ci} = \boldsymbol{x}^\top_{ci}\boldsymbol{\beta} + \gamma_c$ be a linear predictor, and for a known link function $h^*(\cdot)$, $\theta_{ci}(\boldsymbol{\beta}; \boldsymbol{x}_{ci}, \gamma_c) = h^*(\boldsymbol{x}^\top_{ci}\boldsymbol{\beta} + \gamma_c) = h^*(r_{ci})$. Then, the exponential family density of the cluster response y_{ci}, conditional on γ_c, has the form*

$$\mathcal{S}: \quad g^*_{ci}(y_{ci}|\gamma_c) = \exp\left[\{y_{ci}\theta_{ci}(\boldsymbol{\beta};\ \boldsymbol{x}_{ci}, \gamma_c) - b(\theta_{ci}(\boldsymbol{\beta};\ \boldsymbol{x}_{ci}, \gamma_c))\}\varphi + c^*(y_{ci}, \varphi)\right], \tag{18}$$

where $\varphi = 1$ for Poisson and binary data, but, it is a scalar function or parameter for the linear/normal data in which case (18) becomes a two-parameters exponential density. Furthermore, $b(\cdot)$ and $c^*(\cdot)$ in (18), are known functional form for all normal, counts and binary data, yielding the conditional mean and variance of y_{ci} as

$$E\left[Y_{ci}|\gamma_c\right] = \mu^*_{ci}(\boldsymbol{\beta}, \gamma_c) = b^{(1)}(\theta_{ci}(\cdot)) = \frac{\partial b(\theta_{ci}(\cdot))}{\partial \theta_{ci}} \tag{19}$$

$$\text{var}\left[Y_{ci}|\gamma_c\right] = \sigma^*_{c,ii}(\boldsymbol{\beta}, \gamma_c) = \frac{1}{\varphi} b^{(2)}(\theta_{ci}(\cdot)) = \frac{1}{\varphi}\frac{\partial^2 b(\theta_{ci}(\cdot))}{\partial \theta^2_{ci}}. \tag{20}$$

Proof The moments in (19) and (20) may be derived from the moment generating function of the exponential family distribution (18). For details, see for example, Sutradhar (2011, Exercise 4.6, Chapter 4) and Sutradhar and Rao (2001, Lemma 1, Section 3). □

Notice that (19) $\Rightarrow$(1), because for the normal/linear data: $b(\theta_{ci}(\cdot)|\gamma_c) = \frac{1}{2}\theta^2_{ci}(\cdot)$, and $h^*(r_{ci}) = r_{ci}$, with $r_{ci} = \boldsymbol{x}^\top_{ci}\boldsymbol{\beta} + \gamma_c$; for Poisson count data: $\varphi = 1, h^*(r_{ci}) = r_{ci}; b(\theta_{ci}(\cdot)|\gamma_c) = \exp(\theta_{ci}(\cdot))$; and for binary data: $\varphi = 1, h^*(r_{ci}) = r_{ci}; b(\theta_{ci}(\cdot)|\gamma_c) = log(1 + \exp(\theta_{ci}(\cdot)))$. Further notice that conditional on γ_c, the pairwise responses within the c-th cluster are independent. More specifically,

$$\begin{aligned} E[(Y_{ci}Y_{cj})|\gamma_c] &= E[Y_{ci}|\gamma_c]E[Y_{cj}|\gamma_c] = b^{(1)}(\theta_{ci}(\cdot))b^{(1)}(\theta_{cj}(\cdot)) \\ &= \mu^*_{ci}(\boldsymbol{\beta}, \gamma_c)\mu^*_{cj}(\boldsymbol{\beta}, \gamma_c) \Rightarrow \text{ cov }[Y_{ci}, Y_{cj}|\gamma_c] = 0. \end{aligned} \tag{21}$$

Now apply Lemma 1 and find the marginal mean model for the finite population ($\mathcal{F}$) elements along with their pairwise correlation structure as in Lemma 2 below. These unconditional mean, variance and covariances of the first-order hypothetical/imaginary responses will be applied in Sect. 4.1 in order to develop the so-called hypothetical GQL (HGQL) estimating equation for the regression parameter $\boldsymbol{\beta}$. In the same Sect. 4, the second order hypothetical responses and their means, variances and pairwise covariances (involving fourth-order moments) will be exploited to develop a HGQL estimating equation for the cluster variance/correlation parameter σ^2_γ.

Lemma 2 (Unconditional mean, variance and pairwise covariances for $\{y_{ci}, i = 1, \ldots, N_c; c = 1, \ldots, K\} \in \mathcal{F}$). *Under the distributional assumption* (1) *for the random cluster effects* $\{\gamma_c, c = 1, \ldots, K\}$, *i.e., for* $\gamma_c \overset{iid}{\sim} N(0, \sigma^2_\gamma)$, *or equivalently using* $\gamma_c = \sigma_\gamma \gamma^*_c$ *so that* $\gamma^*_c \overset{iid}{\sim} N(0, 1)$ *with standard normal density* $f_N(\gamma^*_c)$, *and using* $\varphi = 1$ *for simpler linear case, the unconditional means, variances, and pairwise covariances for the responses in the cth cluster have the formulas*:

$$\mu_{ci}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right) = E[Y_{ci}] = E_{\gamma_c}[b^{(1)}(\theta_{ci}(\gamma_c))] = \int [b^{(1)}(\theta_{ci}(\sigma_\gamma \gamma^*_c))] f_N(\gamma^*_c) d\gamma^*_c, \tag{22}$$

$$\sigma_{c,ii}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right) = \text{var}\,[Y_{ci}] = \lambda_{c,ii}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right) - \mu^2_{ci}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right), \tag{23}$$

$$\sigma_{c,ij}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right) = \text{cov}\left[Y_{ci}, Y_{cj}\right] = \lambda_{c,ij}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right) - \mu_{ci}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)\mu_{cj}\left(\beta, \sigma^2_\gamma\right), \tag{24}$$

respectively, where

$$\begin{aligned}\lambda_{c,ii}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right) &= \int \left[b^{(2)}(\theta_{ci}(\sigma_\gamma \gamma^*_c)) + \{b^{(1)}(\theta_{ci}(\sigma_\gamma \gamma^*_c))\}^2\right] f_N(\gamma^*_c) d\gamma^*_c \\ &= \int \lambda^*_{c,ii}(\gamma^*_c) f_N(\gamma^*_c) d\gamma^*_c,\end{aligned} \tag{25}$$

$$\begin{aligned}\lambda_{c,ij}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right) &= \int \left[b^{(1)}(\theta_{ci}(\sigma_\gamma \gamma^*_c)) b^{(1)}(\theta_{cj}(\sigma_\gamma \gamma^*_c))\right] f_N(\gamma^*_c) d\gamma^*_c \\ &= \int \lambda^*_{c,ij}(\gamma^*_c) f_N(\gamma^*_c) d\gamma^*_c.\end{aligned} \tag{26}$$

Proof The proof follows from Lemma 1 using conditioning and unconditioning arguments. More specifically, the variances and covariances are computed using the formulas: $\text{var}\,[Y_{ci}] = E_{\gamma_c} E[Y^2_{ci}|\gamma_c] - \mu^2_{ci}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)$, and $\text{cov}\left[Y_{ci}, Y_{cj}\right] = E_{\gamma_c} E[Y_{ci} Y_{cj}|\gamma_c] - \mu_{ci}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)\mu_{cj}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)$, involving conditional and unconditional expectations. □

We suggest solving the integrals in (22)–(26) under Lemma 2, using the well-known binomial approximation to a standard normal integral [e.g., Ten Have and Morabia (1999), Sutradhar (2011, Section 5.1.1)]. More specifically, for a moderately large V, say $V = 10$, first write

$$\gamma^*{}_c(v) = \frac{v - V(1/2)}{\sqrt{V(1/2)(1/2)}}, \quad \text{for } v = 0, 1, 2, \ldots, V;$$

and then approximate the integrals in (22)–(26), as

$$\mu_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) \simeq \sum_{v=0}^{V} b^{(1)}(\theta_{ci}(\sigma_\gamma \gamma_c^*(v)))\binom{V}{v}(1/2)^v(1/2)^{V-v}, \tag{27}$$

$$\begin{aligned}\lambda_{c,ii}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) &= E[Y_{ci}^2] \\ &\simeq \sum_{v=0}^{V}[b^{(2)}(\theta_{ci}(\sigma_\gamma \gamma_c^*(v))) + b^{(1)}(\theta_{ci}(\sigma_\gamma \gamma_c^*(v)))]^2\binom{V}{v}(1/2)^v(1/2)^{V-v},\end{aligned} \tag{28}$$

$$\begin{aligned}\lambda_{c,ij}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) &= E[Y_{ci}Y_{cj}] \\ &\simeq \sum_{v=0}^{V}\left[b^{(1)}(\theta_{ci}(\sigma_\gamma \gamma_c^*(v)))b^{(1)}(\theta_{cj}(\sigma_\gamma \gamma_c^*(v)))\right]\binom{V}{v}(1/2)^v(1/2)^{V-v}.\end{aligned} \tag{29}$$

Notice that even though we have computed the mean, variance and correlation structure of the

$$\text{FP } (\mathcal{F}) : \{y_{ci}|\boldsymbol{x}_{ci}, i = 1, \ldots, N_c;\ c = 1, \ldots, K\}$$

as shown by (22)–(24), we can not, however, use the responses and covariates of this population to estimate the regression($\boldsymbol{\beta}$) and cluster correlation (σ_γ^2) parameters involved in the moments (22)–(24). This is because as explained in Sect. 1, in the FP setup, it is impractical to collect data from the entire which is a large population. Hence, a TSCS (two-stage cluster sample)-based survey sample $s^*(SS(s^*))$ (6) is taken, consisting of $k < K$ clusters chosen from K clusters at the first stage, and then at the second stage, $n_c < N_c$ individuals chosen from N_c individuals of the c-th ($c = 1, \ldots, k$) selected cluster.

More clearly, the main objective of taking the sample s^* (6) from the FP ($\mathcal{F}$) is to estimate the parameters $\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)$ those define the FP through its mean, variance and correlation structure given by (22)–(26). Thus, before we provide such an estimation approach using s^* in Sect. 5, we first show in Sect. 4, how one could estimate these parameters if the FP ($\mathcal{F}$) was available. However, because $\mathcal{F}$ is never available in practice, the estimating equations to be developed in Sect. 4 are preferably referred to as the hypothetical estimating equations (HEE), more specifically as the hypothetical GQL (HGQL) estimating equations.

4 FP-based hypothetical GQL (HGQL) estimation

4.1 HGQL estimating function for β In population total form

From Lemma 2, we write the N_c-dimensional hypothetical response vector from the c-th ($c = 1, \dots, K$) cluster, as $\boldsymbol{y}_{c\in\mathcal{F}} = (y_{c1}, \dots, y_{ci}, \dots, y_{cN_c})^\top$. Using its moment properties from (22)–(24), write the mean vector, variance and covariance matrices as

$$E[\boldsymbol{Y}_{c\in\mathcal{F}}] = \boldsymbol{\mu}_c\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) \\ = (\mu_{c1}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right), \dots, \mu_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right), \dots, \mu_{cN_c}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right))^\top \tag{30}$$

$$\boldsymbol{V}_c\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) = \operatorname{diag}\left[\sigma_{c,ii}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right), \dots, \sigma_{c,ii}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right), \dots, \sigma_{c,ii}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)\right] \tag{31}$$

$$\text{cov}\ [\boldsymbol{Y}_{c\in\mathcal{F}}] = \boldsymbol{\Sigma}_{c,N_c}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) = \boldsymbol{V}_c^{\frac{1}{2}}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)\boldsymbol{R}_{c,N_c}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)\boldsymbol{V}_c^{\frac{1}{2}}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right), \tag{32}$$

where $\boldsymbol{R}_{c,N_c}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)$ is the $N_c \times N_c$ correlation matrix. We denote this and its inverse matrix as

$$\boldsymbol{R}_{c,N_c}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) = (\rho_{c,N_c,ij}) : N_c \times N_c,\ \boldsymbol{R}_{c,N_c}^{-1}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) = (\rho_{c,N_c,ij}^{(-1)}) : N_c \times N_c.$$

If the FP was available, following the well GLMM-based GQL approach (Sutradhar 2004) one could solve the GQL estimating equation given in Lemma 3 below, in order to estimate the regression parameters $\boldsymbol{\beta}$ involved in the mean functions given by (22).

Lemma 3 *Using the notations from (30) to (32), for $\boldsymbol{y}_{c\in\mathcal{F}} \sim \left(\boldsymbol{\mu}_c\left(\boldsymbol{\beta}, \sigma_\gamma^2\right), \boldsymbol{\Sigma}_{c,N_c}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)\right)$, when σ_γ^2 is known, the GQL estimating equation for $\boldsymbol{\beta}$, is given by*

$$\tau_y(\boldsymbol{\beta}) = \sum_{c=1}^{K} \frac{\partial \boldsymbol{\mu}_c^\top\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)}{\partial \boldsymbol{\beta}} \boldsymbol{\Sigma}_{c,N_c}^{-1}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)\left(y_{c\in\mathcal{F}} - \boldsymbol{\mu}_c\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)\right) = 0, \tag{33}$$

where the estimating function in the left hand side of (33) subsequently has the population total form given by

$$\tau_y(\boldsymbol{\beta}) = \sum_{c=1}^{K}\sum_{i=1}^{N_c} \boldsymbol{a}_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)\left(y_{ci} - \mu_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)\right) \equiv \sum_{c=1}^{K}\sum_{i=1}^{N_c} z_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right), \tag{34}$$

say, where, the $p \times 1$ vector $\boldsymbol{a}_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)$, has the formula

$$\boldsymbol{a}_{ci}(\cdot) = \sum_{u=1}^{N_c} \frac{\bar{\sigma}_{c,uu}}{\sigma_{c,uu}^{\frac{1}{2}}\sigma_{c,ii}^{\frac{1}{2}}} r_{c,N_c,ui}^{(-1)} \boldsymbol{x}_{cu} \,:\, p \times 1, \tag{35}$$

with

$$\begin{aligned}
\frac{\partial \boldsymbol{\mu}_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)}{\partial \boldsymbol{\beta}} &= \int \frac{\partial [b^{(1)}(\theta_{ci}(\sigma_\gamma \gamma_c^*))]}{\partial \boldsymbol{\beta}} f_N(\gamma_c^*) d\gamma_c^* \\
&= \boldsymbol{x}_{ci} \sum_{v=0}^{V} b^{(2)}(\theta_{ci}(\sigma_\gamma \gamma_c^*(v))) \binom{V}{v} (1/2)^v (1/2)^{V-v} \\
&= \boldsymbol{x}_{ci} \bar{\sigma}_{c,ii}.
\end{aligned} \tag{36}$$

Proof First of all, for a given cluster c, the GQL estimating Eq. (33) is a generalized form [Sutradhar (2003, Section 3; 2004, Eqn. (3.4))] of the independence assumption-based QL estimating equation studied by Wedderburn (1974). Then, the summation over $c = 1, \ldots, K$, follows from the fact that all K clusters are independent in the FP. □

Next, because it is given in (36) that $\frac{\partial \boldsymbol{\mu}_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)}{\partial \boldsymbol{\beta}} = \boldsymbol{x}_{ci}\bar{\sigma}_{c,ii}$, it then follows that

$$\begin{aligned}
\frac{\partial \boldsymbol{\mu}_c^\top\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)}{\partial \boldsymbol{\beta}} &= \left[\frac{\partial \mu_{c1}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)}{\partial \boldsymbol{\beta}}, \ldots, \frac{\partial \mu_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)}{\partial \boldsymbol{\beta}}, \ldots, \frac{\partial \mu_{cN_c}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)}{\partial \boldsymbol{\beta}}\right] \\
&= (\boldsymbol{x}_{c1}, \ldots, \boldsymbol{x}_{ci}, \ldots, \boldsymbol{x}_{cN_c}) \text{ diag } [\bar{\sigma}_{c,11}, \ldots, \bar{\sigma}_{c,ii}, \ldots, \bar{\sigma}_{c,N_cN_c}] \\
&= (\boldsymbol{x}_{c1}, \ldots, \boldsymbol{x}_{ci}, \ldots, \boldsymbol{x}_{cN_c}) \bar{\boldsymbol{V}}_{c,N_c}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right), \text{ (say)},
\end{aligned} \tag{37}$$

yielding $\boldsymbol{a}_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)$ as in (35), because

$$\begin{aligned}
&\frac{\partial \boldsymbol{\mu}_c^\top\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)}{\partial \boldsymbol{\beta}} \boldsymbol{\Sigma}_{c,N_c}^{-1}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) \\
&= (\boldsymbol{x}_{c1}, \ldots, \boldsymbol{x}_{ci}, \ldots, \boldsymbol{x}_{cN_c}) \bar{\boldsymbol{V}}_{c,N_c}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) \boldsymbol{V}_c^{-\frac{1}{2}}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) \boldsymbol{R}_{c,N_c}^{-1}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) \boldsymbol{V}_c^{-\frac{1}{2}}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) \\
&= (\boldsymbol{x}_{c1}, \ldots, \boldsymbol{x}_{ci}, \ldots, \boldsymbol{x}_{cN_c}) \bar{\boldsymbol{V}}_c\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) \boldsymbol{V}_c^{-\frac{1}{2}}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) (r_{c,N_c,ij}^{(-1)}) \boldsymbol{V}_c^{-\frac{1}{2}}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) \\
&= \left(\boldsymbol{a}_{c1}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right), \ldots, \boldsymbol{a}_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right), \ldots, \boldsymbol{a}_{cN_c}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)\right) \\
&= \boldsymbol{A}_{c,N_c}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) \,:\, p \times N_c, \text{ (say)}.
\end{aligned} \tag{38}$$

4.2 HGQL estimating function for σ_γ^2 in population total form

Because σ_γ^2 is the cluster variance/correlation parameter, for its estimation in the GLMM setup, Jiang (1998), for example, has used the traditional MM (method of moments) approach, whether Sutradhar (2004) has exploited the so-called GQL approach which produces more efficient estimates than the MM approach. To develop such a GQL estimating equation in the present setup, we first provide the formulas for all possible fourth-order moments, in Lemma 4 below. The variances and covariances of the second-order responses are then constructed as in Lemma 5.

Lemma 4

(a) *The conditional fourth-order leading moments have the formulas*

$$\begin{aligned}\lambda^*_{c,iiii}(\gamma^*_c) = E[Y^4_{ci}|\gamma^*_c] &= b^{(4)}(\theta_{ci}(\sigma_\gamma\gamma^*_c)) + 3[b^{(2)}(\theta_{ci}(\sigma_\gamma\gamma^*_c))]^2 + 4\lambda^*_{c,iii}(\gamma^*_c)b^{(1)}(\theta_{ci}(\sigma_\gamma\gamma^*_c)) \\ &- 6b^{(2)}(\theta_{ci}(\sigma_\gamma\gamma^*_c))[b^{(1)}(\theta_{ci}(\sigma_\gamma\gamma^*_c))]^2 - 3[b^{(2)}(\theta_{ci}(\sigma_\gamma\gamma^*_c))]^4,\end{aligned} \tag{39}$$

where

$$\begin{aligned}\lambda^*_{c,iii}(\gamma^*_c) &= E[Y^3_{ci}|\gamma^*_c] \\ &= b^{(3)}(\theta_{ci}(\sigma_\gamma\gamma^*_c)) + 3\lambda^*_{c,ii}(\gamma^*_c)b^{(1)}(\theta_{ci}(\sigma_\gamma\gamma^*_c)) - 2[b^{(1)}(\theta_{ci}(\sigma_\gamma\gamma^*_c))]^3\end{aligned} \tag{40}$$

with $\lambda^*_{c,ii}(\gamma^*_c) = E[Y^2_{ci}|\gamma^*_c] = [b^{(2)}(\theta_{ci}(\sigma_\gamma\gamma^*_c)) + \{b^{(1)}(\theta_{ci}(\sigma_\gamma\gamma^*_c))\}^2]$ as in (25); yielding the unconditional fourth-order leading moments as

$$\lambda_{c,iiii}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right) = \int \lambda^*_{c,iiii}(\gamma^*_c)f_N(\gamma^*_c)d\gamma^*_c. \tag{41}$$

(b) *The formulas for the unconditional fourth-order product moments are given by*

$$\begin{aligned}\lambda_{c,iijj}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right) &= E[Y^2_{ci}Y^2_{cj}] \\ &= \int [\lambda^*_{c,ii}(\gamma^*_c)\lambda^*_{c,jj}(\gamma^*_c)]f_N(\gamma^*_c)d\gamma^*_c, \ i < j = 2, \ldots, N_c;\end{aligned} \tag{42}$$

$$\begin{aligned}\lambda_{c,iijj'}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right) &= E[Y^2_{ci}Y_{cj}Y_{cj'}] \\ &= \int [\lambda^*_{c,ii}(\gamma^*_c)b^{(1)}(\theta_{cj}(\sigma_\gamma\gamma^*_c))b^{(1)}(\theta_{cj'}(\sigma_\gamma\gamma^*_c))]f_N(\gamma^*_c)d\gamma^*_c, \ i < j, i < j'; j \neq j';\end{aligned} \tag{43}$$

$$\begin{aligned}\lambda_{c,ii'jj}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right) &= E[Y_{ci}Y_{ci'}Y^2_{cj}] \\ &= \int [\lambda^*_{c,jj}(\gamma^*_c)b^{(1)}(\theta_{ci}(\sigma_\gamma\gamma^*_c))b^{(1)}(\theta_{ci'}(\sigma_\gamma\gamma^*_c))]f_N(\gamma^*_c)d\gamma^*_c, \ i < j, i' < j'; j = j', i \neq i';\end{aligned} \tag{44}$$

$$\begin{aligned}\lambda_{c,ii'jj'}\left(\boldsymbol{\beta},\sigma^2_\gamma\right) &= E[Y_{ci}Y_{cj}Y_{ci'}Y_{cj'}] \\ &= \int \lambda^*_{c,ij}(\gamma^*_c)\lambda^*_{c,i'j'}(\gamma^*_c)f_N(\gamma^*_c)d\gamma^*_c,\ i<j, i'<j'; i\neq i', j\neq j',\end{aligned} \tag{45}$$

where $\lambda^*_{c,ii}(\gamma^*_c)$ *is as in* (40) [*see also* (25)], *and* $\lambda^*_{c,ij}(\gamma^*_c) = E[(Y_{ci}Y_{cj})|\gamma^*_c] = b^{(1)}(\theta_{ci}(\sigma_\gamma\gamma^*_c))b^{(1)}(\theta_{cj}(\sigma_\gamma\gamma^*_c))$, for example, as in (26).

Proof The formulas under (b), given in from (42) to (45), are obvious. For example, to compute the fourth-order unconditional moment $E[Y^2_{ci}Y_{cj}Y_{cj'}]$ in (43), we observe that the exponential responses within the square brackets are independent conditional on γ^*_c. Thus,

$$\begin{aligned}E[Y^2_{ci}Y_{cj}Y_{cj'}] &= E_{\gamma^*_c}\left[E(Y^2_{ci}|\gamma^*_c)E(Y_{cj}|\gamma^*_c)E(Y_{cj'}|\gamma^*_c)\right] \\ &= E_{\gamma^*_c}\left[\lambda^*_{c,ii}(\gamma^*_c)b^{(1)}(\theta_{cj}(\sigma_\gamma\gamma^*_c))b^{(1)}(\theta_{cj'}(\sigma_\gamma\gamma^*_c))\right]\end{aligned}$$

is obtained, as in (43). However, the formula for the conditional fourth-order leading moment derived under (a) in (39) requires a long but straightforward algebra. More specifically, following Sutradhar and Rao (2001, Lemma 1), for example, we first derive the formulas for $E\left[(Y_{ci}-b^{(1)}(\theta_{ci}(\sigma_\gamma\gamma^*_c)))^3|\gamma^*_c\right]$ and $E\left[(Y_{ci}-b^{(1)}(\theta_{ci}(\sigma_\gamma\gamma^*_c)))^4|\gamma^*_c\right]$, and then unplug the formulas for $\lambda^*_{c,iii}(\gamma^*_c)$ and $\lambda^*_{c,iiii}(\gamma^*_c)$, from their respective equation. □

Using Lemma 4, one may now compute the covariance matrices of the second-order response vectors as in Lemma 5 below.

Lemma 5 *Let the squared and pairwise products of the hypothetical responses are stacked separately to form two vectors as*

$$\begin{aligned}\boldsymbol{p}_{c\in\mathcal{F}} &= (y^2_{c1},\ldots,y^2_{ci},\ldots,y^2_{cN_c})' : N_c\times 1 \\ \boldsymbol{q}_{c\in\mathcal{F}} &= (y_{c1}y_{c2},\ldots,y_{ci}y_{cj},\ldots,y_{c(N_c-1)}y_{cN_c})'; i<j : N_c(N_c-1)/2\times 1,\end{aligned} \tag{46}$$

respectively. Their covariance matrices $\boldsymbol{\Psi}_{c,N_c}(\boldsymbol{\beta},\sigma^2_\gamma)$ *and* $\boldsymbol{\Omega}_{c,N_c}(\boldsymbol{\beta},\sigma^2_\gamma)$, *say, have the formulas given by*

$$\begin{aligned}\text{cov}\,[\boldsymbol{p}_{c\in\mathcal{F}}] &= \boldsymbol{\Psi}_{c,N_c}(\boldsymbol{\beta},\sigma^2_\gamma) : N_c\times N_c \\ &= \begin{cases}\text{var}\,[Y^2_{ci}] = \psi_{c,ii}(\cdot) = \lambda_{c,iiii}-\lambda^2_{c,ii} & \forall\, i=1,\ldots,N_c \\ \text{cov}\,[Y^2_{ci},Y^2_{cj}] = \psi_{c,ij}(\cdot) = \lambda_{c,iijj}-\lambda_{c,ii}\lambda_{c,jj} & \forall\, i\neq j; i,j=1,\ldots,N_c,\end{cases}\end{aligned} \tag{47}$$

and

$$
\text{cov}\,[\boldsymbol{q}_{c\in\mathcal{F}}] = \boldsymbol{\Omega}_{c,N_c}(\boldsymbol{\beta},\sigma^2_\gamma)\,:\,N_c(N_c-1)/2\times N_c(N_c-1)/2
$$
$$
= \begin{cases} \text{var}\,[Y_{ci}Y_{cj}] = \omega_{c,ij,ij}(\cdot) = \lambda_{c,iijj} - \lambda^2_{c,ij} & \forall\, i<j \\ \text{cov}\,[Y_{ci}Y_{cj}, Y_{ci'}Y_{cj'}] = \omega_{c,ij,i'j'} = \lambda_{c,iji'j'} - \lambda_{c,ij}\lambda_{c,i'j'} & \forall\, i\neq j\neq i'\neq j' \\ \text{cov}\,[Y_{ci}Y_{cj}, Y_{ci}Y_{cj'}] = \omega_{c,ij,ij'} = \lambda_{c,iijj'} - \lambda_{c,ij}\lambda_{c,ij'} & \forall\, i=i'<j\neq j' \\ \text{cov}\,[Y_{ci}Y_{cj}, Y_{ci'}Y_{cj}] = \omega_{c,ij,i'j} = \lambda_{c,ii'jj} - \lambda_{c,ij}\lambda_{c,i'j} & \forall\, j=j'>i\neq i', \end{cases} \tag{48}
$$

respectively.

Proof The unconditional higher order expectations derived in Lemma 4 imply the variances and covariances shown under Lemma 5. For example, in (47),

$$
\text{var}\,[Y^2_{ci}] = E[Y^4_{ci}] - [E[Y^2_{ci}]]^2 = \lambda_{c,iiii} - \lambda^2_{c,ii},
$$

where the formula for $\lambda_{c,iiii}$ is derived in (41), along with the formula for $\lambda_{c,ii}$ from (25). Similarly, the covariances in (48) are computed. For example, in (48),

$$
\text{cov}\,[Y_{ci}Y_{cj}, Y_{ci'}Y_{cj'}] = E[Y_{ci}Y_{cj}Y_{ci'}Y_{cj'}] - E[Y_{ci}Y_{cj}]E[Y_{ci'}Y_{cj'}] = \lambda_{c,iji'j'} - \lambda_{c,ij}\lambda_{c,i'j'},
$$

for all $i\neq j\neq i'\neq j'$, where the formula for $\lambda_{c,iji'j'}$ is derived in (45) under Lemma 4(b), along with the formula for $\lambda_{c,ij}$ from (26). □

Given that the covariance matrices of $\boldsymbol{p}_{c\in\mathcal{F}}$ and $\boldsymbol{q}_{c\in\mathcal{F}}$ are derived in (47) and (48), respectively, we use them and construct a GQL estimating equation for σ^2_γ, as in Lemma 6 below.

Lemma 6 *Write*

$$
E[\boldsymbol{p}_{c\in\mathcal{F}}] = \lambda_c\left(\boldsymbol{\beta},\sigma^2_\gamma\right) = (\lambda_{c,11}(\cdot),\ldots,\lambda_{c,ii}(\cdot),\ldots,\lambda_{c,N_cN_c})'\,:\,N_c\times 1, \tag{49}
$$

$$
\begin{aligned} E\left[\boldsymbol{q}_{c\in\mathcal{F}}\right] &= \tilde{\lambda}_c\left(\boldsymbol{\beta},\sigma^2_\gamma\right) \\ &= (\lambda_{c,12}(\cdot),\ldots,\lambda_{c,ij}(\cdot),\ldots,\lambda_{c,N_c-1,N_c})'\,:\,N_c(N_c-1)/2\times 1, \end{aligned} \tag{50}
$$

where $\lambda_{c,ii}(\cdot)$ *is given in* (25), *and* $\lambda_{c,ij}(\cdot)$ in (26). *Use these mean moments and their covariances from* (47)–(48) *and write*

$$
\boldsymbol{p}_{c\in\mathcal{F}} \sim (\lambda_c\left(\boldsymbol{\beta},\sigma^2_\gamma\right), \boldsymbol{\Psi}_{c,N_c}(\boldsymbol{\beta},\sigma^2_\gamma)), \;\text{ and }\; \boldsymbol{q}_{c\in\mathcal{F}} \sim (\tilde{\lambda}_c\left(\boldsymbol{\beta},\sigma^2_\gamma\right), \boldsymbol{\Omega}_{c,N_c}(\boldsymbol{\beta},\sigma^2_\gamma)).
$$

Then, the HGQL estimating equation for σ^2_γ *is given by*

$$\begin{aligned}\tau_y(\sigma^2_\gamma) &= \tau_{y,1}(\sigma^2_\gamma) + \tau_{y,2}(\sigma^2_\gamma)\\ &= \sum_{c=1}^{K} \frac{\partial \boldsymbol{\lambda}^\top_c\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)}{\partial \sigma^2_\gamma} \boldsymbol{\Psi}^{-1}_{c,N_c}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)(\boldsymbol{p}_{c\in\mathcal{F}} - \lambda_c\left(\boldsymbol{\beta}, \sigma^2_\gamma\right))\\ &\quad + \sum_{c=1}^{K} \frac{\partial \tilde{\boldsymbol{\lambda}}^\top_c\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)}{\partial \sigma^2_\gamma} \boldsymbol{\Omega}^{-1}_{c,N_c}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)\left(\boldsymbol{q}_{c\in\mathcal{F}} - \tilde{\lambda}_c\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)\right) = 0.\end{aligned} \tag{51}$$

Proof Similar to the GQL estimating equation for $\boldsymbol{\beta}$ given in (33), the GQL estimating equation for σ^2_γ in (51) follows from Sutradhar (2003, Section 3; 2004, Eqn. (3.4)). The derivatives involved in (51) are presented in Lemma 7 below. □

Lemma 7 *For $\gamma^*_c(v) = [v - V(1/2)]/\sqrt{V(1/2)(1/2)}$ as used in (27), the derivatives involved in (51), more specifically the derivatives* of $\lambda_{c,ii}(\boldsymbol{\beta}, \sigma^2_\gamma)$ [*see* (25)] *and* $\lambda_{c,ij}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)$ [*see* (26)] *with respect* to σ^2_γ, *are given by*

$$\begin{aligned}\frac{\partial \lambda_{c,ii}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)}{\partial \sigma^2_\gamma} = \frac{1}{2\sigma_\gamma} \sum_{v=0}^{V} &\left[\left\{b^{(3)}(\theta_{ci}(\sigma_\gamma\gamma^*_c(v))) + 2b^{(1)}(\theta_{ci}(\sigma_\gamma\gamma^*_c(v)))b^{(2)}(\theta_{ci}(\sigma_\gamma\gamma^*_c(v)))\right\}\right.\\ &\left.\times\ \gamma^*_c(v)\right]\binom{V}{v}(1/2)^v(1/2)^{V-v},\end{aligned} \tag{52}$$

and

$$\begin{aligned}\frac{\partial \lambda_{c,ij}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)}{\partial \sigma^2_\gamma} = \frac{1}{2\sigma_\gamma} \sum_{v=0}^{V} &\left[\left\{b^{(1)}(\theta_{ci}(\sigma_\gamma\gamma^*_c(v)))b^{(2)}(\theta_{cj}(\sigma_\gamma\gamma^*_c(v))) + b^{(2)}(\theta_{ci}(\sigma_\gamma\gamma^*_c(v)))\right.\right.\\ &\left.\left.\times\ b^{(1)}(\theta_{cj}(\sigma_\gamma\gamma^*_c(v)))\right\}\gamma^*_c(v)\right]\binom{V}{v}(1/2)^v(1/2)^{V-v},\end{aligned} \tag{53}$$

respectively, where $\gamma^*_c(v) = [v - V(1/2)]/\sqrt{V(1/2)(1/2)}$, *similar to* (27), *for example*.

Proof Using the formulas for $\lambda_{c,ii}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)$ and $\lambda_{c,ij}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)$ from (25) and (26), respectively, we write

$$\frac{\partial \lambda_{c,ii}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)}{\partial \sigma^2_\gamma} = \int \left[\frac{\partial b^{(2)}(\theta_{ci}(\sigma_\gamma\gamma^*_c))}{\partial \sigma^2_\gamma} + \frac{\partial \{b^{(1)}(\theta_{ci}(\sigma_\gamma\gamma^*_c))\}^2}{\partial \sigma^2_\gamma}\right] f_N(\gamma^*_c)d\gamma^*_c, \tag{54}$$

$$\frac{\partial \lambda_{c,ij}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)}{\partial \sigma^2_\gamma} = \int \frac{\partial\left[b^{(1)}(\theta_{ci}(\sigma_\gamma\gamma^*_c))b^{(1)}(\theta_{cj}(\sigma_\gamma\gamma^*_c))\right]}{\partial \sigma^2_\gamma} f_N(\gamma^*_c)d\gamma^*_c. \tag{55}$$

Now because the continuous function $b(\theta_{ci}\ (\boldsymbol{\beta};\ \boldsymbol{x}_{ci}, \gamma_c))$ involved in the exponential density (18) is a function of

$$\theta_{ci}(\boldsymbol{\beta}; \boldsymbol{x}_{ci}, \gamma_c) = [\boldsymbol{x}_{ci}^{\top}\boldsymbol{\beta} + \gamma_c] = [\boldsymbol{x}_{ci}^{\top}\boldsymbol{\beta} + \sigma_\gamma \gamma_c^*]$$

(see Lemma 2 for the scalar transformation), the derivatives in (54) and (55), reduce to

$$\frac{\partial \lambda_{c,ii}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)}{\partial \sigma_\gamma^2} = \frac{1}{2\sigma_\gamma} \int \Big[\left\{b^{(3)}(\theta_{ci}(\sigma_\gamma \gamma_c^*)) + 2b^{(1)}(\theta_{ci}(\sigma_\gamma \gamma_c^*))b^{(2)}(\theta_{ci}(\sigma_\gamma \gamma_c^*))\right\} \\ \times \gamma_c^*\Big] f_N(\gamma_c^*) d\gamma_c^*, \tag{56}$$

and

$$\frac{\partial \lambda_{c,ij}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)}{\partial \sigma_\gamma^2} = \frac{1}{2\sigma_\gamma} \int \Big[\left\{b^{(1)}(\theta_{ci}(\sigma_\gamma \gamma_c^*))b^{(2)}(\theta_{cj}(\sigma_\gamma \gamma_c^*)) + b^{(2)}(\theta_{ci}(\sigma_\gamma \gamma_c^*))b^{(1)}(\theta_{cj}(\sigma_\gamma \gamma_c^*))\right\} \\ \times \gamma_c^*\Big] f_N(\gamma_c^*) d\gamma_c^*, \tag{57}$$

respectively. Finally, using the transformation $\gamma_c^*(v) = [v - V(1/2)]/\sqrt{V(1/2)(1/2)}$, the integrations in (54) and (55) are replaced by the summations shown in (52) and (53), respectively. $\square$

Next, we apply Lemma 7 and re-express the HGQL estimating equation for σ_γ^2 from (51) under Lemma 6, in the FP total form as in Lemma 8 below.

Lemma 8 *Denote the inverse of the fourth-order covariance matrices in* (47) *and* (48), *by*

$$\begin{aligned} \boldsymbol{\Psi}_{c,N_c}^{-1}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) &\equiv \left(\psi_{c,N_c,ij}^{(-1)}\right) : N_c \times N_c, \quad \text{and} \\ \boldsymbol{\Omega}_{c,N_c}^{-1}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) &\equiv \left(\omega_{c,N_c,ij,i'j'}^{(-1)}\right) : N_c(N_c-1)/2 \times N_c(N_c-1)/2, \end{aligned} \tag{58}$$

respectively. Next use these inverse matrices and the derivatives from Lemma 7, and write

$$\frac{\partial \lambda_c^{\top}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)}{\partial \sigma_\gamma^2}\left(\psi_{c,N_c,ij}^{(-1)}\right) \\ = \left(\xi_{c1}\left(\beta, \sigma_\gamma^2\right), \ldots, \xi_{ci}\left(\beta, \sigma_\gamma^2\right), \ldots, \xi_{cN_c}\left(\beta, \sigma_\gamma^2\right)\right), \text{ (say)}, \tag{59}$$

where $\xi_{ci}\left(\beta, \sigma_\gamma^2\right) = \sum_{u=1}^{N_c} \frac{\partial \lambda_{c,uu}\left(\beta, \sigma_\gamma^2\right)}{\partial \sigma_\gamma^2} \psi_{c,N_c,ui}^{(-1)}$, *and*

$$\frac{\partial \tilde{\lambda}_c^{\top}\left(\beta, \sigma_\gamma^2\right)}{\partial \sigma_\gamma^2}\left(\omega_{c,N_c,ij,i'j'}^{(-1)}\right) = (\tilde{\xi}_{c1}\left(\beta, \sigma_\gamma^2\right), \ldots, \tilde{\xi}_{c,ij}\left(\beta, \sigma_\gamma^2\right), \ldots, \tilde{\xi}_{c,N_c-1,N_c}\left(\beta, \sigma_\gamma^2\right)), \text{ (say) }, \tag{60}$$

where $\tilde{\xi}_{c,ij}\left(\beta, \sigma_\gamma^2\right) = \sum_{u<\ell}^{N_c} \frac{\partial \lambda_{c,u\ell}\left(\beta, \sigma_\gamma^2\right)}{\partial \sigma_\gamma^2} \omega_{c,N_c,u\ell,ij}^{(-1)}$. *The HGQL estimating function/equation in* (51) *then has the FP total form given by*

$$\tau_y(\sigma_\gamma^2) = \tau_{y,1}(\sigma_\gamma^2) + \tau_{y,2}(\sigma_\gamma^2) = \sum_{c=1}^{K}\sum_{i=1}^{N_c} \xi_{ci}\left(\beta, \sigma_\gamma^2\right)\left(y_{ci}^2 - \lambda_{c,ii}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)\right) + \sum_{c=1}^{K}\sum_{i<j}^{N_c} \tilde{\xi}_{c,ij}\left(\beta, \sigma_\gamma^2\right)\left(y_{ci}y_{cj} - \lambda_{c,ij}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)\right) = 0. \tag{61}$$

Proof The result in (61) follows by combining (58), (59), and (60) through a matrix algebra. □

5 Survey sample-based doubly weighted (SSDW) GQL estimation

It is clear from the last section that if the $(\mathcal{F}) \equiv \{(y_{ci}, x_{ci}), i = 1, \ldots, N_c;\ c = 1, \ldots, K\}$ was available, one would then have consistently and efficiently estimated $\boldsymbol{\beta}$ by solving the HGQL estimating equation for $\boldsymbol{\beta}$ given in (34) under Lemma 3, i.e., $\tau_y(\boldsymbol{\beta}) = \sum_{c=1}^{K}\sum_{i=1}^{N_c} \boldsymbol{a}_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)\left(y_{ci} - \mu_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)\right) = 0$, and σ_γ^2 by solving the HGQL estimating equation $\tau_y(\sigma_\gamma^2) = 0$ for σ_γ^2, derived in (61) under Lemma 8. Notice that because $\boldsymbol{a}_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)$ in (34), and $\xi_{ci}\left(\beta, \sigma_\gamma^2\right)$ and $\tilde{\xi}_{c,ij}\left(\beta, \sigma_\gamma^2\right)$ in (61), are constructed based on the inverse correlation/covariance matrices, it is convenient to refer them as the correlation weights for the FP responses involved in the FP total function. However, as the $(\mathcal{F}) \equiv \{(y_{ci}, x_{ci}), i = 1, \ldots, N_c;\ c = 1, \ldots, K\}$ is not available in practice, in this section, we utilize the two-stage sample SS $(s^*) \equiv \{(y_{ci}, x_{ci}), i = 1, \ldots, n_c;\ c = 1, \ldots, k\}$ defined in (6), to construct the SS-based suitable estimating equations for both $\boldsymbol{\beta}$ and σ_γ^2.

For the purpose, it is essential that the s^*-based estimating functions are unbiased for the $(\mathcal{F})$-based estimating functions, namely for $\tau_y(\boldsymbol{\beta})$ (34) for $\boldsymbol{\beta}$ estimation, and for $\tau_y(\sigma_\gamma^2)$ in (61) for σ_γ^2 estimation. To develop such sample-based functions/equations, one needs to accommodate (i) the sampling weights for the selection of s^* from $\mathcal{F}$, and (ii) use the sample-based correlation weights, for example, $\boldsymbol{a}_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)$ for $(c, i) \in s^*$, for $\boldsymbol{\beta}$ estimation. Specific details for the development of SS-based estimating function for $\boldsymbol{\beta}$ estimation and σ_γ^2 estimation are given in Sects. 5.1 and 5.2.

As far as the sampling weights are concerned, let $w_{ci\in s^*}$ denote such sampling design (D_{s^*}) weights for the selection of s^* from $\mathcal{F}$. Based on the first and the second-stage sampling designs given in (9), we will use the sampling weights as

$$w_{(c,i)\in s^*} = \frac{K}{k}\frac{N_c}{n_c}, \tag{62}$$

which is also used by others such as Valliant (1987, Section 2), and Lee et al. (2016, Eqn. (2.1)), especially for the estimation of the FP total ($\tau_y = \sum_{c=1}^{K}\sum_{i=1}^{N_c} y_{ci}$).

5.1 SSDW estimating equation for β in sample total form

As outlined above, we develop the two-stage survey sampling-based doubly weighted (SSDW) GQL estimating equation for $\boldsymbol{\beta}$ as in the following theorem.

Theorem 1 *Re-express the* ($\mathcal{F}$) from (6) as

$$\begin{aligned} \mathcal{F}:\ & \{(y_{ci},\boldsymbol{x}_{ci}), i=1,\dots,N_c;\ c=1,\dots,K\} \\ & \Rightarrow \{(y_{ci},\boldsymbol{a}_{ci}), i=1,\dots,N_c;\ c=1,\dots,K\}, \end{aligned} \tag{63}$$

where $\boldsymbol{a}_{ci}(\cdot)$ *denotes an inverse correlation-based weighted function of covariates in the cth cluster, as shown by* (35). *To reflect this change in* ($\mathcal{F}$), *modify the survey sample* s^* *from* (6) *as*

$$\begin{aligned} SS(s^*):\ & \{(y_{ci},\boldsymbol{x}_{ci}), i=1,\dots,n_c;\ c=1,\dots,k\} \\ & \Rightarrow \{(y_{ci},\boldsymbol{a}_{ci}), i=1,\dots,n_c;\ c=1,\dots,k\}. \end{aligned} \tag{64}$$

We may then exploit the modified $SS(s^*)$ *from* (64) *along with the sampling weights* $w_{(c,i)\in s^*}$ *from* (62), *and estimate* $\boldsymbol{\beta}$ *involved in the* ($\mathcal{F}$)*-based estimating function* $\tau_y(\boldsymbol{\beta})$ in (34), *by solving the SSDW (survey sample-based doubly weighted) estimating equation, given by*

$$\hat{\tau}_y(\boldsymbol{\beta}) = \sum_{c=1}^{k}\sum_{i=1}^{n_c} w_{(c,i)\in s^*}\boldsymbol{a}_{ci}(\cdot)\Big(y_{ci} - \mu_{ci}\Big(\boldsymbol{\beta},\sigma_\gamma^2\Big)\Big) = 0 \Rightarrow \hat{\boldsymbol{\beta}}_{\text{SSDW}}, \tag{65}$$

as we can show that the proposed $\hat{\tau}_y(\boldsymbol{\beta})$ *in* (65) *is design unbiased for the targeted* ($\mathcal{F}$)*-based estimating function*

$$\tau_y(\boldsymbol{\beta}) = \sum_{c=1}^{K}\sum_{i=1}^{N_c} \boldsymbol{a}_{ci}\Big(\boldsymbol{\beta},\sigma_\gamma^2\Big)\Big(y_{ci} - \mu_{ci}\Big(\boldsymbol{\beta},\sigma_\gamma^2\Big)\Big), \tag{66}$$

in (34).

Proof Recall from (13) that under the present two-stage sampling setup, the design expectation of a data function, namely $E_{D_{s^*}}(\cdot)$, is equivalent to two successive expectations, written as

$$E_{D_{s^*}}(\cdot) \equiv E_{p_1}E_{p_{2c}}[(\cdot)|p_1],$$

where p_1 is the first-stage sampling design for clusters selection, and p_{2c} is the second-stage sampling design for individuals selection within a given cluster (c). Thus, we express the design expectation over the SSDWGQL estimating function $\hat{\tau}_y(\boldsymbol{\beta})$, as

$$\begin{aligned} E_{D_{s^*}}\left[\hat{\tau}_y(\boldsymbol{\beta})\right] &= E_{p_1}\left[\sum_{c=1}^{k} E_{p_{2c}}\left\{\sum_{i=1}^{n_c} w_{(c,i)\in s^*}\boldsymbol{a}_{ci}(\cdot)\left(y_{ci}-\mu_{ci}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\right)\right\}|p_1\right] \\ &= \frac{K}{k}\left[E_{p_1}\sum_{c=1}^{k}\frac{N_c}{n_c}E_{p_{2c}}\left\{\sum_{i=1}^{n_c}\boldsymbol{a}_{ci}(\cdot)\left(y_{ci}-\mu_{ci}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\right)\right\}|p_1\right], \end{aligned} \tag{67}$$

by (62).

Next, we define two random indicator variables as follows. First, let $\delta_{1,c}$ be a random indicator variable such that

$$\delta_{1,c} = \begin{cases} 1 & \text{if the } c- \text{ th cluster is in the first-stage sample } s_1^* \\ 0 & \text{otherwise} \end{cases} \tag{68}$$

[Cochran (1977, Section 2.9)], with $E_{p_1}[\delta_{1,c}] = \frac{k}{K}$, following (9(i)). Similarly, let $\delta_{2,i|c}$ be the second random indicator variable such that

$$\delta_{2,i|c} = \begin{cases} 1 & \text{if the } i \text{ th unit from the } c\text{-th cluster is in the sample } s_{2c}^* \\ 0 & \text{otherwise} \end{cases} \tag{69}$$

with $E_{p_{2c}}[\delta_{2,i|c}] = \frac{n_c}{N_c}$, following (9(ii)). We may then re-express the design expectation in (67), as

$$\begin{aligned} E_{D_{s^*}}[\hat{\tau}_y(\beta)] &= \frac{K}{k}\sum_{c=1}^{K}\frac{N_c}{n_c}E_{p_1}\delta_{1,c}\left[\sum_{i=1}^{N_c}E_{p_{2c}}\delta_{2,i|c}\boldsymbol{a}_{ci}(\cdot)\left(y_{ci}-\mu_{ci}\left(\beta,\sigma_\gamma^2\right)\right)\right] \\ &= \frac{K}{k}\sum_{c=1}^{K}\frac{N_c}{n_c}E_{p_1}\delta_{1,c}\left[\sum_{i=1}^{N_c}\frac{n_c}{N_c}\boldsymbol{a}_{ci}(\cdot)\left(y_{ci}-\mu_{ci}\left(\beta,\sigma_\gamma^2\right)\right)\right], \text{ using (69)} \\ &= \frac{K}{k}\sum_{c=1}^{K}\frac{N_c}{n_c}\frac{k}{K}\left[\sum_{i=1}^{N_c}\frac{n_c}{N_c}\boldsymbol{a}_{ci}(\cdot)\left(y_{ci}-\mu_{ci}\left(\beta,\sigma_\gamma^2\right)\right)\right], \text{ using (68)} \\ &= \sum_{c=1}^{K}\sum_{i=1}^{N_c}\boldsymbol{a}_{ci}(\cdot)\left(y_{ci}-\mu_{ci}\left(\beta,\sigma_\gamma^2\right)\right), \end{aligned} \tag{70}$$

which is the same as the HGQL estimating function $\tau_y(\beta)$, in (66). Hence, the SSDW estimating function $\hat{\tau}_y(\beta)$, in (65), is design unbiased for the HGQL estimating function $\tau_y(\beta)$, in (66). □

5.2 SSDW estimating equation for σ^2_γ in sample total form

Notice that unlike the use of the first-order responses for the estimation of $\boldsymbol{\beta}$, one uses second-order responses to estimate this cluster variance/correlation parameter (Jiang 1998; Sutradhar 2004). A reflection of this difference is clear from the HGQL estimating equation (34) for $\boldsymbol{\beta}$, and the HGQL estimating equation (51) for σ_γ^2. More specifically, the HGQL estimating function in the left hand side of (51) consists of two sub-functions. The first sub-function $\tau_{y,1}(\sigma_\gamma^2)$ is based on squared responses, whereas the sub-function $\tau_{y,2}(\sigma_\gamma^2)$ was written separately using the pair-wise product responses. Now because the estimating function $\tau_{y,1}(\sigma_\gamma^2)$ is quite similar to that of the estimating function $\tau_y(\boldsymbol{\beta})$ (34) for $\boldsymbol{\beta}$, the survey sample (s^*) -based unbiased estimating function for $\tau_{y,1}(\sigma_\gamma^2)$ would be similar to that of $\hat{\tau}_y(\boldsymbol{\beta})$ developed in (65). However, for computational convenience, the specific formula for an unbiased function $\hat{\tau}_{y,1}(\sigma_\gamma^2)$, along with a slightly different type of unbiased function $\hat{\tau}_{y,2}(\sigma_\gamma^2)$ for $\tau_{y,2}(\sigma_\gamma^2)$ is provided in Theorem 2 below.

Theorem 2 *In view of the two sub-functions in the FP ($\mathcal{F}$)-based HGQL estimating function for σ_γ^2* in (61) [*see also* (51)], *let the FP ($\mathcal{F}$) from* (6) *be re-expressed as*

$$\begin{aligned} \mathcal{F}:\ & \{(y_{ci}, \boldsymbol{x}_{ci}), i = 1, \dots, N_c; c = 1, \dots, K\} \\ & \Rightarrow \{(y_{ci}^2, \xi_{ci}), i = 1, \dots, N_c; c = 1, \dots, K\}, \end{aligned} \tag{71}$$

as a reflection of the first sub-function in (61), *and as*

$$\begin{aligned} \mathcal{F}:\ & \{(y_{ci}, \boldsymbol{x}_{ci}), i = 1, \dots, N_c;\ c = 1, \dots, K\} \\ & \Rightarrow \{(y_{ci}y_{cj}, \tilde{\xi}_{c,ij}), i < j = 2, \dots, N_c;\ c = 1, \dots, K\}, \end{aligned} \tag{72}$$

as a reflection of the second sub-function in (61). *Also suppose that the survey sample s^* in* (6) *with selected sampled individuals* $(\{1, \dots, i, \dots, n_c;\ c = 1, \dots, k\})$ *from* $\mathcal{F}:\ \{1, \dots, i, \dots, N_c;\ c = 1, \dots, K\}$, *is re-expressed as*

$$\begin{aligned} \mathcal{F}:\ & \{(y_{ci}^2, \xi_{ci}), i = 1, \dots, N_c;\ c = 1, \dots, K\} \\ & \Rightarrow s^*:\ \{(y_{ci}^2, \xi_{ci}), i = 1, \dots, n_c;\ c = 1, \dots, k\}, \end{aligned} \tag{73}$$

corresponding to (71), and as

$$\begin{aligned} \mathcal{F}:\ & \{(y_{ci}y_{cj}, \tilde{\xi}_{c,ij}), i < j = 2, \dots, N_c; c = 1, \dots, K\} \\ & \Rightarrow s^*:\ \{(y_{ci}y_{cj}, \tilde{\xi}_{c,ij}), i < j = 2, \dots, n_c; c = 1, \dots, k\}, \end{aligned} \tag{74}$$

corresponding to (72). *We may then exploit the modified SS(s^*) from* (73)–(74) *along with the sampling weights $w_{(c,i)\in s^*}$ from* (62), *and estimate σ_γ^2 involved in the ($\mathcal{F}$)-based estimating function $\tau_y(\sigma_\gamma^2)$ in* (61), *by solving the SSDW (survey sample-based doubly weighted) estimating equation for σ_γ^2*, given by

$$\begin{aligned}\hat{\tau}_y(\sigma_\gamma^2) = \hat{\tau}_{y,1}(\sigma_\gamma^2) + \hat{\tau}_{y,2}(\sigma_\gamma^2) &= \sum_{c=1}^{k}\sum_{i=1}^{n_c} w_{(c,i)\in s^*}\xi_{ci}\left(\beta,\sigma_\gamma^2\right)\left(y_{ci}^2 - \lambda_{c,ii}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\right)\\ &+ \sum_{c=1}^{k}\sum_{i<j}^{n_c} w_{(c,i)\in s^*}\left(\frac{N_c-1}{n_c-1}\right)\tilde{\xi}_{c,ij}\left(\beta,\sigma_\gamma^2\right)\left(y_{ci}y_{cj} - \lambda_{c,ij}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\right)\\ &= 0 \Rightarrow \hat{\sigma}^2_{\gamma,\mathrm{SSDW}},\end{aligned} \tag{75}$$

which yields the desired SSDW estimator $\hat{\sigma}^2_{\gamma,\mathrm{SSDW}}$ for σ^2_γ, *as we can show that SSDW estimating function* $\hat{\tau}_y(\sigma_\gamma^2)$ *in* (75) *is a design unbiased sample function for the* $(\mathcal{F})$-*based estimating function* $\tau_y(\sigma_\gamma^2)$ in (61).

Proof Notice that the first term in the left hand side of (75), i.e.,

$$\hat{\tau}_{y,1}(\sigma_\gamma^2) = \sum_{c=1}^{k}\sum_{i=1}^{n_c} w_{(c,i)\in s^*}\xi_{ci}\left(\beta,\sigma_\gamma^2\right)\left(y_{ci}^2 - \lambda_{c,ii}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\right)$$

is similar to $\hat{\tau}_y(\boldsymbol{\beta}) = \sum_{c=1}^{k}\sum_{i=1}^{n_c} w_{(c,i)\in s^*}\boldsymbol{a}_{ci}(\cdot)\left(y_{ci} - \mu_{ci}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\right)$ in (65) for $\boldsymbol{\beta}$ estimation. Thus, the design expectation of $\hat{\tau}_{y,1}(\sigma_\gamma^2)$ can be computed by using the same operation as in (70), yielding

$$E_{D_{s^*}}\left[\hat{\tau}_{y,1}(\sigma_\gamma^2)\right] = \sum_{c=1}^{K}\sum_{i=1}^{N_c}\xi_{ci}\left(\beta,\sigma_\gamma^2\right)\left(y_{ci}^2 - \lambda_{c,ii}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\right) = \tau_{y,1}(\sigma_\gamma^2), \tag{76}$$

which is the first term in the HGQL estimating function for σ_γ^2 in (61).

Next to examine the unbiasedness of the second term in the left hand side of (75), using the joint indicator variables from (69) for paired individuals selection, we take the design expectation of this second term as follows.

$$\begin{aligned}E_{D_{s^*}}\left[\hat{\tau}_{y,2}(\sigma_\gamma^2)\right] &= E_{D_{s^*}}\left[\sum_{c=1}^{k}\sum_{i<j}^{n_c} w_{(c,i)\in s^*}\left(\frac{N_c-1}{n_c-1}\right)\tilde{\xi}_{c,ij}\left(\beta,\sigma_\gamma^2\right)\left(y_{ci}y_{cj} - \lambda_{c,ij}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\right)\right]\\ &= \frac{K}{k}\sum_{c=1}^{K}\frac{N_c}{n_c}\left(\frac{N_c-1}{n_c-1}\right)E_{p_1}\delta_{1,c}\left[\sum_{i<j}^{N_c}E_{p_{2c}}\{\delta_{2,i|c}\delta_{2,j|c}\}\tilde{\xi}_{c,ij}\left(\beta,\sigma_\gamma^2\right)\left(y_{ci}y_{cj} - \lambda_{c,ij}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\right)\right]\\ &= \frac{K}{k}\sum_{c=1}^{K}\frac{N_c}{n_c}\left(\frac{N_c-1}{n_c-1}\right)\frac{k}{K}\left[\sum_{i<j}^{N_c}\frac{n_c(n_c-1)}{N_c(N_c-1)}\tilde{\xi}_{c,ij}\left(\beta,\sigma_\gamma^2\right)\left(y_{ci}y_{cj} - \lambda_{c,ij}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\right)\right]\\ &= \sum_{c=1}^{K}\sum_{i<j}^{N_c}\tilde{\xi}_{c,ij}\left(\beta,\sigma_\gamma^2\right)(y_{ci}y_{cj} - \lambda_{c,ij}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)) = \tau_{y,2}(\sigma_\gamma^2),\end{aligned} \tag{77}$$

which is the second term in the left hand side of the targeted $(\mathcal{F})$-based HGQL estimating function in (61) for σ_γ^2. This proves the theorem by combining (76) and (77) together. □

6 Asymptotic properties

6.1 Asymptotic normality of the SSDW estimator of the regression parameter β

Because the regression parameters are of main interest, in this section, we examine the asymptotic distributional behavior of $\hat{\boldsymbol{\beta}}_{\text{SSDW}}$ obtained by solving the SSDW (survey sample-based doubly weighted) estimating equation (65) for $\boldsymbol{\beta}$. Notice that in general this nonlinear estimating equation (65) is solved iteratively by using the large sample ($n = \sum_{c=1}^{k} n_c \to \infty$)-based first-order Taylor series approximation given by

$$\hat{\boldsymbol{\beta}}_{\text{SSDW}} - \boldsymbol{\beta} \simeq - \left[\sum_{c=1}^{k} \sum_{i=1}^{n_c} w_{(c,i)\in s^*} \frac{\partial \boldsymbol{a}_{ci}(\cdot)(y_{ci} - \mu_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right))}{\partial \boldsymbol{\beta}^\top} \right]^{-1} \times \left[\sum_{c=1}^{k} \sum_{i=1}^{n_c} w_{(c,i)\in s^*} \boldsymbol{a}_{ci}(\cdot)(y_{ci} - \mu_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)) \right] + o_p(1/\sqrt{n}), \tag{78}$$

which, for $n \to \infty$, by using the notations $\boldsymbol{y}_{c\in s^*} = (y_{c1}, \ldots, y_{ci}, \ldots, y_{cn_c})^\top$ and $\boldsymbol{\mu}_{c\in s^*}(\cdot) = (\mu_{c1}(\cdot), \ldots, \mu_{ci}(\cdot), \ldots, \mu_{cn_c}(\cdot))^\top$, may be re-expressed as

$$\hat{\boldsymbol{\beta}}_{\text{SSDW}} - \boldsymbol{\beta} = - \left[\frac{1}{k} \sum_{c=1}^{k} \frac{K}{k} \frac{N_c}{n_c} \boldsymbol{A}^*_{c,N_c}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) \frac{\partial \boldsymbol{\mu}_{c\in s^*}(\cdot)}{\partial \boldsymbol{\beta}^\top} \right]^{-1} \times \left[\frac{1}{k} \sum_{c=1}^{k} \frac{K}{k} \frac{N_c}{n_c} \boldsymbol{A}^*_{c,N_c}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) (\boldsymbol{y}_{c\in s^*} - \boldsymbol{\mu}_{c\in s^*}(\cdot)) \right] \tag{79}$$

$$= - \left[\frac{1}{k} \sum_{c=1}^{k} \frac{\partial \boldsymbol{f}_c(\boldsymbol{\beta}|\boldsymbol{y}_c)}{\partial \boldsymbol{\beta}^\top} \right]^{-1} \left[\frac{1}{k} \sum_{c=1}^{k} \boldsymbol{f}_c(\boldsymbol{\beta}|\boldsymbol{y}_c) \right], \text{ (say)}. \tag{80}$$

We remark that in the present two-stage cluster sampling setup, the sample size $n = \sum_{c=1}^{k} n_c \to \infty$, mainly by considering large number of independent clusters such that $k \to K \to \infty$. As far as the cluster size is concerned, it is enough to have n_c's as $n_c \to N_c$, where, as pointed out in Sect. 1, N_c's are large but fixed.

Now, the derivatives in (79), by using similar notations as in (37), may be expressed as

$$\frac{\partial \boldsymbol{\mu}_{c\in s^*}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)}{\partial \boldsymbol{\beta}^\top} = \begin{pmatrix} \frac{\partial \mu_{c1}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)}{\partial \boldsymbol{\beta}^\top} \\ \vdots \\ \frac{\partial \mu_{ci}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)}{\partial \boldsymbol{\beta}^\top} \\ \vdots \\ \frac{\partial \mu_{cn_c}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)}{\partial \boldsymbol{\beta}^\top} \end{pmatrix} : n_c \times p$$
$$= \left[(\boldsymbol{x}_{c1},\ldots,\boldsymbol{x}_{ci},\ldots,\boldsymbol{x}_{cn_c})\ \mathrm{diag}\left[\bar{\sigma}_{c,11},\ldots,\bar{\sigma}_{c,ii},\ldots,\bar{\sigma}_{c,n_cn_c}\right]\right]^\top$$
$$= \left[(\boldsymbol{x}_{c1},\ldots,\boldsymbol{x}_{ci},\ldots,\boldsymbol{x}_{cn_c})\bar{\boldsymbol{V}}_{c,n_c}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\right]^\top. \tag{81}$$

Also, in (79), the inverse correlation-based weight matrix has the form

$$\boldsymbol{A}^*_{c,N_c}\left(\boldsymbol{\beta},\sigma_\gamma^2\right) = \left(\boldsymbol{a}_{c1}\left(\boldsymbol{\beta},\sigma_\gamma^2\right),\ldots,\boldsymbol{a}_{cn_c}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\right) : p \times n_c, \tag{82}$$

which is a sample version of the $\mathcal{F}$-based weight matrix

$$\boldsymbol{A}_{c,N_c}\left(\boldsymbol{\beta},\sigma_\gamma^2\right) = (\boldsymbol{a}_{c1}\left(\boldsymbol{\beta},\sigma_\gamma^2\right),\ldots,\boldsymbol{a}_{ci}\left(\boldsymbol{\beta},\sigma_\gamma^2\right),\ldots,\boldsymbol{a}_{cN_c}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)) : p \times N_c,$$

defined in (38).

Following two lemmas will be applied to derive the asymptotic distribution of $\hat{\boldsymbol{\beta}}_{\mathrm{SSDW}}$. Lemma 9 below deals with the probability convergence of $\sum_{c=1}^k \frac{\partial \boldsymbol{f}_c(\boldsymbol{\beta}|\boldsymbol{y}_c)}{\partial \boldsymbol{\beta}^\top}$ in (80), whereas Lemma 10 below provides the covariance matrix of $\sum_{c=1}^k \boldsymbol{f}_c(\boldsymbol{\beta}|\boldsymbol{y}_c)$ in (80).

Lemma 9 *The inverse correlation-based weighted gradient matrix within the square bracket of the first term in the right hand side of* (80) *converges in probability as*

$$\sum_{c=1}^k \frac{\partial \boldsymbol{f}_c(\boldsymbol{\beta}|\boldsymbol{y}_c)}{\partial \boldsymbol{\beta}^\top} = \sum_{c=1}^k \frac{K}{k}\frac{N_c}{n_c}\boldsymbol{A}^*_{c,N_c}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\frac{\partial \boldsymbol{\mu}_{c\in s^*}}{\partial \boldsymbol{\beta}^\top}$$
$$\to_p \sum_{c=1}^K \boldsymbol{A}_{c,N_c}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\frac{\partial \boldsymbol{\mu}_{c\in\mathcal{F}}}{\partial \boldsymbol{\beta}^\top} = \tilde{\boldsymbol{G}}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\ \text{(say)}, \tag{83}$$

where $\boldsymbol{A}_{c,N_c}(\cdot)$ *is the* $p \times N_c$ *matrix defined in* (38), *and* $\frac{\partial \boldsymbol{\mu}_{c\in\mathcal{F}}}{\partial \boldsymbol{\beta}^\top}$ is the $N_c \times p$ *derivative matrix given by* (37).

Proof Notice that in (83)

$$\sum_{c=1}^k \frac{K}{k}\frac{N_c}{n_c}\boldsymbol{A}^*_{c,N_c}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\frac{\partial \boldsymbol{\mu}_{c\in s^*}}{\partial \boldsymbol{\beta}^\top} = \sum_{c=1}^k\sum_{i=1}^{n_c}\frac{K}{k}\frac{N_c}{n_c}\boldsymbol{a}_{ci}(\cdot)\frac{\partial \mu_{ci}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)}{\partial \boldsymbol{\beta}^\top}. \tag{84}$$

Further notice that for

$$z_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) = \boldsymbol{a}_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)(y_{ci} - \mu_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)), \tag{85}$$

it was shown in (70) that

$$\begin{aligned} &\sum_{c=1}^{k}\sum_{i=1}^{n_c}\frac{K}{k}\frac{N_c}{n_c} z_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) \\ &\quad \rightarrow_p E_{D_{s^*}}\left[\sum_{c=1}^{k}\sum_{i=1}^{n_c}\frac{K}{k}\frac{N_c}{n_c} z_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)\right] = \sum_{c=1}^{K}\sum_{i=1}^{N_c} z_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right). \end{aligned} \tag{86}$$

Hence, replacing $(y_{ci} - \mu_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right))$ in (85) with $\frac{\partial \mu_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)}{\partial \boldsymbol{\beta}^\top}$, and applying (86), we obtain the convergence as

$$\sum_{c=1}^{k}\sum_{i=1}^{n_c}\frac{K}{k}\frac{N_c}{n_c}\boldsymbol{a}_{ci}(\cdot)\frac{\partial \mu_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)}{\partial \boldsymbol{\beta}^\top} \rightarrow_p \sum_{c=1}^{K}\sum_{i=1}^{N_c}\boldsymbol{a}_{ci}(\cdot)\frac{\partial \mu_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)}{\partial \boldsymbol{\beta}^\top}, \tag{87}$$

which is the same as $\sum_{c=1}^{K}\boldsymbol{A}_{c,N_c}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)\frac{\partial \boldsymbol{\mu}_{c\in\mathcal{F}}}{\partial \boldsymbol{\beta}^\top} = \tilde{\boldsymbol{G}}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)$ given by (83) under the lemma. $\square$

In Lemma 10 below, we provide the formula for the covariance matrix of $\sum_{c=1}^{k}\boldsymbol{f}_c(\boldsymbol{\beta}|\boldsymbol{y}_c) = \sum_{c=1}^{k}\sum_{i=1}^{n_c} w_{(c,i)\in s^*}\boldsymbol{a}_{ci}(\cdot)\left(y_{ci} - \mu_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)\right) : p \times 1$, defined in (78)–(80).

Lemma 10 *For* $z_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) = \boldsymbol{a}_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)(y_{ci} - \mu_{ci}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right))$, *suppose that*

$$\boldsymbol{Z}_c = \sum_{i=1}^{N_c} z_{ci}(\cdot),\ \bar{\boldsymbol{Z}}_c = \frac{\boldsymbol{Z}_c}{N_c}, \bar{\boldsymbol{Z}} = \frac{1}{K}\sum_{c=1}^{K}\boldsymbol{Z}_c.$$

Also suppose that

$$\begin{aligned} \boldsymbol{V}_{1.}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) &= \frac{1}{K}\sum_{c=1}^{K}\left\{(\boldsymbol{Z}_c - \bar{\boldsymbol{Z}})(\boldsymbol{Z}_c - \bar{\boldsymbol{Z}})'\right\}, \quad \text{and} \\ \boldsymbol{V}_c\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) &= \frac{1}{N_c}\sum_{i=1}^{N_c} z_{ci}z_{ci}^\top - \frac{2}{N_c(N_c-1)}\sum_{i<j}^{N_c} z_{ci}z_{cj}^\top, \end{aligned} \tag{88}$$

denote the FP-based between and within clustered covariance matrices, respectively. Then,

$$\begin{aligned}\operatorname{cov}\left[\sum_{c=1}^{k}\boldsymbol{f}_c(\boldsymbol{\beta}|\boldsymbol{y}_c)\right] &= \operatorname{cov}\left(\sum_{c=1}^{k}\sum_{i=1}^{n_c} w_{(c,i)\in s^*}\boldsymbol{z}_{ci}(\cdot)\right)\\ &= K^2\left(\frac{K-k}{K}\right)\frac{1}{k}\boldsymbol{V}_{1\cdot}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)+(K/k)\left[\sum_{c=1}^{K}N_c^2\frac{N_c-n_c}{N_c}\frac{1}{n_c}\boldsymbol{V}_c\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\right]\\ &= \boldsymbol{V}_n^*\left(\boldsymbol{\beta},\sigma_\gamma^2\right),\ (\text{say})\,,\end{aligned} \tag{89}$$

which, when sampling fractions are assumed to be negligible, reduces to

$$\boldsymbol{V}_n^*\left(\boldsymbol{\beta},\sigma_\gamma^2\right)=\frac{K^2}{k}\boldsymbol{V}_{1\cdot}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)+(K/k)\left[\sum_{c=1}^{K}\frac{N_c^2}{n_c}\boldsymbol{V}_c\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\right]. \tag{90}$$

Proof Recall from Sect. 5.1 that under the present TSCS setup, the design expectation has the form $E_{D_{s^*}}(\cdot)\equiv E_{p_1}E_{p_{2c}}[(\cdot)|p_1]$. By this token for covariance computation, we use

$$\operatorname{cov}_{D_{s^*}}(\cdot)=\operatorname{cov}_{p_1}[E_{p_{2c}}\{(\cdot)|p_1\}]+E_{p_1}[\operatorname{cov}_{p_{2c}}\{(\cdot)|p_1\}].$$

More specifically, we need to compute

$$\begin{aligned}&\operatorname{cov}_{p_1}\left[E_{p_{2c}}\left\{\left(\sum_{c=1}^{k}\sum_{i=1}^{n_c}w_{(c,i)\in s^*}\boldsymbol{z}_{ci}\right)|p_1\right\}\right]\\ &=\operatorname{cov}_{p_1}\left[\frac{K}{k}\sum_{c=1}^{k}N_cE_{p_{2c}}\left\{\frac{1}{n_c}\sum_{i=1}^{n_c}\boldsymbol{z}_{ci}\right\}\right]\end{aligned} \tag{91}$$

and

$$\begin{aligned}&E_{p_1}\operatorname{cov}_{p_{2c}}\left(\sum_{c=1}^{k}\sum_{i=1}^{n_c}w_{(c,i)\in s^*}\boldsymbol{z}_{ci}\right)\\ &=E_{p_1}\left[(K^2/k^2)\sum_{c=1}^{k}N_c^2\operatorname{cov}_{p_{2c}}\left\{\frac{1}{n_c}\sum_{i=1}^{n_c}\boldsymbol{z}_{ci}\right\}\right].\end{aligned} \tag{92}$$

Now to simplify $E_{p_{2c}}\left\{\frac{1}{n_c}\sum_{i=1}^{n_c}\boldsymbol{z}_{ci}\right\}$ in (91) and $\operatorname{cov}_{p_{2c}}\left\{\frac{1}{n_c}\sum_{i=1}^{n_c}\boldsymbol{z}_{ci}\right\}$ in (92), we first write

$$\frac{1}{n_c}\sum_{i=1}^{n_c}\boldsymbol{z}_{ci}=\frac{1}{n_c}\sum_{i=1}^{N_c}\delta 2,i|c\boldsymbol{z}_{ci}, \tag{93}$$

where $\delta 2,i|c$ is the random indicator variable defined in (69) [see also Cochran (1977, Section 2.9)], with

$$E[\delta_{2,i|c}] = \frac{n_c}{N_c}, \quad \text{var}\,(\delta_{2,i|c}) = \frac{n_c}{N_c}(1 - \frac{n_c}{N_c}), \tag{94}$$

$$\begin{aligned}\text{cov}\,\left(\delta_{2,i|c}, \delta_{2,j|c}\right) &= E\left(\delta_{2,i|c}\delta_{2,j|c}\right) - E\left(\delta_{2,i|c}\right)E\left(\delta_{2,j|c}\right) \\ &= \frac{n_c(n_c - 1)}{N_c(N_c - 1)} - \left(\frac{n_c}{N_c}\right)^2 = -\frac{n_c}{N_c(N_c - 1)}(1 - \frac{n_c}{N_c}).\end{aligned} \tag{95}$$

Hence,

$$E_{p_{2c}}\left\{\frac{1}{n_c}\sum_{i=1}^{n_c} z_{ci}\right\} = \frac{1}{n_c}\sum_{i=1}^{N_c} E[\delta_{2,i|c}]z_{ci} = \frac{1}{N_c}\sum_{i=1}^{N_c} z_{ci}, \tag{96}$$

$$\begin{aligned}\text{cov}_{p_{2c}}\left\{\frac{1}{n_c}\sum_{i=1}^{n_c} z_{ci}\right\} &= \text{cov}_{p_{2c}}\left\{\frac{1}{n_c}\sum_{i=1}^{N_c} \delta_{2,i|c}z_{ci}\right\} \\ &= \frac{1}{n_c^2}\left[\sum_{i=1}^{N_c} z_{ci}z_{ci}^\top\,\text{var}\,(\delta_{2,i|c}) + 2\sum_{i<j}^{N_c} z_{ci}z_{ci}^\top\,\text{cov}\,(\delta_{2,i|c}, \delta_{2,j|c})\right] \\ &= \left(1 - \frac{n_c}{N_c}\right)\frac{1}{n_c}\left[\frac{1}{N_c}\sum_{i=1}^{N_c} z_{ci}z_{ci}^\top - \frac{2}{N_c(N_c - 1)}\sum_{i<j}^{N_c} z_{ci}z_{cj}^\top\right] \\ &= \left(1 - \frac{n_c}{N_c}\right)\frac{1}{n_c}\boldsymbol{V}_c\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)\end{aligned} \tag{97}$$

Now putting (96) in (91) we obtain

$$\begin{aligned}\text{cov}_{p_1}E_{p_{2c}}\left(\sum_{c=1}^{k}\sum_{i=1}^{n_c} w_{(c,i)\in s^*}z_{ci}\right) &= \text{cov}_{p_1}\left[(K/k)\sum_{c=1}^{k}\left\{\sum_{i=1}^{N_c} z_{ci}\right\}\right] = K^2\,\text{cov}_{p_1}\left[\frac{1}{k}\sum_{c=1}^{k}\boldsymbol{Z}_c\right] \\ &= K^2\left(\frac{K-k}{K}\right)\frac{1}{k}\frac{1}{K}\sum_{c=1}^{K}\left\{(\boldsymbol{Z}_c - \bar{\boldsymbol{Z}})(\boldsymbol{Z}_c - \bar{\boldsymbol{Z}})'\right\} \\ &= K^2\left(\frac{K-k}{K}\right)\frac{1}{k}\boldsymbol{V}_{1.}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right),\end{aligned} \tag{98}$$

by using the first matrix formula from (88). Similarly by using (97) in (92), we compute

$$E_{p_1}\operatorname{cov}_{p_{2c}}\left(\sum_{c=1}^{k}\sum_{i=1}^{n_c} w_{(c,i)\in s^*}z_{ci}\right) = E_{p_1}\left[(K^2/k^2)\sum_{c=1}^{k} N_c^2 \operatorname{cov}_{p_{2c}}\left\{\frac{1}{n_c}\sum_{i=1}^{n_c} z_{ci}\right\}\right]$$
$$= E_{p_1}\left[(K^2/k^2)\sum_{c=1}^{k} N_c^2 \frac{N_c-n_c}{N_c}\frac{1}{n_c}\boldsymbol{V}_c\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\right]$$
$$= (K/k)\left[\sum_{c=1}^{K} N_c^2 \frac{N_c-n_c}{N_c}\frac{1}{n_c}\boldsymbol{V}_c\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\right], \tag{99}$$

by using the second matrix formula from (88). The covariance formula in the lemma, more specifically in (89), now follows by adding (99) with (98). □

We now provide the main distributional result in Theorem 3 below.

Theorem 3

$$\lim_{k\to K\to\infty}\hat{\boldsymbol{\beta}}_{\mathrm{SSDW}} \to_d N_p(\boldsymbol{\beta}, \tilde{\boldsymbol{G}}^{-1}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)V_n^*\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\tilde{\boldsymbol{G}}^{-1}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)), \tag{100}$$

where $\tilde{\boldsymbol{G}}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)$ *is the* $p\times p$ *matrix given by* (83) *under Lemma* 9, *and the* $p\times p$ *matrix* $V_n^*\left(\boldsymbol{\beta},\sigma_\gamma^2\right)$ *has the formula as in* (89)–(90) under Lemma 10.

Proof First by using Lemma 9, more specifically by (83), it follows from (80) that

$$\left[\hat{\boldsymbol{\beta}}_{\mathrm{SSDW}}-\boldsymbol{\beta}\right] \to_p -k\tilde{\boldsymbol{G}}^{-1}\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\left[\frac{1}{k}\sum_{c=1}^{k}\boldsymbol{f}_c(\boldsymbol{\beta}|\boldsymbol{y}_c)\right]. \tag{101}$$

Next one may write

$$\bar{\boldsymbol{f}}_n(\boldsymbol{\beta}|\sigma_\gamma^2) = \frac{1}{k}\left[\sum_{c=1}^{k}\boldsymbol{f}_c(\boldsymbol{\beta}|\boldsymbol{y}_c)\right] \sim \left(0, \frac{1}{k^2}V_n^*\left(\boldsymbol{\beta},\sigma_\gamma^2\right)\right), \tag{102}$$

because $\boldsymbol{f}_c(\boldsymbol{\beta}|\boldsymbol{y}_c)$ has the formula given in (79), which by (70) converges to a quantity which is a zero/null vector by (34). This leads to $E_{D_{s^*}}\left(\frac{1}{k}\sum_{c=1}^{k}\boldsymbol{f}_c(\boldsymbol{\beta}|\boldsymbol{y}_c)\right)=0$, and $\operatorname{cov}_{D_{s^*}}\left(\frac{1}{k}\sum_{c=1}^{k}\boldsymbol{f}_c(\boldsymbol{\beta}|\boldsymbol{y}_c)\right)=\frac{1}{k^2}V_n^*\left(\boldsymbol{\beta},\sigma_\gamma^2\right)$ by Lemma 10 (see (89)).

Further because $\boldsymbol{y}_1,\dots,\boldsymbol{y}_c,\dots,\boldsymbol{y}_k$ involved in $\sum_{c=1}^{k}\boldsymbol{f}_c(\boldsymbol{\beta}|\boldsymbol{y}_c)$ are independent vectors from k clusters, we assume that the multivariate version of Lindeberg's condition holds, that is,

$$\lim_{k\to K\to\infty}\boldsymbol{V}^{*-1}_n\sum_{c=1}^{k}\sum_{\{\boldsymbol{f}'_c\boldsymbol{V}^{*-1}_n\boldsymbol{f}_c\}>\epsilon}\boldsymbol{f}_c\boldsymbol{f}'_c p^*(\boldsymbol{f}_c)=0 \tag{103}$$

holds, for all $\epsilon>0$, $p^*(\cdot)$ being the probability distribution of $\boldsymbol{f}_c$. Then, the Lindeberg–Feller central limit theorem [Amemiya (1985). Theorem 3.3.6] implies the following convergence in distribution $(\to_d)$:

$$\boldsymbol{Z}_n = k[\boldsymbol{V}_n^*]^{-\frac{1}{2}}\bar{\boldsymbol{f}}_n(\boldsymbol{\beta}) \rightarrow_d N_p(0, \boldsymbol{I}_p). \tag{104}$$

$\boldsymbol{I}_p$ being the $p \times p$ identity matrix. Next use (104) in (101) and obtain

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\text{SSDW}} - \boldsymbol{\beta} &= \tilde{\boldsymbol{G}}^{-1}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)[\boldsymbol{V}_n^*\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)]^{\frac{1}{2}}[\boldsymbol{V}_n^*\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)]^{-\frac{1}{2}}k\bar{\boldsymbol{f}}_n(\boldsymbol{\beta}|\sigma_\gamma^2)\\ &= \tilde{\boldsymbol{G}}^{-1}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)[\boldsymbol{V}_n^*\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)]^{\frac{1}{2}}\boldsymbol{Z}_n\\ &\rightarrow_d N_p(0, \tilde{\boldsymbol{G}}^{-1}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)\boldsymbol{V}_n^*\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)\tilde{\boldsymbol{G}}^{-1}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)),\end{aligned} \tag{105}$$

which is the same multivariate normal distribution stated in (100) under the theorem. □

6.2 Consistency of $\hat{\boldsymbol{\beta}}_{\text{SSDW}}$

We examine the asymptotic order of convergence of $\hat{\boldsymbol{\beta}}_{\text{SSDW}}$ to $\boldsymbol{\beta}$ in Theorem 4 below. For the purpose, we first notice that $\hat{\boldsymbol{\beta}}_{\text{SSDW}}$ satisfies the convergence relationship given in (101). Hence, its convergence to $\boldsymbol{\beta}$ will depend on two regularity conditions; first condition ($\boldsymbol{C}_1^*$) on $\tilde{\boldsymbol{G}}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)$ in (101), and the second condition ($\boldsymbol{C}_2^*$) will be needed on the covariance matrix of $\frac{1}{k}\sum_{c=1}^{k}\boldsymbol{f}_c(\boldsymbol{\beta}|\boldsymbol{y}_c)$. These conditions are developed as follows.

$\boldsymbol{C}_1^*$. By (38) re-express $\tilde{\boldsymbol{G}}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)$ in (101) (see also (83)) as

$$\begin{aligned}\tilde{\boldsymbol{G}}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) &= \sum_{c=1}^{K}\frac{\partial\boldsymbol{\mu}_c^\top\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)}{\partial\boldsymbol{\beta}}\boldsymbol{\Sigma}_{c,N_c}^{-1}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)\frac{\partial\boldsymbol{\mu}_c\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)}{\partial\boldsymbol{\beta}^\top}\\ &= (\boldsymbol{x}_{c1}, \ldots, \boldsymbol{x}_{ci}, \ldots, \boldsymbol{x}_{cN_c})\bar{\boldsymbol{V}}_{c,N_c}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)\boldsymbol{\Sigma}_{c,N_c}^{-1}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)\bar{\boldsymbol{V}}_{c,N_c}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)(\boldsymbol{x}_{c1}, \ldots, \boldsymbol{x}_{ci}, \ldots, \boldsymbol{x}_{cN_c})^\top\\ &= (\boldsymbol{x}_{c1}, \ldots, \boldsymbol{x}_{ci}, \ldots, \boldsymbol{x}_{cN_c})\Theta_{c,N_c}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)(\boldsymbol{x}_{c1}, \ldots, \boldsymbol{x}_{ci}, \ldots, \boldsymbol{x}_{cN_c})^\top\\ &= \sum_{c=1}^{K}\sum_{i=1}^{N_c}\theta_{c,ij}(\cdot)\boldsymbol{x}_{ci}\boldsymbol{x}_{cj}^\top : p \times p,\end{aligned} \tag{106}$$

and suppose that for an $N = \sum_{c=1}^{K}N_c$ dependent finite and bounded quantity ℓ_N, the fixed design covariates $\{\boldsymbol{x}_{ci}, i = 1, \ldots, N_c; c = 1, \ldots, K\}$ in the FP ($\mathcal{F}$) satisfy the condition given by

$$\frac{1}{\sum_{c=1}^{K}N_c}\left|\sum_{c=1}^{K}\sum_{i=1}^{N_c}\theta_{c,ij}(\cdot)\boldsymbol{x}_{ci}\boldsymbol{x}_{cj}^\top\right| \le \ell_N. \tag{107}$$

Before we write the second condition, in Lemma 11 below, we provide the model-assisted formulas for the between and within clustered covariance matrices defined in (88) under Lemma 10.

Lemma 11 *The model assisted between clustered covariance matrix, say* $\boldsymbol{V}_{(1\cdot)M}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)$, *and the model assisted within cluster covariance matrix, say* $\boldsymbol{V}_{(c)M}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)$, *have the formulas*

$$\boldsymbol{V}_{(1\cdot)M}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) = \frac{1}{K}\left(1 - \frac{1}{K}\right) \sum_{c=1}^{K} \sum_{i,j=1}^{N_c} \sigma_{c,ij}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) \boldsymbol{a}_{ci}(\cdot)\boldsymbol{a}_{cj}^\top(\cdot), \tag{108}$$

$$\begin{aligned} \boldsymbol{V}_{(c)M}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) = {} & \frac{1}{N_c} \sum_{i=1}^{N_c} \sigma_{c,ii}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) \boldsymbol{a}_{ci}(\cdot)\boldsymbol{a}_{ci}^\top(\cdot) \\ & - \frac{2}{N_c(N_c - 1)} \sum_{i<j}^{N_c} \sigma_{c,ij}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right) \boldsymbol{a}_{ci}(\cdot)\boldsymbol{a}_{cj}^\top(\cdot), \end{aligned} \tag{109}$$

where $\sigma_{c,ii}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)$, *and* $\sigma_{c,ij}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)$ *for* $i \neq j$, *are given by* (23) *and* (24), *respectively, and* $\boldsymbol{a}_{ci}(\cdot) : p \times 1$ *is given in* (35).

Proof Proof is available from Appendix A in the supplementary file. □

Now suppose that the above model-assisted clustered covariance matrices satisfy the following regularity condition.

$\boldsymbol{C}_2^*$. For finite and bounded quantities m_1 and m_2, the between and within clustered covariance matrices in (88) satisfy

$$\max\left\{\frac{K^2}{N}|\boldsymbol{V}_{(1\cdot)M}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)|\right\} \leq m_1, \quad \max_c\left\{\frac{KN_c}{N}|\boldsymbol{V}_{(c)M}\left(\boldsymbol{\beta}, \sigma_\gamma^2\right)|\right\} \leq m_2, \tag{110}$$

respectively. We use the aforementioned two regularity conditions and prove the consistency of $\hat{\boldsymbol{\beta}}_{\mathrm{SSDW}}$ as in the following theorem.

Theorem 4 *Suppose that the aforementioned two regularity conditions* $\boldsymbol{C}_1^*$ [*in* (107)] *and* $\boldsymbol{C}_2^*$ [*in* (110)] *hold. The SSDW regression parameter estimator obtained from* (65) [*see also* (78)] *then satisfies the order of convergence, as*

$$\begin{aligned} [\hat{\boldsymbol{\beta}}_{\mathrm{SSDW}} - \boldsymbol{\beta}] & \simeq O\left(N^{-1}\ell_N^{-1}\right) O_p\left(\sqrt{N}\left[m_1\frac{1}{k} + m_2\frac{1}{k}\sum_{c=1}^{K}\frac{1}{n_c}\right]^{\frac{1}{2}}\right) \\ & = O_p\left((1/\sqrt{N})\left[\frac{m_1}{\ell_N^2}\frac{1}{k} + \frac{m_2}{\ell_N^2}\frac{1}{k}\sum_{c=1}^{K}\frac{1}{n_c}\right]^{\frac{1}{2}}\right), \end{aligned} \tag{111}$$

implying that

$$\lim_{n_c \to N_c,\, k \to K \to \infty} (\hat{\boldsymbol{\beta}}_{\mathrm{SSDW}} - \boldsymbol{\beta}) \to_p 0, \tag{112}$$

because ℓ_N, m_1, and m_2, *all are finite and bounded quantities. Hence,* $\hat{\boldsymbol{\beta}}_{\text{SSDW}}$ *is consistent for* $\boldsymbol{\beta}$.

Proof Consider the first-order Taylor series approximation for $\hat{\boldsymbol{\beta}}_{\text{SSDW}}$ given by (78). Notice from Lemma 9 that the derivative matrix within the square bracket in the first term of the right hand side of (78) converges in probability to $\tilde{\boldsymbol{G}}\left(\boldsymbol{\beta},\sigma^2_\gamma\right)$ which by the regularity condition $\boldsymbol{C}^*_1$, more specifically by (107) has the order of convergence

$$|\tilde{\boldsymbol{G}}\left(\boldsymbol{\beta},\sigma^2_\gamma\right)| = |\sum_{c=1}^{K}\sum_{i=1}^{N_c}\theta_{c,ij}(\cdot)\boldsymbol{x}_{ci}\boldsymbol{x}^\top_{cj}| = O(\ell_N N). \tag{113}$$

Furthermore, for the second term in the right hand side of (78), one writes $\left[\sum_{c=1}^{k}\sum_{i=1}^{n_c} w_{(c,i)\in s^*}\boldsymbol{a}_{ci}(\cdot)(y_{ci}-\mu_{ci}\left(\boldsymbol{\beta},\sigma^2_\gamma\right))\right] \to_p 0$, in the order of

$$\begin{aligned}
&\left[\text{cov}\left(\sum_{c=1}^{k}\sum_{i=1}^{n_c} w_{(c,i)\in s^*}\boldsymbol{a}_{ci}(\cdot)(y_{ci}-\mu_{ci}\left(\boldsymbol{\beta},\sigma^2_\gamma\right))\right)\right]^{\frac{1}{2}}\\
&= \left[\frac{K^2}{k}\boldsymbol{V}_{(1\cdot)M}\left(\boldsymbol{\beta},\sigma^2_\gamma\right) + (K/k)\sum_{c=1}^{K}\frac{N_c^2}{n_c}\boldsymbol{V}_{(c)M}(\boldsymbol{\beta},\sigma^2_\gamma)\right]^{\frac{1}{2}}, \text{ by applying Lemma 11 to (89)}\\
&\equiv O_p\left(\sqrt{N}\left[m_1\frac{1}{k}+m_2\frac{1}{k}\sum_{c=1}^{K}\frac{1}{n_c}\right]^{\frac{1}{2}}\right),
\end{aligned} \tag{114}$$

by the regularity condition $\boldsymbol{C}^*_2$ in (108). Finally apply (113) and (114) to (78) and obtain the convergence order as in (111) under the theorem. □

6.3 Consistency of $\hat{\sigma}^2_{\gamma,\text{SSDW}}$ obtained from (75)

Use the first-order Taylor series expansion of the SSDW estimating function (75) for σ^2_γ, and write

$$\begin{aligned}
\left[\hat{\sigma}^2_{\text{SSDW}}-\sigma^2_\gamma\right] = &-\left[\sum_{c=1}^{k}\sum_{i=1}^{n_c} w_{(c,i)\in s^*}\xi_{ci}\left(\beta,\sigma^2_\gamma\right)\frac{\partial\lambda_{c,ii}\left(\boldsymbol{\beta},\sigma^2_\gamma\right)}{\partial\sigma^2_\gamma}\right.\\
&\left.+\sum_{c=1}^{k}\sum_{i<j}^{n_c} w_{(c,i)\in s^*}\left(\frac{N_c-1}{n_c-1}\right)\tilde{\xi}_{c,ij}\left(\beta,\sigma^2_\gamma\right)\frac{\partial\lambda_{c,ij}\left(\boldsymbol{\beta},\sigma^2_\gamma\right)}{\partial\sigma^2_\gamma}\right]^{-1}\left[\sum_{c=1}^{k}\sum_{i=1}^{n_c} w_{(c,i)\in s^*}z^*_{ci}\left(\boldsymbol{\beta},\sigma^2_\gamma\right)\right.\\
&\left.+\sum_{c=1}^{k}\sum_{i<j}^{n_c} w_{(c,i)\in s^*}\left(\frac{N_c-1}{n_c-1}\right)\tilde{z^*}_{c,ij}\left(\beta,\sigma^2_\gamma\right)\right] + o_p(1/\sqrt{n})\\
= &-S_1^{-1}S_{2,y} + o_p(1/\sqrt{n}),
\end{aligned} \tag{115}$$

where $z^*_{ci}\left(\boldsymbol{\beta},\sigma^2_\gamma\right)=\xi_{ci}\left(\beta,\sigma^2_\gamma\right)\left(y^2_{ci}-\lambda_{c,ii}\left(\boldsymbol{\beta},\sigma^2_\gamma\right)\right)$, and $\tilde{z}^*{}_{c,ij}\left(\beta,\sigma^2_\gamma\right)=\tilde{\xi}_{c,ij}\left(\beta,\sigma^2_\gamma\right)\left(y_{ci}y_{cj}-\lambda_{c,ij}\left(\boldsymbol{\beta},\sigma^2_\gamma\right)\right)$, by (75) (see also (61)).

Suppose that the following regularity condition $\boldsymbol{C}^*_3$ holds.

$\boldsymbol{C}^*_3$. For an N-dependent finite and bounded quantity h_N,

$$\frac{1}{\sum_{c=1}^K N_c}\left[\sum_{c=1}^{K}\sum_{i=1}^{N_c}\xi_{ci}\left(\beta,\sigma^2_\gamma\right)\frac{\partial\lambda_{c,ii}\left(\boldsymbol{\beta},\sigma^2_\gamma\right)}{\partial\sigma^2_\gamma} + \sum_{c=1}^{K}\sum_{i<j}^{N_c}\tilde{\xi}_{c,ij}\left(\beta,\sigma^2_\gamma\right)\frac{\partial\lambda_{c,ij}\left(\boldsymbol{\beta},\sigma^2_\gamma\right)}{\partial\sigma^2_\gamma}\right]\le h_N. \tag{116}$$

Lemma 12 $\boldsymbol{C}^*_3\Rightarrow S_1(\text{ in }(115))=O(Nh_N)$.

Proof To prove this convergence rate, write $S_{11}=\sum_{c=1}^k\sum_{i=1}^{n_c}w_{(c,i)\in s^*}\xi_{ci}\left(\beta,\sigma^2_\gamma\right)\frac{\partial\lambda_{c,ii}\left(\boldsymbol{\beta},\sigma^2_\gamma\right)}{\partial\sigma^2_\gamma}$, and $S_{12}=\sum_{c=1}^k\sum_{i<j}^{n_c}w_{(c,i)\in s^*}\left(\frac{N_c-1}{n_c-1}\right)\tilde{\xi}_{c,ij}\left(\beta,\sigma^2_\gamma\right)\frac{\partial\lambda_{c,ij}\left(\boldsymbol{\beta},\sigma^2_\gamma\right)}{\partial\sigma^2_\gamma}$, so that

$$S_1(\text{in}\,(115))=[S_{11}+S_{12}]. \tag{117}$$

Notice from (76)–(77) that $S_{11}\to_p E_{D_{s^*}}[S_{11}]=\sum_{c=1}^K\sum_{i=1}^{N_c}\xi_{ci}\left(\beta,\sigma^2_\gamma\right)\frac{\partial\lambda_{c,ii}\left(\boldsymbol{\beta},\sigma^2_\gamma\right)}{\partial\sigma^2_\gamma}$, and $S_{12}\to_p E_{D_{s^*}}\left[S_{12}\right]=\sum_{c=1}^K\sum_{i<j}^{N_c}\tilde{\xi}_{c,ij}\left(\beta,\sigma^2_\gamma\right)\frac{\partial\lambda_{c,ij}\left(\boldsymbol{\beta},\sigma^2_\gamma\right)}{\partial\sigma^2_\gamma}$, respectively. Apply them to (117), and write

$$S_1\to_p\left[\sum_{c=1}^{K}\sum_{i=1}^{N_c}\xi_{ci}\left(\beta,\sigma^2_\gamma\right)\frac{\partial\lambda_{c,ii}\left(\boldsymbol{\beta},\sigma^2_\gamma\right)}{\partial\sigma^2_\gamma}+\sum_{c=1}^{K}\sum_{i<j}^{N_c}\tilde{\xi}_{c,ij}\left(\beta,\sigma^2_\gamma\right)\frac{\partial\lambda_{c,ij}\left(\boldsymbol{\beta},\sigma^2_\gamma\right)}{\partial\sigma^2_\gamma}\right]=O(Nh_N), \tag{118}$$

by using (116), i.e., $\boldsymbol{C}^*_3$. This proves the lemma. □

Next, as far as the response dependent stochastic function $S_{2,y}$ in (115) is concerned, by applying the so-called stochastic convergence principle [e.g., Bishop et al. (1975, Theorem 14.4-1)], we can write

$$\left[S_{2,y} - E(S_{2,y})\right] = \left[S_{2,y} - E_M E_{D_{s^*}}[S_{2,y}]\right] = O_p\left[\text{ var }[S_{2,y}]\right]^{\frac{1}{2}}. \tag{119}$$

Now because by (76) and (77), one obtains

$$\begin{aligned} E_{D_{s^*}}[S_{2,y}] &= E_{D_{s^*}}\left[\sum_{c=1}^{k}\sum_{i=1}^{n_c} w_{(c,i)\in s^*} z^*_{ci}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right) + \sum_{c=1}^{k}\sum_{i<j}^{n_c} w_{(c,i)\in s^*}\left(\frac{N_c - 1}{n_c - 1}\right)\tilde{z^*}_{c,ij}\left(\beta, \sigma^2_\gamma\right)\right] \\ &= \sum_{c=1}^{K}\sum_{i=1}^{N_c}\left\{\xi_{ci}\left(\beta, \sigma^2_\gamma\right)\left(y^2_{ci} - \lambda_{c,ii}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)\right)\right\} + \sum_{c=1}^{K}\sum_{i<j}^{N_c}\left\{\tilde{\xi}_{c,ij}\left(\beta, \sigma^2_\gamma\right)\left(y_{ci}y_{cj} - \lambda_{c,ij}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)\right)\right\}, \end{aligned} \tag{120}$$

it then follows by using the $\mathcal{F}$-based moment properties from (23)–(24) that $E_M[Y^2_{ci} - \lambda_{c,ii}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)] = 0$, and $E_M[Y_{ci}Y_{cj} - \lambda_{c,ij}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)] = 0$, implying by (120) that

$$E(S_{2,y}) = E_M E_{D_{s^*}}[S_{2,y}] = 0. \tag{121}$$

Hence, by (119), one writes

$$S_{2,y} = O_p[\text{ var }[S_{2,y}]]^{\frac{1}{2}}. \tag{122}$$

The purpose of the following lemma is to derive the formula for this variance, i.e., var $\left[S_{2,y}\right]$ in (122).

Lemma 13 *The model-assisted design* (D_{s^*})*-based variance of* $S_{2,y}$ *is given by*

$$\text{var }[S_{2,y}] = E_M\left[\text{ var}_{D_{s^*}}(S_{2,y})\right] = E_M\left[\text{ var}_{p_1} E_{p_{2c}}[S_{2,y}] + E_{p_1}\text{ var}_{p_{2c}}[S_{2,y}]\right], \tag{123}$$

where

$$E_M\left[\text{ var}_{p_1} E_{p_{2c}}[S_{2,y}]\right] = K^2\left(\frac{K-k}{K}\right)\frac{1}{k}v^{\dagger}{}_{(1\cdot)M}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right), \tag{124}$$

$$\begin{aligned} E_M\left[E_{p_1}\text{ var}_{p_{2c}}[S_{2,y}]\right] &= \frac{K}{k}\sum_{c=1}^{K} N_c^2\left[\left\{\frac{N_c - n_c}{N_c}\frac{1}{n_c}v^*{}_{(c)M}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)\right\}\right. \\ &+ \left\{\frac{1}{n_c N_c}\frac{N_c - 1}{n_c - 1}\left[g_1(n_c, N_c)\Phi_{(1,N_c)M}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right) + g_2(n_c, N_c)\Phi_{(2,N_c)M}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right) + g_3(n_c, N_c)\Phi_{(3,N_c)M}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)\right]\right\} \\ &+ \left.\left\{\frac{2}{n_c N_c}\left[g_4(n_c, N_c)\Phi_{(4,N_c)M}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right) + g_5(n_c, N_c)\Phi_{(5,N_c)M}\left(\boldsymbol{\beta}, \sigma^2_\gamma\right)\right]\right\}\right], \end{aligned} \tag{125}$$

with

$$
\begin{aligned}
g_1(n_c,N_c) &= \left[1 - \frac{n_c(n_c-1)}{N_c(N_c-1)}\right] \\
g_2(n_c,N_c) &= \left[\frac{(n_c-2)(n_c-3)}{(N_c-2)(N_c-3)} - \frac{n_c(n_c-1)}{N_c(N_c-1)}\right] \\
g_3(n_c,N_c) &= \left[\frac{(n_c-2)}{(N_c-2)} - \frac{n_c(n_c-1)}{N_c(N_c-1)}\right] \\
g_4(n_c,N_c) &= \left[\frac{(n_c-2)}{(N_c-2)} - \frac{n_c}{N_c}\right] \\
g_5(n_c,N_c) &= \left(1 - \frac{n_c}{N_c}\right),
\end{aligned} \tag{126}
$$

and

$$
v^{\dagger}{}_{(1\cdot)M}\left(\boldsymbol{\beta},\sigma^2_\gamma\right),\ v^{*}{}_{(c)M}\left(\boldsymbol{\beta},\sigma^2_\gamma\right),\ \text{and}\ \{\Phi_{(u,N_c)M}\left(\boldsymbol{\beta},\sigma^2_\gamma\right); u=1,\ldots,5,\}
$$

as the $\mathcal{F}$-based scalar functions dependent on covariates only.

Proof The formulas for the $\mathcal{F}$ covariate-based scalar functions in the lemma are derived in Appendix B in the supplementary file. □

Now suppose that the following regularity condition hold:

$\boldsymbol{C}^*_4$. For $N = \sum_{c=1}^{K} N_c$-dependent finite and bounded quantities $r_{N,1}$, $r_{N,2}$, $r_{N,3}$, and $r_{N,4}$,

$$
\begin{aligned}
&\max\left\{\frac{K^2}{N} v^{\dagger}_{(1\cdot)M}\left(\boldsymbol{\beta},\sigma^2_\gamma\right)\right\} \le r_{N,1};\ \max_c\left\{\frac{KN_c^2}{N} v^{*}_{(c)M}\left(\boldsymbol{\beta},\sigma^2_\gamma\right)\right\} \le r_{N,2}; \\
&\max_c\left[\frac{KN_c^2}{N}\left\{g_1(n_c,N_c)\Phi_{(1,N_c)M}\left(\boldsymbol{\beta},\sigma^2_\gamma\right) + g_2(n_c,N_c)\Phi_{(2,N_c)M}\left(\boldsymbol{\beta},\sigma^2_\gamma\right)\right.\right. \\
&\quad \left.\left. +\ g_3(n_c,N_c)\Phi_{(3,N_c)M}\left(\boldsymbol{\beta},\sigma^2_\gamma\right)\right\}\right] \le r_{N,3}; \\
&\max_c\left[\frac{KN_c^2}{N}\left\{\frac{2}{N_c}\left(g_4(n_c,N_c)\Phi_{(4,N_c)M}\left(\boldsymbol{\beta},\sigma^2_\gamma\right) + g_5(n_c,N_c)\Phi_{(5,N_c)M}\left(\boldsymbol{\beta},\sigma^2_\gamma\right)\right)\right\}\right] \le r_{N,4}.
\end{aligned} \tag{127}
$$

We can now prove the consistency of $\hat{\sigma}^2_{\gamma,\mathrm{SSDW}}$ for σ^2_γ, as in Theorem 5 below.

Theorem 5 *For finite and bounded quantities h_N and $\{r_{N,1},\ r_{N,2},\ r_{N,3},\ r_{N,4}\}$ used to define the regularity conditions $\boldsymbol{C}^*_3$ and $\boldsymbol{C}^*_4$, the SSDW estimator $\hat{\sigma}^2_{\gamma,\mathrm{SSDW}}$ obtained from (75) satisfies the order of convergence, as*

$$\left[\hat{\sigma}^2_{\gamma,\mathrm{SSDW}} - \sigma^2_\gamma\right] \simeq O(N^{-1}h_N^{-1})$$
$$\times O_p\left(\sqrt{N}\left[r_{N,1}\frac{1}{k} + r_{N,2}\sum_{c=1}^{K}\frac{1}{kn_c} + r_{N,3}\sum_{c=1}^{K}\frac{1}{kn_c(n_c-1)} + r_{N,4}\sum_{c=1}^{K}\frac{1}{kn_c}\right]^{\frac{1}{2}}\right) + o_p(1/\sqrt{n})$$
$$= O_p\left((1/\sqrt{N})\left[\frac{r_{N,1}}{h_N^2}\frac{1}{k} + (r_{N,2}+r_{N,4})\frac{1}{h_N^2}\sum_{c=1}^{K}\frac{1}{kn_c} + \frac{r_{N,4}}{h_N^2}\sum_{c=1}^{K}\frac{1}{kn_c(n_c-1)}\right]^{\frac{1}{2}}\right)$$
$$+ o_p(1/\sqrt{n}), \tag{128}$$

where $n = \sum_{c=1}^{k} n_c$. *Thus, we obtain*

$$\lim_{n_c \to N_c,\, k \to K \to \infty, n=\sum_{c=1}^{k} n_c \to \sum_{c=1}^{K} N_c = N \to \infty} [\hat{\sigma}^2_{\gamma,\mathrm{SSDW}} - \sigma^2_\gamma] \to_p 0, \tag{129}$$

showing that $\hat{\sigma}^2_{\gamma,\mathrm{SSDW}}$ *obtained by solving the survey weighted GQL estimating Eq.* (75) *is consistent for* σ^2_γ.

Proof By (118), $S_1^{-1} \equiv O(N^{-1}h_N^{-1})$, where S_1 is defined in (115). Next, by (122),

$$S_{2,y} = O_p[\ \mathrm{var}\ (S_{2,y})]^{\frac{1}{2}},$$

where the variance,var $(S_{2,y})$, is obtained by putting (124) and (125) into (123) under the Lemma 13. The convergence order in (128) then follows from (115). □

7 Concluding remarks

When the two-stage cluster sample s^* constructed in (6) is treated to be a simple random sample-based single-stage cluster sample, it is shown in this paper, specifically in Sect. 2, that this mis-specification produces biased and hence inconsistent regression estimates.

To remedy the abovementioned inference drawback, i.e., to use the TSCS s^* in (6) correctly, it is important to note that this s^* is chosen from a FP ($\mathcal{F}$), also defined in (6), which involves two types of parameters, namely the regression parameter $\boldsymbol{\beta}$, and the cluster correlation parameter σ^2_γ. The estimation of $\boldsymbol{\beta}$ in the presence of σ^2_γ is new in the $\mathcal{F}$ setup; whereas it is known in an $\mathcal{S}$ setup that these parameters can be consistently and efficiently estimated by using the GLMM-based GQL estimation approach (Sutradhar 2004), for example. This motivated us in this paper to construct first two GQL estimating functions for $\boldsymbol{\beta}$ and σ^2_γ using the FP ($\mathcal{F}$)-based hypothetical data. These two functions were then estimated unbiasedly by using the TSCS s^* from (6), and these correct sample-based estimating equations were referred to as the SSDW (survey sample-based doubly weighted) estimating equations. More specifically, ($\mathcal{F}$)-based hypothetical estimating equations were provided in Sect. 4, and the TSCS s^*-based corresponding SSDW estimating equations were developed in Sect. 5. The resulting estimators were shown, in Sect. 6, to be consistent for their respective parameters. Thus, this paper for the first time provided a theoretical foundation for consistent estimation of the parameters using the TSCS data. The results developed in this paper, therefore, should be useful for practitioners from statistical agencies, for example, dealing with TSCS data of large size.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10463-022-00850-6.

Acknowledgements The author thanks two reviewers for their valuable comments and suggestions that lead to the improvement of the paper.

Funding No fund was used to complete this research.

Declarations

Conflict of interest None.

References

Amemiya, T. (1985). *Advanced econometrics*. Cambridge: Harvard University Press.

Binder, D. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review, 51*, 279–292.

Bishop, Y. M. M., Fienberg, S. E., Holland, P. W. (1975). *Discrete multivariate analysis*: *Theory and practice*. Cambridge: The MIT Press.

Burdick, R. K., Sielken, R. L., Jr. (1979). Variance estimation based on superpopulation model in two-stage sampling. *Journal of American Statistical Association, 74*, 438–440.

Christensen, R. (1984). A note on ordinary least squares methods for two-stage sampling. *Journal of American Statistical Association, 79*, 720–721.

Christensen, R. (1987). The analysis of two-stage sampling data by ordinary least squares. *Journal of American Statistical Association, 82*, 492–498.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York: Wiley.

Fuller, W. A. (2009). *Sampling statistics*. New York: Wiley.

Ghosh, M. (1991). Estimating functions in survey sampling: A review. In: V. P. Godambe (Ed.), *Estimating Functions* (pp. 201–210). Oxford Science Publications.

Godambe, V. P., Thompson, M. E. (1986). Parameters of super-population and survey population: Their relationships and estimation. *International Statistical Review, 54*, 127–138.

Jiang, J. (1998). Consistent estimators in generalized linear mixed models. *Journal of American Statistical Association, 93*, 720–729.

Lee, S. E., Lee, P. R., Shin, K. (2016). A composite estimator for stratified two stage cluster sampling. *Communications for Statistical Applications and Methods, 23*, 47–55.

Lehtonen, R., Veijanen, A. (1998). Logistic generalized regression estimators. *Survey Methodology, 24*, 51–55.

Nelder, J. A., Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of Royal Statistical Society, Series A, 135*, 370–384.

Prasad, N. G. N., Rao, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of American Statistical Association, 85*, 163–171.

Särndal, C.-E., Swensson, B., Wretman, J. (1992). *Model assisted survey sampling*. New York: Springer.

Scott, A. J., Holt, D. (1982). The effect of two-stage sampling on ordinary least squares methods. *Journal of American Statistical Association, 77*, 848–854.

Seber, G. A. F. (1984). *Multivariate observations*. New York: Wiley.

Sutradhar, B. C. (2003). An overview on regression models for discrete longitudinal responses. *Statistical Science, 18*, 377–393.

Sutradhar, B. C. (2004). On exact quasi-likelihood inference in generalized linear mixed models. *Sankhya B: The Indian Journal of Statistics, 66*, 261–289.

Sutradhar, B. C. (2011). *Dynamic mixed models for familial longitudinal data*. New York: Springer.

Sutradhar, B. C. (2020). Multinomial logistic mixed models for clustered categorical data in a complex survey setup. *Sankhya A, Available as online first article.* https://doi.org/10.1007/s13171-020-00215-2.

Sutradhar, B. C., Rao, R. P. (2001). On marginal quasi-likelihood inferences in generalized linear mixed models. *Journal of Multivariate Analysis, 76*, 1–34.

Ten Have, T. R., Morabia, A. (1999). Mixed effects models with bivariate and univariate association parameters for longitudinal bivariate binary response data. *Biometrics, 55*, 85–93.

Valliant, R. (1985). Nonlinear prediction theory and the estimation of proportions in a finite population. *Journal of American Statistical Association, 80*, 631–641.

Valliant, R. (1987). Generalized variance functions in stratified two-stage sampling. *Journal of American Statistical Association, 82*, 499–508.

Wedderburn, R. (1974). Quasi-likelihood functions, generalized linear models, and the Gaussian Newton method. *Biometrika, 61*, 439–447.